ANÁLISIS Y PREDICCIÓN DEL COMPORTAMIENTO DE LA OPERACIÓN DE TRANSPORTE DE CARGA AUTOMOTOR USANDO TÉCNICAS DE MINERÍA DE DATOS

Juan Camilo Estévez Cárdenas
Maestría en Ingeniería Industrial

Universidad Nacional de Colombia
Facultad de Ingeniería
Departamento de Ingeniería de Sistemas e Industrial
Bogotá, D.C.
Febrero de 2016

ANÁLISIS Y PREDICCIÓN DEL COMPORTAMIENTO DE LA OPERACIÓN DE TRANSPORTE DE CARGA AUTOMOTOR USANDO TÉCNICAS DE MINERÍA DE DATOS

Juan Camilo Estévez Cárdenas
Maestría en Ingeniería Industrial

Propuesta presentada para optar al título de
Magister Ingeniería de Industrial

Director
Wilson Adarme Jaimes, Ph.D.
Doctor en Ingeniería Industrial y Organizaciones

Codirector
Fabio Augusto González Osorio, Ph.D.
Ph.D. Computer Science

Línea de investigación
Logística y Sistemas Inteligentes

Grupo de investigación
Grupo de investigación SEPRO

Universidad Nacional de Colombia
Facultad de Ingeniería
Departamento de Ingeniería de Sistemas e Industrial
Bogotá, D.C.

Febrero de 2016

Título en español

Análisis y predicción del comportamiento de la operación de transporte de carga automotor usando técnicas de minería de datos

Title in English

Analysis and prediction of the behavior of the automotor load transport operation using data mining techniques

Resumen: En la actualidad Colombia presenta fuertes falencias en cuanto al manejo de los flujos de información Logística, que se encuentra segmentada, dispersa o ausente. Situación que limita la formulación de políticas públicas y planes de acción específicos enfocados a la optimización del sector logístico desde una visión integral de la cadena de abastecimiento, y como apoyo al incremento de competitividad y productividad. De manera particular no se cuenta con información detallada del funcionamiento de los principales corredores logísticos del país, desconociendo detalles importantes como lo son cantidades y tipos de carga movilizadas, tramos críticos que los componen, frecuencias de viajes dentro de los corredores, etc. Los sistemas de información disponibles presentan fuertes falencias en el apoyo a los procesos de planeación y toma de decisiones, ya que no incorporan herramientas descriptivas ni predictivas de la información que permitan realizar monitoreo y extracción de información relevante de grandes volúmenes de datos. Por lo tanto, herramientas analíticas como la minería de datos se convierten en una excelente alternativa para apoyar los procesos de generación de información para la planeación y la toma de decisiones en proyectos de transporte, impactando de manera directa en la logística y competitividad del país. La presente propuesta presenta el resultado del desarrollo de un sistema prototipo de apoyo a la toma de decisiones basado en técnicas de minería de datos que permite solventar las necesidades de información críticas en torno al comportamiento de los corredores logísticos en Colombia. Para la construcción del prototipo se basó en la metodología CRISP-DM 1.0 y los datos provenientes del sistema de información Registro Nacional de Despacho de Carga (RNDC).

Abstract: Currently Colombia has strong shortcomings in the management of logistics information flows, which is segmented, dispersed or absent. This situation limits the formulation of public policies and specific action plans aimed at optimizing the logistics industry from a holistic view of the supply chain and to support the increased competitiveness and productivity. In particular there is little detailed information on the functioning of the main logistics corridors in the country, ignoring important details such as quantities and types of cargo mobilized, critical sections that compose frequencies travel inside corridors, etc. Information systems available have strong weaknesses in supporting planning processes and decision making, and which do not incorporate predictive and descriptive tools of information that allow for monitoring and extracting relevant information from large volumes of data. Therefore, analytical tools such as data mining become an excellent alternative to support the process of generating information for planning and decision-making on transport projects directly impacting on logistics and competitiveness. This proposal presents the result of the development of a prototype system to support decision-making based on data mining techniques enabling meet the needs of critical information about the behavior of the logistic corridors in Colombia. For the construction of the prototype based on the methodology CRISP-DM 1.0 and data

from the National Register Information System Load Dispatch (RNDC).

Palabras clave: Logística, Minería de datos, transporte de carga

Keywords: Logistic, Data Mining, Transport Load

Nota de aceptación

Trabajo de tesis
Aprobado

"Mención Meritoria o Laureada"

Jurado
Elizabeth León Guzmán

Jurado
Luis Gerardo Astaiza Amado

Director
Wilson Adarme Jaimes

Codirector
Fabio Augusto González Osorio

Bogotá, D.C., Febrero de 2016

Dedicado a

A Dios, a mí amada madre, a mí querida familia y novia, por su paciencia y permanente apoyo.

Agradecimientos

Agradezco a mi alma mater, ***Universidad Nacional de Colombia****, por brindarme la oportunidad de seguir creciendo como persona, académicamente y como ciudadano.*

A mis profesores y directores de tesis por su apoyo permanente para la conclusión de este trabajo.

A mis compañeros del grupo de investigación por su solidaridad y amistad.

Índice general

Índice de tablas

Índice de figuras

Introducción

La de la gestión de la cadena de suministro y su componente articulador de transporte en conjunto con los sistemas de información se convierten en un mecanismo clave para soportan la toma de decisiones y la generación de políticas que incentiven la competitividad y productividad en el país.

La gestión eficiente de la cadena de suministro se ha convertido en un factor clave de éxito para las organizaciones, que tienen una orientación hacia el cliente, desde las cuales se busca la satisfacción de las necesidades de estos, debido a que los clientes tienen una tendencia a variar y a ser exigentes[1].

Las organizaciones como elementos activos dentro de la cadena de suministro (CS), requieren un proceso de mejora continua en torno a la correcta gestión de las interrelaciones entre agentes de cadena, con el fin de cumplir con las metas trazadas. Esta mejora busca dar solución a las expectativas de sus clientes en términos de calidad y cumplimiento, es así, como el manejo de la información juega un papel vital para la correcta coordinación de los diferentes eslabones de la cadena.

La logística son modos y medios representados en flujos de dinero, físicos, energía, información y de conocimiento que generalmente se agrupan en tres grandes procesos compra o aprovisionamiento, almacenamiento y distribución. En este constructo el transporte asociado al flujo físico es determinante[27]. La logística considera de manera integral las dimensiones asociadas a la infraestructura, normatividad y servicios logísticos siendo el centro de todo las personas, ver Imagen 1.

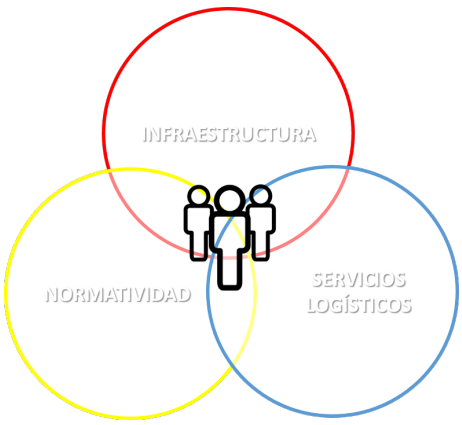


FIGURA 1. Dimensiones de la logística

Todos los eslabones juegan un papel fundamental en la cadena, sin embargo es necesario destacar el proceso de transporte de carga, debido a que permite la movilización de materias primas entre los diferentes agentes que conforman la cadena, es un elemento articulador entre los diferentes eslabones. Esta actividad se ve influenciada permanentemente por diferentes factores externos que afectan su funcionamiento y que en muchos casos son independientes a su operación, como políticas públicas, precios de combustibles y dinámicas económicas de los productos transportados, entre otros. Por lo tanto el uso de sistemas de información como apoyo a la gestión del transporte de carga, brinda facilidades al hacer más eficiente la administración del proceso del transporte al apoyarse con información adecuada para cada uno de los escenarios organizacionales e impactando directamente en la CS, minimizando su incertidumbre y el riesgo.

Desde este punto de vista, la minería de datos se presenta como una herramienta muy útil para obtener información que permita conocer y estimar el comportamiento actual del proceso de transporte de carga, realizar un monitoreo permanente con tecnologías de última generación, y llevar a cabo procesos de regulación, toma de decisiones y prospección a diferentes niveles dentro de la CS.

La mayor cantidad de transporte de carga se realiza por carretera, aproximadamente un 80 % [37], por lo tanto es importante entender las columnas vertebrales de este proceso en Colombia, conociendo cómo se comportan los principales corredores logísticos y su espectro de influencia en el transporte de carga a nivel nacional.

En el contexto nacional existe una permanente necesidad de contar con la información para realizar los procesos de regulación y legislación del transporte, sin embargo, antes del 2013 no existía un conjunto de datos, sistema de información o estudio que permitiera tener una información base para la realización de este estudio. En el 2013 el Ministerio de Transporte (MinTransporte) de Colombia hace un avance implementando el sistema de información Registro Nacional de Despacho de Carga por Carretera (RNDC), el cual permite realizar un registro automatizado de despachos de carga y mejorar el proceso de expedición del manifiesto de carga. Otorgándole la oportunidad a los transportadores de registrar sus movimientos de carga por carrera, bajo la modalidad de libertad vigilada y recolectando información importante acerca del comportamiento del transporte de carga.

A pesar del avance realizado con el RNDC y su importante aceptación dentro del gremio transportador, este no permite realizar un análisis directo y de manera detallada del funcionamiento de los corredores logísticos. No se puede desconocer que cuenta con características importantes, sin embargo estas podrían ser potenciadas mediante la implementación de herramientas analíticas como la minería de datos. El RNDC a través de la minería de datos plantea un escenario propicio para generar información encaminada a entender el comportamiento del transporte de carga en Colombia, y en particular de los principales corredores logísticos carreteros que movilizan carga en el país.

El estudio tiene como propósito encontrar, analizar y presentar información relevante acerca del comportamiento de los corredores logísticos en Colombia, mediante un análisis realizado al sistema de información RNDC y la construcción de un sistema prototipo de apoyo a la toma de decisiones basado en técnicas de minería de datos, para el apoyo en la toma de decisiones y formulación de políticas públicas. Su desarrollo fue basado en la metodología CRISP-DM 1.0 para la construcción de dos modelos: uno descriptivo correspondiente a un modelo de clustering, y otro predictivo correspondiente a un árbol de decisión.

Desde la ingeniería, el desarrollo de la investigación permite involucrar áreas como la logística, la minería de datos y el transporte. En el contexto industrial, los modelos se convierten en herramientas potenciales para ser adoptados no solamente por el Ministerio, sino por las empresas transportadoras, con el objetivo de mejorar sus procesos de planificación.

El documento está estructurado de la siguiente manera: en la primera parte se hace una descripción de la problemática, los objetivos, y la metodología. En la segunda parte se presenta una revisión teórica de temas como logística con sus diferentes técnicas y enfoques; minería de datos y sus aplicados a logística y transporte. En la tercera parte se presenta una caracterización y diagnóstico de la logística haciendo énfasis en el transporte y la implementación de tecnologías. En la cuarta parte se desarrollan los modelos basados en la metodología CRISP 1.0, donde se muestra de forma detallada el proceso de construcción de los modelos, las validaciones, los resultados y el análisis de los resultados. Por último se presentan las conclusiones y el trabajo futuro.

CAPÍTULO 1

Identificación del problema

En los últimos 4 años Colombia ha avanzado poco en términos de competitividad, de acuerdo con el Índice Global de Competitividad (IGC) del Foro Económico Mundial (WEF), pasó del puesto 68 en 2010 al puesto 66 en 2014 entre 144 países[125]. La carencia de un sector de transporte y de una cadena logística que preste servicios competitivos, es una de las grandes restricciones que afectan el desempeño logístico del país. Los bajos niveles de competitividad se evidencian en la *productividad* laboral del sector, la cual es baja en el contexto internacional, evidenciándose su deterioro durante los últimos años. Entre 2005 y 2012, la *productividad* relativa del sector de transporte en Colombia no superó el 17 % de la productividad del mismo sector en Estados Unidos. Durante estos años, la productividad del sector transporte fue inferior al promedio del país, ver figura 1.1.

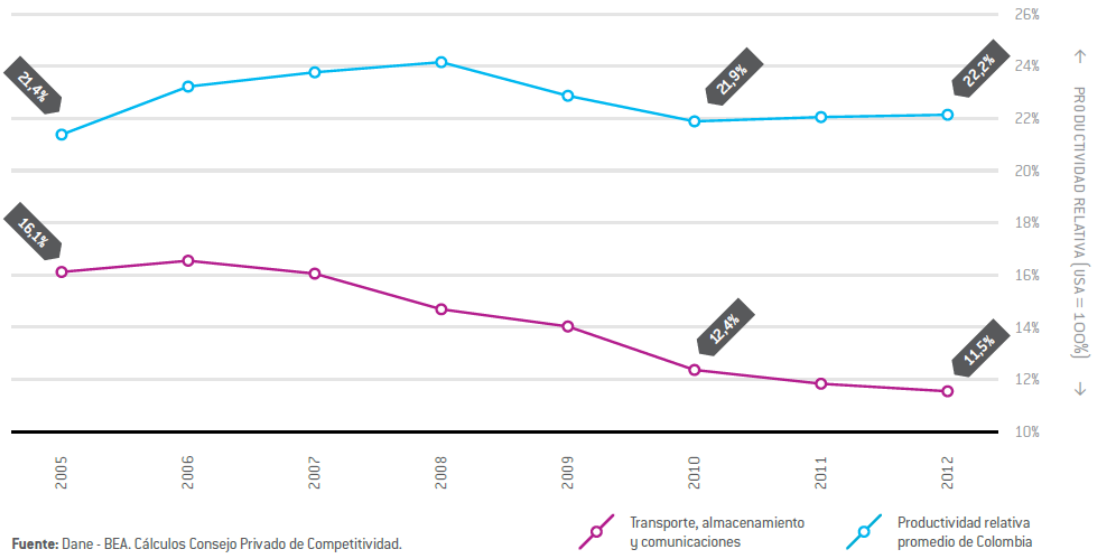


FIGURA 1.1. Productividad laboral relativa de Colombia frente a Estados Unidos (EEUU=100 %), 2005-2012. Fuente: Informe nacional de competitividad 2014-2015.

En el aspecto de infraestructura a Colombia no le va muy bien. Cuando se analiza el subpilar relacionado con infraestructura vial, férrea, portuaria y aeroportuaria, la posición

del país se deteriora aún más. El informe nacional de competitividad indica que si bien es clave contar con una infraestructura de calidad, esta debe estar acompañada de un buen desempeño logístico del país que depende de otros factores los cuales deben ser instrumentados de forma paralela a la agenda de infraestructura. Uno de los factores de rezago para la logística en términos de competitividad, corresponde a la calidad de la infraestructura, que incluye puertos, carreteras y *tecnologías de la información y la comunicación*. Aspectos relacionados con el uso de internet, disponibilidad de las últimas tecnologías, niveles de absorción de las firmas, suscripciones a internet móvil banda ancha, son considerados como ámbitos para el mejoramiento de la competitividad. Sin embargo, durante los últimos cuatro años el país perdió cinco posiciones en el pilar de preparación tecnológica, alcanzando en 2014 el puesto 68 entre los países medidos por el WEF[38].

Algunas de las causas de la baja competitividad en términos logísticos, son las falencias en el manejo de los flujos de información logística, la cual se encuentra segmentada, dispersa y/o ausente [41]. En particular en el ámbito del transporte, la figura 1.2 muestra las causas de los problemas asociados a la recolección de información de transporte[48]. Situación que limita la formulación de políticas públicas y planes de acción específicos enfocados a la optimización del sector logístico del país, desde una visión integral de la CS, y como apoyo al incremento de *competitividad y productividad*[41]. La problemática tiene como posibles causas, la carencia de un observatorio de indicadores que muestre la evolución de la logística del país, imposibilitando la priorización de acciones, la posibilidad de reorientarlas o reformularlas.

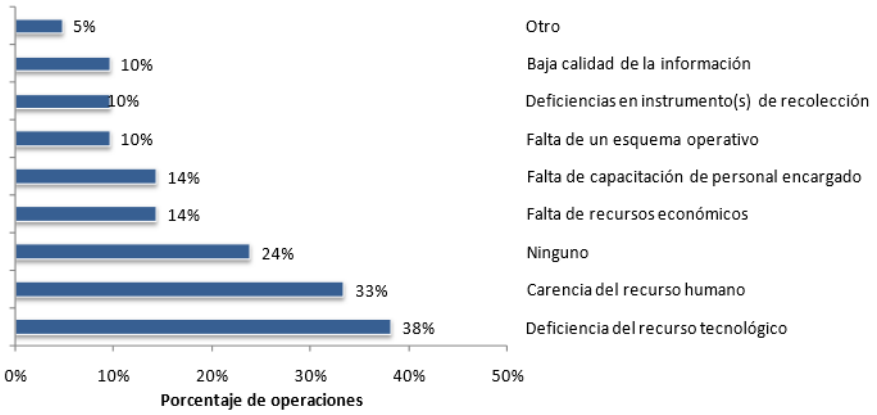


FIGURA 1.2. Causas de los problemas en la recolección de información del transporte. Fuente: DANE

En logística se presenta un uso limitado y escaso aprovechamiento de las Tecnologías de la información y las comunicaciones (TIC), herramientas que constituyen un elemento articulador entre los procesos de producción, distribución y disposición final en la CS. Las TIC son un elemento vinculante entre los diferentes actores involucrados en los flujos de bienes e información en la solicitud y recepción de pedidos, órdenes de servicios, transporte y almacenamiento de los bienes[41].

En el caso gubernamental, los sistemas de información estatales para la planeación logística en Colombia, no cuentan con herramientas analíticas eficientes que faciliten la gestión y la administración de cada una de las operaciones de las empresas y de la red logística que estas involucran; impidiendo la captura, la transferencia y la gestión de la información de una forma adecuada. La figura 1.3 muestra que una de los principales

causas de los problemas en el procesamiento de la información de transporte está asociada a la deficiencia del recurso tecnológico [48]. Lo que dificulta el proceso de administración, disminuye la capacidad de los administradores para la toma de decisiones y aumenta el error asociado a la interacción humana. Sistemas de información en el MinTransporte como el *Registro Nacional de Despacho de Carga por Carretera* (RNDC) y el Registro Único Nacional de Tránsito (RUNT) presentan falencias debido a que no incorporan herramientas descriptivas, ni predictivas de la información como apoyo al proceso de planeación y toma de decisiones, por lo tanto no se extrae información relevante de grandes volúmenes de datos, dificultando que la información pertinente esté disponible en el tiempo correcto para soportar la toma de decisiones estratégicas para el país.

FIGURA 1.3. Causas de los problemas de procesamiento de información del transporte. Fuente: DANE

En este contexto, la Comisión Económica para América Latina y el Caribe (CEPAL) coincide en que la utilización de sistemas inteligentes de transporte así como de otras TICs, constituyen la hebra que conecta y alimenta a una cadena logística cada vez más compleja y extensa, incrementando la competitividad de los participantes y *maximizando la productividad de la infraestructura* disponible.[54].

En el año 2013, MinTransporte implementó el sistema de información RNDC que tiene como finalidad optimizar el flujo de información acerca de la operación de transporte de carga. Sirve como base para el monitoreo de las relaciones económicas por parte de los integrantes del sector de transporte de carga, así como para el control por parte de la autoridad competente, garantizando la seguridad en la prestación del servicio público de transporte de carga, a cargo de aquellos particulares que se encuentran legalmente constituidos y debidamente habilitados por el Ministerio de Transporte[45].

A través del RNDC, las empresas de Servicio Público de Transporte Terrestre Automotor de Carga, deben expedir el manifiesto electrónico de carga y transmitir los datos al Ministerio de Transporte, conforme a lo dispuesto en el artículo 27 del Decreto 173 de 2001 y el artículo 4 del Decreto 1499 de 2009, donde menciona que el Manifiesto de Carga solo aplica a operaciones de transporte intermunicipal[45]. Los datos transmitidos del Manifiesto de Carga son fuente de información estadística de la movilización de carga en el país, y se convierte en uno de los insumos para fijar las políticas del sector con base en los indicadores generados [45].

El RNDC provee información sobre las operaciones que realizan las empresas que prestan el servicio público de transporte de carga por carretera que son insumos para monitorear el comportamiento del mercado conforme lo establece el Decreto 2092 de 2011[45]. El flujograma de recolección de datos en las diferentes etapas del proceso de transporte de carga se muestra en la Figura 1.4

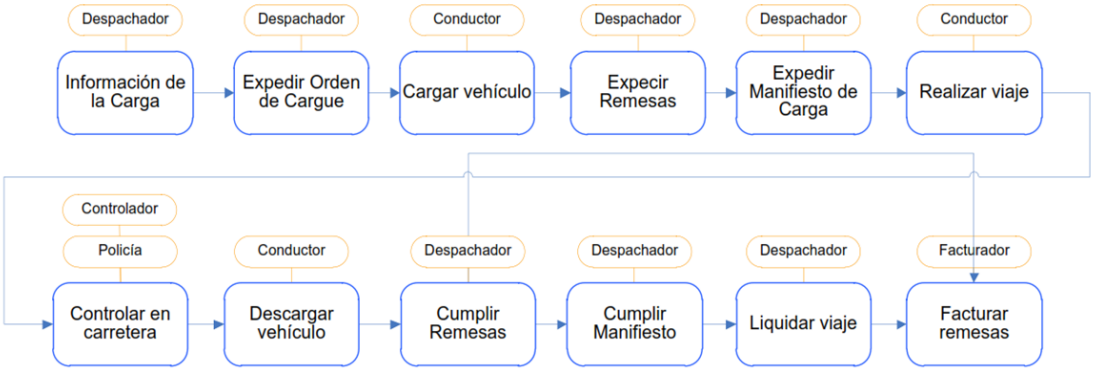


FIGURA 1.4. Diagrama de proceso de transporte de carga por carretera[45]

Este registro provee información a las autoridades encargadas de ejercer el control operativo y administrativo, como la Superintendencia de Puertos y Transporte, Policía Nacional, Dirección de Impuestos y Aduanas Nacionales (DIAN) y la Unidad de Información y Análisis Financiero (UIAF)[45].

Sin embargo, para el MinTransporte los datos en sí mismos no son útiles, es necesario procesarlos para extraer información y conocimiento. La entidad no explota la información de manera permanente, ni aplica técnicas o métodos para procesar y analizar datos de manera paralela al funcionamiento del RNDC. Tampoco se consumen de manera eficiente los datos que este integra para producir información valiosa en términos de indicadores, análisis y prospecciones. Por el contrario, se aplican técnicas convencionales de análisis estadístico de manera poco sistemática y que cumplen una tarea fundamental cuando se quiere obtener un *entendimiento inicial de los datos*, relegando la aplicación de técnicas alternativas como la minería de datos que permiten conocer con un mayor nivel de profundidad el comportamiento de los datos.

En estado actual del RNDC limita el alcance del monitoreo y no explota en su totalidad la riqueza de la información recolectada, incidiendo de forma directa en la calidad de la toma de decisiones y la formulación de políticas públicas. En particular la ausencia de modelos descriptivos dificulta la realización de labores de monitoreo del sistema logístico nacional, obviando relaciones o patrones en los datos de transporte de carga, que a simple vista con la estadística convencional no son visibles. La falta de modelos predictivos influye de manera directa en la calidad de las políticas públicas, debido a que durante su formulación no se tiene en cuenta la evolución de la información y ni las tendencias marcadas por los datos.

El RNDC cuenta con el potencial para generar información de calidad que permita la elaboración de bases técnicamente soportadas, para la formulación de políticas públicas y estrategias público-privadas que faciliten el desarrollo del sistema logístico de carga en los principales ***corredores logísticos de Colombia***, los cuales movilizan por lo menos el 70 % de la carga por carretera a nivel nacional y cuya información en la actualidad

es escasa y dispersa. El Consejo Privado de Competitividad en su Informe Nacional de Competitividad 2013-2014 recomendó expedir un documento Conpes que tuviera en cuenta la prioridad de corredores logísticos y la identificación de los cuellos de botella que limitaran la intermodalidad y la multimodalidad[38].

En consecuencia, se propone la minería de datos en conjunto con el RNDC como una herramienta que permite generar información de los principales *corredores logísticos en Colombia*, apoyando al proceso de toma de decisiones en la formulación de políticas para la logística en transporte con miras a mejorar la competitividad y productividad del país. También facilitará el proceso de gerenciar los corredores logísticos, que, además de velar por el buen mantenimiento y la ejecución de los corredores, busquen formas más eficientes de transportar la carga utilizando diferentes modos de transporte. Dado que la productividad de muchos de los sectores de la economía depende del sector transporte, incrementar su competitividad implicaría mejorar la competitividad del país[38]. Adicionalmente servirá como instrumento para facilitar el monitoreo y regulación de las actividades de transporte de carga realizado por parte del Ministerio de Transporte, y en particular para las entidades que tienen como fuente de información primaria el RNDC.

Dentro de este marco de referencia y teniendo en cuenta las características de los datos almacenados por el RNDC surge un interrogante:

¿Cómo incide el análisis y predicción de la operación del transporte de carga por carretera mediante el uso de técnicas de minería de datos en la mejora de la competitividad y productividad del país?

CAPÍTULO 2

Objetivos

2.1. Objetivo general

Desarrollar un sistema prototipo de análisis y predicción de la operación de transporte de carga terrestre colombiano mediante el uso técnicas de minería de datos, para el apoyo en la toma de decisiones y formulación de políticas públicas.

2.2. Objetivos específicos

1. Construir un *conjunto de datos* de la operación de transporte de carga automotor colombiano, apropiados para el entrenamiento y validación de modelos descriptivos y predictivos de técnicas de minería de datos.

2. Diseñar, elaborar y evaluar un *modelo descriptivo* de la operación de transporte de carga automotor colombiano, basado en técnicas de minería de datos, para determinar el comportamiento actual del sistema.

3. Diseñar, elaborar y evaluar un *modelo predictivo*, basado en técnicas de minería de datos, para estimar el comportamiento de la operación de transporte de carga automotor colombiano.

4. Desarrollar el sistema de análisis y predicción, evaluarlo de manera sistemática con datos actuales, para determinar la calidad de la información arrojada.

5. Evaluar la utilidad de los resultados de los modelos para la toma de decisiones y la formulación de políticas públicas en el sector de transporte de carga terrestre colombiano.

CAPÍTULO 3

Metodología

La naturaleza de la investigación es de tipo exploratorio, se realiza a partir de información primaria recolectada en la base de datos del RNDC. Se usa el software Rapidminer para realizar tareas de limpieza, selección, estandarización, normalización y creación de campos nuevos. Una vez tratada la información se aplicaron técnicas de minería de datos para la creación de los modelos propuestos, los cuales permiten describir y predecir el comportamiento del transporte de carga por carretera asociado a los principales corredores logísticos.

La metodología aplicada para el desarrollo de la investigación abordo cuatro etapas: en primer lugar se realizó una revisión del estado del arte en torno a las aplicaciones de minería de datos en logística, profundizando en las aplicaciones en transporte. En segundo lugar se realizó una selección de la metodología de minería de datos, de acuerdo a las características de los datos, el campo de aplicación y la pertinencia de la metodología. En tercer lugar se realizó la aplicación de la metodología de minería de datos seleccionada. Por último se recopilaron y analizaron los resultados arrojados por el proceso de minería de datos. Finalmente, se desarrolló las conclusiones y el trabajo futuro.

3.1. Metodología de minería de datos

Las metodologías de minería de datos permiten llevar a cabo el proceso de minería de datos en forma sistemática y no trivial, ayudan a las organizaciones a entender el proceso de descubrimiento de conocimiento y proveen una guía para la planificación y ejecución de los proyectos. Hay múltiples metodologías para llevar a cabo tareas de minería de datos, entre las más conocidas se encuentran: KDD[53], CRISP-DM 1.0[20], SEMMA[78].

Para la ejecución de la investigación se seleccionó la metodología CRISP-DM 1.0 debido a que de acuerdo al portal KDnuggets, uno de los principales portales en temas de análisis de datos, es la más utilizada en proyectos de analytics, data mining, o data science. En una encuesta realizada por este portal acerca de las principales metodologías utilizadas[113], CRISP-DM obtuvo el primer lugar con un 43 % del total, ver figura 3.1. Es una metodología fácil de entender, adaptable a diferentes contextos y que especifica cada una de las etapas y actividades por las que tienen que pasar los proyectos de minería de datos.

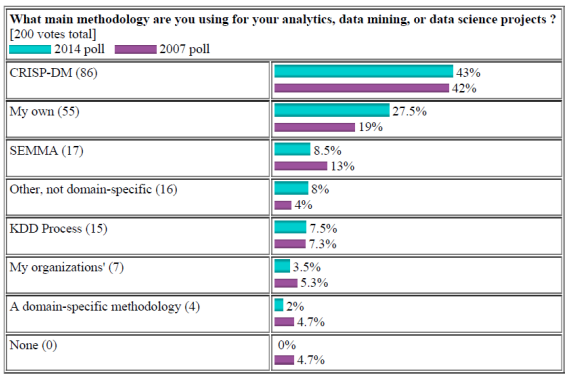


FIGURA 3.1. Encuesta KDnuggets para metodologías en proyectos de analytics, data mining, o data science[113].

Se descartaron metodologías como: SEMMA debido a que utiliza un software en particular de la empresa SAS; KDD es una buena alternativa y es de amplia difusión, pero su principal campo de aplicación es en contextos académicos, lo cual podría llegar a generar inconvenientes debido a la naturaleza de los datos. La tabla 3.1 muestra una comparación entre las principales metodologías de minería de datos[31].

Metodología	Knowledge Discovery in Databases (KDD)	SEMMA	Cross-Industry Standard Process for Data Mining (CRISP-DM)
Autor	Fayyad et al.	SAS Institute	Chapman et al. (IBM)
Dominio de Origen	Académico	Industria	Industria
Número de pasos	9	5	6
Pasos	1. Desarrollar y comprender el dominio de aplicación. 2. Crear un conjunto de datos objetivo. 3. Limpiar los datos y preprocesarlos. 4. Reducción de datos y proyección 5. Escoger la tarea de minería de datos. 6. Escoger el algoritmo de minería de datos. 7. Minado de datos. 8. Interpretar los patrones minados. 9. Consolidar el conocimiento descubierto	1. Muestreo 2. Comprensión 3. Modificación 4. Modelado. 5. Valoración	1. Comprensión del negocio. 2. Comprensión de los datos. 3. Preparación de los datos. 4. Modelamiento. 5. Evaluación. 6. Despliegue.
Notas	El más popular y más citado modelo; provee descripción técnica detallada al respecto del análisis de los datos, pero carece de aspectos empresariales.	Está especialmente enfocada al desarrollo del modelo de minería, y quedan fuera de su alcance otros aspectos del proyecto como el conocimiento del problema en estudio o la planificación de la implementación.	Usa vocabulario fácil de comprender; tiene buena documentación; divide todos los pasos en subpasos que proveen todos los detalles necesarios.
Software que lo soporta	MineSet	SAS Enterprise Miner	Clementine SPSS
Dominios de aplicación reportados	Medicina, ingeniería, producción, e-business, software	Marketing, ventas.	Medicina, ingeniería, marketing, ventas

TABLA 3.1. Comparación de metodologías de minería de datos [31]

3.2. CRISP-DM 1.0

El nombre CRISP-DM surge de las siglas en ingles Cross-Industry Standard Process for Data Mining[20], el cual fue establecido en el año de 1990 por cuatro compañías: Integral Solutions Ltd, un proveedor comercial de soluciones de minería de datos; NCR, un proveedor de bases de datos; DaimlerChrysler, una manufacturera de automóviles; y OHRA, una compañía de seguros. Las últimas dos compañías sirvieron como fuente de datos y casos de estudio[31].

La metodología CRISP-DM 1.0 está compuesta por seis fases (figura 3.2):

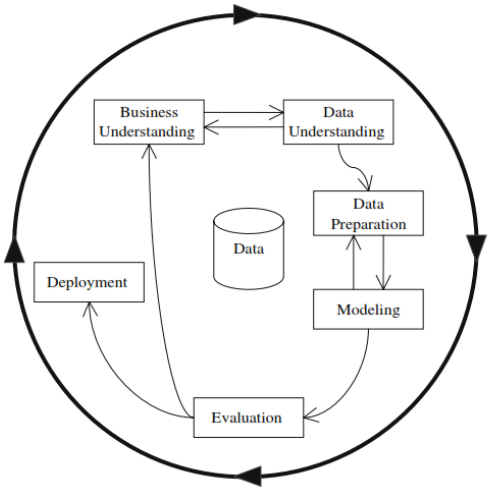


FIGURA 3.2. Fases de la metodología CRISP-DM[20]

- **Comprensión del negocio**: centrado en la comprensión de los objetivos y los requerimientos, para convertir tal conocimiento en la definición de un problema de minería de datos y diseñar así un plan preliminar para alcanzar los objetivos. En esta fase se realizan tareas como: determinar los objetivos del negocio, evaluar la situación, determinar las metas de minería de datos, realizar un plan del proyecto.

- **Comprensión de los datos**: consiste en la recolección y comprensión de los datos, identificando problemas de calidad en los datos y descubriendo señales interesantes en los datos, para plantear hipótesis sobre la información oculta. En esta fase se realizan tareas como: recolectar los datos iniciales, descripción de los datos, exploración de los datos, verificación de la calidad de los datos.

- **Preparación de los datos**: consiste en todas las actividades realizadas para construir el conjunto final de datos que serán utilizados para la formulación del modelo. Incluye labores de guardado y selección de atributos, transformación y limpieza de los datos. En esta fase se realizan tareas como: selección de los datos, limpieza de los datos, construcción de los datos, integración de los datos, aplicación de formato a los datos.

- **Modelamiento**: se realiza la selecciona las técnicas de modelado para la construcción de los modelos, que son aplicados y calibrados de acuerdo a los parámetros mas adecuados. En esta fase se realizan tareas como: seleccionar la técnica de modelamiento, generar una prueba del diseño, construir el modelo, evaluar el modelo.

- **Evaluación**: tiene como entrada el modelo más óptimo realizado en la fase anterior, y como objetivo evaluar si el modelo cumple con los objetivos del negocio, determinar si hay elementos que no fueron considerados en el planteamiento del modelo, y finalmente tomar una decisión acerca del uso final que se les va a dar a los datos generados. En esta fase se realizan tareas como: evaluar los resultados y revisar los procesos, determinar los siguientes pasos.

- **Despliegue**: busca organizar y presentar el conocimiento generado de tal forma que el cliente pueda usarlo. Puede tener como producto desde la generación de un reporte, hasta una implementación compleja en donde pueda repetirse el proceso de minería de datos. En esta fase se realizan tareas como: plan de despliegue, plan de monitoreo y mantenimiento, realizar el reporte final y revisión del proyecto.

CAPÍTULO 4

Estado del arte

4.1. Gestión de la cadena de suministro (SCM)

La cadena de suministro (SC) es una red de organizaciones y procesos de negocio para adquisición de materias primas, la transformación de estos materiales en productos intermedios y terminados, y la distribución de los productos terminados a los clientes. Se relacionan proveedores, fábricas, centros de distribución, puntos de venta y clientes para el suministro de bienes y servicios, desde el origen hasta el consumo. En general, los materiales, la información y el flujo de pagos a través de la SC fluyen en ambas direcciones [93].

La SCM es definida como la gestión de los materiales y los flujos de información dentro y entre las instalaciones, como proveedores, plantas de fabricación y montaje y centros de distribución[135]. Esta abarca desde la planificación y gestión de todas las actividades implicadas en el abastecimiento y adquisición, conversión, hasta todas las actividades de la gestión logística. Es importante destacar que, también incluye la coordinación y la colaboración con los socios de canal, que pueden ser proveedores, intermediarios, proveedores de servicios de terceros y clientes. En esencia, SCM integra la oferta y la gestión de la demanda dentro y fuera de las empresas.

4.1.1. Logística

El Council of Supply Chain Management Professionals define logística como "el proceso de planear, implementar y controlar el flujo y el almacenamiento eficiente y efectivo de los bienes, servicios e información relacionada desde el punto de origen al punto de consumo con el objetivo de satisfacer los requerimientos del cliente" [115]

La logística juega un papel determinante en el costo final de la mercancía, representando entre un 10 % y un 15 % del costo final de un producto elaborado[41], debido a que el abastecimiento, manipulación, almacenamiento y transporte de la carga puede incrementar en gran medida su valor, y en muchas ocasiones genera pérdida de competitividad de los productos y las empresas. Reducir los costos de la logística, mejorar los niveles y la calidad del servicio y aumentar su eficiencia operativa tienen un efecto directo en la generación de valor para las empresas y consumidores.

4.1.2. Teorías y enfoques en SCM

Las teorías y enfoques realizan enfasis en un marco de trabajo sobre el cual se realiza el proceso de SCM, a continuación se presentan algunas de las más representativas en la actualidad.

Green Supply Chain Management: pensamiento ambiental integrado en SCM, que incluye el diseño de productos, el abastecimiento y selección de materiales, procesos de fabricación, entrega del producto al consumidor final, así como la gestión de fin de vida del producto después de su vida útil. El alcance de SCM verde va desde el monitoreo reactivo de los programas de gestión de medio ambiente hasta prácticas más proactivas implementas a través de diversas R como lo son: reducir, reutilizar, retrabajo, refrescar, recuperación, reciclaje, refabricación, logística inversa, etc[130].

Lean Supply Chain Management: es una forma continua de mejora basada en equipos enfocados en identificar y eliminar "gastos". Gastos es una actividad que no agrega valor a la cadena desde el punto de vista del consumidor. En vez de una dieta, lean debe ser pensado como un programa de salud a largo plazo para los negocios. La implementación de lean en la organización genera beneficios como: reducción de costos operativos, mejora del rendimiento operacional y tiempos más cortos entre los ciclos de pedidos de los clientes. Lean hace énfasis en ocho gastos que son necesarios reducir: gastos de inventario, gastos de transporte, gastos de movimiento, gastos de espera, gastos de sobreproducción, gastos de sobreprocesamiento, gastos por errores o defectos y gastos comportamentales o de empleados subutilizados[111].

4.1.3. Técnicas y herramientas

Las técnicas y herramientas apoyan el proceso de gestión de la cadena de suministro y principalmente sirven como apoyo en la toma de decisiones para realizar una adecuada administración de los eslabones que componen la SC.

Métodos Heurísticos: procedimientos que tratan de descubrir una solución factible o muy buena, pero no necesariamente una solución óptima, para un problema específico bajo consideración. El procedimiento debe ser suficientemente eficiente como para manejar problemas grandes. Con frecuencia, el procedimiento es un algoritmo iterativo novedoso, donde cada iteración implica la realización de una búsqueda de una nueva solución que puede ser mejor que la solución que se encontró con anterioridad[73].

Metaheurísticas: son un método de solución general que proporciona tanto una estructura general como criterios estratégicos para desarrollar un método heurístico específico que se ajuste a un tipo particular de problema. Entre las metaheurística más difundidas se encuentran: la búsqueda tabú, templado simulado y algoritmos genéticos[73].

Programación entera: aplicación particular de la programación lineal en donde además de las restricciones propias de cada problema, se añade una restricción que exige que todos los valores de la solución estén dados por valores enteros. Esta es utilizada frecuentemente

para solucionar problemas en la que es necesario asignar a las actividades cantidades enteras de personas máquinas o vehículos[73].

Programación lineal entera mixta : consiste en un método que busca dar una solución matemática optima al problema de ubicación, o almenos una solución de precisión conocida. Su principal beneficio es su capacidad de manejar costos fijos de manera óptima. Ademas permite solucionar problemas en donde se debe encontrar el número, tamaño y ubicaciones de los almacenes en una red de cadena de suministros que minimizarán los costos fijos y variables lineales de desplazar todos los productos a través de la red. Adicionalmente la solución está sujeta a restricciones como: no puede excederse el suministro disponible de las plantas para cada producto, debe cumplirse la demanda para todos los productos, la utilización de cada almacén no puede exceder su capacidad, debe lograrse una utilización mínima de un almacén antes de que éste pueda abrirse y por último todos los productos de un mismo cliente deben atenderse desde el mismo almacén. La programación lineal entera mixta tiene como limitante los largos tiempos de solución del método para manejar problemas de ubicación a gran escala, a pesar de las mejoras notables en procesamiento de datos de los computadores actuales. Tambien presenta una dificultad para manejar funciones no lineales como pueden presentarse en las políticas de inventario, tarifas de transportación y relaciones de ventas y servicio al cliente[11].

Simulación: es una técnica matemática para probar el rendimiento de un sistema dadas unas entradas y/o unas opciones de configuración del sistema con un grado de incertidumbre. El uso de la simulación facilita la secuenciación de las operaciones de producción, flujos de análisis de la producción y el diseño de las instalaciones de fábrica. Una simulación produce distribuciones de probabilidad modelando el comportamiento de un sistema. Una compañía puede producir un modelo de simulación para construir planes de procesos para evaluar el rendimiento de la construcción de un plan bajo múltiples escenarios de demanda [115].

4.1.4. Tecnologías de la información

La necesidad de las empresas y naciones de ser competitivas en el nuevo entorno nacional e internacional, direccionado por la alta competencia y la apertura a nuevos mercados como resultado de la firma de tratados de libre comercio, genera un interés por disminuir costos logísticos y mejorar los niveles de servicio, elementos en donde las tecnologías de la información juegan un papel determinante. Algunas de las tecnologías más difundidas en SCM son las siguientes:

E-commerce: hace referencia al uso del Internet y la Web para realizar transacciones comerciales. El e-commerce está relacionado con la integración de servicios de comunicación electrónicos como el correo y el internet para llevar a cabo transacciones comerciales entre diferentes organizaciones e individuos. Dichas transacciones ocurren en Internet y en la Web. Las transacciones comerciales involucran el intercambio un valor, por ejemplo dinero, a través de las fronteras organizacionales o individuales a cambio de productos y servicios[93]. En el caso de la logística es una tecnología ampliamente usada tanto para la comunicación con el cliente, como para la generación de demanda de los diferentes bienes y servicios comercializados por la organización.

Radio Frequency Identification (RFID): consiste en sistemas que proveen tecnología para rastrear el movimiento de bienes a través de la SC. Los sistemas RFID usan pequeñas etiquetas con microchips embebidos que contienen la información acerca de un artículo y su localización, la cual es transmitida mediante señales de radio a cortas distancias de los lectores RFID. Los lectores RFID pasan entonces los datos a una red o a un computador para su procesamiento. A diferencia de los códigos de barras, las etiquetas RFID no necesitan contacto de línea de visión para ser leídos.

Las etiquetas RFID son programadas electrónicamente con información que puede identificar únicamente a un artículo más otra información asociada al mismo, como lo son ubicación, donde y cuando fue hecho, o su estatus durante la producción. Embebido en la etiqueta hay un microchip para almacenar la información. El resto de la etiqueta es una antena que transmite los datos al lector.

La unidad de lectura consiste en una antena y un radio transmisor con capacidad de decodificación unida a un dispositivo fijo o portátil. El lector emite ondas de radio en rangos que van desde 1 pulgada a 100 pies, dependiendo de su poder de salida, la radio frecuencia empleada, y condiciones ambientales que lo rodea. Cuando la etiqueta RFID entra en el rango del lector, esta se activa y empieza a enviar la información, el lector captura esos datos, los decodifica y envía de vuelta a través de una red cableada o inalámbrica a un computador para posterior procesamiento.

En control de inventario y SCM, los sistemas RFID capturan y gestionan información más detallada sobre los artículos en bodegas y la producción, que los sistemas de códigos de barras. Si un gran número de artículos son enviados juntos, los sistemas RFID permiten realizar seguimiento de cada paleta, lote o inclusive artículo unitario en el envío[93]. Importantes empresas como Unilever, Chevrolet, Ford, Proctor & Gamble y Wal-Mart tienen implementaciones de sistemas RFID[8].

Sistemas de información: el progreso en las tecnologías de la información y el incremento del uso del internet en el día a día de las empresas ha creado oportunidades para el software basado en SCM. Existen varios desarrollos originarios de las diferentes escuelas como ERP, aplicaciones de integración optimización matemática de la SC. Los objetivos básicos siguen siendo los mismos: bajos niveles de inventarios y mejoramiento de los servicios al cliente mediante el mejoramiento de la agilidad de fabricación[71].

Existen varias formas de clasificar el software utilizado en SCM, dependiendo de si se usa interna o externamente a la organización puede ser intrafirma o extrafirma. Desde el punto de vista de gestión de datos puede clasificarse como transaccional en caso de que se utilice para adquirir, procesar y comunicar los datos brutos sobre el pasado y el presente de las operaciones de la red de suministro de la firma, o analítico que permite evaluar y diseminar modelos de decisión basados en las bases de datos de decisión de la SC. Desde el punto de vista de la SC hay dos tipos principales de software: aplicaciones de planeación y aplicaciones de ejecución. En la figura 4.1 se puede observar la variedad de sistemas de información que existen dependiendo de sus características, cada color indica un software con características únicas.

Las principales implementaciones de sistemas de información se han dado en torno al apoyo de la gestión del almacenamiento y transporte, sistemas de planeación empresarial de recursos (ERP), aplicaciones de software para SCM y aplicaciones de integración empresariales[71].

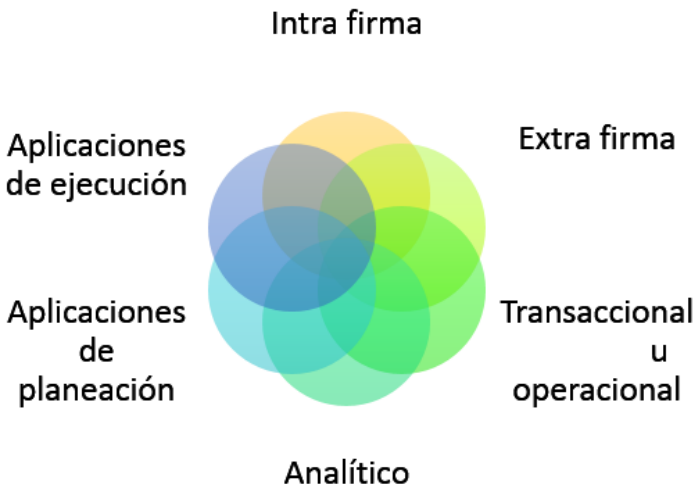


FIGURA 4.1. Sistemas de información en gestión de la cadena de suministro.

Unas de las tecnologías que se están empezando a implementar en el contexto logístico son los sistemas de información de ***inteligencia de negocios***, sistemas que tienen como finalidad realizar una explotación adecuada de la información con el propósito de hacer uso estratégico de la misma, facilitar la toma de decisiones y generar mejoras en la organización. En el contexto logístico una implementación de un sistema de inteligencia de negocios permitirá obtener mejoras en temas como:

- Obtener mejor información de los clientes, proveedores, socios e interesados.
- Añadir valor a los productos, servicios y procesos actuales.
- Crear nuevos productos, servicios, uso, canales ofertas y precios.
- Crear nuevos negocios y modelos de negocio y transformar mercados enteros.
- Realizar de manera continua experimentos de comportamiento del sistema logístico.
- Facilitar la toma de decisiones descentralizada y aplanar las estructuras.
- Potenciar la colaboración interna y premiar la innovación.
- Facilitar la colaboración externa, con proveedores o clientes, o de socios de negocio.
- Maximizar la gestión del talento.

La arquitectura de un sistema de inteligencia de negocios integra diferentes tipos de tecnologías con el fin de sacar el mayor provecho de la información. Una arquitectura típica está compuesta por 5 capas[21]: la primera corresponde a las fuentes de información, que se encarga de recolectar la información operacional de funcionamiento del negocio de fuentes de información que pueden ser homogéneas o heterogéneas y estar en diferentes tipos de formatos de almacenamiento. La segunda es la capa de estandarización o de movimiento de datos, que se encarga, mediante la utilización de un ETL o de motores de procesamiento complejo de eventos, de realizar un tratamiento de la información para unificar y transfórmala a un formato común. La tercera capa se encarga del almacenamiento de

los datos mediante la utilización de base de datos relacionales o motores MapReduce. La cuarta capa corresponde a los servidores de capa media, que consisten en un conjunto de servidores con funciones especializadas para realizar la explotación de la información, se pueden encontrar servidores de minería de datos, servidores OLAP, motores de búsqueda empresariales y servidores de reportes. Por último, la capa de aplicaciones de presentación, que se encargan de presentar la información a los usuarios de la manera más entendible posible utilizando tecnologías como las hojas de cálculo, tableros de mando, consultas ad hoc y buscadores (Figura 4.2).

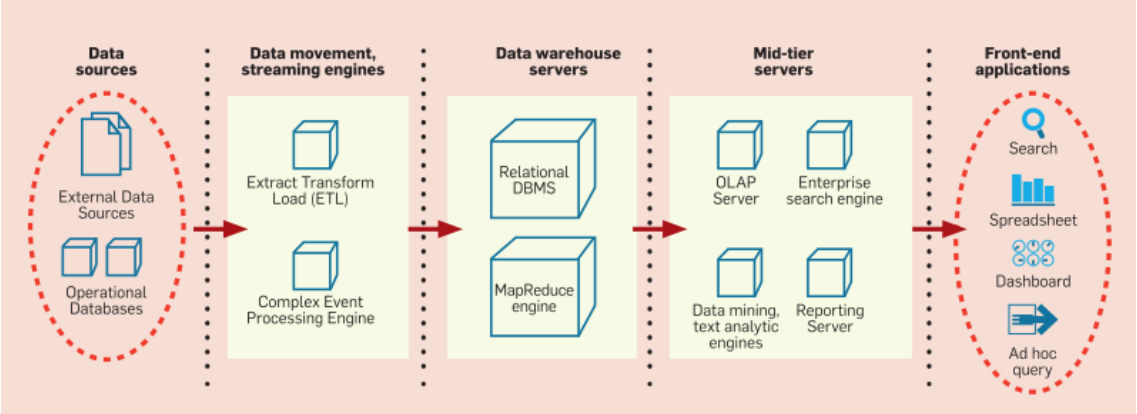


FIGURA 4.2. Arquitectura típica de un sistema de inteligencia de negocios[21].

Es importante resaltar que la capa que más agrega valor a la información procesada por el sistema de inteligencia de negocios, corresponde a la de servidores de capa media, ya que esta permite realizar análisis más sofisticados de la información mediante la utilización de cubos OLAP y de motores de minería de datos.

4.2. Minería de datos

La minería de datos consiste en un conjunto de técnicas y algoritmos que facilitan la búsqueda de información oculta y valiosa, en grandes volúmenes de datos[97]. Muchas organizaciones han recolectado y almacenado gran cantidad de datos sobre clientes actuales, clientes potenciales, proveedores y socios comerciales, sin embargo, se presenta una incapacidad para descubrir información oculta en los datos, impidiendo que las organizaciones transformen esos datos en conocimiento útil y valioso [107].

Como resultado de la aplicación de técnicas de minería de datos, se generan modelos que contienen y generan información depurada, confiable, bien definida y validada. Los modelos sirven para agrupar diferentes atributos dentro de una base de datos, como se realiza con técnicas de reglas de asociación. También permiten predecir la tendencia de una variable dependiente, como sucede con los arboles de decisión. Por último, se puede generar predicciones del comportamiento de las variables, como ocurre con el uso de series de tiempo y la técnica de regresión logística.

Hay dos tipos de tareas en minería de datos, las descriptivas y las predictivas [7][2]. Las tareas descriptivas buscan conocer cómo es el comportamiento pasado y actual de determinado conjunto datos, algunas técnicas representativas son agrupación, resumen, asociación y descubrimiento de secuencias. Por otro lado, las tareas predictivas buscan

hacer pronósticos de cómo se comportarían los datos en determinados escenarios futuros, teniendo en cuenta el registro histórico de los mismos, entre las técnicas más representativas se encuentran clasificación, regresión, análisis de series de tiempo y predicciones.

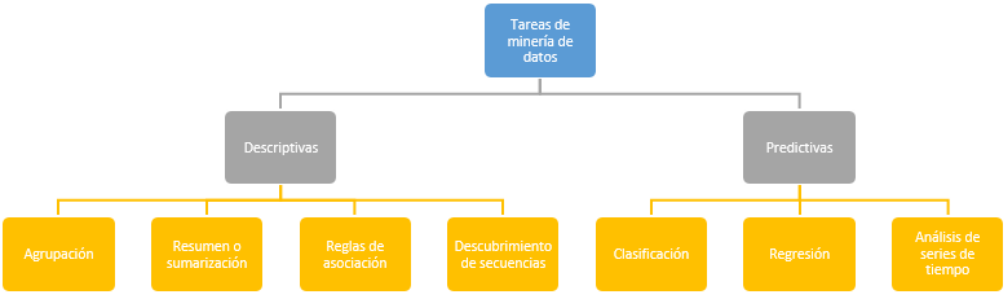


FIGURA 4.3. Tareas en minería de datos.

Dependiendo de si se conoce o no las categorías de salida o clases, las técnicas de minería pueden ser de *aprendizaje supervisado* o *no supervisado* [31] [81]. Las técnicas no supervisadas también son conocidas con el nombre de técnicas de descubrimiento de conocimiento y se utilizan para la detección de patrones ocultos, que pueden ser utilizados para la toma de decisiones. Cuando se intenta explorar repositorios de gran tamaño y complejidad se recomienda el uso de técnicas no supervisadas como reglas de asociación[3].

Las técnicas de aprendizaje supervisadas generan modelos a partir de datos que relacionan el valor de una etiqueta dependiente, con otros valores descriptivos independientes. Su principal uso está asociado a la predicción y clasificación de datos con etiquetas desconocidas. De forma general, la creación de un modelo supervisado se da en dos fases: la primera fase llamada de entrenamiento, consiste en la generación del modelo utilizando un algoritmo supervisado y un conjunto de datos de entrenamiento. En esta fase se imprime en el modelo mediante el algoritmo las características propias de cada conjunto de datos. La segunda fase llamada de validación, verifica la calidad del modelo mediante la utilización de un nuevo conjunto de datos de prueba. La aplicación de una red neuronal es un ejemplo de técnicas supervisadas, en donde se conoce los datos de entrada y la clasificación o predicción final que se quiere obtener.

Existe la posibilidad de combinar algoritmos de minería de datos, de tal forma que la salida de un modelo, se convierte en la entrada para un segundo algoritmo, generando así un segundo modelo con un mayor nivel de complejidad.

En la actualidad hay gran cantidad de técnicas y aplicaciones en minería de datos en áreas como marketing, finanzas, telecomunicaciones, manufactura, entretenimiento, gobierno, empresas farmacéuticas y ciencias de la salud [59][97][105][107][128]. Entre los principales métodos de minería de datos se encuentran: generalización, caracterización, clasificación, agrupamiento, asociación, evolución, pattern matching, visualización de datos y minado guiado de metadatos, los cuales hacen uso de diferentes herramientas de computación como: computación granular, neurocomputación, computación evolutiva y la vida artificial [148]. Asimismo, los algoritmos más utilizados en minería de datos son: C4.5, k-Means, SVM, Apriori, EM, PageRank, AdaBoost, kNN, Naive Bayes, and CART [149].

4.2.1. Asociación

La asociación es un método no supervisado de agrupamiento, es decir, no se tiene control directo de como las asociaciones entre los diferentes datos son generados. Son un ejemplo de minería de datos descriptiva, ya que su principal objetivo es determinar relaciones ocultas entre las diferentes variables dentro de una base de datos. La extracción de patrones frecuentes, llamadas reglas de asociación, conduce a patrones de datos con un nivel predefinido de la regularidad[2].

Una regla de asociación es una afirmación lógica que relaciona dos o más variables y son el producto del descubrimiento de relaciones de asociación. Los algoritmos de reglas de asociación descubren patrones de la forma "Si X entonces Y"[103] en donde X corresponde a una premisa, mientras Y corresponde a la conclusión. Las reglas de asociación generan una relación de causalidad a partir del cálculo de la frecuencia de ocurrencias entre dos o más atributos.

La búsqueda de patrones mediante reglas de asociación es compleja, ya que no es posible partir de una hipótesis, es una tarea netamente exploratoria, sin embargo, al delimitar el problema y restringir el conjunto de datos se puede obtener información acerca de la presunta relación de los atributos o campos de una base de datos. Una asociación entre dos o más atributos ocurre cuando la frecuencia conjunta de ocurrencia de los atributos es considerada alta.

Medidas como el soporte y la confianza permiten la evaluación de la calidad de los patrones extraídos en términos de la fuerza estadística del patrón [31]. En términos generales se busca la obtención de un número reducido de reglas con altos valores para el soporte y la confianza[103]. El soporte consiste en la frecuencia de aparición de un elemento o conjunto de elementos en determinado conjunto de datos poblacional, mientras que la confianza indica el porcentaje de elementos que cumplen la regla propuesta dentro del conjunto de datos soporte. Un criterio para la elección de reglas de asociación consiste en seleccionar las reglas con un grado de confianza y soporte mayor al mínimo especificado por el usuario[15].

Desde el punto de vista práctico la aplicación de técnicas de asociación tienen ventajas como: la facilidad de interpretación de las reglas y la posibilidad de aplicarla a grandes volúmenes de datos. En cuanto a las limitaciones se trabaja únicamente con variables categóricas, forzando la conversión de variables continuas a variables categóricas; segundo, el tiempo de procesamiento puede ser muy largo dependiendo del número de variables involucradas en el análisis, y por último, el número de reglas puede ser muy grande, razón por la cual es necesario realizar un adecuado análisis y priorización de las reglas[104].

Este tipo de técnica es aplicada con gran frecuencia en tareas de marketing como análisis de la canasta de compras, predicción de compras, personalización de productos, diseño de catálogos de productos, agregación de ventas, distribución de las tiendas y segmentación de clientes basado en los patrones de compra. Un ejemplo es el sistema de personalización y recomendación de Amazon, que basado en el historial de consumo y técnicas de asociación presenta a los clientes sugerencias de productos [31]. Entre los algoritmos más difundidos se encuentran el Apriori, Eclat y FP- Growth.

4.2.1.1. Algoritmo FP-Growth

Algoritmo creado por Jiawei Han [69], que calcula el conjunto de ítems frecuentes de un conjunto de datos mediante el uso de una estructura de datos de tipo FP-tree y un conjunto de datos de tipo binominal. FP-Growth se caracteriza por ser escalable y presentar mejor rendimiento que otros algoritmos similares como el Apriori o Tree Projection[69].

Un conjunto de ítems frecuentes son grupos de ítems o clases que aparecen juntos de manera regular en los datos. El algoritmo surge basado en el problema de análisis de la canasta de mercado en donde se describen relaciones de muchos a muchos entre los diferentes tipos de objetos o artículos. Los ítems o productos, son representados como atributos de tipo binominal y por lo tanto toman dos valores unicamente, "true.º "false", en donde true implica que el elemento hace parte de la canasta. Las canastas o también llamadas transacciones, corresponden a cada observación o cliente del conjunto de datos.

El problema de los ítems frecuentes busca encontrar el conjunto de ítems que aparecen juntos en un umbral mínimo de transacciones. Ese umbral es llamado también criterio de mínimo soporte y corresponde al número de veces que un conjunto de ítems aparece en un conjunto de datos dividido por el número total de ejemplos.

El descubrimiento de ítems frecuentes es visto con frecuencia como descubrimiento de reglas de asociación, aunque estas últimas son una forma más compleja de caracterización de los datos cuyo descubrimiento depende fundamentalmente del descubrimiento de ítems frecuentes[5]. Debido al desempeño en el procesamiento de grandes conjuntos de datos se considera el algoritmo FP-Growth mejor en comparación al algoritmo A priori, también utilizado para hallar conjunto de ítems frecuentes[5].

4.2.2. Clasificación

La clasificación permite reconocer los patrones que describen a un grupo de elementos. Usa los elementos existentes que han sido clasificados para inferir un conjunto de reglas y realizar la clasificación, es utilizada para realizar análisis descriptivos y predictivos. La clasificación se realiza teniendo en cuenta un conjunto de etiquetas o clases determinados de manera previa, implica asignar etiquetas a los registros de datos inéditos sobre la base de los conocimientos extraídos de los datos históricos.

Algunos métodos de clasificación son template maching, nearest mean classifier, subspace method, 1-nearest neighbor rule, k-nearest neighbor rule, bayes plug-in, Logistic classifier, parzen classifier, fisher linear discriminant, binary decision tree, perceptron, multi-layer perceptron, radial basis network y support vector classifier[81]. Dentro de los algoritmos mas utilizados de clasificación se encuentran árboles de decisión[147], redes neuronales[94], árboles C4.5, CART y clasificadores bayesianos [4] [28][29]].

4.2.3. Agrupación

Su funcionamiento es muy similar a la clasificación, la principal diferencia radica en que no hay grupos o etiquetas definidos. La segmentación o clustering divide a una población en subpoblaciones más pequeñas con un comportamiento similar. Las técnicas de agrupación buscan maximizar la homogeneidad dentro de los elementos pertenecientes a los grupos y a la vez la heterogeneidad entre los grupos identificados. Las tareas de descripción se

centran en la explicación de las relaciones entre los datos[2], mientras que las tareas de predicción se centran en determinar la validez de los grupos formados.

Un algoritmo de agrupamiento identifica conjuntos de elementos similares de acuerdo con una métrica predefinida. Dentro de los algoritmos de agrupación se encuentran: k-means, k-medoids, DBSCAN, algoritmos jerárquicos, reconocimiento de patrones, estadística bayesiana y las redes neuronales. Existen múltiples aplicaciones de la agrupación en áreas como genética[77], medicina[120], marketing[106], telecomunicaciones[109] e informática[13].

4.2.4. Minería de datos en logística

Existe una considerable tasa de producción científica en lo relacionado a minería de datos en SCM. La aplicación de técnicas y métodos de minería de datos es variado entre las diferentes actividades y procesos de SCM. De acuerdo a una revisión realizada en la herramienta bibliográfica Scopus la producción científica, a excepción del año 2004, ha tenido un crecimiento constante. El crecimiento en el número de publicaciones tiene como punto de inicio principios de la década del 2000, y alcanza su punto máximo en el año 2013, ver figura 4.4.

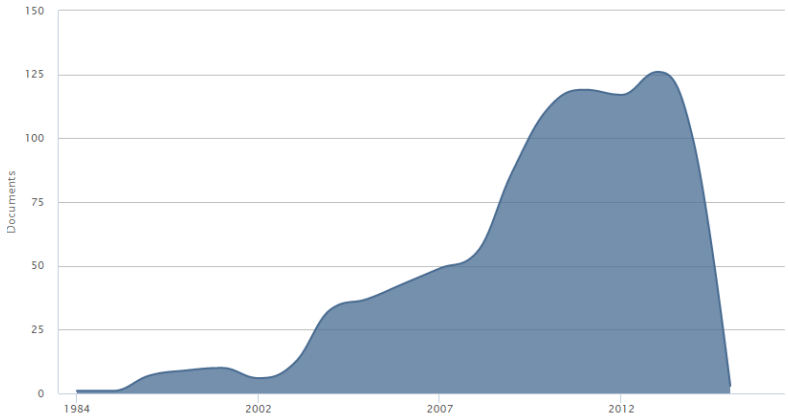


FIGURA 4.4. Tendencias de investigación de minería de datos en gestión Logística. Fuente: Elsevier SCOPUS

Se clasificaron los artículos de acuerdo a la técnica de minería de datos, actividad y proceso de SCM para evidenciar cómo se comporta la producción de literatura. Se seleccionaron cuatro grandes procesos que agrupan diferentes actividades relacionadas con su aplicación dentro de la SC, estos son: suministro, gestión de existencias, fabricación y distribución. En la tabla 4.1 se muestra los resultados resumidos de la clasificación de la literatura.

Métodos	Literatura	Técnica	Actividad	Proceso
	[98]	Reglas de asociación	Selección de proveedores	Suministro
	[82]	Reglas de asociación difusas		
	[137]	Reglas de asociación y algoritmos genéticos	Manejo de inventarios	Gestión de existencias
	[68]	Reglas de asociación		
	[25]	Agrupamiento y reglas de asociación	Ordenamiento de lotes basado en la demanda	Distribución

	[83]	Reglas de asociación difusas	Evaluación de la agilidad en la cadena de suministro	
	[129]	Algoritmos de asociación	Configuración óptima del producto	Fabricación
	[96]	Algoritmo apriori y agrupación	Desarrollo de nuevos productos	
	[2]	Algoritmos de asociación, agrupamiento, clasificación	Diseño de familias de productos	
	[114]	Reglas de asociación	Configuración dinámica de la cadena	
	[92]	Reglas de asociación difusas	Satisfacción del usuario	
	[112]	Reglas de asociación, redes neuronales, arboles de decisión, k - means	CRM	Distribución
	[26]	Reglas de asociación	Agrupación por lotes para el proceso de distribución	
Clasificación	[89]	Lógica difusa, redes neuronales, árboles de clasificación y técnicas de agrupamiento	Pronóstico en ventas	Distribución
	[112]	Reglas de asociación, redes neuronales, árboles de decisión, k - means	CRM	Distribución
	[137]	Reglas de asociación y algoritmos genéticos	Manejo de inventarios	Gestión de existencias
	[2]	Algoritmos de asociación, agrupamiento, clasificación	Diseño de familias de productos	Fabricación
	[94]	Redes neuronales	Selección de proveedores	Suministro
	[147]	Redes neuronales y programación entera		
	[4]	Redes neuronales		
	[28]	Redes neuronales		
	[29]	Redes neuronales		
	[151]	Redes neuronales y k-means		
	[95]	Teoría de conjuntos ásperos y análisis relacional gris		
	[55]	Redes bayesianas y lógica difusa		
	[89]	Lógica difusa, redes neuronales, árboles de clasificación y técnicas de agrupamiento	Pronóstico en ventas	Distribución
	[112]	Reglas de asociación, redes neuronales, árboles de decisión, k - means	CRM	

Agrupación

	[2]	Algoritmos de asociación, agrupamiento, clasificación	Diseño de familias de productos	Fabricación
	[96]	Algoritmo apriori y agrupación	Desarrollo de nuevos productos	
	[62]	Agrupamiento por K-means	Gestión de movimiento de mercancías	Gestión de existencias
	[74]	Algoritmos de agrupamiento	Selección de proveedores	Suministro
	[22]	Agrupamiento por K-means		
	[151]	Aplicación de redes neuronales y k-means		

TABLA 4.1. Tabla de clasificación de literatura de minería de datos aplicada a logística. Elaboración propia

La tasa de producción es proporcional con respecto a la aplicación de los tres métodos de minería: asociación, clasificación y agrupación. Respecto a las técnicas de minería hay una clara predilección por técnicas como reglas de asociación, redes neuronales, y algoritmos de agrupamiento, ver tabla 4.1.

Se evidenciaron aplicaciones de reglas de asociación en diferentes actividades de SCM como son: selección de proveedores [98, 82], manejo de inventarios [137, 68], ordenamiento de lotes basado en la demanda [25], evaluación de la agilidad en la cadena de suministro [83], configuración óptima del producto [129], desarrollo de nuevos productos[96], diseño de familias de productos [129], configuración dinámica de la cadena[114], satisfacción del usuario[92], CRM[112][Ngai 2009], agrupación por lotes para el proceso de distribución [26]. En cuanto a métodos de clasificación se encontraron aplicaciones relacionadas con las actividades de: pronóstico en ventas [89], CRM[112], manejo de inventarios [89], diseño de familias de productos[89] y selección de proveedores [94, 147, 4, 28, 29, 151, 95, 55, 98].

En actividades relacionadas con métodos de agrupamiento, estos se aplicaron a actividades como pronóstico de ventas[89], CRM[112], diseño de familias de productos[2], desarrollo de nuevos productos [96], gestión del movimiento de mercancías [62] y selección de proveedores [74, 22, 151]. Dentro de los documentos recolectados hay una clara tendencia hacia la aplicación de técnicas de minería en el proceso de selección de proveedores [98, 83, 94, 147, 4, 28, 151, 95, 55, 74, 22]. Además se identificó una clara tendencia hacia la utilización de redes neuronales como herramienta de selección de proveedores[98] [94, 147, 4, 28, 29, 151, 95, 55]

Los resultados de la revisión muestran que la minería de datos ha sido ampliamente usada en SCM y es un área muy activa en producción científica. Las principales técnicas de minería de datos aplicadas son: reglas de asociación, redes neuronales y agrupamiento. Por otro lado, las actividades de la cadena que más presentan aplicaciones son: selección de proveedores, diseño de productos, manejo de inventarios y ordenamiento por lotes.

4.2.5. Minería de datos en transporte

Mediante una revisión a la herramienta bibliográfica SCOPUS y las principales bases de datos de información científica, se encontró que la cantidad de información disponible

acerca de aplicaciones de minería de datos en temas de transporte en la actualidad es muy limitada, ver tabla 4.2.

Método	Técnica	Autor	Objetivo
Asociación	Secuencial Pattern Data Mining	[19]	Determinar el estado del vehículo cuando está fuera de servicio, manejando en contra de las regulaciones de tráfico o cuando se desvía de la ruta.
	Association Data Mining Rules	[67]	Identificar por que suceden los accidentes de tranvía y carros eléctricos en una red de transporte público en República Checa.
	Association Rules	[12]	Encontrar asociaciones entre un conjunto de vías y diferentes tipos de accidentes.
	Rough sets theory and the theory of association rules	[138]	Determinar las posibles causas de accidentes en las vías utilizando técnicas de minería de datos.
Clasificación	Bayesian binary classification method	[58]	Minar patrones de transporte personales los cuales proveen señales para el diseño de sistemas de información de viajes personalizados para cada usuario. Los datos fueron recolectados mediante sistemas de recaudo automático del sistema de buses de Lisboa (Portugal). Se formula un algoritmo de clasificación para identificar la combinación de características con más alta capacidad de predicción.
Agrupación	Clustering and Text mining	[12]	Identificar las características comunes en diferentes tipos de accidentes.
	Dendrogram clustering	[91]	Evaluar métodos para predecir tiempos de viajes personalizados para los usuarios del sistema del metro de Londres, y clasificar las estaciones basado en futuros patrones de movilidad, con el fin de identificar el conjunto de estaciones que son de mayor interés para el usuario y así proveer actualizaciones útiles de los viajes.

TABLA 4.2. Tabla de clasificación de literatura minería de datos aplicada a transporte. Elaboración propia

Se encontraron aplicaciones asociadas al uso de sistemas de diagnóstico automático de vehículos[117]. Sistemas que permiten determinar de una manera rápida y eficiente las causas de las fallas, minimizando la intervención de un humano. También investigaciones en manejo del tráfico y congestión por carretera, monitoreo de conductores somnolientos, análisis de accidentes en carretera, gestión de información del pavimento, sistemas de información geográfica para datos de transporte, datos GPS, registros de carreteras de vídeo, datos espaciales y el análisis de datos de rugosidad de carretera utilizando herramientas de minería de datos para identificar la compleja relación entre la naturaleza de datos de mundo lógico, físico, real y virtual. Se identificaron tambien algunas aplicaciones potenciales de minería de datos al transporte como lo son[12]:

Gestión del tráfico: aplicaciones como clasificación de los vehículos en diferentes categorías, clasificación de los tipos de accidentes en las vías críticas, identificación de los sitios

de alto riesgo basado en la frecuencia de los accidentes, asociaciones entre vías y los tipos de accidentes, agrupación de características comunes para tipos de accidentes.

Gestión de la información del pavimento: la correcta administración de la información de vías pavimentadas, permitirá a las entidades correspondientes tomar mejores decisiones acerca del mantenimiento y rehabilitación de las vías. Variables como tipo de vía, tipo de material de construcción, años de funcionamiento de la vía, tipos de vehículos (pasajeros o de carga), entre otras, analizadas con técnicas de minería de datos presentan un gran potencial para convertirse en herramientas para la toma de decisiones [6].

Sistemas de información geográfico para transporte: debido a que estos sistemas de información pueden modelar de una forma muy cercana diferentes tipos de información contenida por capas de elementos del mismo tipo, la minería de datos se puede llegar a convertir en una excelente herramienta para determinar las complejas relaciones entre las diferentes capas.

Datos espaciales de las carreteras: mediante sistemas de video y el uso de vehículos se puede recolectar información valiosa acerca de las carreteras, la cual puede ser explotada para obtener patrones y referencias geométricas espaciales inventariadas de las vías y carreteras, que permitirán realizar una mejor planeación de su uso y mantenimiento.

Análisis de los datos de rugosidad de la vía: medidas de la rugosidad de la vía permitirá obtener una mejor idea de la calidad de un pavimento, satisfacción del usuario de la vía y costos operativos de los vehículos. La minería de datos puede ser una buena herramienta para determinar relaciones entre la localización de la vía, satisfacción del uso de la vía por parte del usuario y rugosidad de la misma.

De esta manera, se presenta el estado del arte sobre el cual se desarrolla el trabajo de investigación. Se presentó las bases teóricas en torno a logísticas y minería de datos, y una recopilación de la literatura más sobresaliente en cuanto a la aplicación de minería de datos en logística y transporte.

CAPÍTULO 5

Caracterización y diagnóstico

El transporte de carga está altamente concentrado en las carreteras, por donde circula aproximadamente el 80 % de la carga movilizada en el país[37]. Según el Plan Nacional de Desarrollo 2014-2018 el transporte de carga representa en el país un 23 % del producto interno bruto[43]. Se describirá a continuación cómo es la distribución de la población Colombiana por departamentos y ciudades, con el fin de identificar cuáles son los principales centros de origen y destino de la carga.

5.1. Demografía

La población colombiana se distribuye principalmente en la región andina, siendo Bogotá la ciudad más poblada, seguida por los departamentos de Antioquia, Valle del Cauca y Cundinamarca, ver figura 5.1.

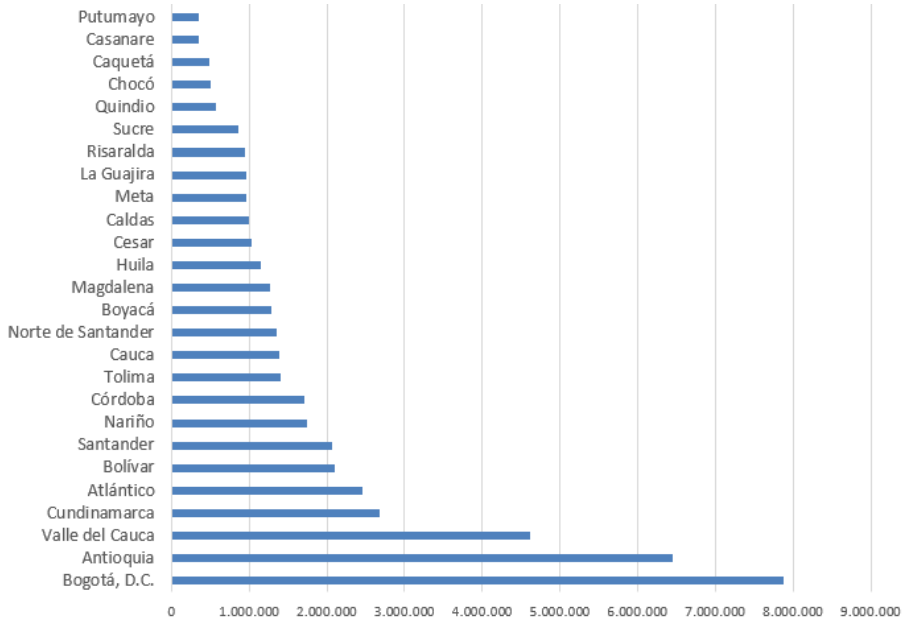


Figura 5.1. Distribución poblacional por departamento. Fuente: DANE

En cuanto la distribución de la población por municipios esta se ubica en su mayoría en la ciudad de Bogotá con casi 8 millones de habitantes, seguido por Medellín, Cali, Barranquilla y Cartagena, ver figura 5.2.

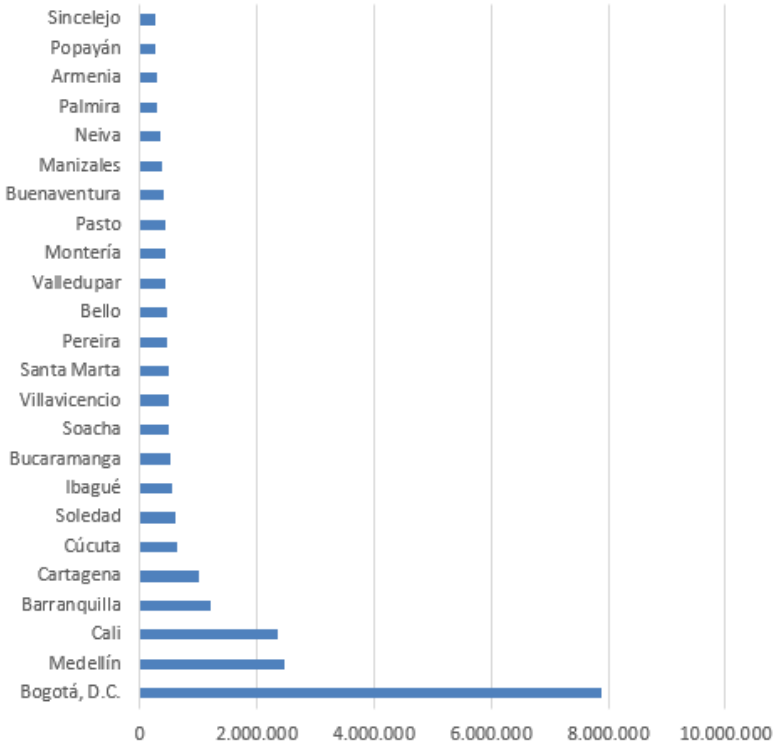


FIGURA 5.2. Distribución poblacional por municipios. Fuente: DANE

5.2. Infraestructura

El transporte terrestre de carga en Colombia enfrenta inconvenientes que se enmarcan en diferentes contextos, se puede resaltar el rezago en materia de infraestructura, donde la red secundaria y terciaria suma aproximadamente el 90 % de las vías colombianas y existe un retraso amplio en comparación con América Latina y más del 87 % de los países del mundo[150]. El flujo de recursos para la construcción, rehabilitación y mantenimiento de las vías secundarias y terciarias, no es estable, y por esta razón el mantenimiento no es preventivo como debería ser, sino correctivo. Se evidencia deficiencia en las condiciones técnicas de la malla vial, un bajo número de puentes, túneles y doble calzados que dificulta el movimiento de mercancías desde los orígenes hasta los destinos. Un país de las características de Colombia debería tener 26 % más de kilómetros de carreteras, es decir, existe un déficit de 45 mil kilómetros aproximadamente[150]. Según la encuesta nacional de logística 3 de las 4 barreras que impactan la logística de los prestadores de servicios logísticos están asociados a temas de infraestructura[44], ver figura 5.3.

Tambien se presenta obsolescencia del parque automotor debido a retrasos en incentivos y acciones regulatorias, donde el 80 % de los vehículos que transportan carga tienen más de 20 años de antigüedad[102]. Según cálculos del sector, en el país circulan 250.000 camiones, de ellos, 23.000 movilizan la carga relacionada con el comercio exterior. Se estima que un 73,9 % de estos 23.000 camiones tienen alrededor de 25 años de antigüedad[49]. Tampoco

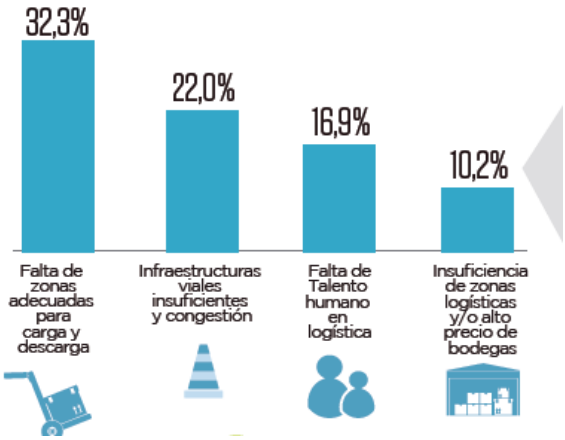


FIGURA 5.3. Principales barreras que impactan la logística de los prestadores de servicios logísticos. Fuente: Encuesta nacional de logística 2015

se ha evidenciado tarifas de transporte diferenciadoras que beneficien o promueva el uso de nueva flota, generando problemas no solamente en términos de la calidad del servicio, sino también a nivel ambiental.

De otro lado, hay un alto costo en las tarifas de los peajes, siendo el tercer país más costoso de América Latina despues de Uruguay y Perú; excesivo número de peajes, con 33.825 kilómetros de vías concesionadas existen 462 peajes, es decir, hay un peaje cada 73 kilómetros [79]. También influyen factores como barreras topográficas propias del relieve colombiano, problemas de seguridad debido a los conflictos políticos internos, altos costos de combustible y problemas de educación vial. Estos inconvenientes se ven reflejados en el índice de desempeño logístico del banco mundial, el cual evalúa diversos factores que inciden en el comportamiento de la logística de comercio exterior[41]. Como resultado del análisis Colombia pasó de ocupar el puesto 64 entre 155 países en el estudio realizado en el año de 2012[10] a ocupar el puesto 97 entre 160 en el año 2014[9].

En este sentido, Colombia ha iniciado la construcción de carreteras llamadas de cuarta generación o 4G, con una inversión aproximada de 47 billones de pesos con lo que se debería contemplar no solo la modernización de las vías, sino también la construcción de carriles exclusivos para el transporte de carga con el fin de potenciar la competitividad, facilitar la productividad e incentivar el sector con miras a aprovechar los trece acuerdos comerciales vigentes con deferentes países.

5.2.1. Corredores logísticos

El Banco Interamericano de Desarrollo (BID) [119] define bajo el nombre de corredor, a un eje funcional logístico, asociado a un movimiento comercial de mercancías, y compuesto físicamente por un conjunto de tramos, por ejemplo, carreteras, y nodos de infraestructura de diferentes tipos, como puntos de descargue de mercancía tales como puertos, aeropuertos, etc. Para este trabajo, se define *tramo* como un segmento del corredor que tiene al inicio (origen) y al final (destino) cualquiera de los municipios que componen el corredor.

En el contexto colombiano, el CONPES 3547 [41], define corredores logísticos como aquellos que articulan de manera integral orígenes y destinos en aspectos físicos y funcio-

nales como la infraestructura de transporte, los flujos de información y comunicaciones, las prácticas comerciales y de facilitación del comercio.

En Colombia, los corredores logísticos unen los principales centros de producción con los de consumo interno y los nodos de transferencia de comercio exterior (como puertos, aeropuertos y pasos de frontera). Estos distribuyen la gran mayoría de la carga tanto de comercio exterior como interno y están estrechamente relacionados con el patrón de desarrollo vial, incluyendo los diferentes modos de transporte[41].

Teniendo en cuenta los corredores propuestos por el CONPES 3547, ver figura 5.4, y un contraste realizado con los datos del RNDC se identificó como los principales corredores de transporte de carga automotor a: Bogotá-Barranquilla, Bogotá-Bucaramanga, Bogotá-Cali, Bogotá-Medellín, Medellín-Barranquilla, Medellín-Bucaramanga, Medellín-Cali. Cada corredor tiene asociado un conjunto de municipios intermedios entre su origen y destino principal y un conjunto de municipios ramales por los cuales se distribuye la mercancía a municipios aledaños.

A pesar de que los corredores logísticos están claramente identificados, en la actualidad no existe un método o sistema para la recolección y análisis de información relacionada con el comportamiento de los corredores. Esta coyuntura impide realizar adecuadamente los procesos regulatorios, de monitoreo y vigilancia sobre los corredores con información actualizada que refleje el comportamiento actual del movimiento de carga.

FIGURA 5.4. Principales corredores logísticos definidos por el CONPES 3547

5.2.2. Infraestructura tecnológica

Adicional al retraso de la infraestructura vial, sobresale también el retraso en torno a la *infraestructura tecnológica* que permite la comunicación efectiva entre las entidades

administradoras de las vías, los conductores de los camiones y las entidades reguladoras, como por ejemplo sistemas para reportar condiciones de concurrencia, calidad y accidentes en las vías.

La encuesta nacional del logística de 2015 muestra que solamente del 31.8 % de las empresas cuentan con un sistema de trazabilidad implementado, de estos 63 % realizan trazabilidad hacia atrás, 83.3 % realizan trazabilidad interna y 59.3 % trazabilidad hacia adelante[44].

Se observa que no solamente Colombia, sino también en Latinoamérica la adopción de herramientas tecnológicas está altamente atomizado. Esta situación representa un enorme desafío para las autoridades sectoriales, ya que cada vez más la competitividad de los puertos depende de la calidad de su interconexión al interior del territorio y los servicios logísticos que provee. Es por ello que se debe resolver temas de infraestructura, junto con la mejora de la conexión del puerto con su territorio, iniciando, en primer término con una adecuada asociación entre los participantes de la cadena logística, con el fin de aunar esfuerzos y disponer de un diagnóstico claro de los desafíos y problemas que se deben resolver conjuntamente y donde la introducción de nuevas tecnologías puede actuar como facilitador del proceso[54].

Un ciclo logístico se caracteriza por la interacción de diferentes elementos entre los que se encuentran: la gestión de la cadena suministro, la infraestructura de conexión de transporte y los flujos de información; aportando en el desempeño del intercambio de bienes y servicios de un área o país en particular. Es así, como una eficiente planificación de flujos, servicios e información, influyen de manera directa en los costos logísticos de distribución y posicionamiento competitivo de los bienes[41]. La figura 5.5 muestra que la deficiencia en la infraestructura tecnológica corresponde a la tercera de las 4 principales barreras que impactan la logística de los usuarios de servicios logísticos[44].

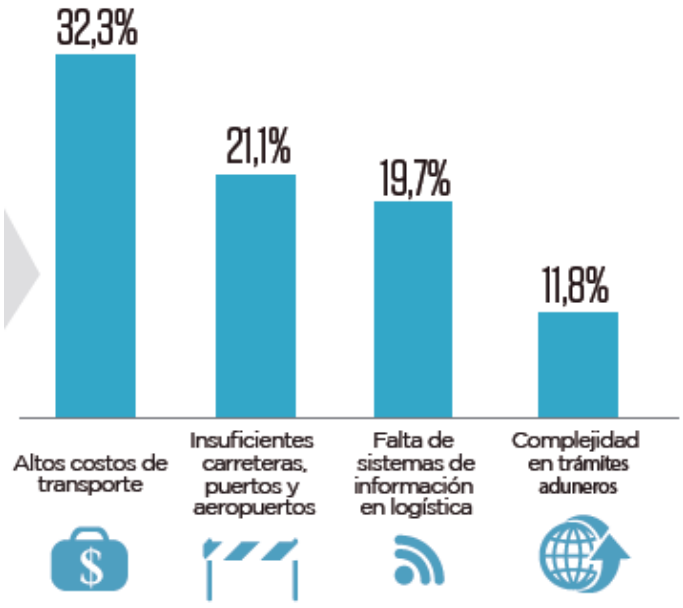


FIGURA 5.5. Barreras logísticas que impactan a los usuarios de servicios logísticos. Fuente: Encuesta Nacional de Logística 2015

Desde los años ochenta las inversiones en computadores y tecnologías de la información han cobrado gran valor dentro de las organizaciones. Se estima que en la década del 2000,

las inversiones dentro de las organizaciones han tenido un valor aproximado del 50 % de todo el capital invertido [143]. En grandes operadores logísticos mundiales, se considera que la estrategia tecnológica es la clave diferenciadora en el concierto mundial de competidores y muchas firmas de capacidad global o de altos grados de integración invierten hasta el 40 % de sus utilidades en sistemas de información y tecnologías[118].

Como parte del desarrollo de un Sistema Logístico Nacional, las tecnologías de la información y las comunicaciones juegan un papel fundamental, en la medida que permiten las gestión y la administración de cada una de las operaciones de las empresas y de la red de logística que estas involucran, facilitando la captura, la transferencia y la gestión de la información de forma adecuada. Generando un mejor y más eficiente proceso de administración y fundamentalmente, un aumento de la capacidad de los administradores para la toma de decisiones, gracias a la disponibilidad de la información y a la eliminación del error asociado a la interacción humana.

Temas como mejoramiento de los tiempos de decisión, número de variables de decisión, posibles opciones o estrategias y disponibilidad de información en tiempo real, son factores que influyen en la adopción de tecnologías de la información, las cuales mejoran la capacidad operacional de la empresa optimizándola y mejorando los procesos de comunicación en tiempo real, y disminuyendo la incertidumbre y complejidad de algunos problemas para los tomadores de decisiones.

Las tecnologías de la información y las comunicaciones juegan un papel fundamental en el desarrollo de las industrias y en particular en la mejora de los procesos logísticos de las compañías y el Sistema Logístico Nacional, ya que en todo momento permiten tener disponible la información de todos los procesos, facilitando la gestión de las operaciones. Sin embargo el uso de las tecnologías de la información asociadas a los procesos logísticos en las empresas Colombianas es bajo, como se muestra en la figura 5.6, las empresas Colombianas conocen las tecnologías de la información pero no las aplican [118] relegando esta inversión casi de manera exclusiva a las empresas grandes.

Tecnologías de la información	Disponibilidad
Optimización, planeación y control de transporte (TMS)	57.9%
Gestión de centros de distribución (WMS)	28.1%
Sistema de gestión de distribución (DMS)	28.1%
Gestión de transacciones comerciales/ pedidos (OMS)	28.1%
TMS WMS integrados	15.8%
Software de gestión y planeación de la demanda	24.6%
Interfaces ERP	33.3%
Software para la gestión de flotas	45.6%
Sistema de códigos de barras	24.6%
Sistema de radio frecuencia	15.8%
Sistemas para facturación/auditoria	49.1%
Sistema de rastreo y trazabilidad en tiempo real	73.7%
Acceso vía internet para el cliente	63.2%
Sistema de intercambio electrónico de datos (EDI)	24.6%
Sistema de optimización del picking	15.8%

FIGURA 5.6. Disponibilidad de tecnologías de información de los operadores logísticos. Fuente: Encuesta Nacional de Logística 2015

Las tecnologías de información más difundidas por los operadores logísticos corresponden a sistemas de trazabilidad en tiempo real, accesos vía internet para el cliente y sistemas de optimización, planeación y control de transporte[44]. Se evidencia una visible tendencia hacia el uso de sistemas de tipo operacional, haciendo de lado los sistemas analíticos o de apoyo a la toma de decisiones.

La adopción de sistemas de información como apoyo a la toma de decisiones en logística, se hace en etapas tardías y supeditadas a la realización de una primera etapa de adopción. Hay un proceso de evolución en cuanto al proceso de adopción de sistemas de información en el área de logística, el cual está compuesto por dos etapas: la primera etapa consiste en la solución de las necesidades de integración transaccional, que en la mayoría de los casos se realiza mediante sistemas ERP que incorporan módulos de logística. Esta etapa se caracteriza por que requiere fuertes inversiones en tiempo, dinero y mano de obra por parte de las compañías implementadoras. En una segunda etapa, que pocas empresas alcanzan, se buscan sistemas de información más sofisticados enfocados en la toma de decisiones, incorporando sofisticadas herramientas matemáticas que exigen la disponibilidad de datos históricos de calidad[118].

5.3. Institucionalidad

5.3.1. Ministerio de transporte de Colombia

De acuerdo la Constitución Política Nacional, la Ley 489 de 1998, la Ley 790 de 2002 y normas vigentes, los ministerios son, junto con la presidencia de la república y los departamentos administrativos, los organismos principales de la administración pública nacional y hacen parte del sector central de la rama ejecutiva del poder público. Los ministerios tienen como objetivos primordiales la formulación y adopción de las políticas, planes generales, programas y proyectos del sector administrativo que dirigen, ver anexo 3.

El Ministerio de Transporte de Colombia (MinTransporte), como lo establece el Decreto 087 de 2011, es el organismo del gobierno nacional encargado de formular y adoptar las políticas, planes, programas, proyectos y regulación económica del transporte, el tránsito y la infraestructura, en los modos carretero, marítimo, fluvial, férreo y aéreo del país. El MinTransporte es la cabeza del sector transporte, el cual está constituido por el Instituto Nacional de Vías (INVIAS), la Agencia Nacional de Infraestructuras (ANI), la Unidad Administrativa Especial de Aeronáutica Civil (AEROCIVIL) y la Superintendencia de Puertos y Transporte SUPERTRANSPORTE)[101].

5.3.2. Registro nacional de despacho de carga - RNDC

El Registro Nacional de Despacho de Carga (RNDC) es un sistema de información del Ministerio de Transporte que permite recibir, validar y transmitir la información generada en las operaciones del Servicio Público de Transporte de Carga por Carretera.

Tiene como finalidad optimizar el flujo de información acerca de la operación de transporte de carga. Dicha información servirá como base para el monitoreo de las relaciones económicas por parte de los integrantes del sector de transporte de carga, a través del modelo para la Regulación del Transporte de Carga por Carretera SIRTCC, así como

para el control por parte de la autoridad competente, garantizando la seguridad en la prestación del servicio público de transporte de carga, a cargo de aquellos particulares que se encuentran legalmente constituidos y debidamente habilitados por el Ministerio de Transporte.[45]

Según la resolución 377 del 15 de febrero de 2013, expedida por el Ministerio de Transporte, a partir del 15 de marzo de mismo año se hace obligatorio para todas las empresas de transporte de carga por carretera, la utilización del sistema RNDC para el registro del manifiesto electrónico de carga y otra información relacionada con la operación de transporte terrestre automotor de carga. Ver anexo 2.

5.3.3. Normatividad

El uso de las tecnologías de la información ha tomado tanta importancia dentro del Gobierno Nacional, que ha formulado una serie de políticas para incentivar su uso a todo nivel, incluido el logístico. La aplicación de estas políticas es un insumo importante para la formulación de políticas en el Sistema Logístico Nacional. Algunas de las políticas más importantes se mencionan a continuación:

Plan Nacional de Desarrollo 2010-2014(PND)[42] : plantea la puesta en marcha del *Observatorio Nacional de Logística de Cargas* (ONLC), que busca consolidar, administrar y procesar la información del sector, reflejando la evolución de la logística del país. El observatorio facilitará el diseño de políticas y la priorización de acciones para la eficiencia de las cadenas de abastecimiento, permitiendo hacer un seguimiento efectivo, determinar la necesidad de reorientarlas o reformularlas. Se espera establecer el *nivel de penetración de TIC y SIT* y definir las de acciones requeridas para su uso, masificación y aprovechamiento, con el fin de mejorar la operatividad, articular los sistemas de información y telecomunicaciones a procesos e infraestructuras logísticas.

Plantea la realización de un análisis de las *ventajas y beneficios* que el *uso de TIC* ofrece sobre el *control y monitoreo* de los bienes con el objetivo de promover acciones de facilitación comercial que se analicen e implementen en los nodos de comercio exterior, velando así por la seguridad del movimiento de carga local y extranjera, incrementando y asegurando un nivel de confiabilidad en la cadena de abastecimiento. Se propone el desarrollo de un plan maestro para *Sistemas Inteligentes de Transporte* (SIT) por parte del MinTransporte para el apoyo de los servicios de transporte y logística en el país. El sistema es definido como un conjunto de soluciones tecnológicas de las telecomunicaciones y la informática, diseñadas para hacer más eficiente, seguro, cómodo y sostenible el tránsito y la movilidad en general.

Teniendo en cuenta la eficiencia que incorporan las TIC en los trámites y el acceso a la información al ciudadano, el PND establece que el Gobierno Nacional ofrecerá la totalidad de los *trámites para el sector transportador en línea*, brindará una política de uso de peajes electrónicos, y ofrecerá la información al ciudadano sobre el estado de las carreteras. Así mismo, y en el marco de compartición de infraestructura sectorial, se desarrollará la priorización del desarrollo conjunto de *carreteras y redes de fibra óptica.* Finalmente propone el diseño e implementación del *Registro Nacional de Despacho de Carga* (RNDC) por parte del Ministerio de Transporte, como parte del sistema RUNT y como mejora del manifiesto de carga electrónico.

CONPES 3547 Política Nacional Logística [41]: define el *Sistema Logístico Nacional* como la sinergia de todos y cada uno de los involucrados en la adquisición, el movimiento, el almacenamiento de mercancías y el control de las mismas, así como todo el *flujo de información* asociado a través de los cuales se logra encauzar rentabilidad presente y futura en términos de costos y efectividad en el uso, prestación y facilitación de servicios logísticos y de transporte. Dentro de sus objetivos específicos plantea numerales relacionados con la promoción de temas tecnológicos como: el incentivo de la cultura de la información y promoción el uso de las TIC al servicio de la logística.

Como parte de su plan de acción se proponen actividades como: implementar tecnología para facilitar el comercio exterior, los procesos de control e inspección de la mercancía y su desaduanaje e integración de las TIC en la logística mediante la integración de sistemas de información de almacenamiento, seguimiento y posicionamiento, sistemas de información web y sistemas conducentes.

CONPES 3527 Política Nacional de Competitividad y Productividad [40]: menciona, dentro de la matriz de productos y actividades para la infraestructura de transporte y logística, el desarrollo y puesta en marcha del *Sistema de información de Infraestructura, Logística y Transporte*, el cual deberá estar articulado con los demás sistemas utilizados por otras entidades.

CONPES 3489 Política Nacional de Transporte Público Automotor de Carga [39]: el documento planteado por el departamento nacional de planeación presenta un análisis de la situación actual, el deber ser, y estrategias a abordar para el transporte de carga terrestre en Colombia. Está orientado a fomentar la competitividad de los productos colombianos en los mercados nacionales e internacionales mediante un sistema conformado por la infraestructura y el servicio.

Se define la cadena productiva del transporte de carga como la constituida por socios logísticos y comerciales, en condiciones y relaciones económicas de tal manera que todos los eslabones de la cadena reciban un beneficio que les garantice su desarrollo económico y social, aplicando los principios de alianza estratégica, colaboración y mentalidad empresarial. El documento establece las acciones que se deben ejecutar y los plazos de los responsables a fin de garantizar el cumplimiento de los principios de política establecidos. Las acciones propuestas por el documento CONPES son:

- El diseño y publicación del Índice de Precios del Transporte - IPT.
- Actualizar y/o modificar la normatividad sobre regulación del transporte público de carga, acorde a las políticas y estrategias señaladas en este documento.
- *Implementar un sistema de información para el monitoreo y regulación económica del transporte de carga por carretera.*
- Diseñar y establecer una metodología dinámica de análisis de oferta actual y futura del parque automotor y de la demanda de servicios de transporte.
- Actualizar la normativa de las especificaciones técnicas y de seguridad de los vehículos de transporte de carga.

- Diseño e implementación de un programa integral de reposición del parque automotor de carga. Así como la evaluación de la pertinencia de canalizar los recursos provenientes de la garantía bancaria o póliza de seguro establecida en el decreto 3525 de 2005, a través del Fondo Nacional de Garantías.
- Solicitud de diferimientos arancelarios encaminada a la reducción arancelaria de equipos de transporte destinados a la reposición del parque automotor.
- Evaluación de la situación actual y determinación de un plan de acción para el fortalecimiento de la institucionalidad del sector en los temas de inspección, vigilancia y control.
- Desarrollo de un estudio para evaluar el modelo empresarial del servicio público de transporte terrestre de carga y hacer las recomendaciones respectivas para su optimización.
- Desarrollar un programa de capacitación en procedimientos administrativos y operativos para los integrantes de la cadena de transporte.
- Determinación de estrategias para lograr el efectivo reconocimiento y cumplimiento de los derechos laborales y de seguridad social integral de los conductores, derivados de los contratos laborales conforme a la ley vigente en el territorio nacional y fuera de él en cumplimiento de la actividad del conductor.
- Desarrollo de acuerdos binacionales en materia de transporte y tránsito, con el fin de mejorar las actuales condiciones de operación de transporte internacional, interfronterizo o transfronterizo.
- Estudio relacionado con el mejoramiento de las condiciones de aseguramiento del sector.

5.4. Formación académica en servicios logísticos

5.4.1. Formación en logística

Colombia cuenta con 186 programas activos relacionados con logística, de los cuales en su mayoría son ofertados por instituciones privadas, principalmente universidades, ver tabla 5.1.

	Sector		
Carácter Académico	Oficial	Privada	Total
Universidad	26	68	94
Institución universitaria/escuela tecnológica	15	42	57
Institución tecnológica	10	17	27
Institución técnica profesional	3	5	8
Total general	54	132	186

TABLA 5.1. Programas académicos asociados al campo de la logística. Fuente: SNIES Ministerio de Educación

La formación se lleva a cabo principalmente en Bogotá y en departamentos como Antioquia, Atlántico, Valle del Cauca, Bolívar y Santander, prevaleciendo nuevamente las instituciones de carácter privado, ver tabla 5.2.

	Sector		
Departamento de oferta	Oficial	Privada	Total
Bogotá D.C	13	32	45
Antioquia	9	20	29
Atlántico	7	13	20
Valle del cauca	9	11	20
Bolívar	3	15	18
Santander	3	15	18
Cundinamarca	1	5	6
Magdalena	1	4	5
Norte de Santander	1	4	5
Risaralda	1	4	5
Tolima	2	2	4
Caldas	2	1	3
San Andrés y Providencia	2		2
Casanare		1	1
Cesar		1	1
Huila		1	1
Meta		1	1
Nariño		1	1
Quindío		1	1
Total general	54	132	186

TABLA 5.2. Distribución de oferta de programas de logística por departamentos. Fuente: SNIES Ministerio de Educación

En cuanto al nivel de formación prevalecen principalmente programas de formación técnicos, tecnológicos y de especialización en su gran mayoría con registro calificado, ver tabla 5.3.

	Reconocimiento del Ministerio			
Nivel de Formación	No aplica	Reg. Alta calidad	Reg. Calificado	Total
Tecnológica	2		65	67
Especialización	1		63	64
Formación técnica profesional	2		28	30
Universitaria		1	14	15
Maestría	1		8	9
Doctorado			1	1
Total general	6	1	179	186

TABLA 5.3. Niveles de formación en logística. Fuente: SNIES Ministerio de Educación

5.4.2. Formación en transporte

La formación académica en temas de transporte cuenta con 40 programas de transporte a nivel nacional los cuales se dividen de forma equitativa entre instituciones privadas y públicas, predominando principalmente las Universidades como establecimiento de formación, ver tabla 5.4.

La mayoría de los programas son ofertados en Bogotá y los departamentos de Antioquia, Santander y Valle del Cauca. Ver tabla 5.5.

	Sector		
Carácter Académico	Oficial	Privada	Total
Universidad	14	17	31
Institución universitaria/escuela tecnológica	4	3	7
Institución técnica profesional	1		1
Institución tecnológica	1		1
Total general	20	20	40

TABLA 5.4. Programas académicos asociados al campo del transporte. Fuente: SNIES Ministerio de Educación

	Sector		
Departamento de oferta	Oficial	Privada	Total
Bogotá D.C	5	7	12
Antioquia	3	3	6
Santander	3	1	4
Valle del cauca	4		4
Bolívar		3	3
Atlántico		2	2
Boyacá	2		2
Caldas	2		2
Córdoba	1	1	2
Magdalena		1	1
Meta		1	1
Risaralda		1	1
Total general	20	20	40

TABLA 5.5. Distribución de oferta de programas de transporte por departamentos. Fuente: SNIES Ministerio de Educación

La mayoría de los programas de transporte cuenta con registro calificado reconocido por el Ministerio de Educación y corresponden principalmente a especializaciones, seguidos por programas tecnológicos y técnicos, ver tabla 5.6.

	Reconocimiento del Ministerio			
Nivel de Formación	Reg. Alta Calidad	Reg. Calificado	No aplica	Total
Especialización		14	3	17
Tecnológica		11	1	12
Formación técnica profesional		6	1	7
Maestría		1	2	3
Universitaria	1			1
Total general	1	32	7	40

TABLA 5.6. Niveles de formación en transporte. Fuente: SNIES Ministerio de Educación

5.4.3. Formación del personal en logística y transporte

	Bachillerato	Técnico	Tecnólogo	Universitario	Especialista	Máster	Doctorado	TOTAL
a) Procesamiento de pedidos de clientes	42,7%	22,0%	19,8%	11,8%	3,0%	0,6%	0,1%	100,0%
b) Planeación y reposición de inventarios	30,7%	25,2%	23,9%	11,9%	6,9%	1,2%	0,3%	100,0%
c) Compras y manejo de proveedores	16,6%	29,5%	31,0%	12,6%	8,3%	1,9%	0,2%	100,0%
d) Almacenamiento	63,3%	16,6%	16,5%	2,3%	1,0%	0,2%	0,1%	100,0%
e) Transporte y distribución	64,3%	18,3%	13,7%	3,0%	0,2%	0,4%	0,1%	100,0%
f) Logística de reversa	64,8%	5,7%	10,2%	11,9%	5,0%	1,2%	1,2%	100,0%
g) Comercio exterior	19,2%	21,4%	26,5%	23,3%	6,7%	2,2%	0,8%	100,0%

FIGURA 5.7. Composición del personal en logística según procesos y niveles de escolaridad. Fuente: Encuesta Nacional Logística 2015

Se evidencia una importante brecha en términos formación debido a que aproximadamente 80 % del personal asociado a actividades relacionadas con la logística tienen un nivel de formación entre bachillerato, técnico y tecnólogo. El 20 % restante se divide entre formación universitaria, especialistas, master y doctorados, ver figura 5.7. En lo relacionado con transporte el tema es aún más crítico ya que 96.3 % del personal corresponde a un nivel de formación entre bachillerato, técnico y tecnólogo.

Se cierra este capítulo habiendo presentado de forma general cómo es la distribución de la población colombiana asociado a los principales centros poblacionales, los problemas de infraestructura que enfrenta el país y la importancia de los corredores logísticos como elementos estratégicos para el transporte en Colombia. De manera adicional se presentó una revisión del contexto institucional del transporte y los procesos de formación en servicios logísticos en Colombia.

CAPÍTULO 6

Comprensión de los datos

6.1. Levantamiento de los datos

El estudio se realizó sobre una copia de los registros de la base de datos del RNDC facilitada por el MinTransporte que incluye datos comprendidos en el periodo de febrero de 2013 y diciembre del año 2013. Los registros evidencian la masificación progresiva en el uso del sistema de información RNDC, pasando de una baja cantidad de registros en el mes marzo, a un número elevado en el mes de diciembre.

6.2. Descripción de los datos

Los datos se recolectan diariamente y contienen información relacionada con los vehículos, el viaje, origen y destino de la remesa, de actores que intervienen en la operación, fechas y horas de cita para cargues y descargues, tiempos pactados y cumplidos.

Para el análisis descriptivo se analizaron 180.000 observaciones aproximadamente. De las 35 variables que conforman el dataset, la mayoría de tipo categórico y nominal, se seleccionaron 22 consideradas relevantes para la investigación. A continuación se realizará una descripción de la composición de las variables consideradas pertinentes para el estudio:

Variable	Descripción	Valores	Tipo
Naturaleza carga	Define la naturaleza de la carga según la clasificación y características de la misma.	Carga normal, carga peligrosa, carga extradimensionada, carga extrapesada, desechos peligrosos, semovientes, refrigerada.	Nominal
Descripción corta producto	Campo abierto designado por la empresa de transporte para el nombre corto del producto.	Campo abierto.	Nominal
Cantidad cargada	Cantidad cargada del producto a transportar.	Enteros positivos y 0	Numérica
Origen de la remesa	Origen de la remesa	Códificación [Ciudad Departamento]	Nominal

Variable	Descripción	Valores	Tipo
Destino de la remesa	Destino de la Remesa	Códificación [Ciudad Departamento]	Nominal
Horas pacto carga	Horas totales pactadas para el cargue. Incluye horas de espera, horas de cargue y documentación.	Numéricos iguales o mayores que 0.	Numérica
Minutos pacto carga	Minutos del total de tiempo pactado para el cargue. Incluye minutos de espera, minutos de cargue y documentación.	Numéricos entre 0 y 59.	Numérica
Fecha llegada cargue	Fecha de llegada al cargue.	Fechas con el formato dd/mm/yyyy.	Fecha
Hora llegada cargue remesa	Hora de llegada al cargue.	Horas con formato militar hh:mm	Tiempo
Fecha entrada cargue	Fecha de entrada al cargue.	Fechas con el formato dd/mm/yyyy.	Fecha
Hora entrada cargue remesa	Hora entrada del cargue	Horas con formato militar hh:mm	Tiempo
Fecha salida cargue	Fecha de salida del cargue.	Fechas con el formato dd/mm/yyyy.	Fecha
Hora salida cargue remesa	Hora de salida del cargue. Corresponde a la diferencia en tiempo de la fecha y hora de entrada con la fecha y hora de salida	Horas con formato militar hh:mm	Tiempo
Rem hora real carga	Horas reales de carga	Numéricos enteros positivos	Tiempo
Rem minutos real cargue	Minutos reales de carga	Numéricos enteros positivos	Tiempo
Fecha capacitada cargue	Fecha pactada para el cargue	Fechas con el formato dd/mm/yyyy.	Fecha
Hora cita pactada cargue	Hora pactada para la cita del cargue	Formato de horas militar hh:mm	Tiempo
Fecha cita pactada descargue	Fecha pactada para el descargue	Fechas con el formato dd/mm/yyyy.	Fecha
Hora cita pactada descargue remesa	Hora pactada para la cita del descargue	Horas con formato militar hh:mm	Tiempo
Código tipo empaque	Unidad de empaque de la carga a transportar.	Paquetes, bulto, granel liquido, Contenedor de 20 pies, dos contenedores de 20 pies, contenedor de 40 pies, cilindros, granel solido, varios, no aplica, carga estibada	Nominal
Unidad medida capacidad	Unidad de medida de la carga registrada. Para contenedor vacío, este campo no es reportado.	Kilogramos, galones, vacio.	Nominal

Tabla 6.1. Tabla descriptiva de atributos seleccionados del RNDC

Para mayor información acerca de la naturaleza de los datos ver el Anexo 2.

6.3. Exploración de los datos

Se seleccionaron como datos exploratorios los registros correspondientes al mes de diciembre, en principio debido a que son los más recientes y reflejan un comportamiento actual, segundo porque de acuerdo a la evolución experimentada por el sistema de información, deben ser los más estables en cuento al comportamiento de la información. Sobre los datos se realizaron tareas exploratorias, verificación del comportamiento y la calidad de los datos. A continuación se describirá el comportamiento en los atributos seleccionados.

6.3.1. Naturaleza de la carga

Atributo de tipo nominal, se encuentra dividido en 7 categorías: carga normal, carga peligrosa, refrigerada, desechos peligrosos, carga extradimensionada, carga extrapesada, semovientes. La mayoría de los datos se encuentran agrupados en la categoría de carga normal, como se evidencia en la tabla 6.2, siendo estos aproximadamente el 92 % de los registros, seguido por la categoría carga peligrosa con un 7 % de la carga.

El hecho de que la mayoría de los usuarios hayan registrado sus cargas como carga normal, no refleja la realidad de manera cercana, ya que en un análisis comparativo realizado con el atributo de descripción corta del producto se evidenció que habían productos que fueron mal clasificados por los usuarios, por ejemplo: transporte de ganado en varios casos lo clasifican como carga normal, en vez de semovientes. Esta incorrecta clasificación por parte de los usuarios se podría deber a la falta de conocimiento de cuál debería ser la categoría adecuada para su tipo de carga transportada.

Debido a que la mayoría de los datos son ubicados como carga normal, se evidencia un sesgo en la información a la hora de aplicar algún modelo de minería de datos, debido que siempre será una categoría opcionada a estar en todos los modelos y no refleja el comportamiento real del transporte de carga.

Categoría	Frecuencia	Porcentaje %
Carga Normal	153445	0,918
Carga Peligrosa	12414	0,074
Refrigerada	1006	0,006
Desechos Peligrosos	143	0,001
Carga Extradimensionada	138	0,001
Carga Extrapesada	43	0,000
Semovientes	1	0,000

TABLA 6.2. Distribución de datos dentro del atributo naturaleza de la carga

6.3.2. Descripción corta del producto

Es un campo digitado por el usuario, lo que implica un alto nivel de dispersión en su comportamiento, ya que se ve afectado por factores como uso de mayúsculas, tildes, número de caracteres, etc. En concordancia con lo mencionado en el análisis realizado al atributo se detectaron 9.201 categorías diferentes para 171.954 registros aproximadamente.

En la figura 6.1 se observa que los productos más transportados en principio corresponden a petróleo y sus diferentes derivados y materiales para construcción, especialmente cemento en sus diferentes variedades, productos alimenticios como pollo y bebidas de diferentes tipos.

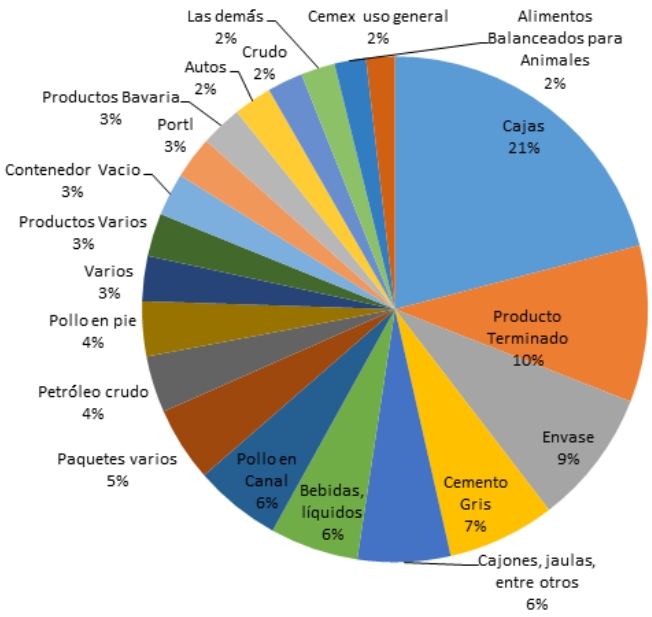


FIGURA 6.1. Distribución de categorías para el atributo descripción corta del producto

Debido a la dispersión de los datos, para lograr una verdadera utilidad del campo es necesario realizar una agregación utilizando diccionarios de datos para agrupar los registros de acuerdo a sectores y subsectores productivos. Con este ajuste a los datos el número de categorías será más manejable y se podrán establecer tendencias del tipo de carga transportada.

6.3.3. Unidad medida de capacidad

Es el campo que muestra la unidad de medida de la cantidad de carga, esta agrupada en dos categorías: Kilogramos (1) y Galones (2). De acuerdo a la figura 6.2 se observa que los productos más transportados en Colombia corresponden a cargamentos sólidos (kilogramos) siendo aproximadamente el 93 % de la carga, el 7 % restante corresponde a cargamento líquido (galones).

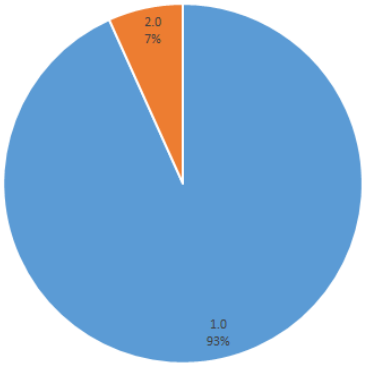


FIGURA 6.2. Unidad de medida capacidad

6.3.4. Origen de la remesa

Campo que relaciona la ciudad y departamento de origen de la carga. Este atributo tiene un número de clases muy alto debido a que Colombia cuenta con aproximadamente 1.000 municipios.

De acuerdo con la figura 6.3 se aprecia que Bogotá D.C. es la ciudad que más genera carga, seguido por Buenaventura, Barranquilla y Cartagena principales puertos marítimos del país. En quinto lugar se ubica la ciudad de Medellín.

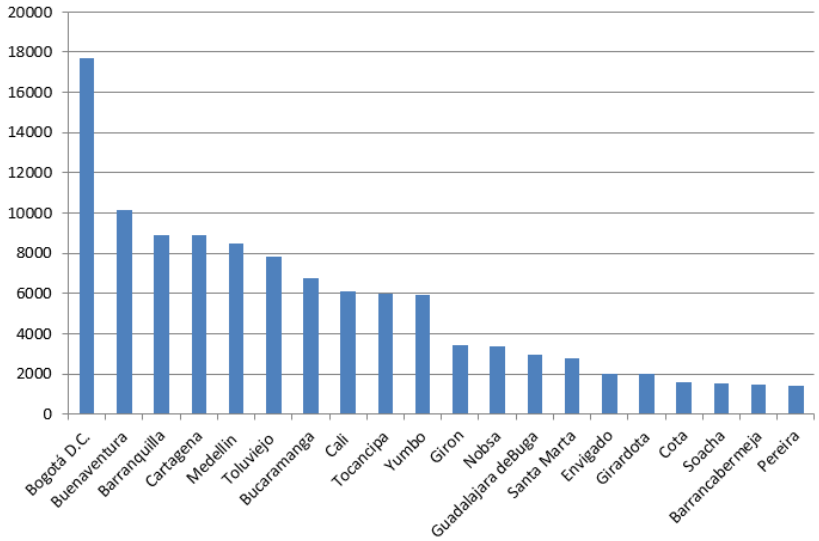


FIGURA 6.3. Orígenes de la remesa

6.3.5. Destino de la remesa

Campo que relaciona los destinos de la carga, al igual que el atributo origen de la carga tiene aproximadamente 1.000 categorías presentes. De acuerdo con la figura 6.4 se observa que el principal destino es Bogotá D.C., seguido de otras importantes capitales como lo son: Medellín, Cali, Barranquilla, Cartagena y Bucaramanga.

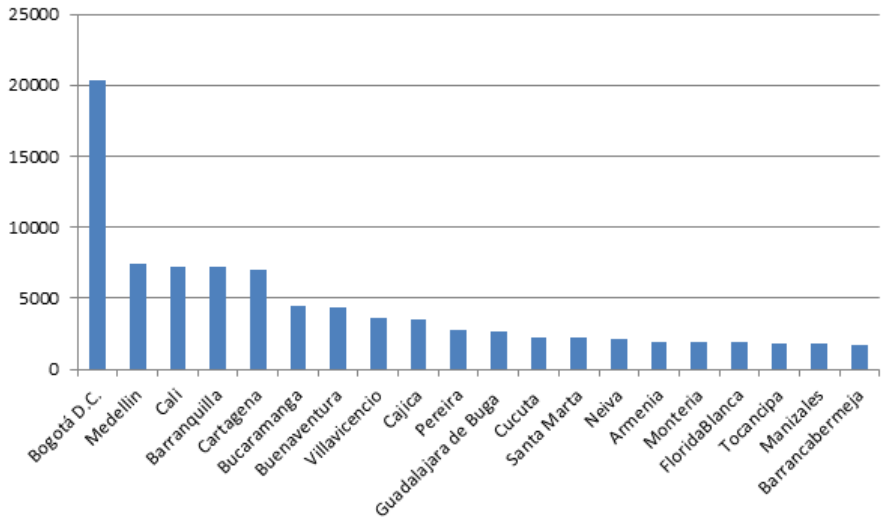


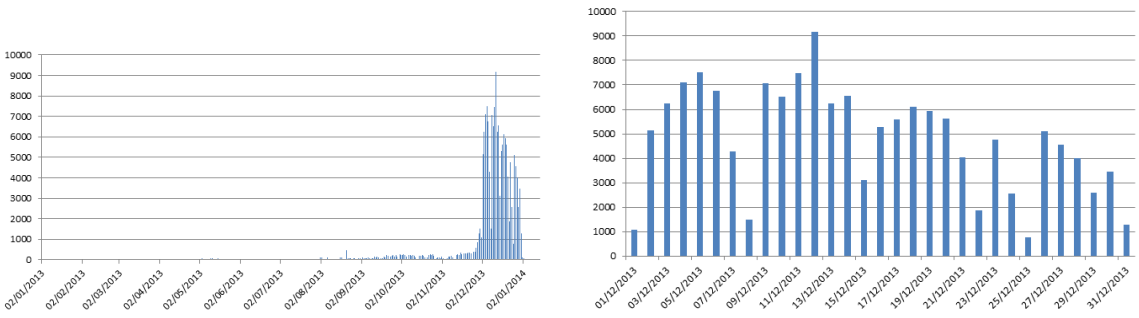
FIGURA 6.4. Destinos de la remesa

Comparando los origenes y destinos de la carga, se observa que la cantidad de viajes que entran a Buenaventura versus la cantidad que sale, muestra un énfasis en Buenaventura como puerto importador.

6.3.6. Fecha cita pactada cargue

En este atributo se observa que la mayoría de las fechas pactadas de cargue están presentes en el mes de diciembre o meses contiguos, sin embargo, como situación especial, se observan registros con fechas de pacto de cargue muy distantes al mes analizado, como por ejemplo registros con fechas en los meses de mayo, junio y julio, ver figura 6.5a.

También se observa que hay registros con datos correspondientes al mes de enero de 2014, lo cual indica que algunas empresas programan el proceso de cargue de mercancía con suficiente antelación.



(A) Completo (B) Diciembre

FIGURA 6.5. Fecha cita pactada cargue

La ventana de tiempo del mes de diciembre, ver figura 6.5b, muestra un comportamiento cíclico decreciente de forma semanal, en donde se aumenta de forma progresiva el número de citas pactadas desde el lunes llegando a su pico en la mitad de la semana, y

disminuye progresivamente el fin de semana. De igual manera, se nota el decrecimiento de la cantidad de registros desde mediados hasta fin de mes.

Durante la exploración de los datos se encontró también que varios registros presentaban datos faltantes, fechas incongruentes, por ejemplo 01/12/1989, la fecha pactada cargue en muchos casos era posterior a la fecha pactada descargue entre otros.

6.3.7. Horas pactadas cargue

En la figura 6.6 se observa que los tiempos de cargue pactados en su mayoría corresponden a 12 horas, sin embargo, también se presenta una alta afluencia de ocurrencias entre 0 y 2 horas. En un porcentaje menor se encuentran los tiempos de carga con más de 12 horas de duración.

Se encontró que había una gran cantidad de datos faltantes para este campo. Además la existencia de tiempos pactados con valores muy altos, por ejemplo 700 horas y que corresponde probablemente a un dato atípico.

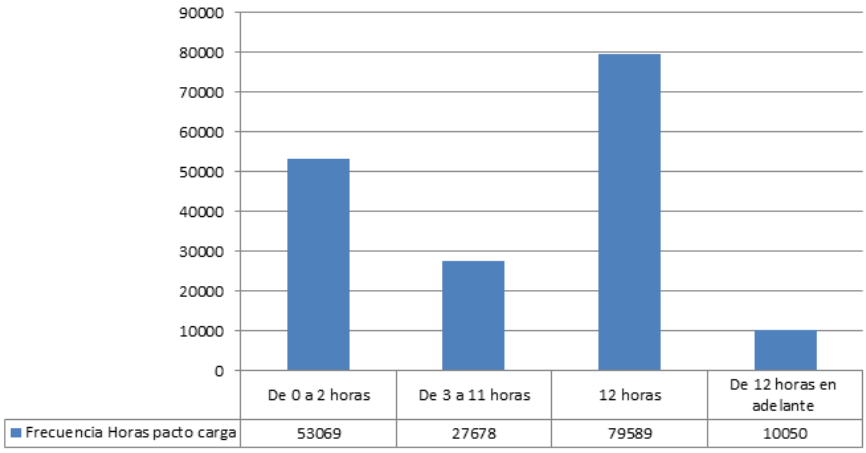


FIGURA 6.6. Horas pactadas para el cargue

6.3.8. Horas reales de cargue de la remesa

Este campo se construye sumando el campo de horas. Como se observa en la figura 6.7, la mayoría de tiempos de carga se distribuyen en su mayoría entre 0 y 2 horas. En una proporción menor se distribuyen entre 3 y más de 5 horas.

Es un campo calculado por el RNDC, no obstante, se encontraron inconsistencias en algunos datos donde el resultado final correspondía a valores negativos, los cuales son observaciones inconsistentes. También se observó la presencia de datos faltantes, que hasta cierto punto es tolerable teniendo en cuenta que pueden haber procesos de cargue que duren únicamente días o minutos.

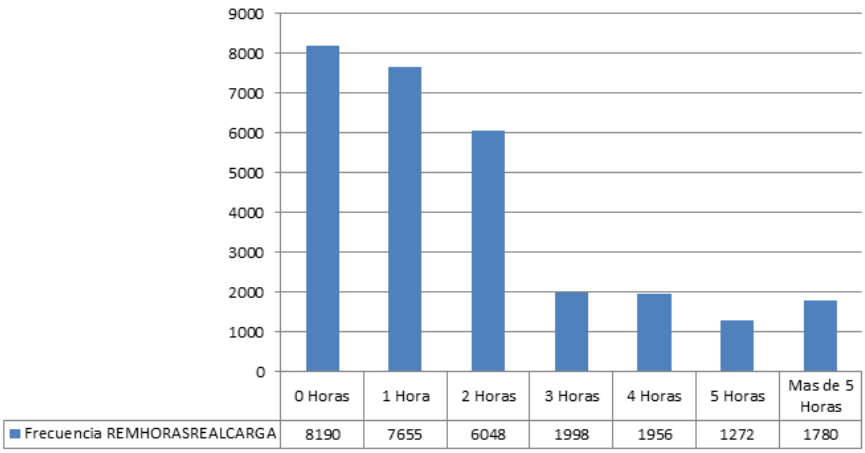


FIGURA 6.7. Horas reales de cargue de la remesa

6.3.9. Minutos reales de cargue de la remesa

En la figura 6.8 se observa que la mayoría de los tiempos se distribuyen en el intervalo entre 1 y 30 minutos, siendo prevalente los tiempos en minutos correspondientes a 30 minutos. En el campo se encontraron inconsistencias en una pequeña proporción de datos en donde el valor correspondía a valores negativos. También se observó la presencia de datos faltantes.

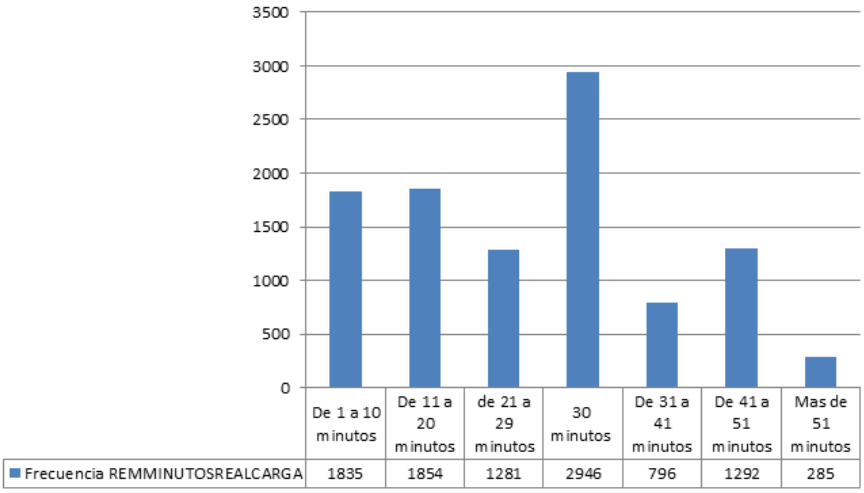
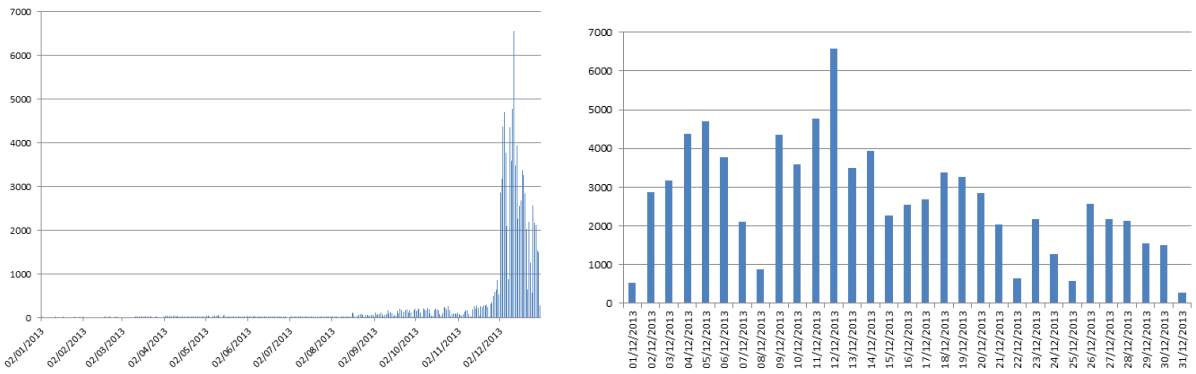


FIGURA 6.8. Minutos reales de cargue de la remesa

6.3.10. Fecha de entrada a cargue

En la figura 6.9a se observa que la mayoría de los campos se ubican en el mes de diciembre y meses cercanos como octubre y noviembre. Sin embargo, se observan datos con fechas de cargue con varios meses de antelación, inclusive desde febrero de 2013, lo cual pone en duda la calidad de estos datos.



(A) Completo (B) Diciembre

FIGURA 6.9. Fecha de entrada a cargue

En los datos correspondientes al mes de diciembre, ver figura 6.9b, se puede notar un patrón de tipo senoidal semana a semana, la cual tiene gran cantidad de movilización de carga a principio e intermedio de la semana y hacia el final disminuye de manera considerable. Se observa una clara disminución entre la cantidad de carga que entra a cargue a principio, respecto a la de fin de mes.

Respecto al campo se encontró la presencia de datos faltantes y de fechas de entrada posteriores a las fechas de entrada a descargue, además de valores incongruentes como 1899/12/30.

6.3.11. Hora de llegada al cargue de la remesa

La figura 6.10 muestra que las horas más frecuentes de llegada al cargue están entre 9 am y 12 m, de hecho se podría decir que la distribución de los datos tiende a ser simétrica con respecto al intervalo de 9:00 a.m. a 12 m. También se identifica que existe una baja tendencia en entrar a cargue durante las primeras horas del día y las altas horas de la noche, es decir entre 9:00 p.m. y las 3:00 a.m.

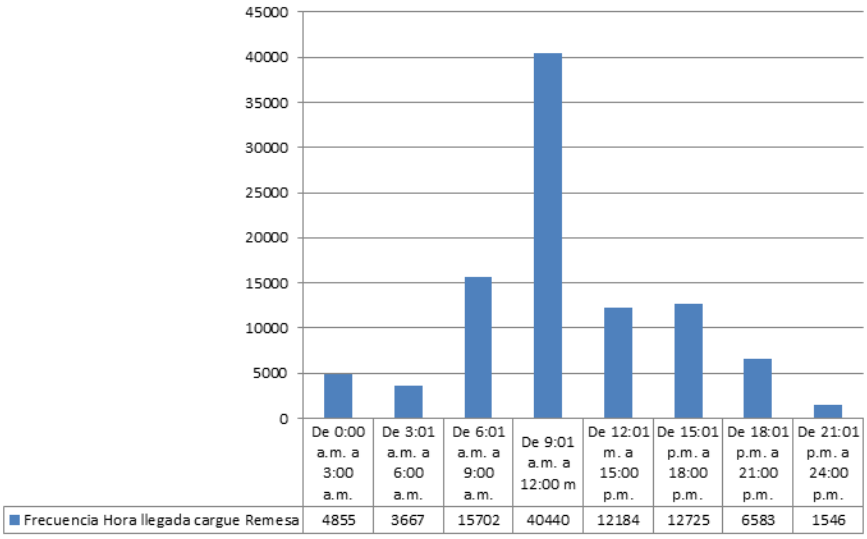


FIGURA 6.10. Hora de llegada al cargue de la remesa

6.3.12. Minutos pactados para el cargue

En la figura 6.11 se observa que la frecuencia se encuentra distribuida de forma uniforme, a excepción de la frecuencia para 67 minutos que es la más alta. Una posible explicación es que la mayoría de los minutos pactos de carga tienden a valores mayores a 67.

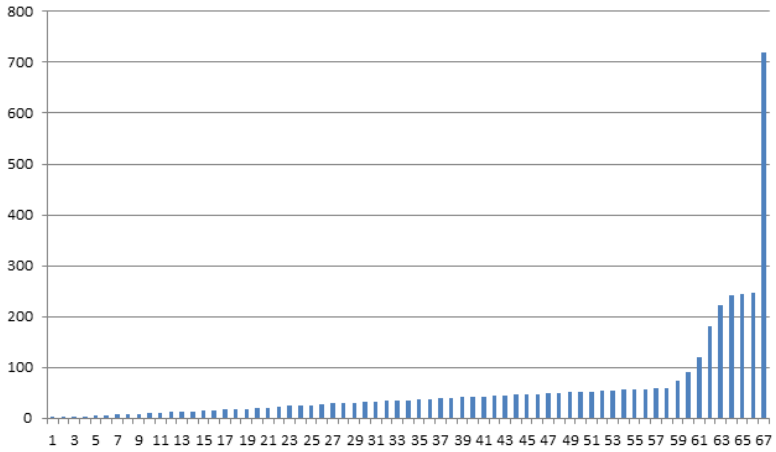


FIGURA 6.11. Minutos pactados para el cargue

6.3.13. Cantidad cargada

El comportamiento de la cantidad cargada se observa en la figura 6.12, donde su distribución es asimétrica hacia la derecha. La cantidad mas transportada de carga oscila entre 0 y 11.000 unidades, ya sean kilogramos o galones. Se identificaron valores con cantidades de carga negativos, valores en 0 para camiones cargados, valores faltantes y presencia de datos atípicos.

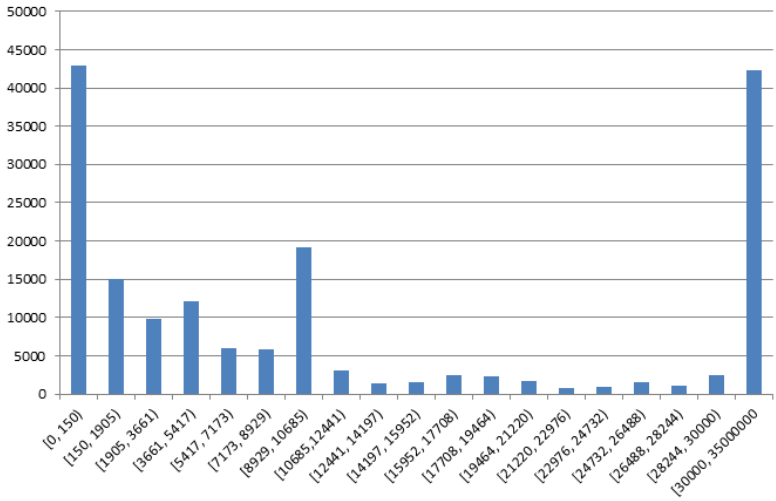
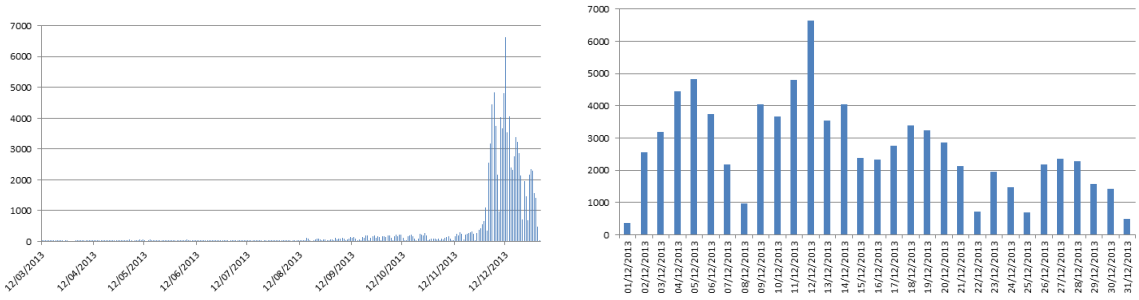


FIGURA 6.12. Cantidad cargada

6.3.14. Fecha salida cargue

En la figura 6.13a se puede apreciar que la mayoría de los datos están agrupados en el mes de diciembre, sin embargo hay varios registros con meses anteriores que llevan a pensar en la existencia de retrasos en la entrega de carga o en transporte de carga que recorre grandes distancias. La existencia de fechas de salida cargue muy antiguas lleva a pensar en la existencia de valores atípicos, o de valores corruptos o mal diligenciados dentro de la base de datos.



(A) Completo (B) Diciembre

FIGURA 6.13. Fecha de salida de cargue

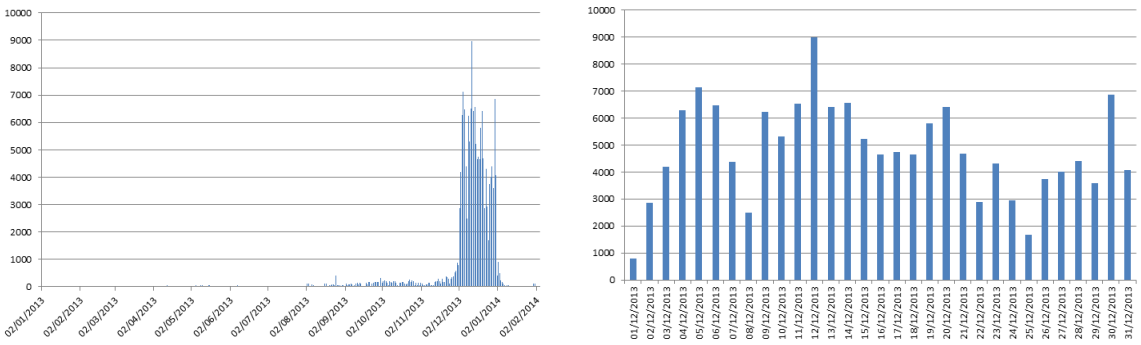
En la figura 6.13b se presentan los datos correspondientes al mes de diciembre, en donde se observa la existencia de un posible ciclo que se repite cada semana, aumentando de forma progresiva desde el lunes llegando a su pico en la mitad de la semana, y disminuyendo nuevamente hasta alcanzar el fin de semana. Se evidencia también que el día pico donde más se dio salida de mercancía corresponde al día 12 de diciembre y además que en las dos primeras semanas de diciembre es donde hay más movimiento y en las semanas siguientes tiende a disminuir progresivamente hasta final de mes.

En cuanto a inconvenientes encontrados en el campo, se presentaron fechas con valores incongruentes por ejemplo 1989/12/30 y valores faltantes. También, las fechas de salida cargue en muchos casos son posteriores a la fecha de entrada a descargue.

6.3.15. Fecha cita pactada descargue

Los datos se agrupan en su mayoría en el mes de diciembre ver figura 6.14a. Para los casos en donde la fecha de descargue es anterior al mes de diciembre, se puede intuir un posible retardo en el cargue o en la entrega de la carga o valores incongruentes o mal diligenciados. Por otro lado, las fechas posteriores al mes de diciembre se cuentan como datos válidos, ya que se puede pactar una fecha de descargue para meses posteriores al de la realización del envió.

En la figura 6.14b se observa la ventana de tiempo de diciembre en donde este atributo tiene un comportamiento similar al de las fechas pactadas para el cargue, sin embargo se observa un claro agrupamiento de los datos hacia las tres primeras semanas del mes y un claro descenso del movimiento de carga en días domingos y festivos. El día que más presento movimiento de descargue corresponde al día 12 de diciembre, el que menos registra, es el correspondiente al 25 de diciembre.



(A) Completo (B) Diciembre

FIGURA 6.14. Fecha pactada para el descargue

En cuanto a la calidad de los datos, se encontró la presencia de datos faltantes y de fechas de salida de descargue anteriores a las fechas de entrada a cargue. Además valores incongruentes como 1899/12/30.

6.3.16. Cumplimiento en llegadas a cargue

Variable construida a partir de los datos presentes del RNDC y corresponde al cumplimiento de la empresa transportadora en torno a la cita pactada para la realización del cargue de la remesa. Para la construcción de esta variable se tuvo en cuenta la diferencia entre los atributos tiempo de cargue pactado y tiempo de llegada a cargue.

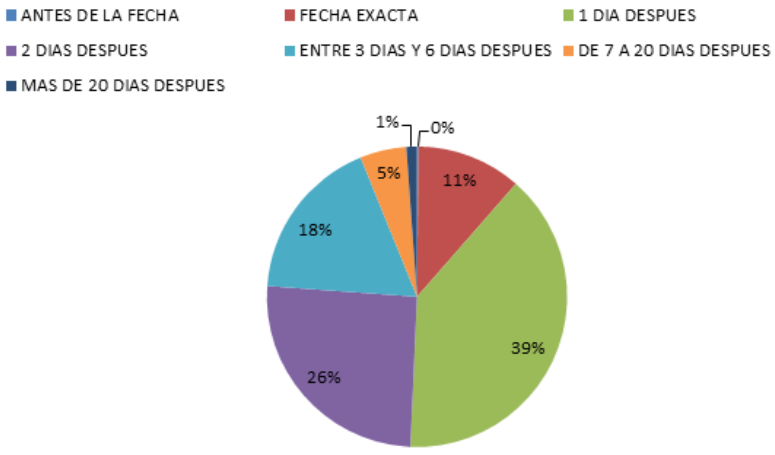


FIGURA 6.15. Cumplimiento en la llegada a cargue

En la figura 6.15 se puede advertir que en la mayoría de las ocasiones, con un 65 % de las observaciones, el transportador llega a realizar el cargue uno a dos días después del pactado, lo cual genera un impacto directo en la posterior entrega de la mercancía. También se evidencia que la frecuencia para las fechas de llegadas a tiempo es muy baja, solamente con un 11 %. De acuerdo a todo lo anterior es claro que el cumplimiento de

las empresas transportadoras es muy bajo, generando retrasos en los demas procesos de transporte.

6.3.17. Tiempos de cargue

La figura 6.16 se presenta los tiempos de cargue de mercancía, que corresponde a la diferencia entre el tiempo de salida del cargue y el tiempo de entrada del cargue. Como se observa, la mayoría de los tiempos de cargue se realizan en un periodo de tiempo máximo una hora. También observa que con el paso de las horas la frecuencia tiende a ser menor.

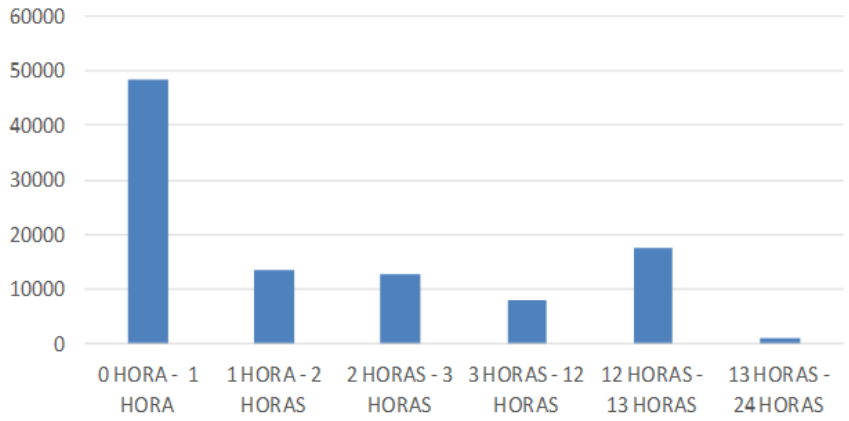


FIGURA 6.16. Tiempos de cargue

6.4. Verificación de la calidad de los datos

Con el análisis exploratorio de los datos se evidenció que debido a como está estructurado el proceso de recolección de los datos se presenta gran cantidad de datos faltantes entre los diferentes campos, situación que genera problemas en los procesos de análisis de la información. También se observa la necesidad de ajustar algunas variables, como por ejemplo el origen y el destino de la carga que deben ser desagregadas en departamento y en ciudad.

Otro error común en los datos es la falta de estandarización de algunos campos, generando gran dispersión en el comportamiento de los datos. Un ejemplo es el campo descripción corta de producto, que es un campo abierto para la escritura por parte de usuario. También se identificaron errores de semántica entre los campos, como en la relación entre la variable de tiempo de llegada de la carga y el tiempo de salida de la carga en donde en algunos casos se presentan errores de concordancia, como por ejemplo tiempos fechas y horas iguales en la variable de hora de salida del cargue y hora de llegada del cargue, o inclusive más extraños en donde la carga llegaba en una fecha anterior a que fuera cargada. A pesar de los ajustes que es necesario realizar a los datos, estos cuentan con un nivel de calidad aceptable para el desarrollo del estudio.

CAPÍTULO 7

Preparación de los datos

Durante esta fase se realizó la limpieza de un dataset compuesto por 480.000 registros aproximadamente. Se desarrolló un modelo de ETL para realizar las tareas de selección, limpieza, estandarización, creación de nuevos campos e integración de los datos.

7.1. Selección de los datos

De acuerdo con la revisión realizada en la exploración de los datos, se determinó que las variables con mejores condiciones de calidad y que mejor describen el proceso de transporte de carga por carretera son: *origen de la remesa, destino de la remesa, descripción corta del producto, cantidad de carga, tipo de empaque, fecha de salida de cargue, fecha de llegada a descargue.*

7.2. Limpieza de los datos

7.2.1. Reducción del dataset

Se realizó un filtro a los atributos de origen y destino de tal forma que un municipio tenga al menos 30 viajes, ya sea como origen o destino de la carga. La medida permite seleccionar los datos que sean más representativos para este campo y disminuir el nivel de dispersión.

El filtro permitió disminuir el número de clases de aproximadamente 1000 municipios, a trabajar solo con 260 municipios conservando un número de registros representativo y mitigar la maldición de la dimensionalidad. La anterior medida facilitará el proceso de modelamiento, ya que en algunos casos se requiere realizar una binarización de los datos en donde cada clase se convierte en un atributo, por lo que es claro que el trabajo con 260 atributos es mucho más eficiente que con 1000 atributos.

También se seleccionaron los registros que no tenían datos faltantes para los campos origen y destino de la carga y que pertenecían a alguno de los siete corredores logísticos.

El número de registros pasó de 480.000 a 365.294, número razonable para realizar la construcción de cualquier modelo, conservando la representatividad de los datos y teniendo en cuenta las capacidades computacionales disponibles.

7.2.2. Normalización

Se realizó la normalización entre 0 y 1 y eliminación de valores negativos para el campo cantidad de carga.

7.2.3. Detección de datos atípicos

Se realizó la detección y tratamiento de datos atípicos del campo cantidad de carga, eliminando 62 valores que agregaban ruido a el conjunto de datos.

7.3. Construcción de datos

7.3.1. Construcción del campo tramo

Un tramo corresponde a la concatenación del municipio de origen y destino. En transporte representa una porción de vía o ruta por la cual va a circular la carga. De esta manera una viaje que va desde San Gil a Bucaramanga pertenecerá al mismo tramo que un viaje que va desde Bucaramanga hacia San Gil, ver figura 7.1.

FIGURA 7.1. Tramo San Gil - Bucaramanga

El número total de tramos que puede tener un corredor corresponde a la combinatoria de tamaño 2 del número de municipios que lo conforman.

7.3.2. Construcción del campo actividad productiva

Se construye el campo actividad productiva entre otras razones debido a que el campo naturaleza de la carga tiene un sesgo claro en torno al tipo de carga normal por el mal

diligenciamiento de los usuarios, el proceso de ajustar los datos es un proceso muy complejo y se decidió como alternativa trabajar con el campo descripción corta del producto.

El campo descripción corta producto es diligenciado de forma directa por el usuario, por lo cual presenta una gran dispersión de información. Para solucionar el problema de información irrelevante, altamente dispersa y entrópica de este campo, se creó una clasificación de tipo de carga con respecto a las actividades de los sectores productivos que se presenta en la tabla 7.1. Las actividades fueron deducidas a partir de la frecuencia de ocurrencia de los diferentes términos del dataset.

Sector	Descripción	Actividad
Sector primario / Sector Agropecuario	Extracción directa de la naturaleza, se encuentran los factores económicos fuertes como son la agricultura, la ganadería, la pesca entre otros. En este se sector se extraen los bienes directamente de la naturaleza, de igual manera la minería hace parte de este sector porque de allí se extraen todas las piedras preciosas que hay en Colombia. Estas actividades ocupan el 7 % de la población activa en Colombia y el 3 % de lo que Colombia produce en total.	Agricultura
		Ganadería
		Pesca
		Minería
		Madera
		Avicultura
Sector secundario / Sector Industrial	Transformación de alimentos y materias primas, este sector se ocupa de actividades como son la industria metalúrgica, energética, textil, alimentación, química, madera y de igual forma la minería. Este sector se ocupa de las actividades destinadas a la transformación de alimentos y materia primas, en Colombia ocupa el 63 % de la población activa que se dedica a la transformación de materia en productos para el consumo y beneficio de todos los colombianos.	Construcción
		Metalúrgica
		Muebles y electrodomésticos
		Automotriz
		Cosméticos y aseo
		Envases y empaques
		Vidrios y cerámicas
		Químicos
		Alimentos y bebidas
		Alimento animales
		Petrolera
Sector terciario / Sector Servicios	Sector dedicado a ofrecer servicios a la sociedad, a las personas y a las empresas. Este sector se ocupa de los servicios de comercio, transporte, turismo entre otros. Se encarga de las actividades que cubren la demanda, por ejemplo de los servicios que al pueblo deben ser prestados, el trasporte que impulsa al comercio y posteriormente enriquece el turismo, que es en Colombia una de las actividades que más impulsa y que aumenta cada vez más el comercio a los lugares más concurridos por las personas, tanto del país como las personas del exterior.	Contenedores
		Varios

Tabla 7.1. Tabla de actividades productivas identificadas en el RNDC

Para la homogenización de la información se usó un diccionario de datos en el cual se agruparon todas las categorías pertenecientes a cada una de las actividades productivas. El diccionario agrupa en categorías más grandes las clases contenidas en el campo, de tal forma que en la actividad productiva agricultura se agrupan por ejemplo, maíz, papa, arroz, etc. El diccionario completo utilizado se encuentra en los anexos.

Las actividades emergentes que se pudieron observar después de un análisis de los datos fueron: *alimentos y bebidas, petrolera, envases y empaques, varios, agricultura, avicultura, construcción, metalúrgica, automotriz, contenedores, alimento animales, cosméticos y aseo, muebles y electrodomésticos, minería, vidrios y cerámicas, madera, químicos.* Así por ejemplo los términos papa, maíz, arroz pertenecen a una actividad *agrícola*, los términos cemento, ladrillos y tejas a la actividad *construcción*, etc. La distribución de la clases pasó de tener 1.000 clases aproximadamente a tener solo 17 despues de la estandarización. El diccionario completo aplicado al dataset se puede consultar en el anexo de la pagina 125.

7.3.3. Construcción de los corredores logísticos

Para la construcción del campo asociado a cada uno de los corredores se tuvo en cuenta la opinión de expertos, consulta de documentos del ministerio de transporte y el análisis de los datos de frecuencia de los viajes del RNDC.

Un viaje hace parte de un corredor cuando es transportado por al menos uno de los tramos que hacen parte del corredor.

El corredor Bogotá-Barranquilla está compuesto por 19 municipios, dos ramales, uno asociado a los municipios cercanos a la ciudad de Bogotá, y otro de municipios cercanos a la ciudad de Barranquilla y un conjunto de municipios intermedios entre las dos ciudades principales, ver tabla 8.6.

El corredor Bogotá-Bucaramanga está compuesto por 23 municipios, dos ramales, uno asociado a los municipios cercanos a la ciudad de Bogotá, y otro de municipios cercanos a la ciudad de Bucaramanga y un conjunto de municipios intermedios entre las dos ciudades principales, ver tabla 8.7.

El corredor Bogotá-Cali está compuesto por 20 municipios, dos ramales, uno asociado a los municipios cercanos a la ciudad de Bogotá y otro de municipios cercanos a la ciudad de Cali y un conjunto de municipios intermedios entre las dos ciudades principales, ver tabla 8.8.

El corredor Bogotá-Medellín está compuesto por 18 municipios, dos ramales, uno asociado a los municipios cercanos a la ciudad de Bogotá y otro de municipios cercanos a la ciudad de Medellín y un conjunto de municipios intermedios entre las dos ciudades principales, ver tabla 8.9.

El corredor Medellín-Barranquilla está compuesto por 16 municipios, dos ramales, uno asociado a los municipios cercanos a la ciudad de Medellín, y otro de municipios cercanos a la ciudad de Barranquilla y un conjunto de municipios intermedios entre las dos ciudades principales, ver tabla 8.10.

El corredor Medellín-Bucaramanga está compuesto por 17 municipios, dos ramales, uno asociado a los municipios cercanos a la ciudad de Medellín y otro de municipios cercanos a

la ciudad de Bucaramanga y un conjunto de municipios intermedios entre las dos ciudades principales, ver tabla 8.11.

El corredor Medellín-Cali está compuesto por 21 municipios, dos ramales, uno asociado a los municipios cercanos a la ciudad de Medellín y otro de municipios cercanos a la ciudad de Cali y un conjunto de municipios intermedios entre las dos ciudades principales, ver tabla 8.12.

7.3.4. Desagregación de los campos origen y destino de la remesa

Se realizó la separación de municipio y departamento para los atributos origen y destino de la remesa, lo anterior con el fin de hallar posibles patrones teniendo en cuenta las relaciones municipales y departamentales. La medida anterior permitió pasar de tener 2 atributos a tener 4 atributos potencialmente útiles para la realización del proceso de minería de datos.

7.4. Integración de los datos

Como resultado de los procesos de selección, limpieza y construcción se creó un dataset final con 365.387 registros y con 17 atributos entre los cuales se encuentran: municipio de origen de la carga, municipio de destino de la carga, departamento de origen de la carga, departamento de destino de la carga, descripción corta del producto, actividad productiva asociada a la carga, cantidad de carga, fecha de salida de cargue, fecha de llegada a descargue, tramo, corredor Bogotá-Barranquilla, corredor Bogotá-Bucaramanga, corredor Bogotá-Cali, corredor Bogotá-Medellín, corredor Medellín-Barranquilla, corredor Medellín-Bucaramanga, corredor Medellín-Cali.

7.5. Aplicación de formato a los datos

7.5.1. Binarización de campos

Se ajustaron los datos a un formato apropiado para probar algoritmos como FP-Growth y K-means. Se realizó una binarización de los atributos asociados a los corredores logísticos, departamentos de origen y destino de la carga, y actividades productivas.

CAPÍTULO 8

Transporte de carga en Colombia

En este capítulo se realiza una síntesis de la información nueva y relevante generada a partir del preprocesamiento de la información. Su importancia radica en que por una parte permitirá entender de una mejor manera el comportamiento del transporte de carga, por donde circula aproximadamente el 80 % de la carga movilizada en el país[37], y por otro lado permitirá mejorar el entendimiento de los resultados arrojados por los modelos de minería de datos.

En relación con los productos transportados, el sector que mayor moviliza carga corresponde al sector de alimentos y bebidas con un 20.9 % de los viajes, seguido por construcción con un 14.3 %, productos de la industria petrolera con 11 %, agricultura con 9.1 % y envases y empaques con un 8.9 %, ver tabla 8.1. Como se puede observar 10 actividades agrupan el 90.3 % de los viajes, el restante 10 % es repartido por otras actividades en menor proporción.

Actividades	Soporte
Alimentos y bebidas	20.9 %
Construcción	14.3 %
Petrolera	11.0 %
Agricultura	9.1 %
Envases y empaques	8.9 %
Avicultura	7.9 %
Automotriz	5.5 %
Minería	4.8 %
Contenedores	4.3 %
Alimento animales	3.6 %
Total	90.3 %

Tabla 8.1. Distribución de las actividades productivas

La distribución de los orígenes de la carga por departamento, presenta con el 16.9 % de la carga originada al departamento del Valle del Cauca, seguido por Cundinamarca con 13.6 %, Santander con 11 %, Bogotá con un 10 % y Antioquia con un 8 %. Como se observa en la tabla 8.2, nueve departamentos y Bogotá originan el 84 % de la carga transportada.

Departamento	Soporte
Valle del Cauca	16.9 %
Cundinamarca	13.6 %
Santander	11.0 %
Bogotá D.C.	10.0 %
Atlántico	8.2 %
Antioquia	8.0 %
Sucre	6.3 %
Boyacá	5.5 %
Bolivar	4.5 %
Total	84.0 %

TABLA 8.2. Distribución de departamentos que originan carga

La distribución de los destinos de la carga por departamento, presenta que el 14 % de la carga tiene como destino el departamento del Valle del Cauca, seguido por Cundinamarca con un 11.3 %, Santander con 11 %, Bogotá con 10.9 % y Antioquia con 9.1 %. Como se observa en la tabla 8.3, ocho departamentos y Bogotá reciben el 74.1 % de la carga transportada.

Departamento	Soporte
Valle del Cauca	14 %
Cundinamarca	11.3 %
Santander	11.0 %
Bogotá D.C.	10.9 %
Antioquia	9.1 %
Bolivar	6.7 %
Atlántico	6.4 %
Meta	4.7 %
Total	74.1 %

TABLA 8.3. Distribución de departamentos que destino de la carga

En cuanto al comercio entre departamentos se identificó que 7.7 % del transporte de carga por carretera a nivel nacional se lleva a cabo de forma interna en el departamento del Valle del Cauca, de la misma forma el 6.3 % de la carga se comercia en Santander y el 4.1 % entre Cundinamarca y Bogotá. En cuanto a la construcción, el principal originador de este tipo de carga es el departamento de Sucre y el principal receptor es el departamento de Cundinamarca. En el sector de alimentos y bebidas, el principal originador de la carga es el departamento de Valle del Cauca con 5.4 %, seguido de Cundinamarca con 4.7 %. Adicionalmente, se observa que el 3.5 % de la carga movilizada tiene como origen Santander, y corresponde a productos avícolas.

8.1. Corredores logísticos

Los corredores que más movilizan carga en orden descendente serían: Bogotá-Barranquilla (21.3 %), Bogotá-Cali (19.6 %), Bogotá-Bucaramanga (16.2 %), Medellín-Cali

(13.1 %), Bogotá-Medellín (10.7 %), Medellín-Barranquilla (7.9 %), Medellín-Bucaramanga (7.8 %). Cabe aclarar que para el caso anterior, los porcentajes no suman 100 % debido a que hay tramos que son compartidos por más de un corredor, ver tabla 8.4.

Corredor	Porcentaje
Bogotá-Barranquilla	21.3 %
Bogotá-Cali	19.6 %
Bogotá-Bucaramanga	16.2 %
Medellín-Cali	13.1 %
Bogotá-Medellín	10.7 %
Medellín-Barranquilla	7.9 %
Medellín-Bucaramanga	7.8 %

TABLA 8.4. Distribución de las frecuencias de viajes en los corredores logísticos

La tabla 8.5) presenta las intersecciones entre corredores con mayor nivel de concurrencia en número de viajes. En primer lugar aparece los corredores Bogotá-Bucaramanga y Bogotá-Barranquilla con 7.2 % de los viajes a nivel Nacional, Bogotá-Cali y Medellín-Cali con 7.0 %, Bogotá-Barranquilla y Bogotá-Medellín con 6.9 %, Bogotá-Bucaramanga y Bogotá-Medellín con 6.9 %, Bogotá-Cali y Bogotá-Barranquilla con 6.8 %. Como hipótesis se observa que Bogotá y municipios aledaños (ramales) movilizan aproximadamente 6.8 % de la carga a nivel nacional.

Frecuencia	Corredor 1	Corredor 2
7.2 %	Bogotá-Barranquilla	Bogotá-Bucaramanga
7,0 %	Bogotá-Cali	Medellín-Cali
6.9 %	Bogotá-Barranquilla	Bogotá-Medellín
6.9 %	Bogotá-Bucaramanga	Bogotá-Medellín
6.8 %	Bogotá-Barranquilla	Bogotá-Cali
6.8 %	Bogotá-Cali	Bogotá-Bucaramanga
6.8 %	Bogotá-Cali	Bogotá-Medellín
5.7 %	Bogotá-Bucaramanga	Medellín-Bucaramanga

TABLA 8.5. Distribución de las frecuencias de los viajes en las intersecciones de los corredores

8.1.1. Corredor Bogotá-Barranquilla

El corredor Bogotá-Barranquilla esta compuesto por los municipios presentados en la tabla 8.6.

Corredor	Municipios Intermedios	Ramales Bogotá	Ramales Barranquilla
Bogotá-Barranquilla	Tunja, Barrancabermeja	Bogotá, Cota, Villavicencio, Mosquera, Funza, Chía, Cajicá, Tocancipá, Zipaquirá.	Aguachica, Valledupar, Fundación, Santa Marta, Cartagena, Turbaco, Toluviejo, Barranquilla.

TABLA 8.6. Municipios pertenecientes al corredor Bogotá-Barranquilla

FIGURA 8.1. Corredor Bogotá-Barranquilla

De acuerdo a los datos recolectados por el RNDC el transporte de mercancías por el corredor Bogotá-Barranquilla es principalmente de alimentos y bebidas (20 %), elementos para la construcción (18 %), elementos varios (14 %), envases y empaques (13 %) y productos de la industria petrolera (12 %), ver figura 8.2.

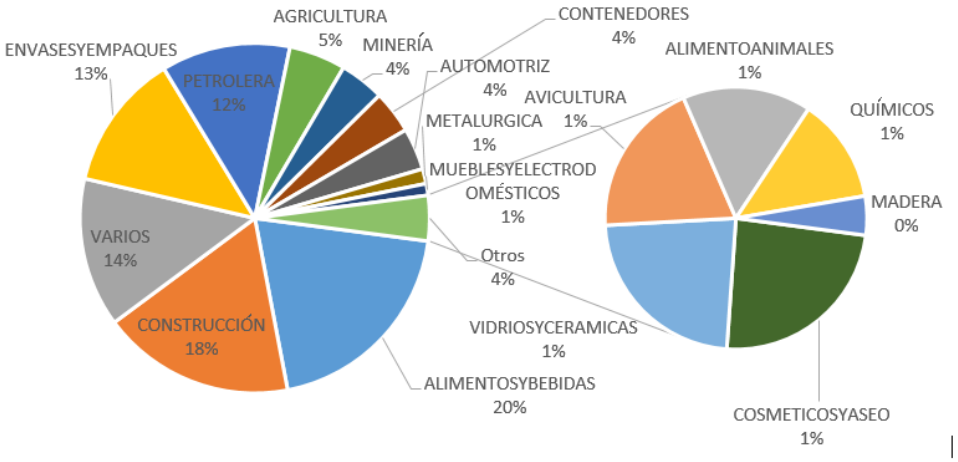


FIGURA 8.2. Productos transportados en el corredor Bogotá-Barranquilla

Los principales productos transportados por este corredor, que corresponden a alimentos y bebidas y elementos de construcción, representan el 4.9 % y 4.4 % de los viajes

realizados a nivel nacional. En cuanto a los departamentos que originan la carga por este corredor, aparece Bogotá con 5.3 %, seguido de Cundinamarca con 5.0 % y Sucre con 3.5 % del total de viajes realizados a nivel nacional. Como principales departamentos de destino de la carga se encuentran Cundinamarca con 5.8 %, Bogotá con 5.4 % y Bolívar con 3.6 % del total de viajes realizados a nivel nacional.

Respecto a la utilización de tramos de vías, la figura 8.3 muestra los 20 tramos más concurridos del corredor Bogotá-Barranquilla.

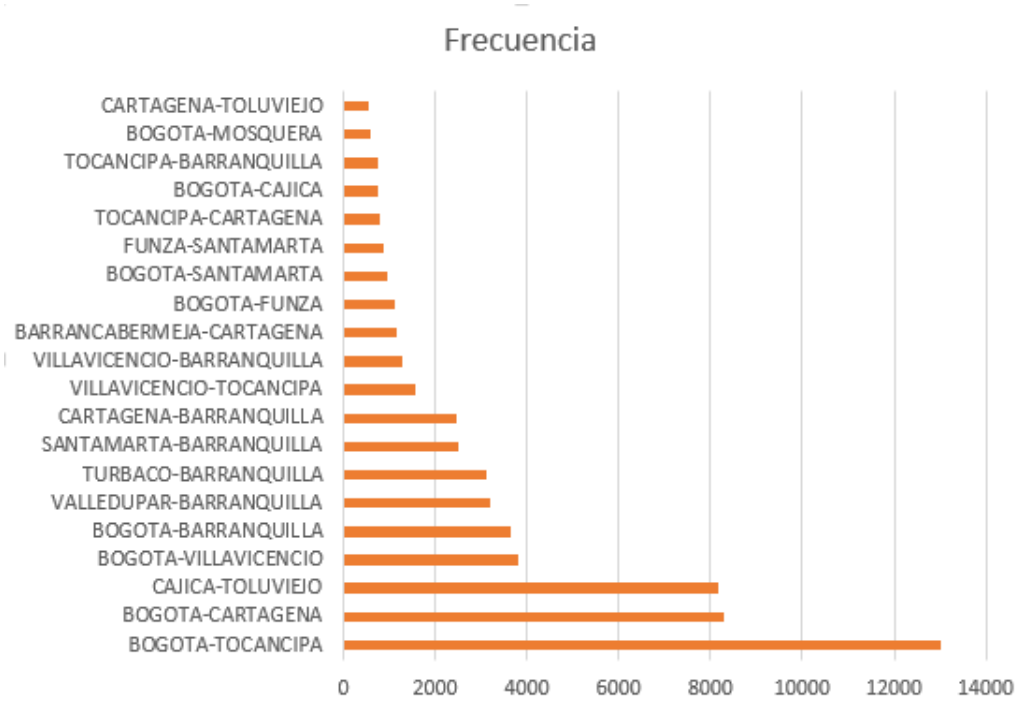


FIGURA 8.3. Tramos mas concurridos del corredor Bogotá-Barranquilla

El corredor Bogotá-Barranquilla muestra que el principal departamento de origen de la carga para sus intersecciones entre los corredores Bogotá-Bucaramanga, Bogotá-Medellín y Bogotá Cali corresponde a Cundinamarca. En cuento a la principal ciudad de destino en sus intersecciones con los corredores Bogotá-Cali, Bogotá-Bucaramanga y Bogotá-Medellín corresponde a la ciudad de Bogotá. Adicionalmente el principal origen y destino perteneciente al corredor Bogotá-Barranquilla es Cundinamarca-Bogotá con 3.6 %, una de las actividades productivas que más moviliza carga corresponde a la construcción y tiene como origen de la carga el departamento de Sucre.

8.1.2. Corredor Bogotá-Bucaramanga

El corredor Bogotá-Bucaramanga está compuesto por 23 municipios, dos ramales, uno asociado a los municipios cercanos a la ciudad de Bogotá, y otro de municipios cercanos a la ciudad de Bucaramanga y un conjunto de municipios intermedios entre las dos ciudades principales. Los municipios que lo componen son presentados en la tabla 8.7.

Corredor	Municipios Intermedios	Ramales Bogotá	Ramales Bucaramanga
Bogotá-Bucaramanga	Tunja, Chiquinquirá, Duitama, Sogamoso,	Bogotá, Cota, Villavicencio, Mosquera, Funza, Chía, Cajicá, Tocancipá,	Socorro, San Gil, Los Santos, Piedecuesta, Floridablanca, Girón, Lebrija, Rionegro, Barrancabermeja, Cúcuta, Bucaramanga.

TABLA 8.7. Municipios pertenecientes al corredor Bogotá-Bucaramanga

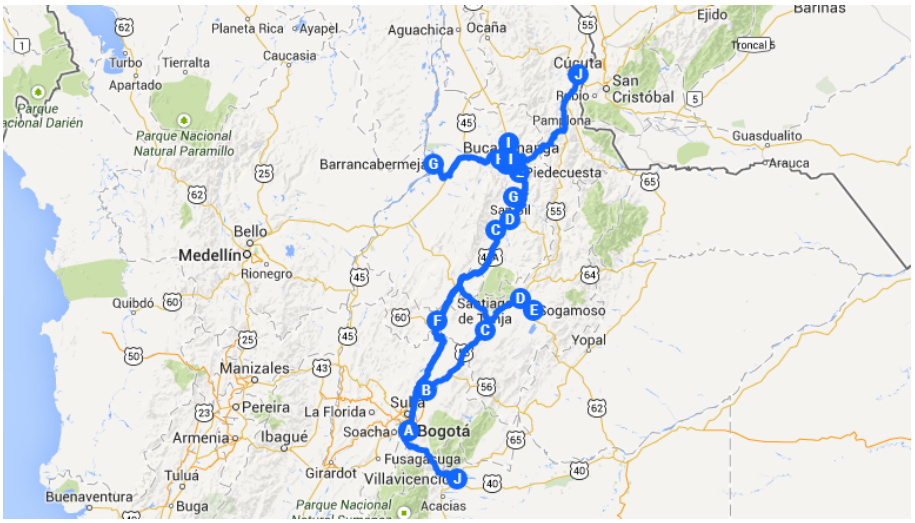


FIGURA 8.4. Corredor Bogotá-Bucaramanga

El transporte de mercancías por el corredor Bogotá-Bucaramenga es principalmente de alimentos y bebidas (27 %), envases y empaques (15 %), productos relacionados con la avicultura (14 %), elementos varios (14 %), alimento para animales (9 %) y productos de la industria petrolera (9 %), ver figura 8.5.

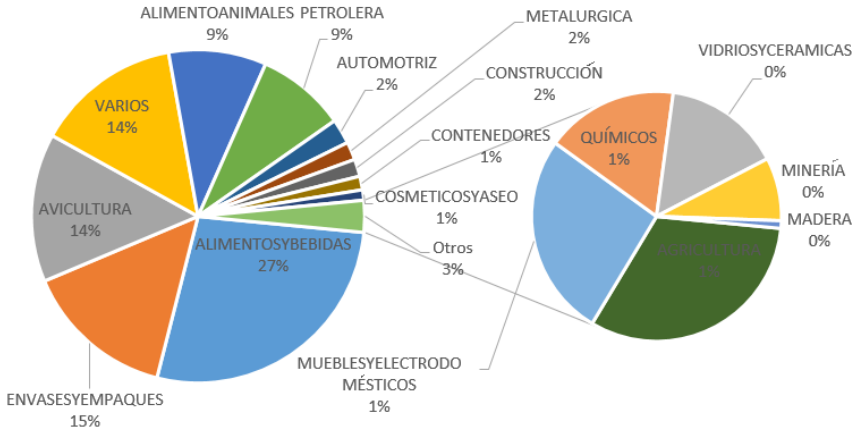


FIGURA 8.5. Productos transportados en el corredor Bogotá-Bucaramanga

El corredor Bogotá-Bucaramanga moviliza el 5.2 % de los viajes a nivel nacional correspondiente a alimentos y bebidas. Los principales departamentos generadores de carga son Santander con 7.3 %, Cundinamarca con 4.9 % y Bogotá con 3.5 %. Por otro lado los principales departamentos destino son Santander con 6.5 % y Bogotá con 4.6 %.

La figura 8.6 muestra los 20 tramos mas concurridos del corredor Bogotá-Bucaramanga.

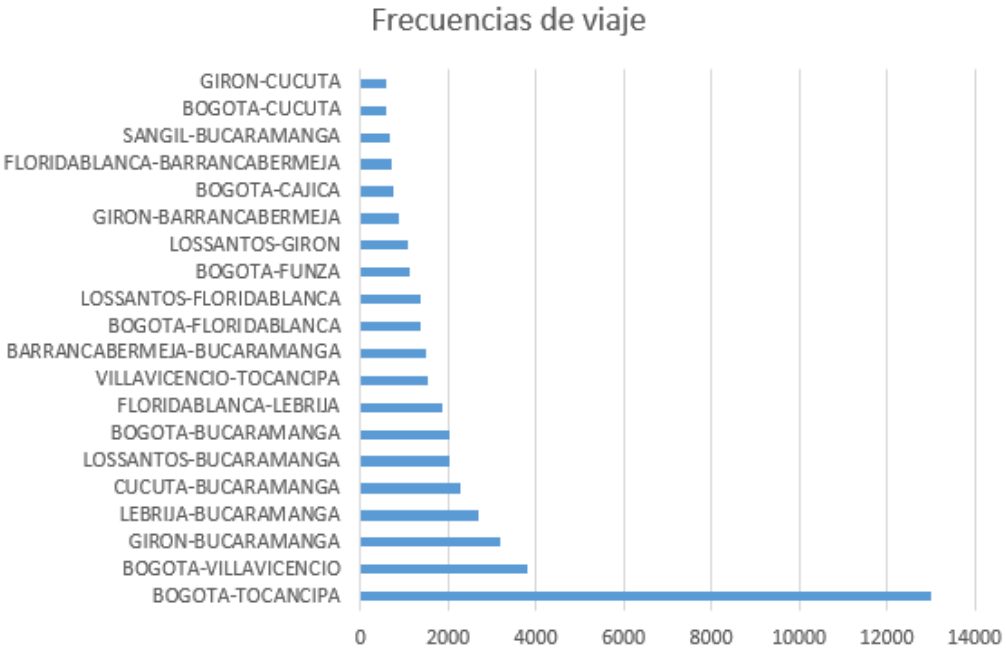


FIGURA 8.6. Tramos mas concurridos del corredor Bogotá-Bucaramanga

El corredor Bogotá-Bucaramanga presenta como principal generador y receptor de carga en su intersección con el corredor Bogotá-Medellín al departamento de Santander. Se observa también que el 5.5 % de la carga a nivel nacional se moviliza de manera interna en el departamento de Santander. El comercio entre Cundinamarca y Bogotá también ocupa un renglón importante dentro del corredor Bogotá-Bucaramanga, ya que dentro de este se moviliza el 3.6 % de los viajes a nivel nacional.

8.1.3. Corredor Bogotá-Cali

El corredor Bogotá-Cali está compuesto por 20 municipios, dos ramales, uno asociado a los municipios cercanos a la ciudad de Bogotá y otro de municipios cercanos a la ciudad de Cali y un conjunto de municipios intermedios entre las dos ciudades principales, ver tabla 8.8.

Corredor	Municipios Intermedios	Ramales Bogotá	Ramales Cali
Bogotá-Cali	Soacha, Fusagasugá, Melgar, Girardot, Ibagué, Armenia.	Bogotá, Villavicencio, Tocancipá, Chía, Mosquera, Funza, Cota.	Tuluá, Guadalajara de Buga, Buenaventura, Palmira, Yumbo, Jamundí, Cali.

TABLA 8.8. Municipios pertenecientes al corredor Bogotá-Cali

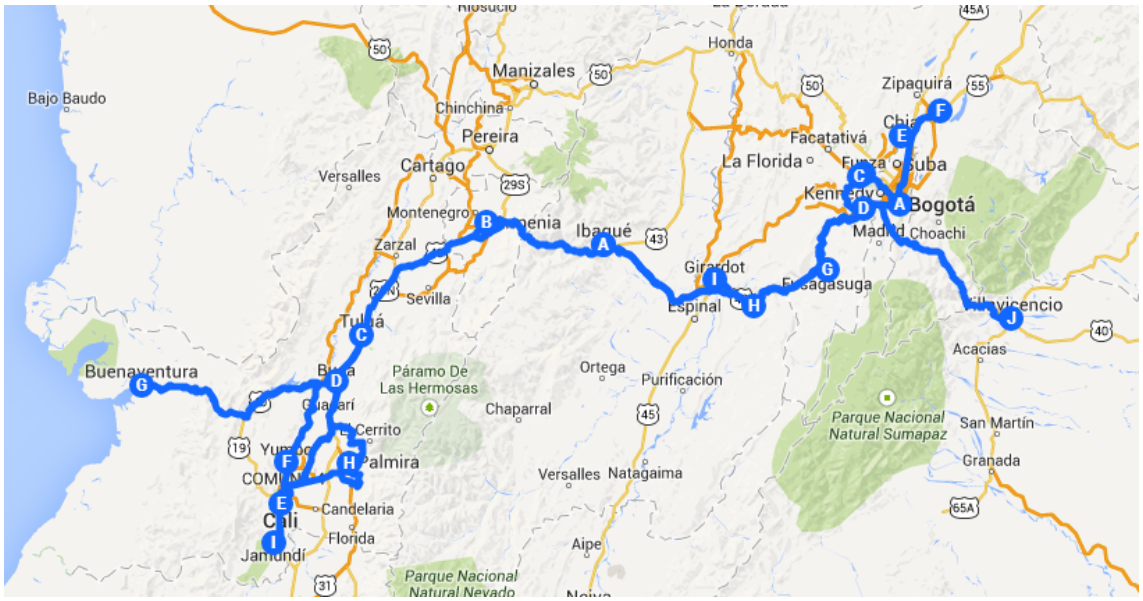


FIGURA 8.7. Corredor Bogotá-Cali

El transporte de mercancías por el corredor Bogotá-Cali corredor es principalmente de alimentos y bebidas (29 %), elementos varios (21 %), envases y empaques (11 %), productos agrícolas (9 %), productos de la industria petrolera (9 %), carga contenerizada (5 %), productos relacionados con la avicultura (5 %), productos industria automotriz (5 %), ver figura 8.8.

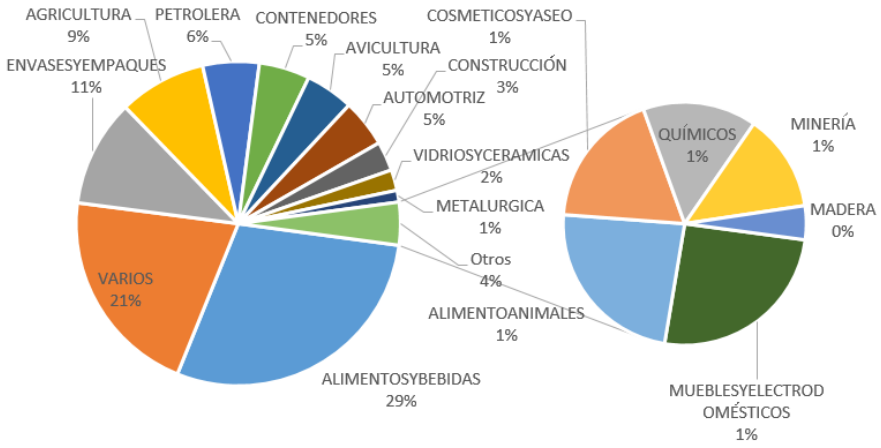


FIGURA 8.8. Productos transportados en el corredor Bogotá-Cali

El corredor Bogotá-Cali genera el 7.2 % de los viajes a nivel nacional con el transporte de alimentos y bebidas. Los principales departamentos que originan carga para este corredor son: Valle del Cauca con 9.2 %, Cundinamarca con 5.6 % y Bogotá con 4.4 % de los viajes a nivel nacional. Los principales departamentos destinos de la carga del corredor son Valle del Cauca con un 8.4 % y Bogotá con 5.1 % de los viajes a nivel nacional.

La figura 8.9 muestra los 20 tramos mas concurridos del corredor Bogotá-Cali.

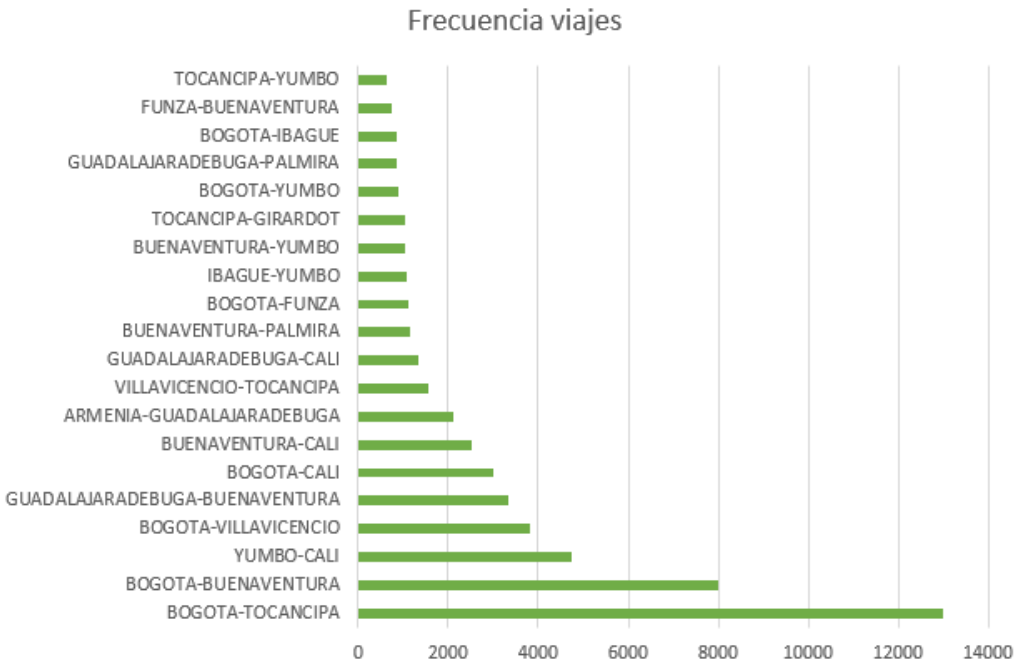


FIGURA 8.9. Tramos mas concurridos del corredor Bogotá-Cali

El corredor Bogotá-Cali muestra que el principal generador y receptor de carga es el departamento de Valle del Cauca, seguido por el comercio interno entre Bogotá y Cundinamarca. El principal generador y receptor de la carga en la intersección entre el corredor Bogotá-Cali y Medellín-Cali, es el departamento de Valle del Cauca. En la intersección con el corredor Bogotá-Bucaramanga, muestra que la movilización de carga se realiza principalmente entre el departamento de Cundinamarca y Bogotá. La intersección con el corredor Bogotá-Medellín tiene como principal departamento de origen Cundinamarca, y como principal destino Bogotá.

8.1.4. Corredor Bogotá-Medellín

El corredor Bogotá-Medellín está compuesto por 18 municipios, dos ramales, uno asociado a los municipios cercanos a la ciudad de Bogotá y otro de municipios cercanos a la ciudad de Medellín y un conjunto de municipios intermedios entre las dos ciudades principales, ver tabla 8.9.

Corredor	Municipios Intermedios	Ramales Bogotá	Ramales Medellín
Bogotá-Medellín	Guaduas, Honda, La Dorada, Marinilla.	Bogotá, Cota, Villavicencio, Mosquera, Funza, Chía, Cajicá, Tocancipá.	Rionegro, Bello, Itagüí, Envigado, Sabaneta, Medellín.

TABLA 8.9. Municipios pertenecientes al corredor Bogotá-Medellín

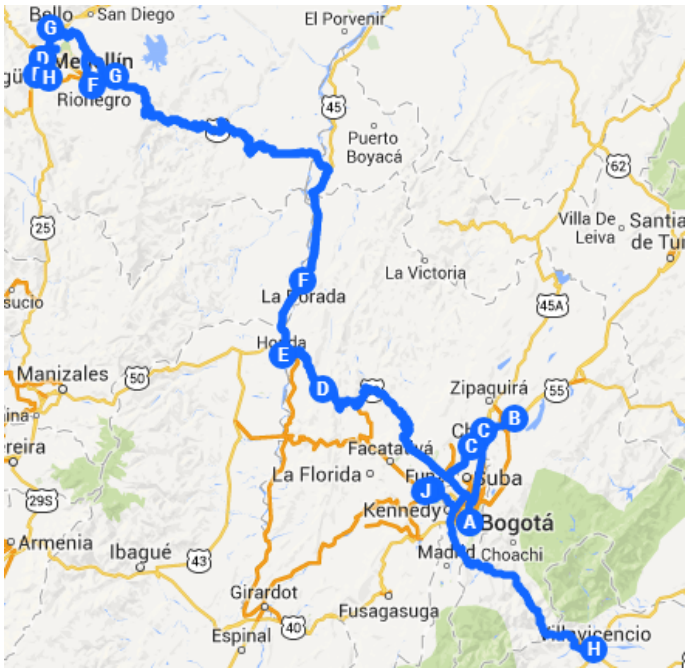


FIGURA 8.10. Corredor Bogotá-Medellín

El transporte de mercancías por el corredor Bogotá-Medellín es principalmente de alimentos y bebidas (29 %), elementos varios (22 %), envases y empaques (16 %), productos de la industria petrolera (10 %) y productos industria automotriz (8 %), ver figura 8.11.

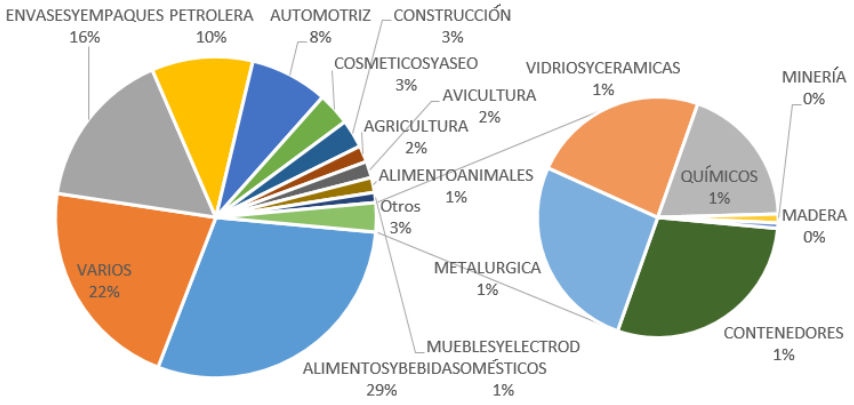


FIGURA 8.11. Productos transportados en el corredor Bogotá-Medellín

En corredor Bogotá-Medellín soporta el 4.4 % de los viajes a nivel nacional, y tiene como su principal destino la ciudad de Bogotá. Cundinamarca con 4.8 % de los viajes a nivel nacional, corresponde al principal generador de carga en este corredor y la actividad productiva que más moviliza viajes corresponde a la de alimentos y bebidas con un 4.0 % del total de viajes a nivel nacional.

La figura 8.12 muestra los 20 tramos mas concurridos del corredor Bogotá-Medellín.

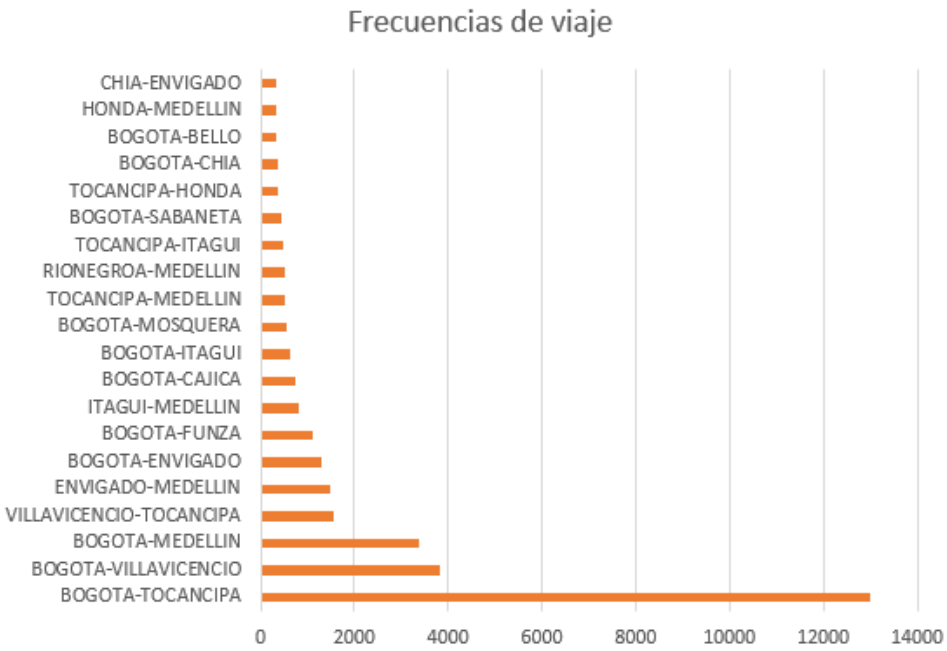


FIGURA 8.12. Tramos mas concurridos del corredor Bogotá-Medellín

8.1.5. Corredor Medellín-Barranquilla

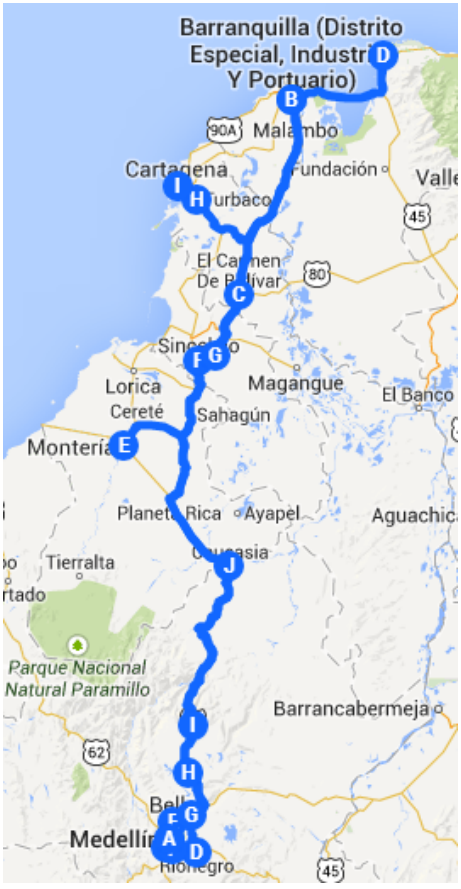


FIGURA 8.13. Corredor Medellín-Barranquilla

El corredor Medellín-Barranquilla está compuesto por 16 municipios, dos ramales, uno asociado a los municipios cercanos a la ciudad de Medellín, y otro de municipios cercanos a la ciudad de Barranquilla y un conjunto de municipios intermedios entre las dos ciudades principales, ver tabla 8.10.

Corredor	Municipios Intermedios	Ramales Medellín	Ramales Barranquilla
Medellín-Barranquilla	Girardota, Santa Rosa de Osos, Yarumal, Montería, Sincelejo, Toluviejo	Medellín, Sabaneta, Rionegro, Envigado, Itagüí, Bello.	Turbaco, Cartagena, Santa Marta, Barranquilla.

TABLA 8.10. Municipios pertenecientes al corredor Medellín-Barranquilla

El transporte de mercancías por el corredor Medellín-Barranquilla es principalmente de alimentos y bebidas (25 %), elementos varios (19 %), envases y empaques (14 %), elementos para la construcción (7 %), productos agrícolas (7 %), carga contenerizada (5 %) y productos industria automotriz (6 %), ver figura 8.14.

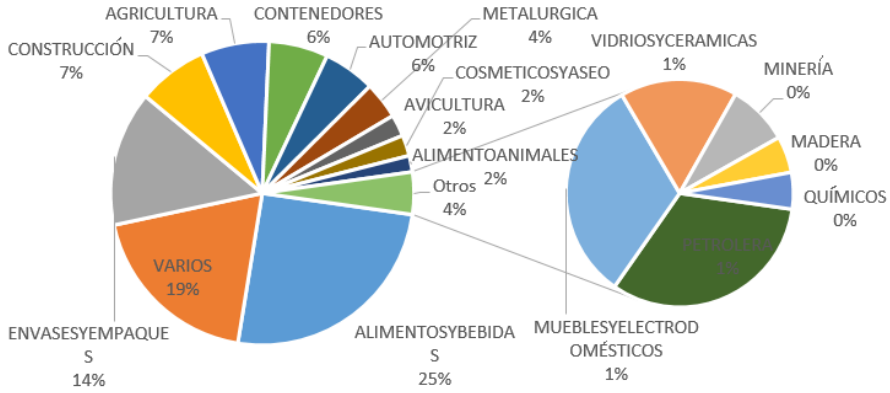


FIGURA 8.14. Productos transportados en el corredor Medellín-Barranquilla

La figura 8.15 muestra los 20 tramos mas concurridos del corredor Medellín-Barranquilla.

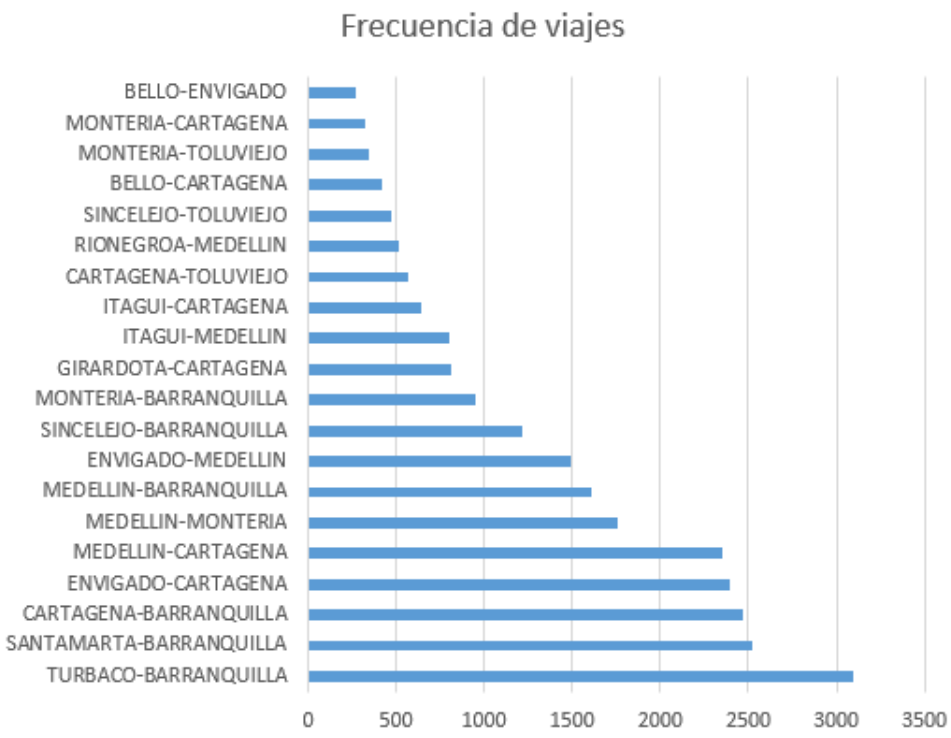


FIGURA 8.15. Tramos mas concurridos del corredor Medellín-Barranquilla

8.1.6. Corredor Medellín-Bucaramanga

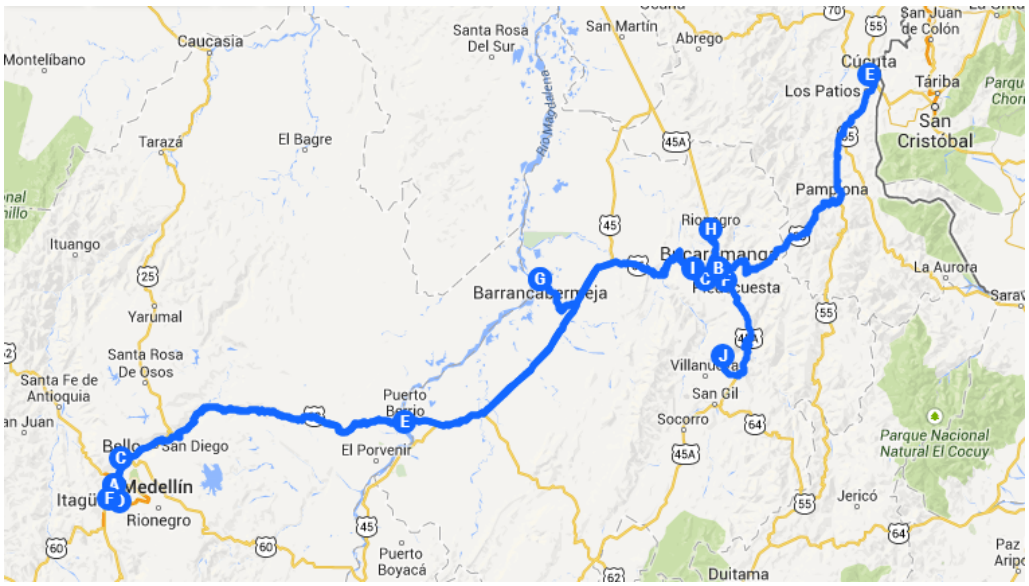


FIGURA 8.16. Corredor Medellín-Bucaramanga

El corredor Medellín-Bucaramanga está compuesto por 17 municipios, dos ramales, uno asociado a los municipios cercanos a la ciudad de Medellín y otro de municipios cercanos a la ciudad de Bucaramanga y un conjunto de municipios intermedios entre las dos ciudades principales, ver tabla 8.11.

Corredor	Municipios Intermedios	Ramales Medellín	Ramales Bucaramanga
Medellín-Bucaramanga	Girardota, Puerto Berrío, Itagüí, Barrancabermeja.	Medellín, Sabaneta, Rionegroa, Envigado, Itagüí, Bello.	Lebrija, Girón, Los Santos, Floridablanca, Rionegro, Cúcuta, Bucaramanga.

TABLA 8.11. Municipios pertenecientes al corredor Medellín-Bucaramanga

El transporte de mercancías por el corredor Medellín-Bucaramanga es principalmente de productos avícolas (26 %), seguido de alimentos y bebidas (20 %), elementos varios (19 %), alimento para animales (18 %), envases y empaques (6 %) y productos de la industria automotriz (3 %), ver figura 8.17.

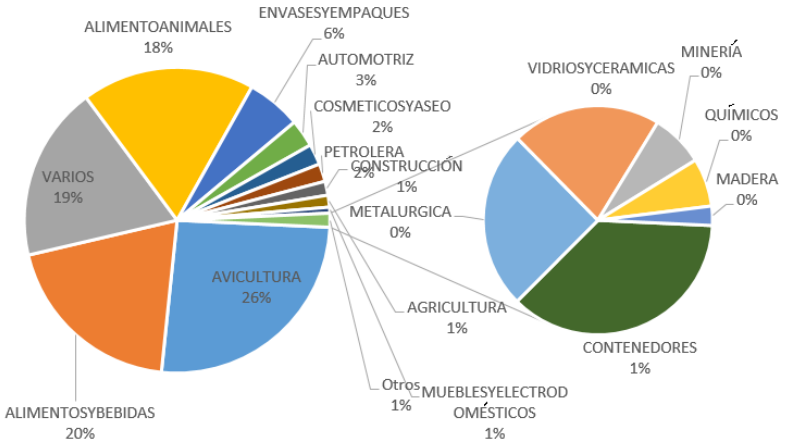


FIGURA 8.17. Productos transportados en el corredor Medellín-Bucaramanga

La figura 8.18 muestra los 20 tramos mas concurridos del corredor Medellín-Bucaramanga

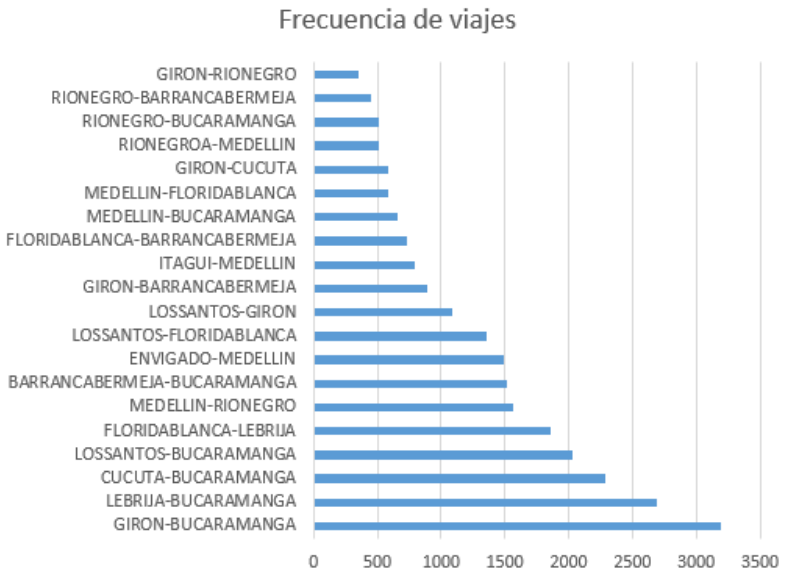


FIGURA 8.18. Tramos mas concurridos del corredor Medellín-Bucaramanga

En el corredor Medellín-Bucaramanga como rasgo interesante se identificó a Santander como principal departamento de origen y destino de la carga con 5.3 % y 5.0 % de los viajes movilizados a nivel nacional.

8.1.7. Corredor Medellín-Cali

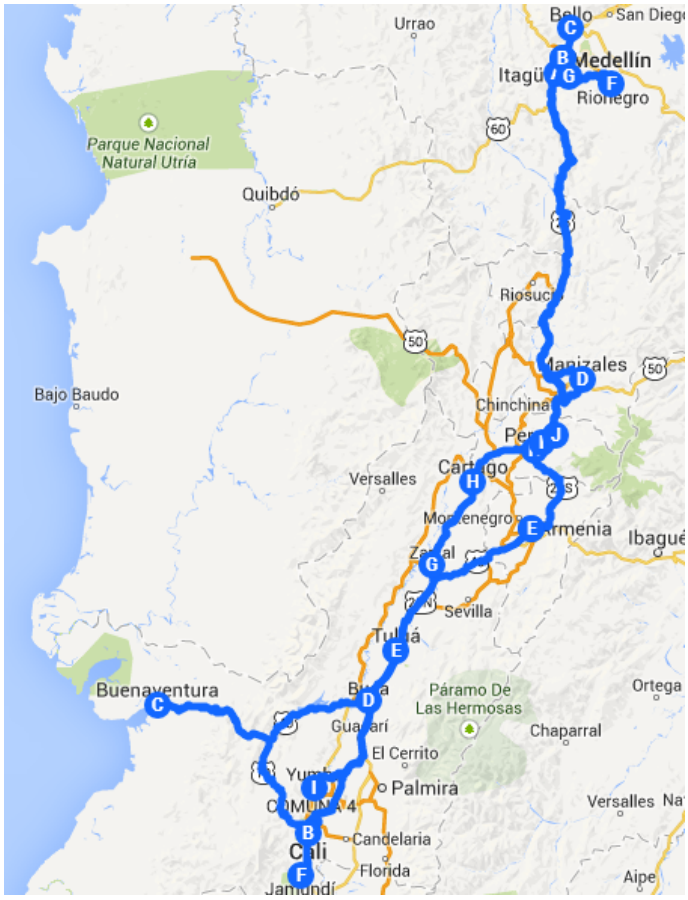


FIGURA 8.19. Corredor Medellín-Cali

El corredor Medellín-Cali está compuesto por 21 municipios, dos ramales, uno asociado a los municipios cercanos a la ciudad de Medellín y otro de municipios cercanos a la ciudad de Cali y un conjunto de municipios intermedios entre las dos ciudades principales, ver tabla 8.12.

Corredor	Municipios Intermedios	Ramales Medellín	Ramales Cali
Medellín-Cali	Zarzal, Cartago, Armenia, Pereira, Dosquebradas, Santa Rosa de Cabal, Manizales, Caldas.	Rionegroa, Envigado, Itagüí, Bello, Sabaneta, Medellín.	Cali, Palmira, Guadalajara de Buga, Buenaventura, Yumbo, Jamundí, Tuluá.

TABLA 8.12. Municipios pertenecientes al corredor Medellín-Cali

El transporte de mercancías por este corredor es principalmente de elementos varios (26 %), seguido de alimentos y bebidas (25 %), productos avícolas (11 %), productos agríco-

las (11 %), productos de la industria automotriz (6 %), y carga contenerizada (6 %), ver figura 8.20.

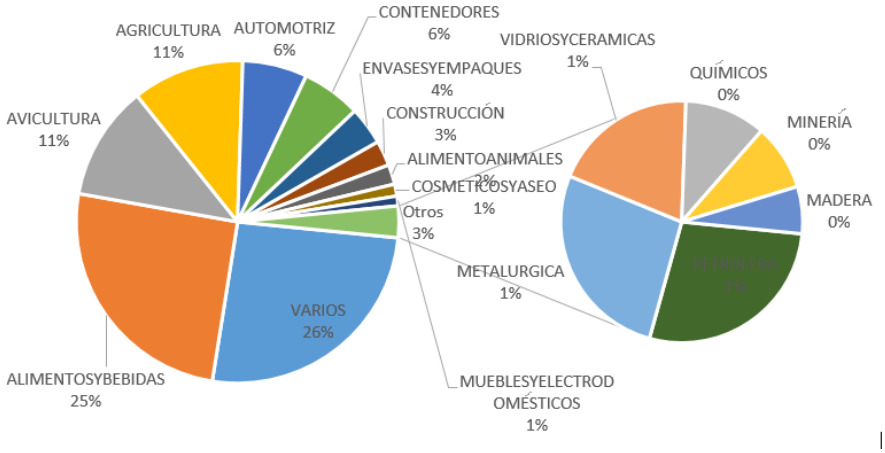


FIGURA 8.20. Productos transportados en el corredor Medellín-Cali

El corredor logístico Medellín-Cali presenta comercio principalmente de alimentos y bebidas con 4.5 %, siendo el principal origen y destino de la carga el Valle del Cauca con 9.3 % y 7.7 % de los viajes movilizados a nivel nacional.

La figura 8.21 muestra los 20 tramos mas concurridos del corredor Medellín-Cali.

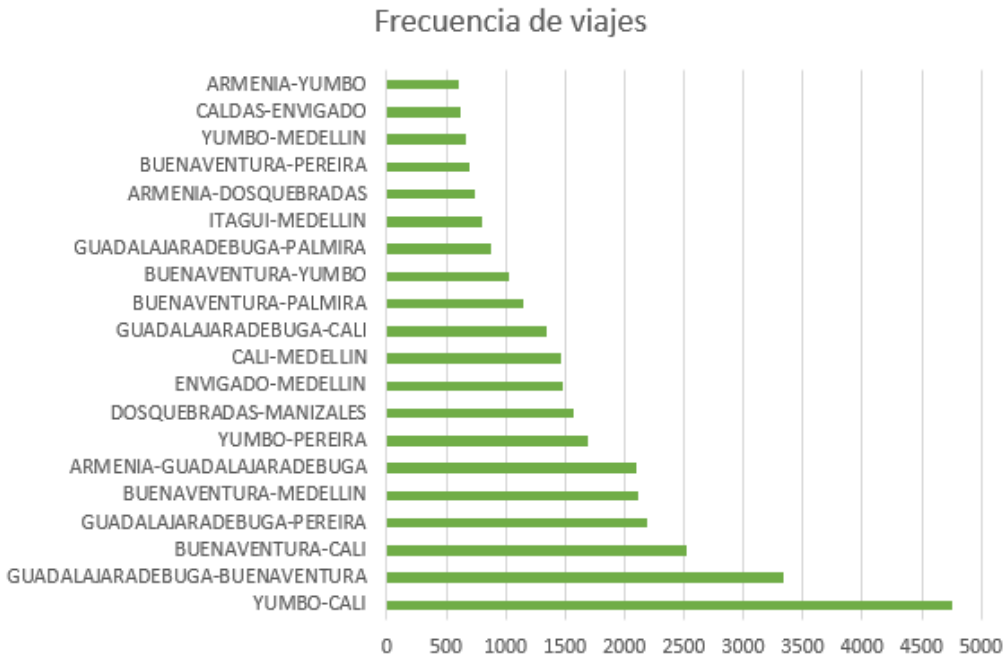


FIGURA 8.21. Tramos mas concurridos del corredor Medellín-Cali

CAPÍTULO 9

Modelamiento

9.1. Modelo descriptivo

9.1.1. Selección de la técnica para el modelo descriptivo

Teniendo en cuenta las características de los datos, se consideró pertinente usar un modelo de clustering para describir los principales clusters de transporte de carga y su comportamiento en términos de producto transportado, localización geográfica, corredores y tramos.

9.1.2. Construcción del modelo

El modelo está compuesto por cuatro fases claramente definidas, ver figura 9.1. La primera etapa, consiste en la adquisición de la información, que básicamente recupera la información de los registros del RNDC. La segunda etapa, corresponde al preprocesamiento en donde se realizan tareas de limpieza de datos, selección de atributos, construcción y selección de tramos, detección de datos atípicos, binarización y por finalmente integración de los datos, ver figura 9.2 y 9.3. En la última etapa se lleva a cabo el proceso de modelamiento, donde se aplica el algoritmo K-means, se realiza su validación y el almacenamiento de los resultados, ver figura 9.4.

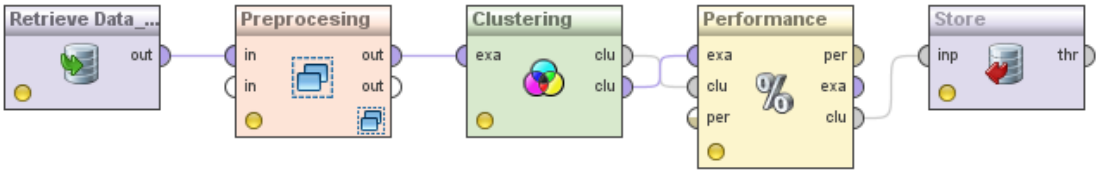


FIGURA 9.1. Modelo de clustering

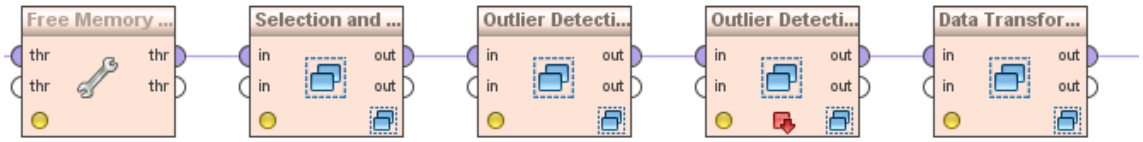


FIGURA 9.2. Preprocesamiento del modelo de clustering

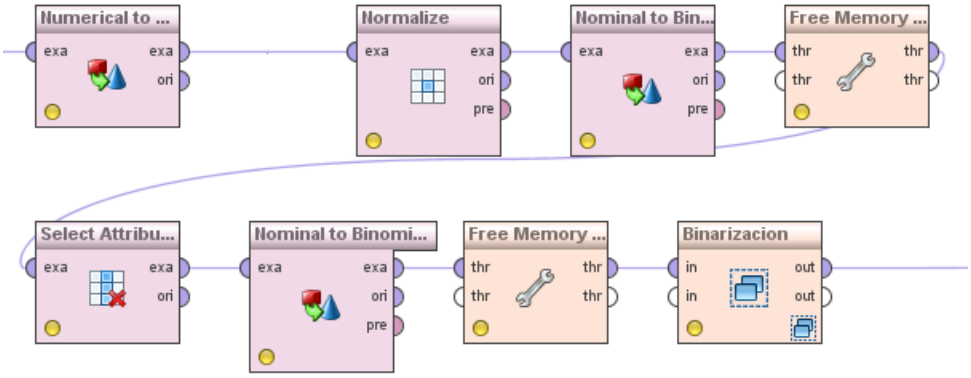


FIGURA 9.3. Preprocesamiento del modelo de clustering (Transformación de datos)

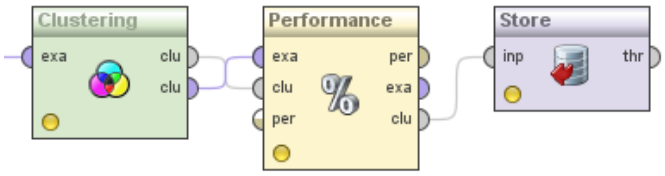


FIGURA 9.4. Modelamiento de clusters

9.1.2.1. Datos de entrada

Como datos de entrada al modelo se ingresaron 100.215 registros, con los campos asociados a los corredores logísticos, la cantidad cargada, el departamento de origen y destino, la actividad productiva, el tipo de empaque y el tramo por el que circula la carga, ver tabla 9.1.

Atributo	Rango de valores	Tipo de variable	Significado
Bogotá-Barranquilla	true, false	Binomial	Indica si el viaje circula por el corredor Bogotá-Barranquilla
Bogotá-Bucaramanga	true, false	Binomial	Indica si el viaje circula por el corredor Bogotá-Bucaramanga
Bogotá-Cali	true, false	Binomial	Indica si el viaje circula por el corredor Bogotá-Cali
Bogotá-Medellín	true, false	Binomial	Indica si el viaje circula por el corredor Bogotá-Medellín
Medellín-Barranquilla	true, false	Binomial	Indica si el viaje circula por el corredor Medellín-Barranquilla.

Medellín-Bucaramanga	true, false	Binomial	Indica si el viaje circula por el corredor Medellín-Bucaramanga.
Medellín-Cali	true, false	Binomial	Indica si el viaje circula por el corredor Medellín-Cali.
Cantidad cargada	Valores normalizados entre 0 y 1	Numerica	Indica cual es la cantidad de carga transportada en cada viaje.
Departamento origen	Enteros positivos	Nominal	Indica cual es el departamento de origen de la carga.
Departamento destino	Enteros positivos	Nominal	Indica cual es el departamento de destino de la carga.
Tipo de empaque	Valores en el intervalo entre (0,19).	Nominal	Indica a cual es el tipo de empaque con el cual se transporta la mercancia.
Actividad	Alimentos y bebidas, petrolera, envases y empaques, agricultura, avicultura, construcción, automotriz, contenedores, alimento animales, minería.	Polinominal	Indica a cual actividad productiva pertenece la carga transportada
Tramo	Origen-Destino	Polinominal	Indica a cual tramo pertenece el viaje. Esta compuesto por la ciudad de origen y destino de la mercancia.

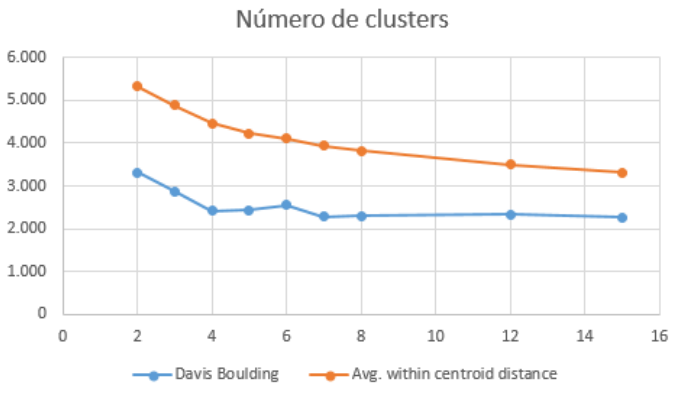
TABLA 9.1. Datos de entrada modelo de clustering

9.1.2.2. Configuración de los parámetros

Para determinar el número adecuado de clusters se realizó una evaluación utilizando los algoritmos Davis Boulding y el cálculo del promedio de las distancias al centroide de cada cluster. Como resultado se pudo determinar que un número adecuado de clusters es 6 debido a que corresponde a un punto de inflexión para ambos algoritmos, ver figuras 9.5a y 9.5b.

k	Davis Boulding	Avg. within centroid distance
2	3.321	5.325
3	2.866	4.892
4	2.421	4.467
5	2.429	4.228
6	2.553	4.103
7	2.279	3.942
8	2.291	3.824
12	2.333	3.494
15	2.272	3.306

(A) Tabla



(B) Grafica

FIGURA 9.5. Número de clusters

9.1.3. Resultados de la aplicación del modelo

Los resultados de la aplicación del modelo se muestran en las figuras 9.10a a la 9.9, en donde los colores más cercanos al color rojo indican mayor actividad de la clase dentro de cada cluster y los más cercanos a color verde indican lo contrario.

Atributo	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6
BOGOTA-BARRANQUILLA	0	0,003	0,353	1	0,962	0
BOGOTA-BUCARAMANGA	0,997	0,026	1	0	0,952	1
BOGOTA-CALI	0	0	0	0,964	0,001	0,764
BOGOTA-MEDELLIN	0,507	0	0,016	0,996	0,012	0
MEDELLIN-BARRANQUILLA	0,195	0	0,961	0	0,026	0
MEDELLIN-BUCARAMANGA	0,314	0,688	0	0	0	0
MEDELLIN-CALI	0,413	0	0	0	0	0,604
CANTIDADCARGADA	0,124	0,2	0,422	0,12	0,366	0,255
DEPARTAMENTO DESTINO = ANTIOQUIA	0,453	0,027	0,276	0	0	0,076
DEPARTAMENTO DESTINO = ATLANTICO	0,021	0	0,079	0	0,141	0
DEPARTAMENTO DESTINO = BOGOTA D, C,	0,146	0,079	0	0,615	0,173	0,155
DEPARTAMENTO DESTINO = BOLIVAR	0,005	0	0,288	0	0,185	0
DEPARTAMENTO DESTINO = BOYACA	0	0,043	0	0,002	0,041	0
DEPARTAMENTO DESTINO = CALDAS	0,039	0	0	0	0	0,029
DEPARTAMENTO DESTINO = CESAR	0	0	0,022	0	0,041	0
DEPARTAMENTO DESTINO = CORDOBA	0,006	0	0,132	0	0	0
DEPARTAMENTO DESTINO = CUNDINAMARCA	0,025	0,008	0	0,144	0,331	0,073
DEPARTAMENTO DESTINO = MAGDALENA	0,006	0	0,102	0	0,037	0
DEPARTAMENTO DESTINO = META	0,012	0,003	0	0,237	0,041	0,006
DEPARTAMENTO DESTINO = NORTE DE SANTANDER	0,015	0,105	0	0	0,001	0
DEPARTAMENTO DESTINO = QUINDIO	0,013	0	0	0	0	0,059
DEPARTAMENTO DESTINO = RISARALDA	0,063	0	0	0	0	0,071
DEPARTAMENTO DESTINO = SANTANDER	0,066	0,734	0	0,002	0,008	0
DEPARTAMENTO DESTINO = SUCRE	0,003	0	0,085	0	0,001	0
DEPARTAMENTO DESTINO = TOLIMA	0,013	0	0,016	0	0	0,044
DEPARTAMENTO DESTINO = VALLE DEL CAUCA	0,113	0	0	0	0	0,486
DEPARTAMENTO ORIGEN = ANTIOQUIA	0,752	0,002	0,213	0	0	0
DEPARTAMENTO ORIGEN = ATLANTICO	0	0	0,435	0	0,106	0
DEPARTAMENTO ORIGEN = BOGOTA D, C,	0,098	0,097	0	0,279	0,219	0,101
DEPARTAMENTO ORIGEN = BOLIVAR	0	0	0,263	0	0,14	0
DEPARTAMENTO ORIGEN = BOYACA	0	0,023	0	0	0,012	0
DEPARTAMENTO ORIGEN = CALDAS	0,037	0	0	0	0	0,005
DEPARTAMENTO ORIGEN = CESAR	0	0	0	0	0,06	0
DEPARTAMENTO ORIGEN = CUNDINAMARCA	0,07	0,038	0,016	0,712	0,073	0,064
DEPARTAMENTO ORIGEN = MAGDALENA	0	0	0,02	0	0,065	0
DEPARTAMENTO ORIGEN = META	0	0	0	0,008	0,011	0,001
DEPARTAMENTO ORIGEN = NORTE DE SANTANDER	0,004	0,017	0	0	0,004	0
DEPARTAMENTO ORIGEN = QUINDIO	0,005	0	0	0	0	0,022
DEPARTAMENTO ORIGEN = RISARALDA	0,022	0	0	0	0	0,056
DEPARTAMENTO ORIGEN = SANTANDER	0,011	0,821	0	0	0,042	0
DEPARTAMENTO ORIGEN = SUCRE	0	0	0,047	0	0,267	0

FIGURA 9.6. Mapa de calor parte 1

DEPARTAMENTO ORIGEN = TOLIMA	0	0	0	0	0	0,01
DEPARTAMENTO ORIGEN = VALLE DEL CAUCA	0	0	0	0	0	0,741
EMPAQUE = BULTO	0,226	0,254	0,13	0,042	0,335	0,118
EMPAQUE = CARGA ESTIBADA	0,067	0,125	0,416	0,661	0,045	0,139
EMPAQUE = CILINDROS	0,014	0,004	0,008	0,002	0,009	0,004
EMPAQUE = CONTENEDOR 20 PIES	0,015	0	0,048	0,003	0,028	0,044
EMPAQUE = CONTENEDOR 40 PIES	0,03	0,002	0,141	0,008	0,075	0,124
EMPAQUE = GRANEL LIQUIDO	0,007	0,03	0,011	0,127	0,159	0,066
EMPAQUE = GRANEL SOLIDO	0,028	0,013	0,104	0,031	0,19	0,087
EMPAQUE = NO APLICA	0,012	0,001	0,004	0,002	0,005	0,008
EMPAQUE = PAQUETES	0,473	0,487	0,072	0,107	0,108	0,354
EMPAQUE = VARIOS	0,128	0,083	0,064	0,019	0,044	0,054
SECTORES = AGRICULTURA	0,032	0,013	0,08	0,007	0,094	0,138
SECTORES = ALIMENTOANIMALES	0,016	0,18	0,005	0,019	0,004	0,019
SECTORES = ALIMENTOSYBEBIDAS	0,132	0,177	0,272	0,44	0,058	0,243
SECTORES = AUTOMOTRIZ	0,159	0,01	0,101	0,017	0,032	0,06
SECTORES = AVICULTURA	0,017	0,266	0,011	0,009	0,005	0,118
SECTORES = CONSTRUCCION	0,04	0,012	0,092	0,036	0,337	0,045
SECTORES = CONTENEDORES	0,039	0,007	0,111	0,005	0,053	0,073
SECTORES = COSMETICOSYASEO	0,047	0,008	0,01	0,007	0,008	0,006
SECTORES = ENVASESYEMPAQUES	0,069	0,097	0,19	0,288	0,036	0,057
SECTORES = METALURGICA	0,022	0,008	0,033	0,005	0,028	0,018
SECTORES = MINERIA	0,003	0,004	0,004	0	0,075	0,009
SECTORES = MUEBLESYELECTRODOMESTICOS	0,017	0,012	0,008	0,001	0,014	0,011
SECTORES = PETROLERA	0,006	0,026	0,007	0,12	0,145	0,011
SECTORES = VARIOS	0,383	0,173	0,061	0,037	0,089	0,156
SECTORES = VIDRIOSYCERAMICAS	0,008	0,005	0,006	0,004	0,016	0,026
TRAMO = ARMENIA-DOSQUEBRADAS	0,001	0	0	0	0	0,012
TRAMO = ARMENIA-GUADALAJARADEBUGA	0	0	0	0	0	0,036
TRAMO = ARMENIA-YUMBO	0	0	0	0	0	0,01
TRAMO = BARRANCABERMEJA-BUCARAMANGA	0	0,05	0	0	0	0
TRAMO = BARRANCABERMEJA-CARTAGENA	0	0	0	0	0,032	0
TRAMO = BELLO-CARTAGENA	0	0	0,018	0	0	0
TRAMO = BELLO-ENVIGADO	0,011	0	0	0	0	0
TRAMO = BELLO-RIONEGRO	0,011	0	0	0	0	0
TRAMO = BELLO-SANTAMARTA	0	0	0,009	0	0	0
TRAMO = BOGOTA-BARRANQUILLA	0	0	0	0	0,093	0
TRAMO = BOGOTA-BELLO	0,013	0	0	0	0	0
TRAMO = BOGOTA-BUCARAMANGA	0	0,064	0	0	0	0
TRAMO = BOGOTA-BUENAVENTURA	0	0	0	0	0	0,135
TRAMO = BOGOTA-CAJICA	0	0	0	0,028	0,003	0
TRAMO = BOGOTA-CALI	0	0	0	0	0	0,052

FIGURA 9.7. Mapa de calor parte 2

TRAMO = BOGOTA-CARTAGENA	0	0	0	0	0,214	0
TRAMO = BOGOTA-CHIA	0	0	0	0,017	0	0
TRAMO = BOGOTA-COTA	0	0	0	0,012	0	0
TRAMO = BOGOTA-CUCUTA	0	0,015	0	0	0,002	0
TRAMO = BOGOTA-ENVIGADO	0,049	0	0	0	0	0
TRAMO = BOGOTA-FLORIDABLANCA	0	0,041	0	0	0	0
TRAMO = BOGOTA-FUNZA	0	0	0	0,051	0	0
TRAMO = BOGOTA-GIRON	0	0,016	0	0	0	0
TRAMO = BOGOTA-IBAGUE	0	0	0	0	0	0,015
TRAMO = BOGOTA-ITAGUI	0,025	0	0	0	0	0
TRAMO = BOGOTA-MEDELLIN	0,131	0	0	0	0	0
TRAMO = BOGOTA-MOSQUERA	0	0	0	0,027	0	0
TRAMO = BOGOTA-RIONEGRO	0,003	0,013	0	0	0	0
TRAMO = BOGOTA-SABANETA	0,015	0	0	0	0	0
TRAMO = BOGOTA-SANTAMARTA	0	0	0	0	0,025	0
TRAMO = BOGOTA-TOCANCIPA	0	0	0	0,59	0	0
TRAMO = BOGOTA-TOLUVIEJO	0	0	0	0	0,01	0
TRAMO = BOGOTA-TUNJA	0	0	0	0	0,009	0
TRAMO = BOGOTA-VILLAVICENCIO	0	0	0	0,169	0	0
TRAMO = BOGOTA-YUMBO	0	0	0	0	0	0,015
TRAMO = BOGOTA-ZIPAQUIRA	0	0	0	0	0,012	0
TRAMO = BUENAVENTURA-CALI	0	0	0	0	0	0,045
TRAMO = BUENAVENTURA-MEDELLIN	0,03	0	0	0	0	0,023
TRAMO = BUENAVENTURA-PALMIRA	0	0	0	0	0	0,019
TRAMO = BUENAVENTURA-PEREIRA	0	0	0	0	0	0,012
TRAMO = BUENAVENTURA-YUMBO	0	0	0	0	0	0,018
TRAMO = CAJICA-SANTAMARTA	0	0	0	0	0,013	0
TRAMO = CAJICA-SOGAMOSO	0	0	0	0	0,009	0
TRAMO = CAJICA-TOLUVIEJO	0	0	0	0	0,211	0
TRAMO = CALDAS-ENVIGADO	0,025	0	0	0	0	0
TRAMO = CALI-ENVIGADO	0,01	0	0	0	0	0
TRAMO = CALI-MEDELLIN	0,028	0	0	0	0	0,013
TRAMO = CALI-PEREIRA	0	0	0	0	0	0,01
TRAMO = CARTAGENA-BARRANQUILLA	0	0	0,101	0	0,001	0
TRAMO = CARTAGENA-TOLUVIEJO	0	0	0	0	0,014	0
TRAMO = CHIA-ENVIGADO	0,012	0	0	0	0	0
TRAMO = COTA-MEDELLIN	0,011	0	0	0	0	0
TRAMO = CUCUTA-BUCARAMANGA	0	0,073	0	0	0	0
TRAMO = DOSQUEBRADAS-MANIZALES	0,029	0	0	0	0	0,013
TRAMO = ENVIGADO-CARTAGENA	0,001	0	0,099	0	0	0
TRAMO = ENVIGADO-MEDELLIN	0,06	0	0	0	0	0
TRAMO = FLORIDABLANCA-BARRANCABERMEJA	0	0,023	0	0	0	0

FIGURA 9.8. Mapa de calor parte 3

TRAMO = FLORIDABLANCA-LEBRIJA	0	0,058	0	0	0	0
TRAMO = FUNZA-BUENAVENTURA	0	0	0	0	0	0,012
TRAMO = FUNZA-MEDELLIN	0,012	0	0	0	0	0
TRAMO = FUNZA-SANTAMARTA	0	0	0	0	0,022	0
TRAMO = GIRARDOTA-CARTAGENA	0	0	0,031	0	0	0
TRAMO = GIRON-BARRANCABERMEJA	0	0,028	0	0	0	0
TRAMO = GIRON-BUCARAMANGA	0	0,102	0	0	0	0
TRAMO = GIRON-CUCUTA	0	0,018	0	0	0	0
TRAMO = GIRON-RIONEGRO	0	0,011	0	0	0	0
TRAMO = GUADALAJARADEBUGA-BUENAVENTURA	0	0	0	0	0	0,056
TRAMO = GUADALAJARADEBUGA-CALI	0	0	0	0	0	0,023
TRAMO = GUADALAJARADEBUGA-PALMIRA	0	0	0	0	0	0,016
TRAMO = GUADALAJARADEBUGA-PEREIRA	0	0	0	0	0	0,038
TRAMO = HONDA-MEDELLIN	0,012	0	0	0	0	0
TRAMO = IBAGUE-YUMBO	0	0	0	0	0	0,018
TRAMO = ITAGUI-CARTAGENA	0	0	0,027	0	0	0
TRAMO = ITAGUI-MEDELLIN	0,03	0	0	0	0	0
TRAMO = LEBRIJA-BUCARAMANGA	0	0,086	0	0	0	0
TRAMO = LOSSANTOS-BUCARAMANGA	0	0,065	0	0	0	0
TRAMO = LOSSANTOS-FLORIDABLANCA	0	0,044	0	0	0	0
TRAMO = LOSSANTOS-GIRON	0	0,035	0	0	0	0
TRAMO = MANIZALES-MEDELLIN	0,018	0	0	0	0	0
TRAMO = MEDELLIN-BARRANQUILLA	0,012	0	0,055	0	0	0
TRAMO = MEDELLIN-BUCARAMANGA	0,023	0,003	0	0	0	0
TRAMO = MEDELLIN-CARTAGENA	0,004	0	0,091	0	0	0
TRAMO = MEDELLIN-CUCUTA	0,01	0	0	0	0	0
TRAMO = MEDELLIN-FLORIDABLANCA	0,012	0,009	0	0	0	0
TRAMO = MEDELLIN-MONTERIA	0,006	0	0,066	0	0	0
TRAMO = MEDELLIN-RIONEGRO	0,059	0,003	0	0	0	0
TRAMO = MEDELLIN-TOLUVIEJO	0	0	0,01	0	0	0
TRAMO = MONTERIA-BARRANQUILLA	0	0	0,04	0	0	0
TRAMO = MONTERIA-CARTAGENA	0	0	0,013	0	0	0
TRAMO = MONTERIA-TOLUVIEJO	0	0	0,014	0	0	0
TRAMO = MOSQUERA-FUNZA	0	0	0	0,015	0	0
TRAMO = MOSQUERA-SANTAMARTA	0	0	0	0	0,013	0
TRAMO = PALMIRA-YUMBO	0	0	0	0	0	0,01
TRAMO = PEREIRA-MEDELLIN	0,019	0	0	0	0	0
TRAMO = PIEDECUESTA-GIRON	0	0,01	0	0	0	0
TRAMO = RIONEGROA-MEDELLIN	0,019	0	0	0	0	0
TRAMO = RIONEGRO-BARRANCABERMEJA	0	0,014	0	0	0	0
TRAMO = RIONEGRO-BUCARAMANGA	0	0,016	0	0	0	0
TRAMO = SANGIL-BUCARAMANGA	0	0,019	0	0	0	0

FIGURA 9.9. Mapa de calor parte 4

Descripción del cluster 1: el cluster está compuesto por viajes que incluye el transporte por el corredor Bogotá-Bucaramanga y las intersecciones entre los corredores Bogotá-Medellín, Medellín- Cali, Medellín-Bucaramanga, Medellín Barranquilla. Tiene una fuerte presencia de Antioquia como origen y destino de la carga y en una menor proporción de Bogotá y Valle del Cauca. El transporte en su mayoría es de paqueteo y en una menor

proporción en bultos. Incluye transporte de sectores varios, automotriz y alimentos y bebidas, ver figura 9.10b.

De acuerdo con los resultados arrojados por el algoritmo, este cluster de transporte esta formado principalmente por la carga movilizada en su gran mayoría al interior del departamento de Antioquia y corresponde en su gran mayoría al transporte de paquetes de mercancía varia, carros y alimentos y bebidas.

(A) Mapa

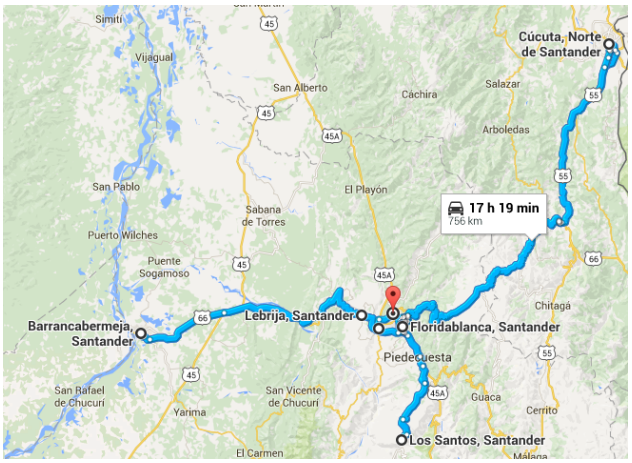
Atributo	Centroide
Bogotá-Bucaramanga	0,997
Departamento origen = Antioquia	0,752
Bogotá-Medellín	0,507
Empaque = Paquetes	0,473
Departamento destino = Antioquia	0,453
Medellín-Cali	0,413
Sectores = Varios	0,383
Medellín-Bucaramanga	0,314
Empaque = Bulto	0,226
Medellín-Barranquilla	0,195
Sectores = Automotriz	0,159
Departamento destino = Bogotá D.C.	0,146
Sectores = Alimentos y bebidas	0,132
Tramo = Bogotá-Medellín	0,131
Empaque = Varios	0,128
Cantidad cargada	0,124
Departamento destino = Valle del Cauca	0,113
Departamento origen = Bogotá D.C.	0,098

(B) Tramos principales

FIGURA 9.10. Mapa cluster 1

Descripción del cluster 2: el cluster tiene como principal departamento de origen y destino el departamento de Santander y en una menor proporción destinos como los departamentos de Norte de Santander, Bogotá D.C. y Boyacá. El cluster tiene tambien una importante participación dentro de la movilización de carga realizada por el corredor Medellín-Bucaramanga.

(A) Mapa



(B) Tramos principales

FIGURA 9.11. Mapa cluster 2

El empaque de la carga transportada principalmente corresponde a paquetes y bultos, y en una menor proporción a carga estibada. Moviliza una importante cantidad de carga de sectores como el avícola, alimento para animales, alimentos y bebidas y varios. Los tramos más representativos para la movilización de la carga son: Girón-Bucaramanga, Lebrija-Bucaramanga, Cúcuta-Bucaramanga, Los Santos-Bucaramanga, Bogotá-Bucaramanga, Floridablanca-Lebrija, Barrancabermeja-Bucaramanga, Los Santos-Floridablanca y Bogotá-Floridablanca, ver figura 9.12.

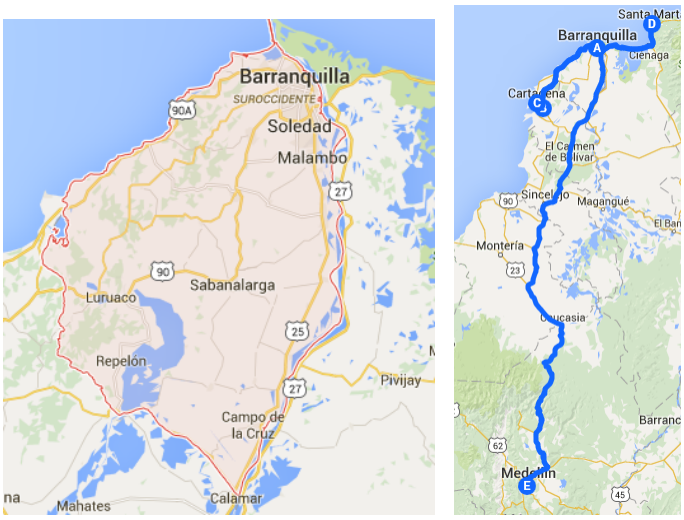
El análisis de los resultados permite inferir que el segundo clúster de transporte está conformado principalmente por el comercio interno desarrollado en el departamento de Santander y se caracteriza principalmente por el transporte de carga avícola y el transporte de alimento para animales, muy probablemente asociado al primero. Los principales municipios generadores y receptores de carga de este cluster corresponden a Girón, Lebrija, Los Santos y Bucaramanga.

Atributo	Centroide
Departamento origen = Santander	0,821
Departamento destino = Santander	0,734
Medellín-Bucaramanga	0,688
Empaque = Paquetes	0,487
Sectores = Avicultura	0,266
Empaque = Bulto	0,254
Cantidad cargada	0,2
Sectores = Alimento animales	0,18
Sectores = Alimentos y bebidas	0,177
Sectores = Varios	0,173
Empaque = Carga estibada	0,125
Departamento destino = Norte de Santander	0,105
Tramo = Girón-Bucaramanga	0,102
Sectores = Envases y empaques	0,097
Departamento origen = Bogotá D.C.	0,097
Tramo = Lebrija-Bucaramanga	0,086
Empaque = Varios	0,083
Departamento destino = Bogotá D.C.	0,079
Tramo = Cúcuta-Bucaramanga	0,073
Tramo = Los Santos-Bucaramanga	0,065
Tramo = Bogotá-Bucaramanga	0,064
Tramo = Floridablanca-Lebrija	0,058
Tramo = Barrancabermeja-Bucaramanga	0,05
Tramo = Los Santos-Floridablanca	0,044
Departamento destino = Boyacá	0,043
Tramo = Bogotá-Floridablanca	0,041

FIGURA 9.12. Cluster 2

Descripción del cluster 3: el cluster moviliza carga principalmente por el corredor Bogotá-Bucaramanga y Medellín-Barranquilla, y en una menor proporción por el corredor Bogotá-Barranquilla. El departamento de origen de la carga es principalmente el Atlántico, seguido en una menor proporción por Bolívar y Antioquia. Se caracteriza por altos volúmenes de carga movilizada de manera estibada, contenedores de 40 pies, bultos y granel sólido. Los principales departamentos de destino son Bolívar, Antioquia y en una menor proporción Córdoba y Magdalena. La carga movilizada corresponde principalmente al sector de

alimentos y bebidas, envases y empaques, carga contenerizada y automotriz. La carga es movilizada principalmente por los tramos Turbaco-Barranquilla, Cartagena-Barranquilla, Envigado-Cartagena, Santa Marta-Barranquilla, Medellín-Cartagena, ver figura 9.14.



(A) Mapa (B) Tramos principales

FIGURA 9.13. Mapa cluster 3

Atributo	Centroide
Bogotá-Bucaramanga	1
Medellín-Barranquilla	0,961
Departamento origen = Atlántico	0,435
Cantidad cargada	0,422
Empaque = Carga estibada	0,416
Bogotá-Barranquilla	0,353
Departamento destino = Bolívar	0,288
Departamento destino = Antioquia	0,276
Sectores = Alimentos y bebidas	0,272
Departamento origen = Bolívar	0,263
Departamento origen = Antioquia	0,213
Sectores = Envases y empaques	0,19
Empaque = Contenedor 40	0,141
Departamento destino = Córdoba	0,132
Empaque = Bulto	0,13
Tramo = Turbaco-Barranquilla	0,127
Sectores = Contenedores	0,111
Empaque = Granel solido	0,104
Departamento destino = Magdalena	0,102
Sectores = Automotriz	0,101
Tramo = Cartagena-Barranquilla	0,101
Tramo = Envigado-Cartagena	0,099
Tramo = Santa Marta-Barranquilla	0,096

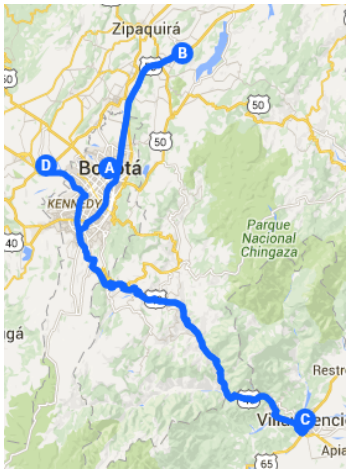
FIGURA 9.14. Cluster 3

El análisis de los resultados permite inferir que el tercer cluster de transporte se caracteriza por el movimiento de carga en el departamento del Atlántico, muy probablemente debido a la existencia del puerto de Barranquilla, la carga generalmente es sólida y

transportada de manera contenerizada. El transporte se realiza principalmente entre las ciudades de Barranquilla, Turbaco, Cartagena, Santa Marta y Medellín.

Descripción del cluster 4: el cluster incluye las intersecciones entre Bogotá-Barranquilla, Bogotá-Medellín y Bogotá-Cali. Tiene como principal generador de carga el departamento de Cundinamarca, seguido por Bogotá D.C. Los principales destinos de la carga son en su orden: Bogotá D.C., departamento del Meta y Cundinamarca.

(A) Mapa



(B) Tramos principales

Atributo	Centroide
Bogotá-Barranquilla	1
Bogotá-Medellín	0,996
Bogotá-Cali	0,964
Departamento origen = Cundinamarca	0,712
Empaque = Carga estibada	0,661
Departamento destino = Bogotá D.C.	0,615
Tramo = Bogotá-Tocancipá	0,59
Sectores = Alimentos y bebidas	0,44
Sectores = Envases y empaques	0,288
Departamento origen = Bogotá D.C.	0,279
Departamento destino = Meta	0,237
Tramo = Bogotá-Villavicencio	0,169
Departamento destino = Cundinamarca	0,144
Empaque = Granel liquido	0,127
Sectores = Petrolera	0,12
Cantidad cargada	0,12
Empaque = Paquetes	0,107
Tramo = Villavicencio-Tocancipá	0,068
Tramo = Bogotá-Funza	0,051

(C) Cluster 4

FIGURA 9.15. Mapa cluster 4

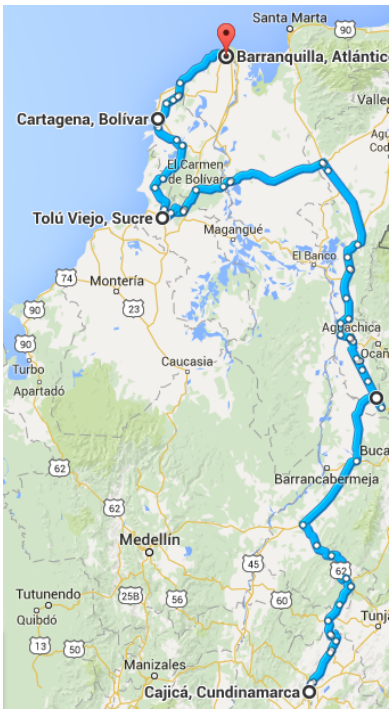
La carga movilizada corresponde principalmente a alimentos y bebidas, seguido de envases y empaques y productos derivados del petróleo. Dicha carga esta empacada principalmente mediante el uso de estibas, contenedores de granel líquido y paquetes. Los tramos mas representativos por donde se moviliza la carga corresponden a Bogotá-Tocancipá, Bogotá-Villavicencio, Villavicencio-Tocancipá y Bogotá- Funza, ver figura 9.15c.

Se puede inferir que el cuarto cluster de transporte tiene como epicentro el comercio de Bogotá con el interior del país, en especial con departamentos como Cundinamarca y Meta, muy probablemente de alimentos y bebidas con el primero y de productos derivados del petróleo con el segundo.

Descripción del cluster 5: el cluster moviliza carga por las intersecciones entre los corredores Bogotá-Barranquilla y Bogotá-Bucaramanga. Los productos transportados pertenecen en gran medida a sectores como: construcción, petrolero y agricultura. El empaque de la carga se realiza en bultos, granel sólido, granel líquido y paquetes. Los departamentos de destino de la carga son Cundinamarca, Bolívar y Bogotá D.C. Los departamentos de origen de la carga son Sucre, Bogotá y Atlántico. Los tramos más representativos son Bogotá-Cartagena, Cajicá-Toluviejo y Bogotá-Barranquilla, ver figura 9.16b.

Se puede inferir que el quinto cluster de transporte corresponde a la movilización de carga entre la intersección los corredores Bogotá-Bucaramanga y Bogotá-Barranquilla.

En su gran mayoría carga dirigida al departamento de Cundinamarca asociada al sector construcción y petrolera. El transporte se realiza principalmente entre las ciudades Bogotá, Cartagena, Cajicá, Toluviejo y Barranquilla.



(A) Mapa

Atributo	Centroide
Bogotá-Barranquilla	0,962
Bogotá-Bucaramanga	0,952
Cantidad cargada	0,366
Sectores = Construcción	0,337
Empaque = Bulto	0,335
Departamento destino = Cundinamarca	0,331
Departamento origen = Sucre	0,267
Departamento origen = Bogotá D.C.	0,219
Tramo = Bogotá-Cartagena	0,214
Tramo = Cajicá-Toluviejo	0,211
Empaque = Granel solido	0,190
Departamento destino = Bolívar	0,185
Departamento destino = Bogotá D.C.	0,173
Empaque = Granel liquido	0,159
Sectores = Petrolera	0,145
Departamento destino = Atlántico	0,141
Departamento origen = Bolívar	0,140
Empaque = Paquetes	0,108
Departamento origen = Atlántico	0,106
Sectores = Agricultura	0,094
Tramo = Bogotá-Barranquilla	0,093

(B) Cluster 5

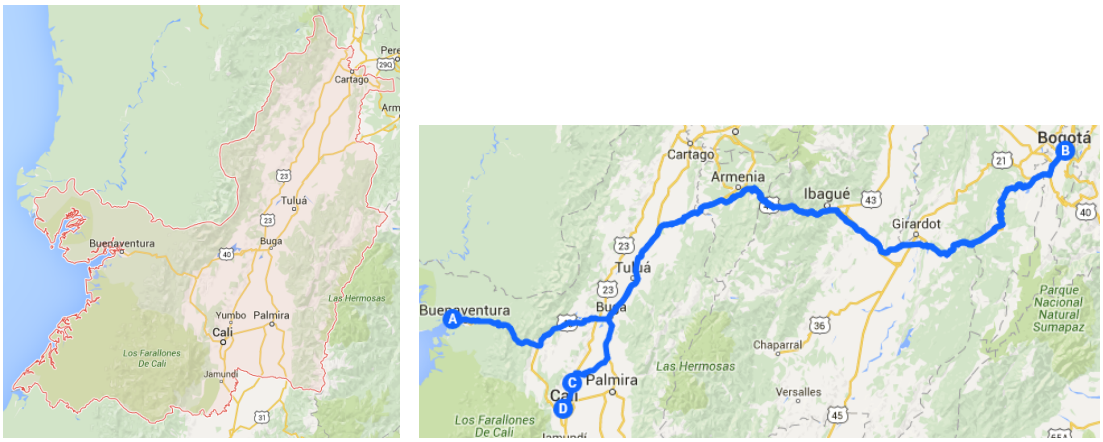
FIGURA 9.16. Mapa cluster 5

Descripción del cluster 6: el cluster moviliza carga por las intersecciones de los corredores Bogotá-Bucaramanga, Bogotá-Cali y Medellín-Cali. El principal departamento de origen de la carga corresponde al Valle del Cauca, seguido de lejos por Bogotá. Entre los departamentos destino de la carga se encuentran en primer lugar departamento del Valle del Cauca, seguido por Bogotá D.C. y Antioquía. La carga es transportada en su gran mayoría en paquetes, seguido de carga estibada, contenedores de 40 pies, bultos y granel sólido. La carga transportada pertenece a sectores como alimentos y bebidas, varios, agricultura, avicultura y contenedores. Los tramos más representativos son Bogotá-Buenaventura y Yumbo-Cali, ver figura 9.17.

El cluster de transporte número seis se caracteriza por el transporte realizado en la zona pacifica del país, relacionado particularmente con el puerto de Buenaventura. El transporte se da principalmente en las vía aledañas a Cali y Buenaventura y corresponde a carga contenerizada y transporte de alimentos.

Atributo	Centroide
Bogotá-Bucaramanga	1,000
Bogotá-Cali	0,764
Departamento origen = Valle del Cauca	0,741
Medellín-Cali	0,604
Departamento destino = Valle del Cauca	0,486
Empaque = Paquetes	0,354
Cantidad cargada	0,255
Sectores = Alimentos y bebidas	0,243
Sectores = Varios	0,156
Departamento destino = Bogotá D.C.	0,155
Empaque = Carga estibada	0,139
Sectores = Agricultura	0,138
Tramo = Bogotá-Buenaventura	0,135
Empaque = Contenedor 40	0,124
Sectores = Avicultura	0,118
Empaque = Bulto	0,118
Departamento origen = Bogotá D.C	0,101
Empaque = Granel solido	0,087
Tramo = Yumbo-Cali	0,082
Departamento destino = Antioquia	0,076
Sectores = Contenedores	0,073

FIGURA 9.17. Cluster 6



(A) Mapa (B) Tramos principales

FIGURA 9.18. Mapa cluster 6

9.1.4. Evaluación de los resultados

El modelo crea un conjunto de clusters de transporte y permite generalizar el comportamiento de los mismos. Su definición se sobrepone con la definición realizada de cada uno de los corredores logísticos. Para su formulación fue necesario la estandarizaron campos que antes se encontraban abiertos para lograr un mayor nivel de generalidad respecto al tipo de carga transportada.

La información generada por el modelo es estratégica, ya que al definir los cluster de transporte permite soportar la toma de decisiones encaminadas a focalizar y fortalecer la

inversión en infraestructura en los puntos carreteros críticos y estratégicos para el país. Permite soportar con información fiable políticas que promuevan el mejoramiento de las condiciones viales por donde movilizan carga los sectores productivos estratégicos para el país, promoviendo mejoras en términos de productividad y competitividad del país. Desde el punto de vista regulatorio se genera la posibilidad de brindar incentivos a los cluster de transporte y productivos para promover la eficiencia de sus procesos logísticos y de transporte.

La información original del RNDC es información operacional y no soporta eficientemente la toma de decisiones para los niveles tácticos y estratégicos. En ese sentido, la información generada por el modelo y los procesos de minería de datos se adapta mejor al tipo de información y al nivel de impacto de la decisión. Además de permitir comparar, analizar, predecir experimentar y generar información de calidad para la entidad, ver figura 9.19.

FIGURA 9.19. Información, nivel de decisión y herramientas para la toma de decisiones. Adaptado:[36]

La adopción de modelos y herramientas similares permitirá desde el punto de vista estratégico al MinTransporte avanzar en la de adopción de herramientas de análisis continuo propuestas por Gartner[56], ver Figura 9.20, el cual tiene beneficios claros en su implementación como lo son la disponibilidad y pertinencia de la información durante el desarrollo de los procesos estratégicos y de planeación.

9.1.5. Revisión de procesos

Algunas de las mejoras que se podrían realizar al modelo son: involucrar otras variables que no se consideraron durante el proceso, mejorar la capacidad de procesamiento para incluir un mayor volumen de datos de entrenamiento y un número mayor de clusters para un mejor nivel de generalización. También se podrían crear modelos derivados para evaluar de manera separada cómo es el comportamiento de cada uno de los clusters.

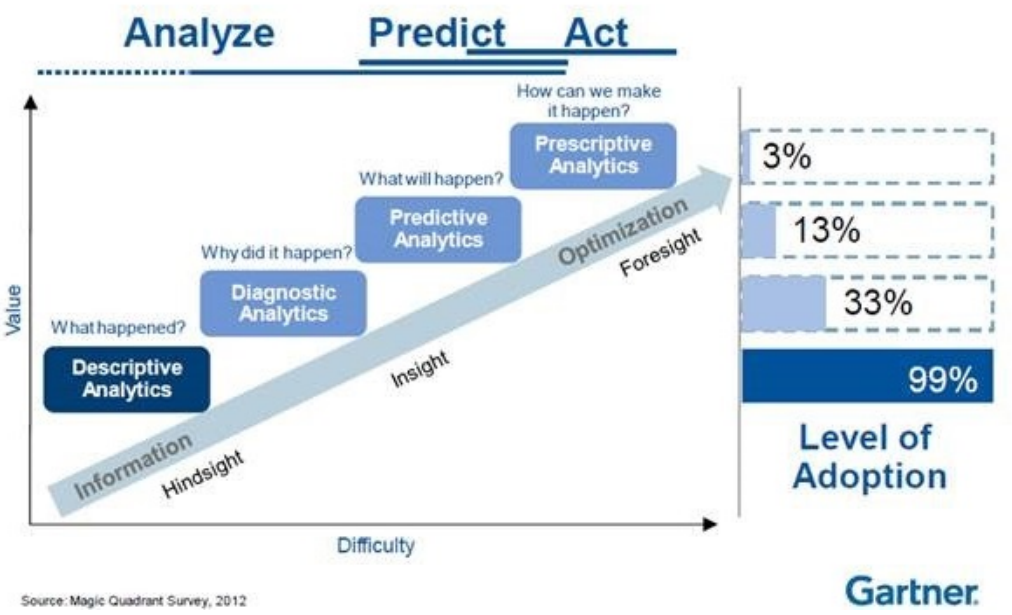


FIGURA 9.20. Modelo de madurez de la adopción de analítica continua.[60]

9.2. Modelo predictivo

9.2.1. Selección de la técnica para el modelo predictivo

Se observa la necesidad de conocer y categorizar el producto transportado adecuadamente es de vital importancia en términos regulatorios, planeación y de normatividad, ya que permitiría acciones como: controles a los vehículos dependiendo del producto a ser transportado, especialización de ciertas calzadas para el transporte de un producto determinado, planes de incentivo para el uso de determinadas calzadas dependiendo del producto transportado, planeación y mejoramiento de la infraestructura existente asociada al transporte de un producto de particular interés para el país, planeación en la construcción de nueva infraestructura, planes de inversión en tramos críticos de los corredores logísticos, planeación y construcción de la infraestructura logística adecuada para soportar de manera eficiente el transporte de determinados productos.

El estado actual del RNDC impide modelar, predecir y describir el comportamiento de la mercancía que se está transportando. Observando esta necesidad se construyó un modelo predictivo por cada una de las actividades productivas que movilizan carga por carretera. Los resultados permitirán hacer estimaciones más acertadas acerca del comportamiento del transporte y análisis más precisos que involucran el producto transportado. Asimismo, también servirá como herramienta para la completar campos faltantes asociados a carga transportada.

Se seleccionó el algoritmo de árbol C45, teniendo en cuenta que es uno de los algoritmos más difundidos y que presenta una mayor precisión para el dataset. Su carácter es mixto, por lo tanto permite realizar análisis descriptivos y a la vez inferencias predictivas.

9.2.2. Generación del test de diseño

Debido a que se trata de una tarea de clasificación, se estableció una tasa de error para los modelos que no sea superior al 15 % de los datos del dataset de prueba. El dataset de prueba corresponde a una muestra estratificada y la calidad de los resultados se verifica mediante una validación con una muestra del 30 % de los datos.

9.2.3. Construcción del modelo de árbol de decisión

Se evaluó árboles para todas las actividades productivas con una profundidad máxima de 20 niveles. Los modelos se desarrollaron en tres etapas: la primera de preprocesamiento, la segunda de entrenamiento, y una tercera de validación, ver figura 9.21.

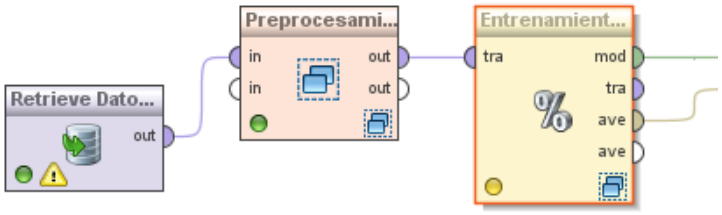


FIGURA 9.21. Modelamiento árbol de decisión

En la etapa de preprocesamiento se seleccionan los atributos de entrada para el entrenamiento, se realizó un filtrado para eliminar registros con datos faltantes, teniendo en cuenta que no afectará en gran medida el dataset. Se realizó una binarización de los campos nominales y una normalización y limpieza de outliers de los campos numéricos. Por último se realiza una conversión del campo tipo de empaque codificado en formato numérico, para que sea tomado como un campo de tipo nominal y se asigna el rol de atributo clase al campo actividad productiva, ver figura 9.22.

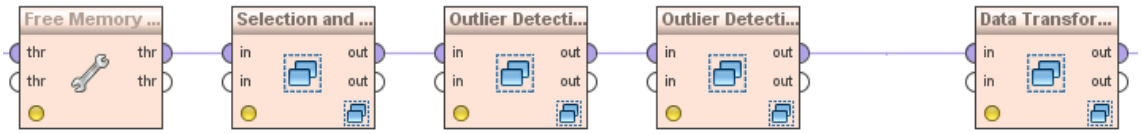


FIGURA 9.22. Preprocesamiento árbol de decisión

En la segunda etapa se realiza la parametrización y construcción del árbol de decisión utilizando el 70 % de los datos de dataset. Por último, en la tercera etapa se realiza la aplicación del modelo y la validación utilizando el 30 % de los datos del dataset, ver figura 9.23.

9.2.3.1. Datos de entrada

Como datos de entrada al modelo se ingresaron 100.215 registros en donde se incluyen los campos: corredores logísticos, cantidad cargada, departamento de origen, departamento de destino, actividad productiva, tipo de empaque y tramo, ver tabla 9.2.

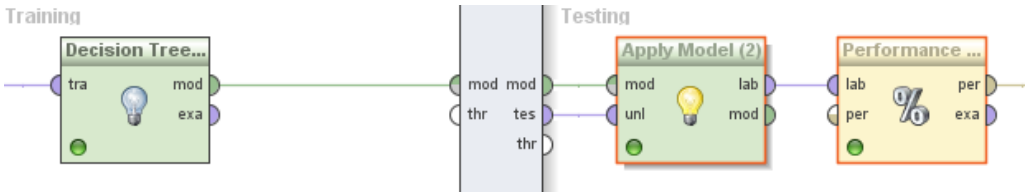


FIGURA 9.23. Modelamiento y validación del árbol de decisión

Atributo	Rango de valores	Tipo de variable	Significado
Bogotá-Barranquilla	true, false	Binomial	Indica si el viaje circula por el corredor Bogotá-Barranquilla
Bogotá-Bucaramanga	true, false	Binomial	Indica si el viaje circula por el corredor Bogotá-Bucaramanga
Bogotá-Cali	true, false	Binomial	Indica si el viaje circula por el corredor Bogotá-Cali
Bogotá-Medellín	true, false	Binomial	Indica si el viaje circula por el corredor Bogotá-Medellín
Medellín-Barranquilla	true, false	Binomial	Indica si el viaje circula por el corredor Medellín-Barranquilla.
Medellín-Bucaramanga	true, false	Binomial	Indica si el viaje circula por el corredor Medellín-Bucaramanga.
Medellín-Cali	true, false	Binomial	Indica si el viaje circula por el corredor Medellín-Cali.
Cantidad cargada	Valores normalizados entre 0 y 1	Numérica	Indica cual es la cantidad de carga transportada en cada viaje.
Departamento origen	Enteros positivos	Nominal	Indica cual es el departamento de origen de la carga.
Departamento destino	Enteros positivos	Nominal	Indica cual es el departamento de destino de la carga.
Tipo de empaque	Valores en el intervalo entre (0,19).	Nominal	Indica cual es el tipo de empaque de la carga.
Alimentos y bebidas	true, false	Binomial	Indica si el viaje transporta alimentos y bebidas
Petrolera	true, false	Binomial	Indica si el viaje transporta productos del petróleo
Envases y empaques	true, false	Binomial	Indica si el viaje transporta envases y empaques
Agricultura	true, false	Binomial	Indica si el viaje transporta productos agrícolas
Avicultura	true, false	Binomial	Indica si el viaje transporta productos avícolas
Construcción	true, false	Binomial	Indica si el viaje transporta elementos asociados a la actividad de la construcción
Automotriz	true, false	Binomial	Indica si el viaje transporta elementos asociados a la actividad automotriz
Contenedores	true, false	Binomial	Indica si el viaje transporta productos contenerizados

Alimento animales	true, false	Binomial	Indica si el viaje transporta alimentos para animales
Minería	true, false	Binomial	Indica si el viaje transporta productos de la actividad minera
Tramo	Origen-Destino	Polinominal	Indica a cual tramo pertenece el viaje. Está compuesto por la ciudad de origen y destino de la carga.

Tabla 9.2. Datos de entrada árbol de decisión

9.2.3.2. Configuración de los parametros

La configuración e interpretación de los parámetros del modelo se presentan en la tabla 9.3.

Parámetro	Valor	Explicación
Criterio	Ganancia de información	Corresponde al criterio seleccionado para dividir cada una de las ramas del árbol. El criterio de ganancia de información consiste en la selección de los atributos con la mínima entropía calculada para que sean divididos.
Máxima profundidad	20	Consiste en el número máximo de niveles que puede tener el árbol.
Aplicar poda	Si	Aplica una poda al árbol generado para evitar el sobre ajuste a los datos.
Confianza	0.25	Especifica el nivel de confianza usado para el cálculo pesimista del error durante la poda.
Aplicar prepoda	Si	Especifica si se aplicará o no prepoda al árbol de decisión.
Ganancia mínima	0.1	Consiste en la ganancia de información calculada antes de dividir el nodo. El nodo es dividido si la ganancia es mayor que la mínima ganancia. Un alto valor de la ganancia mínima resulta en una menor división de los nodos y por lo tanto un árbol más pequeño.
Tamaño mínimo de la hoja	2	El tamaño de un nodo hoja es el número de ejemplos en el subconjunto. El árbol es generado de tal manera que cada subconjunto de nodo hoja tiene al menos un mínimo número de instancias para el tamaño de la hoja.
Tamaño mínimo para dividir	4	El tamaño del nodo es el número de ejemplos en su conjunto. El tamaño del nodo raíz es igual al total del número de ejemplos en el conjunto de ejemplo. Solamente se puede dividir cuando esa talla es mayor o igual al mínimo tamaño para dividir.
Número de prepodas	10	Como la prepoda se ejecuta paralelo al proceso de generación del árbol, esta puede prevenir la división en ciertos nodos cuando la división no agrega un poder discriminativo al árbol entero. Este parámetro ajusta el número de veces que se aplica la prepoda.

Tabla 9.3. Configuración de parámetros árbol de decisión

9.2.4. Resultado del modelo

Los modelos que mejor desempeño presentaron son los relacionados a la actividad petrolera, contenedores, minería y envases y empaques, ver tabla 9.5. Como se evidencia en la tabla 9.4 los resultados arrojados por los arboles de decisión son bastante buenos, un ajuste a cada modelo puede potenciar aún más la capacidad predictiva de cada uno.

El modelo que mejor ejemplifica la capacidad predictiva de los árboles de decisión la actividad petrolera. Esta actividad es uno de los renglones principales de la economía colombiana y presentó una de las mejores capacidades predictivas. La exactitud para determinar si un viaje pertenecía a la actividad petrolera se predijo con una precisión del 96.19 %, en cuanto a su capacidad para determinar que no pertenencia, presento una exactitud del 99.79 %, ver tabla 9.4. Para consultar la evaluación realizada a las otras actividades ver Anexo 4.

El árbol presentado en la figura 9.24 muestra que la actividad petrolera está fuertemente asociada con los empaques para el transporte de líquido a granel. La cantidad de galones transportados generalmente es superior a 13.046 galones y siendo los tramos Santa Marta- Cartagena, Buga-Palmira, Medellín-Santa Marta, Girardota- Santa Marta y Medellín-Bucaramanga algunos de los más representativos respecto a la movilización. Cuando la cantidad de carga transportada es menor a 13.046 galones, es necesario verificar si el viaje no pertenece a otras actividades productivas como alimentos y bebidas, automotriz, minería, agricultura y químicos.

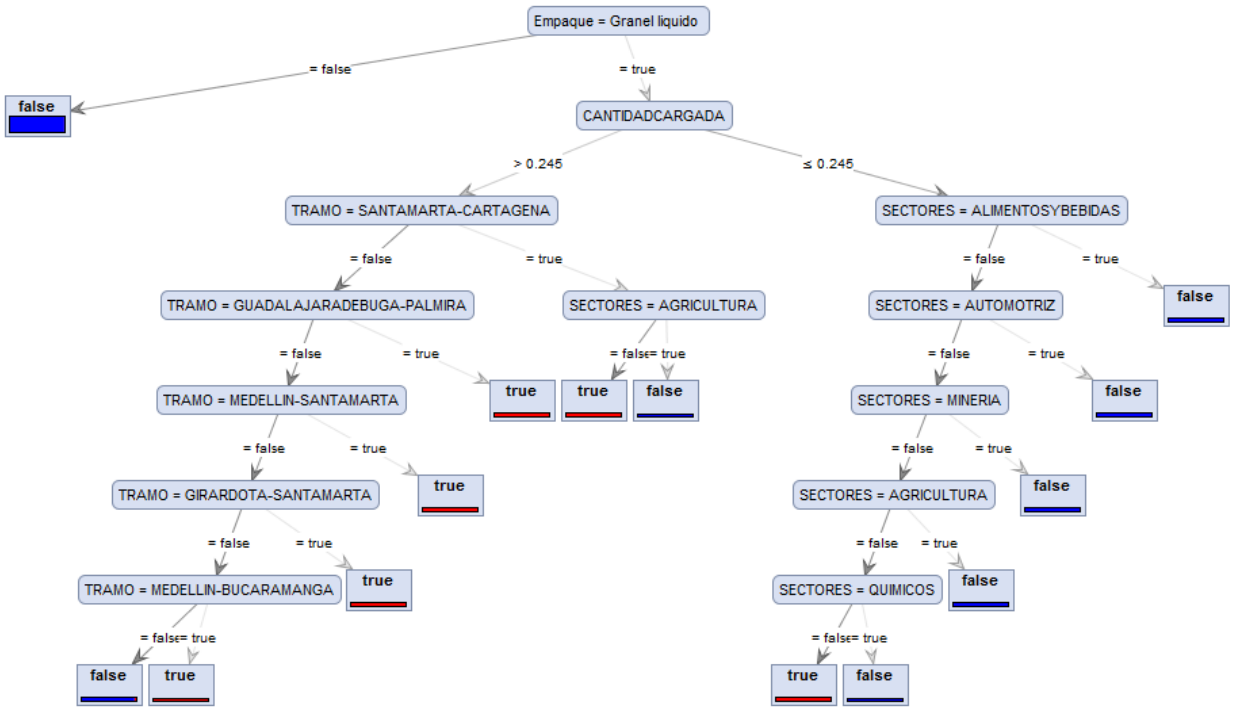


FIGURA 9.24. Árbol de decisión de la actividad petrolera

Con el fin de determinar la precisión adecuada para el árbol de la actividad petrolera, se realizó una evaluación con diferentes niveles de profundidad, y como resultado se determinó que 8 era la profundidad que mejor se ajustaba entre la capacidad predictiva del

árbol y un número de niveles apropiado para realizar un análisis descriptivo y predictivo, ver figura 9.25

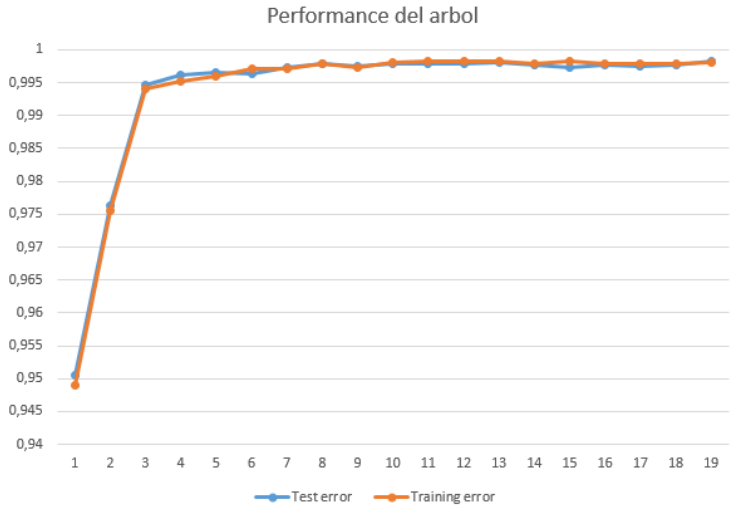


FIGURA 9.25. Performance de predicción del arbol de decisión

9.2.5. Evaluación del modelo

Para la evaluación de los modelos se seleccionó un dataset de prueba mediante muestreo estratificado correspondiente al 30 % de los datos. Como se observa en la tabla 9.4, la exactitud de los modelos disminuye dependiendo del tamaño de la muestra y de la existencia o no de un conjunto de datos balanceado.

	Sin balanceo		Con balanceo		
Actividad productiva	Class precision	Class recall	Class precision	Class recall	Tamaño balanceado
Petrolera	99,79 %	96,19 %	99,25 %	99,29 %	5000
Contenedores	99,39 %	88,56 %	90.94 %	89.70 %	5100
Minería	99,60 %	80,03 %	87.96 %	87.96 %	1800
Envases y empaques	97,54 %	77,28 %	82.82 %	79.17 %	10200
Alimento para animales	98,68 %	68,02 %	97.69 %	97.57 %	3900
Construcción	95,98 %	61,27 %	75.07 %	66.55 %	10000
Avicultura	95,35 %	45,79 %	99.64 %	99.63 %	8200
Automotriz	96,56 %	41,52 %	99.78 %	99.77 %	5900
Cosméticos y aseo	99,21 %	38,16 %	100 %	100 %	1200
Maderas	98,89 %	37,74 %	100 %	100 %	200
Vidrios y cerámicas	99,07 %	37,10 %	100 %	100 %	1300
Agricultura	95,01 %	31,29 %	90.30 %	89.20 %	7400
Alimentos y bebidas	82,55 %	20,24 %	100 %	100 %	20800
Metalurgia	98,25 %	11,47 %	100 %	100 %	1900
Varios	85,38 %	3,85 %	99.69 %	99.69 %	15000
Químicos	99,64 %	1,83 %	98.20 %	98.45	400
Muebles y electrodomésticos	98,95 %	0,32 %	100 %	100 %	1100

TABLA 9.4. Precisión de los arboles por actividad

En el caso de los arboles balanceados con una capacidad de predicción del 100 %, aunque su capacidad es buena, muestran una clara dependencia de su predicción a otras

actividades productivas, como se muestra en la figura 9.26, y a nivel descriptivo no aportan información importante. Por lo tanto una mejora para dichos arboles corresponde a la eliminación de las otras actividades productivas lo cual produce un árbol más equilibrado en términos descriptivos y predictivos, ver figuras 9.27 y 9.28.

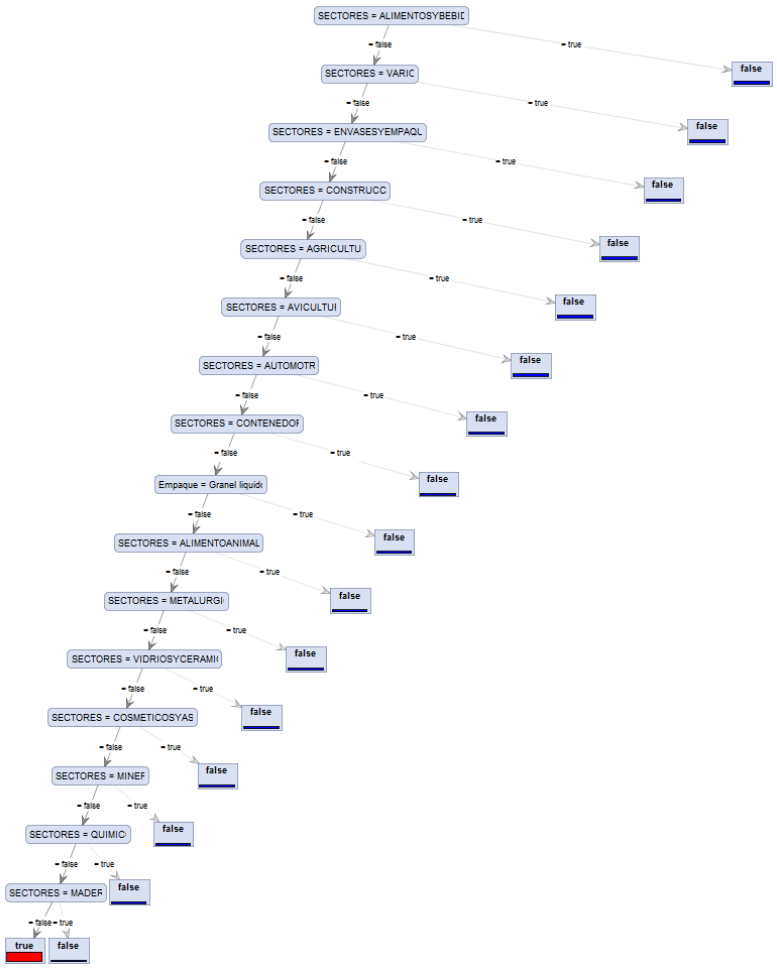


FIGURA 9.26. Árbol de decisión balanceado de la actividad muebles y electrodomésticos

accuracy: 80.67%

	true false	true true	class precision
pred. false	194	8	96.04%
pred. true	108	290	72.86%
class recall	64.24%	97.32%	

FIGURA 9.27. Tabla de contingencia de la mejora al modelo de muebles y electrodomesticos

Como resultado de la construcción de los modelos predictivos, se evidencio que los árboles de decisión son una herramienta apropiada para la toma de decisiones y la planificación estratégica. Estos permiten generalizar en categorías definidas el comportamiento del transporte de carga por carretera con una buena precisión.

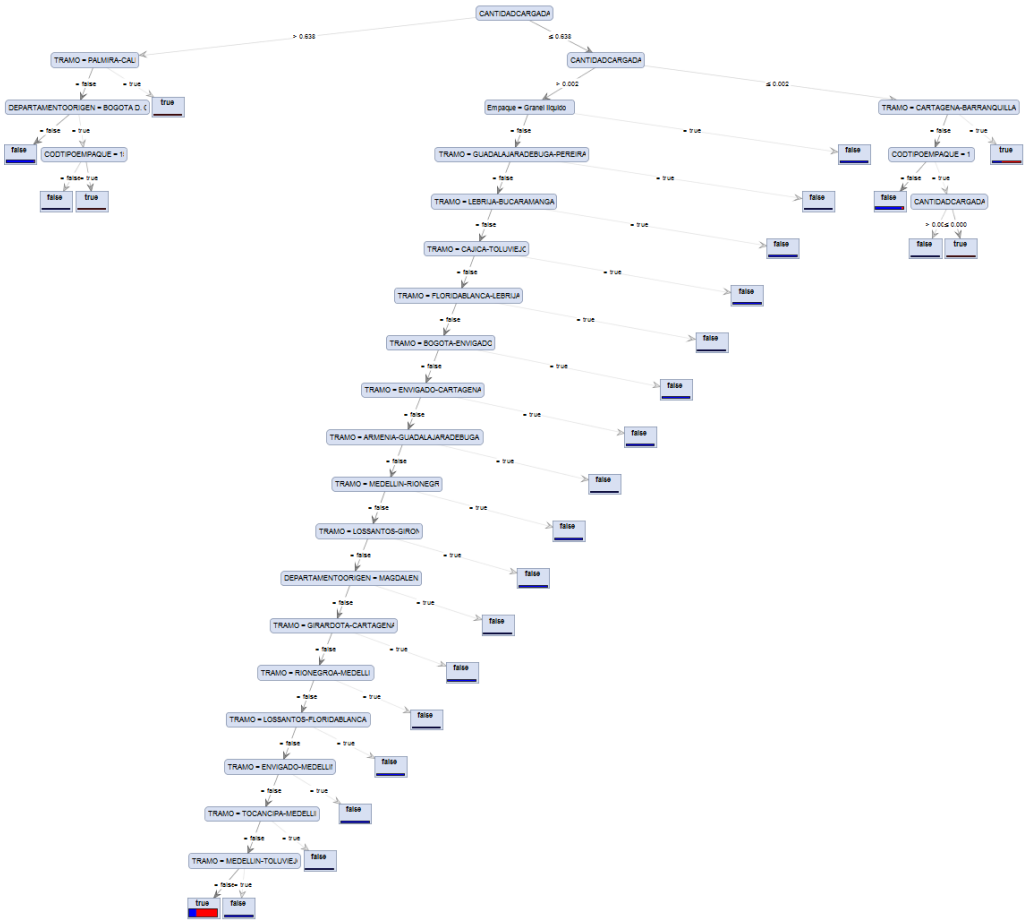


FIGURA 9.28. Árbol de decisión ajustado: actividad muebles y electrodomésticos

La exploración de nuevos árboles que tengan en cuenta otros campos del RNDC, permitirá obtener información importante respecto a la operación del transporte de carga por carretera como: distribución de las actividades productivas por corredor logístico, distribución de las actividades por cada uno de los tramos más transitados, la distribución de actividades productivas por departamentos y municipios, principales tramos utilizados para movilizar determinados tipos de carga, distribución de vehículos utilizados para transportar carga de las actividades productivas, entre otras.

El modelo construido genera nueva información a partir de la existente, permite generalizar el comportamiento de la información gracias a la creación y estandarización de algunos campos y puede ser ejecutado con diferentes datasets y ventanas de tiempo, facilitando la trazabilidad del comportamiento de los sectores productivos y la detección de variaciones inesperadas en la operación del transporte de carga.

La información generada es estratégica, ya que brinda soporte para la toma de decisiones encaminadas a fortalecer la inversión en infraestructura para mejorar o construir nuevas vías que favorezcan las condiciones viales de los principales corredores logísticos, sus tramos críticos y los sectores productivos estratégicos para el país. El modelo facilita las labores de monitoreo y regulación del comportamiento del transporte de carga de determinada actividad productiva, detectando variaciones importantes que requieran una atención prioritaria. A partir del conocimiento de los tramos y corredores más concurri-

dos en términos de volumen y frecuencia, se podrá evaluar donde construir o potenciar los principales centros de intercambio intermodal, especializar o mejorar los centros portuarios de acuerdo a los criterios de oferta y demanda, mejorar los tiempos portuarios y su influencia directa en los costos de transporte de mercancía. A nivel ciudad, permitiría ubicar de manera estratégica centros de cargue y descargue de mercancía dependiendo del volumen y tipo de carga movilizada, minimizando el impacto interno de movilidad para cada una de las ciudades.

	Categoría											
Predicción	Agricultura	Alimento animales	Alimentos bebidas	Automot.	Avicultura	Construc	Contened	Envases empaques	Minería	Petrolera	Varios	Class precision
Agricultura	6325	111	656	176	38	516	0	77	151	41	0	78.17 %
Alimento para animales	92	2411	138	62	500	115	0	15	21	16	0	71.54 %
Alimentos y bebidas	556	130	12073	191	434	906	0	6865	40	123	0	56.63 %
Automotriz	59	47	95	2957	8	189	0	90	14	3	0	85.41 %
Avicultura	124	2	868	305	6162	140	0	821	17	4	0	72.98 %
Construcción	1113	662	873	532	452	13670	0	254	777	255	0	73.54 %
Contenedores	0	0	0	0	0	0	2819	0	0	0	0	100.00 %
Envases y empaques	81	26	84	92	29	146	0	806	5	3	0	63.36 %
Minería	136	27	20	2	0	367	0	14	3274	29	0	84.62 %
Petrolera	60	4	245	12	4	44	0	2	38	11449	0	96.55 %
Varios	0	0	0	0	0	0	0	0	0	0	12605	100.00 %
Class recall	74.01 %	70.50 %	80.21 %	68.31 %	80.79 %	84.94 %	100.00 %	9.01 %	75.49 %	96.02 %	100.00 %	

Tabla 9.5. Tabla de contingencia de la evaluación del modelo de árbol de decisión multiclase

Conclusiones

La investigación realizada se enmarca dentro de un área de conocimiento multidisciplinar, en donde se involucran temas de áreas como: logística, transporte y minería de datos.

Se presentó cuáles han sido los principales métodos, teorías y herramientas de la logística, haciendo énfasis en el componente de sistemas de información, específicamente en los sistemas de minería de datos. Por su parte el estado del arte, mostró el crecimiento que ha tenido la producción científica en torno a la aplicación de minería de datos a logística y el transporte. Se realizó una caracterización en donde se presenta el estado actual de la logística y el transporte, teniendo como ejes principales de análisis: la demografía, la infraestructura, la institucionalidad y la formación académica de las áreas.

Se cumplieron todos los objetivos específicos de la mano con el objetivo general. Primero, construyó un conjunto de datos adecuado para llevar a cabo las tareas de minería de datos, mediante el desarrollo de un modelo de preprocesamiento de datos. Segundo, se construyó un modelo descriptivo de minería de datos, en este caso el modelo clustering. Tercero, se construyó un modelo predictivo del comportamiento del transporte de carga para las actividades productivas, aplicando la técnica de árboles de decisión. Ambos modelos tanto el descriptivo como el predictivo fueron evaluados de manera sistemática para determinar la calidad de la información arrojada, por último, se evaluó la utilidad de los resultados para la formulación de políticas públicas en el sector de transporte de carga terrestre colombiano.

Con los resultados de los modelos se obtuvo información que anteriormente no se disponía, se entendió de una mejor manera como es el comportamiento de los corredores logísticos en Colombia. Se definió con claridad cada uno de los corredores logísticos en términos de los municipios que los componen, carga que movilizan, y principales ciudades generadoras y receptoras de carga.

Por otro lado, la investigación mostró la posibilidad de utilizar nuevas técnicas y metodologías como la minería de datos para el análisis del proceso de transporte de carga por carretera. Los resultados obtenidos del presente ejercicio permiten entender las relaciones estadísticas entre el espacio geográfico y su impacto en la planeación del transporte al poder identificar tramos de vías y los corredores más importantes, para que sean atendidos con proyectos de inversión en infraestructura para su mejoramiento y el mejoramiento de la política de infraestructura vial.

La información generada se convierte en insumo para la planeación en proyectos de desarrollo supraregional, posibilitando que las decisiones sobre la construcción de nueva

infraestructura de transporte estén basadas en criterios sociales o económicos y con el rigor técnico deseable por parte de los modeladores de transporte. Así se disminuye la posibilidad de que las decisiones tomadas sean de carácter político y no responda las necesidades de sectores específicos de alguna región.

Respecto a la pregunta planteada en la problemática, con el desarrollo de la investigación se evidencio, que la minería de datos se convierte en una importante herramienta que puede ser utilizada para optimizar la información de los procesos de decisión estatales, ayudando de manera significativa en la generación de información relevante, depurada y en un tiempo razonable. El uso de información adecuada promueve una correcta toma de decisiones, impactando directamente en una mejor productividad y competitividad del país.

En un contexto gubernamental la información producto de la minería de datos, toma un importante valor en el proceso de formular y adoptar las políticas, planes, programas, proyectos y regulación económica del transporte, el tránsito y la infraestructura del país. Tiene potencial para jugar un papel importante en la puesta en marcha del *observatorio nacional de logística* y en el *plan maestro para sistemas inteligentes de transportes* propuesto por el plan nacional de desarrollo 2010 - 2014, también en el *sistema de información de Infraestructura, Logística y Transporte* propuesto por el CONPES 3527. Por último también se presenta la oportunidad de pertenecer al *sistema de información para el monitoreo y regulación económica del transporte de carga por carretera* propuesto en el CONPES 3489. Todos los anteriores proyectos a la fecha continúan sin ser implementados por parte del gobierno nacional. De igual manera, se presenta como una opción importante de mejora para el RNDC como parte de un componente analítico que permite explotar la información contenida en el RNDC.

Con la inclusión del componente analítico dentro de los sistemas de información presentes y futuros, se generará información de calidad para alimentar el sistema logístico nacional definido en el CONPES 3547, facilitando la administración de los flujos de información y encausando la rentabilidad presente y futura en términos de costos, efectividad en el uso, prestación y facilitación de servicios logísticos y de transporte.

Herramientas analíticas como la minería de datos en conjunto con la integración de sistemas de información de almacenamiento, seguimiento y posicionamiento, sistemas de información web y sistemas conducentes facilitan el análisis de información del comercio exterior, de los procesos de control e inspección de la mercancía y su desaduanaje. También se presenta como una opción para generar, depurar la información y realizar el seguimiento a las tareas de regulación de la renovación de la flota del gremio transportador.

Información de calidad en temas clave como infraestructura, impactan directamente en el proceso de toma de decisiones en cuanto a políticas a nivel nacional, mejorando la productividad y competitividad. El impacto de la minería de datos en relación a la infraestructura, componente fundamental para el desarrollo logístico y transporte, es valioso, ya que su uso adecuado la convierte en una herramienta que genera información de calidad para llevar a cabo procesos de monitoreo, planeación y descubrimiento de información.

A nivel organizacional, las herramientas analíticas facilitan la gestión y la administración de cada una de las operaciones de las empresas y de la red de logística que estas involucran, permitiendo la captura, transferencia y gestión de la información de forma adecuada. Temas como mejoramiento de los tiempos de decisión, número de variables de decisión, posibles opciones o estrategias y disponibilidad de información en tiempo real,

son factores que influyen en la adopción de tecnologías de la información. Estos factores mejoran la capacidad operacional de la organización optimizándola y mejorando los procesos de comunicación en tiempo real, y disminuyendo la incertidumbre y complejidad de algunos problemas para los tomadores de decisiones. Una eficiente planificación de flujos, servicios e información, influyen de manera directa en los costos logísticos de distribución y posicionamiento competitivo de los bienes[41].

Debido a que las tecnologías de información más difundidas por los operadores logísticos corresponden a sistemas de trazabilidad en tiempo real, accesos vía internet para el cliente y sistemas de optimización, planeación y control de transporte, sistemas de tipo operacional, herramientas de tipo analítico se convierten en una opción para realizar la explotación de la información recolectada.

Finalmente, como producto de la investigación, se realizó una ponencia en el segundo congreso Internacional Industria y Organizaciones, el cual se llevó a cabo en el mes de agosto de 2015 en la Universidad Nacional de Colombia.

Trabajo futuro y perspectivas

La investigación desarrollada marca un precedente en el país en torno al estado del arte de las aplicaciones de minería de datos a logística y transporte. Los modelos por su parte, se convierten en una herramienta con gran potencial para ser adaptados a las industrias para la planeación de las rutas más óptimas, zonificación de actividades productivas, como para su integración al sistema de información RNDC por parte del Ministerio de Transporte de Colombia, permitiendo generar información estratégica para la planeación y regulación del transporte de carga por carretera a nivel nacional.

Es necesario resaltar que la minería de datos aplicada al sistema de información RNDC, tiene potenciales aplicaciones para la planeación regional, departamental o regional, para la determinación de las mejores ubicaciones de centros de consolidación de carga y mejores rutas a seguir dentro de cada locación dependiendo del tipo de carga a transportar. Su uso permitiría obtener información base para sustentar el fortalecimiento y/o especialización de algunos tramos viales, en términos de infraestructura dependiendo de la frecuencia o el tipo de carga a transportar.

Otra aplicación potencial de las herramientas de minería de datos en conjunto con los datos del RNDC, corresponde al desarrollo de una serie de tiempo asociada a la frecuencia o a la cantidad de carga movilizada por carretera, lo cual permitirá determinar patrones cíclicos en el comportamiento del transporte de carga, estimar el impacto en el transporte de determinadas anomalías, como por ejemplo paros de transporte, y además, planear de una forma precisa la construcción de infraestructura teniendo en cuenta la frecuencia de algunos tramos. El nivel de pertinencia de la información presente en el RNDC, puede ser potenciada no solamente integrando herramientas analíticas, sino también con herramientas de simulación e información recolectada en sistemas de información geográfico pertenecientes al estado colombiano.

Con la investigación desarrollada se evidenció la posibilidad de desarrollar futuras investigaciones no solamente aplicando otros tipos de algoritmos de minería de datos al RNDC, sino también explotando el potencial dado por la recolección de tiempos que realiza el sistema de información, mediante el uso de técnicas de simulación.

APÉNDICE

Anexo 1: Selección de la herramienta de modelado

Para la selección de la herramienta de minería de datos se realizó una revisión de tres estudios de gran relevancia que comparan las principales tecnologías de minería de datos:

Magic quadrant for advanced analytics platforms: es un estudio realizado por Gartner, empresa especializada en consultoría e investigación de las tecnologías de la información. Los estudios están enfocados en presentar las tecnologías líderes en diferentes ambientes negocios a nivel mundial y así facilitar la toma de decisiones informadas sobre tecnologías clave. En el estudio, Gartner expone una a una y de manera detallada las fortalezas y las debilidades de cada una de las plataformas evaluadas. Como producto final Gartner ubica a todos los competidores en un "cuadrante mágico" en donde los categoriza como retadores, líderes, jugadores de nicho y visionarios[72].

En este estudio se tuvieron en cuenta las siguientes características de las plataformas:

1. Data Access.
2. Visualization and Exploration/Discovery.
3. Data Filtering and Manipulation.
4. Advanced Descriptive Analytics.
5. Predictive Analytics.
6. Optimization.
7. Simulation.
8. Further Advanced Analytics.
9. Analytical Business Use Cases.
10. Delivery, Integration and Deployment.
11. Platform and Project Management.
12. User Experience.
13. Performance and Scalability.

Como se observa en la Figura 29, las tecnologías más sobresalientes son SAS, IBM Rapid-Miner y Knime las cuales son consideradas como líderes y visionarias.

FIGURA 29. Cuadrante mágico de Gartner para plataformas analíticas avanzadas

Encuesta Kdnuggets 2013: KDnuggets es un sitio líder en Business Analytics, Big Data, Data Mining, Data Science. En una encuesta realizada a 1880 líderes de tecnología en diferentes partes del mundo, a cada participante se le preguntó: "¿Qué software para Business Analytics, Big Data, Data Mining, Data Science han usado en los últimos 12 meses como parte de un proyecto real?".

Los resultados muestran que las tecnologías líderes aplicadas en proyectos de la vida real son tecnologías libres en su mayoría, tales como RapidMiner, R, Weka y Pentaho, ver figuras 30 y 31. Las plataformas coloreadas de verde corresponden a las plataformas libres, mientras las de rojo son plataformas privativas o licenciadas. Dentro de las plataformas licenciadas se destacan Excel y SAS en el tercer y el séptimo puesto. Como se puede observar, la plataforma de mayor difusión corresponde a RapidMiner/ RapidAnalytics con un 39,2 % del mercado y con un crecimiento entre 2012 y 2013 de un 13 % aproximadamente [86].

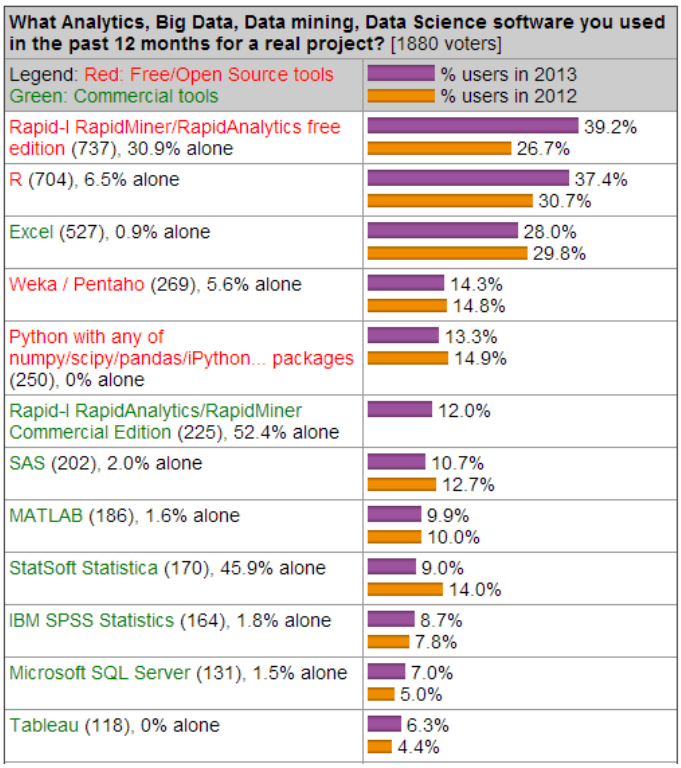


FIGURA 30. Resultados de la encuesta de plataformas líderes en Analytics, Big Data, Data Mining, Data Science en 2013 (Parte 1)[86]

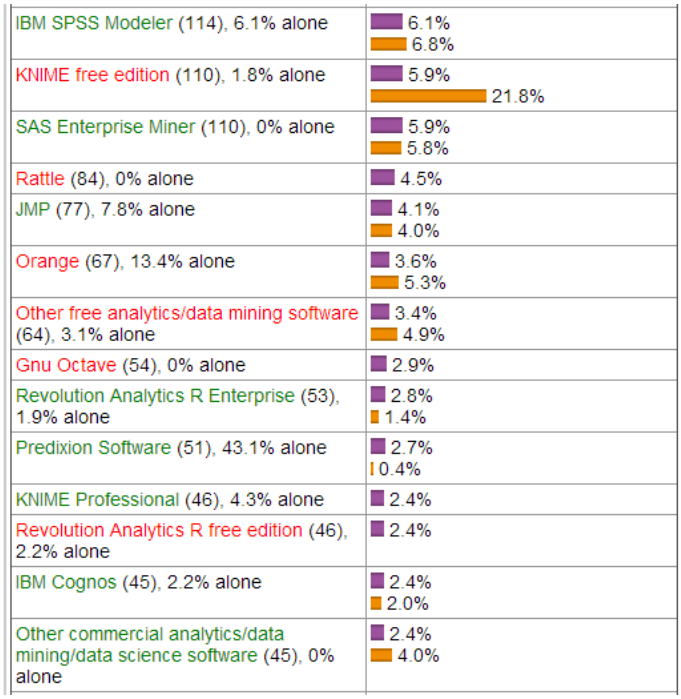


FIGURA 31. Resultados de la encuesta de plataformas líderes en Analytics, Big Data, Data Mining, Data Science en 2013 (Parte 2). [86]

ENCUESTA KDNUGGETS 2014: en el año 2014 nuevamente la página KDnuggets, en una encuesta realizada a 3285 líderes de tecnología en diferentes partes del mundo, muestra una clara predilección por la herramienta Rapidminer. Los resultados de la encuesta se muestran en la figura 32.

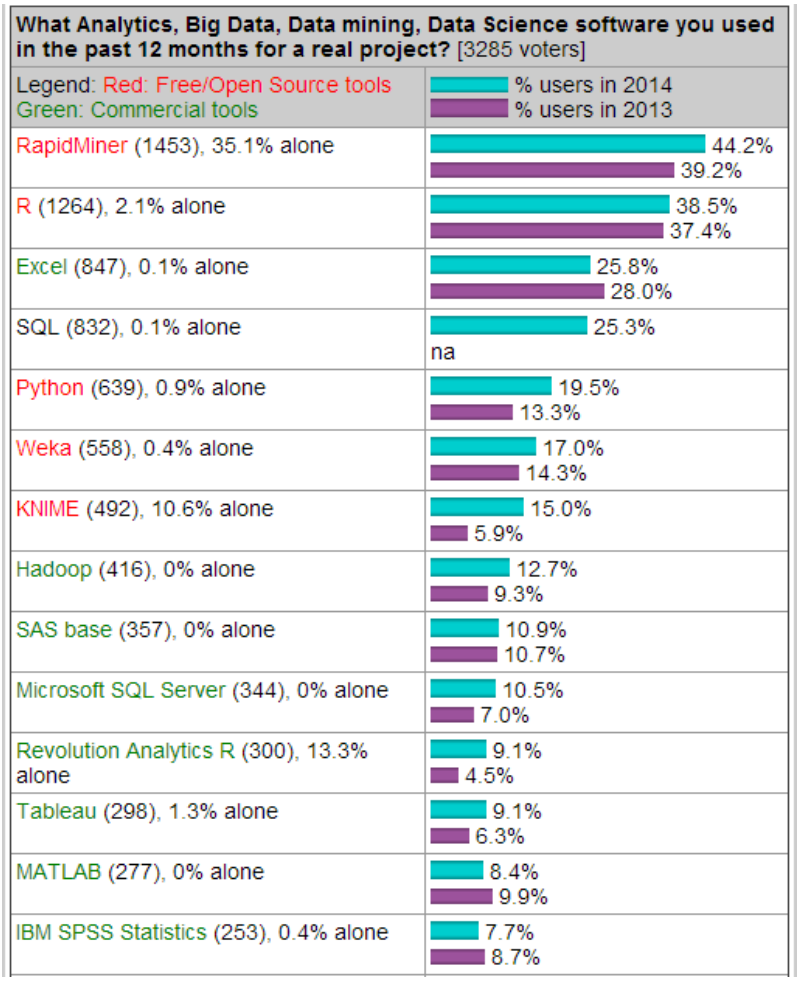


FIGURA 32. Encuesta de plataformas líderes en Analytics, Big Data, Data Mining, Data Science en 2014[87]

Las plataformas coloreadas de verde corresponden a las plataformas libres, mientras las de rojo son plataformas privativas o licenciadas. Los resultados muestran que las tecnologías líderes aplicadas en proyectos de la vida real son tecnologías libres en su mayoría, tales como RapidMiner, R, Weka y Phyton. Dentro de las plataformas licenciadas se destacan Excel y Hadoop en el tercer y el octavo puesto. Como se puede observar, la plataforma de mayor difusión corresponde a RapidMiner con un 44,2 % del mercado y con un crecimiento entre 2013 y 2014 de un 5 % aproximadamente [86].

.1. Descripción de la tecnología seleccionada

De acuerdo con la revisión de los estudios tenidos en cuenta, se identificó que la plataforma tecnológica RapidMiner es la mejor opción para la implementación de los modelos propuestos teniendo en cuenta hace parte de las plataformas líderes y visionarias dentro

del estudio realizado por Gartner, además de ser una de las más difundidas en el ambiente empresarial de acuerdo a la encuesta realizado por Kdnuggets. Adicionalmente, esta plataforma es una de las que mejor se adapta a los requerimientos del proyecto teniendo en cuenta que meneja conexión a diferentes orígenes de datos, cuenta con herramientas de estandarización de datos o labores ETL, generación de reportes y tableros de mando para el monitoreo de la información y herramientas analíticas avanzadas para la realización de tareas de análisis de datos tanto descriptivas como predictivas.

.1.1. Rapidminer Studio

RapidMiner Studio es una plataforma de software que provee un entorno integrado para el aprendizaje automático, minería de datos, minería de textos, análisis predictivo y análisis de negocio. Desarrollado por la empresa que lleva el mismo nombre, es un software multiplataforma y está licenciado como AGPL/Propietary, lo cual implica que las últimas versiones del software son de carácter propietario, mientras que las anteriores son liberadas para la comunidad de software libre.

RapidMiner es compatible con todos los pasos de la minería de datos, incluyendo los resultados de proceso de visualización, validación y optimización. Se ejecuta en un entorno de programación libre de código y cuenta con herramientas de conexión a base de datos, ETL, generación de reportes, creación de tareas programadas, algoritmos de clasificación, correlación, asociación, regresión, detección de anomalías y minería de textos.

Permite hacer análisis predictivo, es decir, análisis de datos históricos y actuales para hacer predicciones estadísticas precisas sobre hechos futuros. Un ejemplo de esto es el análisis del comportamiento histórico de un cliente, que permite predecir qué clientes tienen más probabilidades de salir de un contrato de teléfonos inteligentes. Otro ejemplo es el análisis de mantenimiento y averías registros de una planta de fabricación. Este permite predecir cuándo es probable que falle una máquina en otra parte de los operadores de las plantas, con ello se pueden diseñar los programas de mantenimiento y reducir las averías disruptivas.

Tiene un conector para trabajar con Hadoop (Big Data) y escalas para trabajar con grandes bases de datos relacionales. Es también la única solución de análisis predictivo en el mercado que puede ejecutar procesos analíticos en memoria, en la base de datos y en Hadoop.

Para la realización de tableros de mando, RapidMiner permite generar gráficos en 3-D, matrices de dispersión y mapas autoorganizativos. Además, permite convertir los datos en gráficos exportables, totalmente personalizables con soporte para zoom y panorámica, y reescalar para el máximo impacto visual.

En la actualidad, cuenta con diversas implementaciones en diferentes sectores tales como negocios, industrias, investigación, educación, creación rápida de prototipos y desarrollo de aplicaciones con un alto nivel de aceptación por parte de los usuarios de acuerdo con la encuesta de KDNUGGETS de 2013 y 2014.

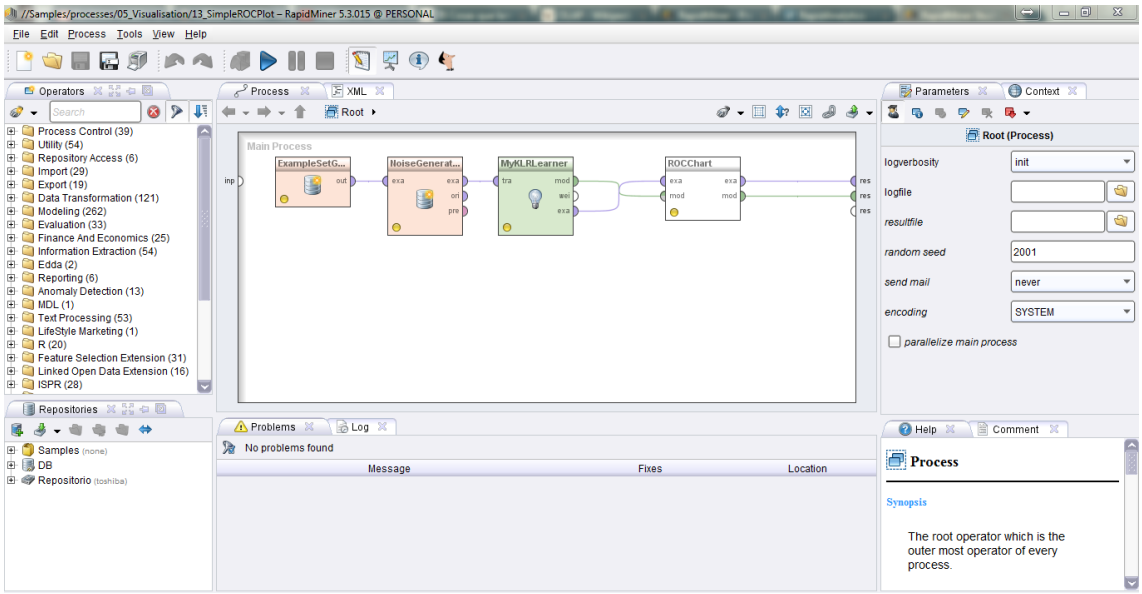


FIGURA 33. Entorno de trabajo de Rapidminer Studio

.1.2. Beneficios de la tecnología

Dado que el entorno de desarrollo de RapidMiner Studio es libre de código, el procesamiento y la generación de modelos descriptivos, de diagnóstico y predictivos en este programa es más fácil de hacer y puede desarrollarse en tiempos muy cortos.

La tecnología es adaptable a cualquier otra plataforma de bases de datos o mediante archivos de texto con formatos comunes para esta tarea, tales como CSV, excel, txt, entre otros.

Los modelos descriptivos permiten aplicar filtros y hacer análisis de sensibilidad con los datos en tiempo real, lo cual facilita la toma de decisiones.

Cuenta con herramientas para hacer la limpieza de los datos, las cuales permiten eliminar información incongruente y mostrar únicamente la información relevante.

Con esta herramienta es posible realizar modelos más sofisticados de minería de datos de acuerdo con la disponibilidad y calidad de la información, por ejemplo de clasificación, correlación, asociación, regresión, detección de anomalías, minería de textos y series de tiempo.

Como se mencionó anteriormente RapidMiner tiene un conector para trabajar con Hadoop (Big Data) y escalas para trabajar con grandes bases de datos relacionales lo cual permite que desde ya la herramienta sea compatible con las tendencias emergentes de almacenamiento de información.

APÉNDICE

Anexo 2: Registro nacional de despacho de carga

La fuente principal para hacer una evaluación de los denominados mercados relevantes, tiene su sustento en la información que las empresas reportan a través del registro de las operaciones de despacho de carga y que bajo ese contexto, la herramienta RNDC, está construida con parámetros y validaciones en línea, que van a permitir que se generen controles sobre la información, de la configuración de vehículos, del viaje, origen-destino, de actores que intervienen en la operación, del valor a pagar y de la variable de fechas y horas de cita para cargues y descargues, tiempos pactados y cumplidos, incluyendo de forma adicional una interfaz de reportes integrada al modelo denominado SIRTCC "Sistema de Información para la Regulación del Transporte de Carga por Carretera".

De esta manera, el MinTransporte cuenta con un instrumento idóneo para garantizar la transparencia y la formalidad que requiere el país y los actores del sector que prestan el Servicio Público de Transporte de Carga por Carretera.

Este instrumento es un elemento crucial de la política de transporte pues equilibra los intereses entre los diferentes actores del proceso. La implementación de esta herramienta busca mejorar la competitividad del país, con la que todos puedan obtener justos beneficios.

El RNDC es el medio para registrar los datos de la actividad transportadora de carga terrestre por carretera y evidencia la evolución de la información de esta operación asímismo informa a las entidades del Estado, encabezadas por el Ministerio de Transporte y la Superintendencia de Puertos y Transporte, a fin de que puedan ejercer sus actividades de control y planificación. Por ejemplo, le permite a la Policía Nacional obtener información en tiempo real para el control del transporte de carga en carretera y generar acciones que disminuyan o eviten la ocurrencia de delitos como robo de mercancías y vehículos. Otras entidades del Estado, como la DIAN, podrán usar esta herramienta tanto para controlar los impuestos del sector y desestimular la informalidad, como para identificar la participación de empresas ilegales en la actividad transportadora. La rigurosidad para generar y trasmitir estos datos es la clave del éxito para que los principios enunciados de control y planificación se cumplan.

Las Empresas de Transporte habilitadas por el MinTransporte para prestar el servicio el Servicio Público de Transporte de Carga por Carretera en el territorio nacional deben registrar de manera obligatoria sus operaciones. Para tal efecto, se dispone de una aplicación, de nombre RNDC (Registro Nacional de Despacho por Carga), que está al alcance

de usuarios debidamente autorizados. Opera en Internet y se puede usar con cualquier navegador.

Por otra parte, las Empresas de Transporte que disponen de su propio sistema de información pueden interactuar con el RNDC a través del uso de los Web Services.

Independiente de cual aplicación utilice la Empresa de Transporte, la transmisión de la información es en tiempo real, por lo tanto, debe hacerse de manera oportuna y rigurosa. Disponer de esta información en línea permite optimizar los controles por parte de las autoridades competentes.

A través del RNDC, las empresas de Servicio Público de Transporte Terrestre Automotor de Carga, deben expedir el manifiesto electrónico de carga y transmitir los datos al Ministerio de Transporte, conforme a lo dispuesto en el artículo 27 del Decreto 173 del 5 de febrero del 2001 y el artículo 4 del Decreto 1499 del 29 de Abril de 2009, donde menciona que el Manifiesto de Carga solo aplica a operaciones de transporte intermunicipal.

Además, los datos transmitidos del Manifiesto de Carga son fuente de información estadística de la movilización de carga en el país y se convierte en uno de los insumos para fijar las políticas del sector con base en los indicadores generados.

De otra parte, el RNDC provee información sobre las operaciones que realizan las empresas que prestan el servicio público de transporte terrestre automotor de carga que son insumos para monitorear el comportamiento del mercado conforme lo establece el Decreto 2092 del 14 de junio de 2011.

Así mismo, este registro provee información a las autoridades encargadas de ejercer el control operativo y administrativo, tales como la Superintendencia de Puertos y Transporte, Policía Nacional, Dirección de Impuestos y Aduanas Nacionales - DIAN y la Unidad de Información y Análisis Financiero - UIAF.

.2. Usuarios del Sistema de Información

Los actores que participan en el proceso de transporte de carga son los siguientes:

Usuario	Descripción
Propietario de la carga	Es la persona natural o jurídica que es propietaria de la mercancía que se transporta.
Generador de carga	Es el Remitente o el Destinatario de la carga cuando este último haga parte del contrato de transporte.
Remitente	Empresa, entidad o persona natural que hace de remitente de la mercancía en nombre del Generador de Carga (su agente, su proveedor, agentes de carga, de aduana, operadores logísticos, etc.). Hace referencia al sitio donde se carga y despacha la mercancía. Puede ser a su vez el mismo Generador de Carga.
Destinatario	Empresa, entidad o persona natural que hace de destinatario de la mercancía en nombre del Generador de Carga (su agente o su cliente). Hace referencia al sitio donde se descarga y recibe la mercancía. Puede ser a su vez el mismo Generador de Carga.
Empresa de transporte	Es quien legalmente cuenta con el permiso concedido por el Ministerio de Transporte para prestar el Servicio Público de Transporte de Carga. Debe contar con los recursos para realizar el movimiento en forma segura. El recurso principal es el vehículo de servicio público que puede ser de su propiedad o de terceros.

Representante legal	Persona que representa a la Empresa de Transporte. Es quien responde ante las autoridades del país por el actuar de su organización en la explotación del negocio del transporte de carga por carretera.
Despachador	Empleado de la Empresa de Transporte encargado de realizar el despacho mediante la expedición del Manifiesto de Carga y las Remesas. A su vez registra el cumplido de los documentos expedidos.
Titular del manifiesto de carga	Es el propietario, poseedor o tenedor de un vehículo de Servicio Público de Transporte de Carga a quien se le debe el Valor a Pagar por la prestación del servicio de transporte.
Conductor	Persona que conduce el Vehículo de Carga y que debe cumplir con los requerimientos legales para poder manejar el tipo de vehículo. Puede estar autorizado para representar al Propietario y/o Tenedor del Vehículo.
Propietario y/o tenedor del vehículo	Es el transportador de hecho al servicio de una empresa de transporte, mediante un contrato de vinculación permanente o temporal del equipo. No hace parte del contrato de transporte, pero sí lo es de la operación necesaria para su ejecución.

.3. Proceso de registro de información

El RNDC está centrado en la evolución de la información que la Empresa de Transporte genera para administrar y controlar su servicio de transporte de carga. Clasifica la recolección de información en las siguientes etapas.

Etapa	Descripción
Información de Carga	Primer paso del proceso de registro que contiene la información del Remitente y la información previa de la carga a transportar. Su principal función es describir la mercancía que se va a cargar, determinar en dónde va a ser cargada y reportar los tiempos pactados de cargue para la mercancía. También sirve para registrar el lugar y los tiempos pactados de descargue de la mercancía.
Información de Viaje	Segundo paso del proceso de registro que contiene la información del vehículo y del conductor que va a efectuar el transporte de carga, y que asocia la Información de Carga de cada una de las mercancías a transportar en el mismo viaje. Una vez una Información de Carga ha sido asociada en este paso, no puede ser asociada a otro viaje. Con esta información la Empresa de Transporte da la orden al Conductor de ir a cargar toda la mercancía a transportar en el viaje. También sirve para totalizar los tiempos pactados reportados en cada Información de Carga.
Orden de Cargue	Documento que autoriza al conductor de un vehículo para recoger una mercancía a nombre de la Empresa de Transporte. Contiene la información del Conductor, del Vehículo y de la Carga a recoger. La Orden de Cargue es almacenada en el RNDC de acuerdo a los registros de Información de Carga y de Vehículo y Conductor de la Información de Viaje. Esto hace que no sea necesario el registro de esta información nuevamente.
Remesa Terrestre de Carga	Documento oficial y obligatorio que representa cada carga durante el viaje y que registra la información restante de la mercancía transportada, el Remitente y el Destinatario. La información de las Remesas proviene de la Información de Carga, aunque puede ser complementada en este paso. Su nueva función en el RNDC es la de registrar los tiempos pactados de cargue y descargue. También sirve para registrar los tiempos ejecutados de cargue, si el cargue ya se hizo efectivo.

.4. Categorías de información en el RNDC

El RNDC está centrado en la evolución de la información que la empresa de transporte genera para administrar y controlar su servicio de transporte de carga. La información contenida en el RNDC se puede agrupar en las siguientes categoría:

Categoría	Descripción
Información de la carga	Primer paso del proceso de registro que contiene la información del Remitente y la información previa de la carga a transportar. Su principal función es describir la mercancía que se va a cargar, determinar en dónde va a ser cargada y reportar los tiempos pactados de cargue para la mercancía. También sirve para registrar el lugar y los tiempos pactados de descargue de la mercancía.
Información del viaje	Segundo paso del proceso de registro que contiene la información del vehículo y del conductor que va a efectuar el transporte de carga, y que asocia la Información de Carga de cada una de las mercancías a transportar en el mismo viaje. Una vez una Información de Carga ha sido asociada en este paso, no puede ser asociada a otro viaje. Con esta información la Empresa de Transporte da la orden al Conductor de ir a cargar toda la mercancía a transportar en el viaje. También sirve para totalizar los tiempos pactados reportados en cada Información de Carga.
Orden de Cargue	Documento que autoriza al conductor de un vehículo para recoger una mercancía a nombre de la Empresa de Transporte. Contiene la información del Conductor, del Vehículo y de la Carga a recoger. La Orden de Cargue es almacenada en el RNDC de acuerdo a los registros de Información de Carga y de Vehículo y Conductor de la Información de Viaje. Esto hace que no sea necesario el registro de esta información nuevamente.
Remesa Terrestre de Carga	Documento oficial y obligatorio que representa cada carga durante el viaje y que registra la información restante de la mercancía transportada, el Remitente y el Destinatario. La información de las Remesas proviene de la Información de Carga, aunque puede ser complementada en este paso. Su nueva función en el RNDC es la de registrar los tiempos pactados de cargue y descargue. También sirve para registrar los tiempos ejecutados de cargue, si el cargue ya se hizo efectivo.
Manifiesto de Carga	Documento oficial y obligatorio que registra la información completa del Titular del Manifiesto de Carga, del Vehículo, del Conductor, del Valor a Pagar por el viaje y relaciona todas las Remesas de la mercancía que está siendo transportada. Una vez una Remesa ha sido asociada en este paso, no puede ser asociada a otro viaje. También sirve para totalizar los tiempos de cargue y descargue reportados en la Remesa. Representa un viaje de vehículo. Un vehículo puede llevar más de un Manifiesto de Carga si ha sido despachado por más de una Empresa de Transporte.
Cumplido de Remesa Terrestre de Carga	Quinto paso del proceso de registro. Contiene información que identifica la Remesa que se cumple, los tiempos pactados y ejecutados de cargue y descargue de la mercancía. Permite a la Empresa de Transporte obtener la información necesaria para generar la factura a cobrar al Generador de Carga.
Cumplido del Manifiesto de Carga	Sexto paso del proceso de registro. Contiene información que identifica el Manifiesto de Carga que se cumple y el Valor a Pagar por el viaje al Titular del Manifiesto. También incluye los tiempos totales pactados y ejecutados de cargue y descargue de todas las Remesas asociadas al Manifiesto de Carga. Permite a la empresa de Transporte obtener la información necesaria para generar el pago por el viaje al Titular del Manifiesto.

Tabla 8. Agrupación de la información dentro del RNDC

.5. Características de los datos - Empresa de transporte

Las empresas de transporte deben registrarse para poder hacer uso del RNDC, en donde se solicita la siguiente información:nit de la empresa, código de la empresa, datos de la empresa, nombre, dirección, teléfono, municipio, usuario y password para acceder a la aplicación, nombres del representante legal, dirección de correo electrónico del representante.

.6. Características de los datos - Tiempos logísticos

El RNDC, registra los datos de la carga, del vehículo, del viaje, y los tiempos logísticos, es decir, los tiempos empleados en realizar las operaciones de transporte, tanto los pactados como los ejecutados.

El tiempo de cargue es el tiempo transcurrido desde que el vehículo se anuncia a la llegada de las instalaciones del Remitente hasta que sale cargado del lugar. Hay tres horas clave para registrar y poder calcular este tiempo: hora de llegada a las instalaciones, hora de entrada a las instalaciones y hora de salida de las instalaciones. Si hay más de un sitio de cargue, habrá tantos tiempos de cargue como lugares de cargue haya. El tiempo total de cargue será la suma de todos los tiempos de cargue.

Al igual que el tiempo de cargue, el de descargue es el tiempo transcurrido desde que el vehículo se anuncia a la llegada de las instalaciones del Destinatario hasta que sale descargado del lugar. Los tres eventos claves para registrar y poder calcular este tiempo son los mismos: hora de llegada a las instalaciones, hora de entrada a las instalaciones y hora de salida de las instalaciones. De igual forma, si hay más de un sitio de descargue, habrá tantos tiempos de descargue como lugares de descargue haya. El tiempo total de descargue será la suma de todos los tiempos de descargue.

También el RNDC registra la fecha del cumplido del viaje, comprendido por los cumplidos de Remesas y el cumplido del Manifiesto de Carga. La fecha del cumplido quedará registrada para determinar el momento a partir del cual se medirá el tiempo de pago del viaje al Titular del Manifiesto de Carga.

El RNDC, como su nombre lo indica, registra todos los datos de estos eventos y los guarda en su base de datos discriminada para cada Empresa de Transporte. Esta base de datos será la fuente de información para todos los efectos que persigue la política de transporte.

.7. Datos Externos

Para facilitar la validación de datos y el registro de despacho, el RNDC está articulado con el Registro Nacional de Conductores (RNC), que corresponde a una base de datos externa que almacena la información de los conductores a nivel nacional, incluyendo: datos personales, número de licencia y fecha de vencimiento de la misma. De la misma forma, está articulado con el Registro Nacional Automotor (RNA) y el Registro Nacional de Remolques (RNR). El primero almacena la información correspondiente a los vehículos, tales como marca, línea, configuración, peso vacío, capacidad, serie, tipo de combustible, tipo de carrocería, modelo, número de SOAT y fecha de vencimiento, propietario y tenedor. Por

otro lado, el RNR almacena la información correspondiente a los remolques a nivel nacional, incluyendo datos como marca, configuración, modelo, peso vacío, tipo de carrocería, propietario y tenedor.

APÉNDICE

Anexo 3: Objetivos del Ministerio de Transporte

El decreto 087 del 17 de enero de 2011 establece las funciones del MinTransporte. Asímismo los objetivos, funciones e integración del Sector Transporte:

Artículo 1°. Objetivo. El Ministerio de Transporte tiene como objetivo primordial la formulación y adopción de las políticas, planes, programas, proyectos y regulación económica en materia de transporte, tránsito e infraestructura de los modos de transporte carretero, marítimo, fluvial, férreo y aéreo y la regulación técnica en materia de transporte y tránsito de los modos carretero, marítimo, fluvial y férreo.

Artículo 2°. Funciones. Corresponde al Ministerio de Transporte cumplir, además de las funciones que determina el artículo 59 de la Ley 489 de 1998, las siguientes[101]:

2.1 Participar en la formulación de la política, planes y programas de desarrollo económico y social del país.

2.2. Formular las políticas del Gobierno Nacional en materia de transporte, tránsito y la infraestructura de los modos de su competencia.

2.3. Establecer la política del Gobierno Nacional para la directa, controlada y libre fijación de tarifas de transporte nacional e internacional en relación con los modos de su competencia, sin perjuicio de lo previsto en acuerdos y tratados de carácter internacional.

2.4. Formular la regulación técnica en materia de tránsito y transporte de los modos carretero, marítimo, fluvial y férreo.

2.5. Formular la regulación económica en materia de tránsito, transporte e infraestructura para todos los modos de transporte.

2.6.1 Establecer las disposiciones que propendan por la integración y el fortalecimiento de los servicios de transporte.

2.7. Fijar y adoptar la política, planes y programas en materia de seguridad en los diferentes modos de transporte y de construcción y conservación de su infraestructura.

2.8. Establecer las políticas para el desarrollo de la infraestructura mediante sistemas como concesiones u otras modalidades de participación de capital privado o mixto.

2.9. Apoyar y prestar colaboración técnica a los organismos estatales en los planes y programas que requieran asistencia técnica en el área de la construcción de obras y de

infraestructura física, con el fin de contribuir a la creación y mantenimiento de condiciones que propicien el bienestar y desarrollo comunitario.

2.10. Elaborar el proyecto del plan sectorial de transporte e infraestructura, en coordinación con el Departamento Nacional de Planeación y las entidades del sector y evaluar sus resultados.

2.11. Elaborar los planes modales de transporte y su infraestructura con el apoyo de las entidades ejecutoras, las entidades territoriales y la Dirección General Marítima, Dimar.

2.12. Coordinar, promover, vigilar y evaluar las políticas del Gobierno Nacional en materia de tránsito, transporte e infraestructura de los modos de su competencia.

2.13. Diseñar, coordinar y participar en programas de investigación y desarrollo científico, tecnológico y administrativo en las áreas de su competencia.

2.14. Impulsar en coordinación con los Ministerios competentes las negociaciones internacionales relacionadas con las materias de su competencia.

2.15. Orientar y coordinar conforme a lo establecido en el presente decreto y en las disposiciones vigentes, a las entidades adscritas y ejercer el control de tutela sobre las mismas.

2.16. Coordinar el Consejo Consultivo de Transporte y el Comité de Coordinación Permanente entre el Ministerio de Transporte y la Dirección General Marítima, Dimar.

2.17. Participar en los asuntos de su competencia, en las acciones orientadas por el Sistema Nacional de Prevención y Atención de Desastres.

2.18. Las demás que le sean asignadas.

Parágrafo 1°. Exceptuase de la Infraestructura de Transporte, los faros, boyas y otros elementos de señalización para el transporte marítimo, sobre los cuales tiene competencia la Dirección General Marítima, Dimar.

Parágrafo 2°. El Instituto Nacional de Concesiones, INCO, y el Instituto Nacional de Vías en relación con lo de su competencia, para el desarrollo de las actividades del modo de Transporte marítimo, serán asesorados por la Dirección General Marítima, Dimar, en el área de su competencia.

Artículo 3°. Dirección. La Dirección del Ministerio de Transporte estará a cargo del Ministro, quien la ejercerá con la inmediata colaboración de los Viceministros.

Artículo 4°. Integración del Sector Transporte. El Nivel Nacional del Sector Transporte está constituido, en los términos de la Ley 105 de 1993, por el Ministerio de Transporte y sus entidades adscritas. Siendo estas:

(a) Instituto Nacional de Vías, Invías.

(b) Instituto Nacional de Vías, Invías.

(c) Agencia Nacional de Infraestructura, ANI.

(d) Unidad Administrativa Especial de Aeronáutica Civil, Aerocivil.

(e) Superintendencia de Puertos y Transporte, Supertransporte.

APÉNDICE

Anexo 4: Selección de árbol de decisión

accuracy: 95.11%			
	true false	true true	class precision
pred. false	27927	1467	95.01%
pred. true	2	668	99.70%
class recall	99.99%	31.29%	

FIGURA 34. Performance árbol de agricultura

accuracy: 98.70%			
	true false	true true	class precision
pred. false	28852	386	98.68%
pred. true	5	821	99.39%
class recall	99.98%	68.02%	

FIGURA 35. Performance árbol de alimento para animales

accuracy: 83.27%			
	true false	true true	class precision
pred. false	23758	5023	82.55%
pred. true	8	1275	99.38%
class recall	99.97%	20.24%	

FIGURA 36. Performance árbol de alimentos y bebidas

accuracy: 96.61%

	true false	true true	class precision
pred. false	28328	1010	96.56%
pred. true	9	717	98.76%
class recall	99.97%	41.52%	

FIGURA 37. Performance árbol de actividad automotriz

accuracy: 95.52%

	true false	true true	class precision
pred. false	27580	1345	95.35%
pred. true	3	1136	99.74%
class recall	99.99%	45.79%	

FIGURA 38. Performance árbol de actividad avicultura

accuracy: 96.21%

	true false	true true	class precision
pred. false	27125	1137	95.98%
pred. true	3	1799	99.83%
class recall	99.99%	61.27%	

FIGURA 39. Performance árbol de actividad construcción

accuracy: 99.33%

	true false	true true	class precision
pred. false	28500	176	99.39%
pred. true	26	1362	98.13%
class recall	99.91%	88.56%	

FIGURA 40. Performance árbol de actividad contenedores

accuracy: 99.21%

	true false	true true	class precision
pred. false	29680	235	99.21%
pred. true	4	145	97.32%
class recall	99.99%	38.16%	

FIGURA 41. Performance árbol de actividad cosméticos y aseo

accuracy: 97.71%

	true false	true true	class precision
pred. false	27053	683	97.54%
pred. true	5	2323	99.79%
class recall	99.98%	77.28%	

FIGURA 42. Performance árbol de actividad envases y empaques

accuracy: 99.89%

	true false	true true	class precision
pred. false	30010	33	99.89%
pred. true	1	20	95.24%
class recall	100.00%	37.74%	

FIGURA 43. Performance árbol de actividad maderera

accuracy: 98.25%

	true false	true true	class precision
pred. false	29469	525	98.25%
pred. true	2	68	97.14%
class recall	99.99%	11.47%	

FIGURA 44. Performance árbol de actividad metalúrgica

accuracy: 99.60%

	true false	true true	class precision
pred. false	29470	118	99.60%
pred. true	3	473	99.37%
class recall	99.99%	80.03%	

FIGURA 45. Performance árbol de actividad minería

accuracy: 98.95%

	true false	true true	class precision
pred. false	29748	315	98.95%
pred. true	0	1	100.00%
class recall	100.00%	0.32%	

FIGURA 46. Performance árbol de actividad muebles y electrodomésticos

accuracy: 99.79%

	true false	true true	class precision
pred. false	28511	59	99.79%
pred. true	4	1490	99.73%
class recall	99.99%	96.19%	

FIGURA 47. Performance árbol de actividad petrolera

accuracy: 99.64%

	true false	true true	class precision
pred. false	29954	107	99.64%
pred. true	1	2	66.67%
class recall	100.00%	1.83%	

FIGURA 48. Performance árbol de actividad de químicos

accuracy: 85.45%

	true false	true true	class precision
pred. false	25514	4370	85.38%
pred. true	5	175	97.22%
class recall	99.98%	3.85%	

FIGURA 49. Performance árbol de actividad varios

accuracy: 99.07%

	true false	true true	class precision
pred. false	29620	278	99.07%
pred. true	2	164	98.80%
class recall	99.99%	37.10%	

FIGURA 50. Performance árbol de actividad vidrios y cerámicas

APÉNDICE

Anexo 5: Diccionario de datos

Actividad	Descripción
AGRICULTURA	PAPA
AGRICULTURA	LECHE
AGRICULTURA	AZUCAR
AGRICULTURA	MAIZ
AGRICULTURA	MaizAmarillo
AGRICULTURA	MAIZAMARIILLO
AGRICULTURA	TRIGO
AGRICULTURA	ARROZ
AGRICULTURA	MALTA
AGRICULTURA	MAIZBRASILERO
AGRICULTURA	TRIGOYMORCAJOOTRANQUILLON
AGRICULTURA	MaizBrasilero
AGRICULTURA	FERTILIZANTES
AGRICULTURA	SoyaSolla
AGRICULTURA	SOYA-T
AGRICULTURA	FRIJOLSOYA
AGRICULTURA	ARROZBLANCO
AGRICULTURA	CAFE
AGRICULTURA	TrigoGrano
AGRICULTURA	VERDURA
AGRICULTURA	CEBADA
AGRICULTURA	SoyaFrijolGrano
AGRICULTURA	TrigoSalvado
AGRICULTURA	SEMILLASYFRUTOSOLEAGINOSOS,SEMILLAS YFRUTOS-DIVERSOS,P
AGRICULTURA	MaizAmericano
AGRICULTURA	ABONO
AGRICULTURA	MaizGluten
AGRICULTURA	FRUTODEPALMAAFRICANA
AGRICULTURA	FRUTASVERDURAS
AGRICULTURA	MAQUINAS,APARATOSYARTEFACTOSAGRICOLAS, HOR-TICOLASOSI
GANADERIA	CARNECONGELADA
GANADERIA	CARNE

Actividad	Descripción
GANADERIA	CERDOSGORDOS
PESCA	PREPARACIONESYCONSERVADEPRESCADO
PESCA	CAVIARYSUSSUCEDANEO
PESCA	PREPARACIONESYCONSERVASDEPESCADO;PESCA
MINERIA	CLINKER
MINERIA	Carbon
MINERIA	CARBON
MINERIA	COQUE
MINERIA	CARBONES
MINERIA	CarbonCoque
MINERIA	CARBONCOQUE
MINERIA	CARBON
MINERIA	CARBONCOKE
MINERIA	CARB?"N"
MINERIA	CARBONCOKE
MINERIA	CARBONMINERAL
MINERIA	CARBONMETALURGICO
MINERIA	CARBONESACTIVADOS
MINERIA	MATERIASMINERALESNATURALESACTIVADAS
MINERIA	CARBONTERMICO
MINERIA	CARBONESACTIVADOS
MINERIA	PremezclaMineral
MADERA	MADERA
AVICULTURA	PREPARACIONESDELTIPODELASUTILIZADAS PARALAA-LIMENTACIO
AVICULTURA	GALLOS
AVICULTURA	GALLINAS
AVICULTURA	PATOS
AVICULTURA	GANZOS
AVICULTURA	PAVOS;YPINTADASDELAESP
AVICULTURA	CARNEYDESPOJOSCOMESTIBLESDEAVES DELAPARTI-DA01,05,
AVICULTURA	CARNEYDESPOJOSCOMESTIBLESDEAVES DELAPARTI-DA01.05,
AVICULTURA	POLLOCONGELADO
AVICULTURA	POLLO
AVICULTURA	HUEVOS
AVICULTURA	POLLITO
AVICULTURA	CARNEYDESPOJOSCOMESTIBLES DEAVESDELAPARTI-DA01.05,
CONSTRUCCION	TUBO
CONSTRUCCION	portls50kg
CONSTRUCCION	CEMEXUSOGENERAL
CONSTRUCCION	CEMEXGRANEL
CONSTRUCCION	EspConcrGra
CONSTRUCCION	CEMEX
CONSTRUCCION	CEMENTO
CONSTRUCCION	CEMENTOGRISESTRUCTURAL42.5KgARGOS2C
CONSTRUCCION	CEMENTOGRISPORTLANDTINTC25Kg
CONSTRUCCION	CEMENTOGRISPORTLANDTI
CONSTRUCCION	CEMENTOGRISPORTLANDTI42.5Kg
CONSTRUCCION	CEMENTOGRISPORTLANDTINTC40Kg

Actividad	Descripción
CONSTRUCCION	TUBOSYACCESORIOSDETUBERIA (POREJEMPLOJUNTAS,CODOSO
CONSTRUCCION	ASFALTO
CONSTRUCCION	CEMENTOGRIS
CONSTRUCCION	CEMENTOPORTLAND
CONSTRUCCION	PRODUCTOSPVCYSUSDERIVADOS
CONSTRUCCION	TUBOSCORTOS
CONSTRUCCION	CEMENTOEMPACADO
CONSTRUCCION	PortlS425
CONSTRUCCION	ACERO
CONSTRUCCION	LADRILLOS
CONSTRUCCION	TUBERIA,HIERRO,PERFILES
CONSTRUCCION	TUBERIAPVC
CONSTRUCCION	TEJAS
CONSTRUCCION	GRAVAS
CONSTRUCCION	ABRASIVOSNATURALESOARTIFICIALES ENPOLVOOENGRANULOSCO
CONSTRUCCION	CEMENTOPORTLANDPOR50KL
CONSTRUCCION	CONCRETO
CONSTRUCCION	EspConcrBig
CONSTRUCCION	CEMENTOBLANCOPORTLANDTINTC20Kg
CONSTRUCCION	TUBERIA
CONSTRUCCION	HERRAMIENTAS
CONSTRUCCION	CEMEXUSOGENERALS50
CONSTRUCCION	ARENA
CONSTRUCCION	MAQUINARIA
CONSTRUCCION	CEMENTOGRANEL
CONSTRUCCION	CEMENTOPORTLAND
CONSTRUCCION	MATERIALPARALACONSTRUCCION
CONSTRUCCION	MAQUINAS
CONSTRUCCION	MATERIALESPARACONSTRUCCION
CONSTRUCCION	PIEDRACALIZA
CONSTRUCCION	CEMENTOGRISPORTLANDTINTCBIGBAG
CONSTRUCCION	CEMENTOGRISGRANEL
CONSTRUCCION	CEMENTOAGRANEL
CONSTRUCCION	HERRAMIENTA
CONSTRUCCION	TB.ACC.PVC
CONSTRUCCION	TUBERIAYACCESORIOSPVC.
CONSTRUCCION	PINTURAS
CONSTRUCCION	PortlS50kg
CONSTRUCCION	PLACASDEYESO
METALURGICA	ALUMINIO
METALURGICA	LAMINA
METALURGICA	CHATARRA
METALURGICA	HOGARFERRETERIACACHARRERIA
METALURGICA	ROLLOSDELAMINA
METALURGICA	HIERROVARILLA
METALURGICA	LAMINADEACERO
METALURGICA	VARILLAS
METALURGICA	ALAMBRE
METALURGICA	TAPAS
METALURGICA	LAMINADOS

Actividad	Descripción
METALURGICA	PRODUCTOSLAMINADOSPLANOS DEHIERROODEACEROSINALEAR,
METALURGICA	PERFILESDEHIERROODEACEROSINALEAR
METALURGICA	ACEROSPLANOS
METALURGICA	HIERRO
METALURGICA	LAMINAS
METALURGICA	VARILLA
METALURGICA	FERRETERIA
METALURGICA	LASDEMASMANUFACTURAS DEHIERROODEACERO
PETROLERA	petrolero
PETROLERA	ACPM
PETROLERA	NAFTA
PETROLERA	COMBUSTIBLE
PETROLERA	ACEITESCRUDOSDEPETROLEO
PETROLERA	ACEITESCRUDOSDEPETROLEOO DEMINERIADEBITUMINOSOS
PETROLERA	EQUIPOPETROLERO
PETROLERA	COMBUSTOLEO
PETROLERA	PETROLEO
PETROLERA	BIODIESEL
PETROLERA	BIOCOMBUSTIBLES
PETROLERA	AGUADEPRODUCCI?”N”
PETROLERA	CRUDO-MORICHE
PETROLERA	GASOLINA
PETROLERA	MATERIAPETROLERO
PETROLERA	1203-3-COMBUSTIBLEPARAMOTORESoGASOLINA
PETROLERA	CRUDODEPETROLEO
PETROLERA	COMBUSTIBLEPARAMOTORESoGASOLINA
PETROLERA	ACEITESDEPETROLEOODEMINERALES BITUMINOSOS,EXCEPTOLOS
PETROLERA	DESTILADOSDEPETRSLEO,N.E.P.o PRODUCTOSDEPETRSLEO,N.E.
PETROLERA	BIODISELAL10 %
PETROLERA	COMBUSTIBLES
PETROLERA	CRUDO
PETROLERA	PETROLEO
PETROLERA	HIDROCARBUROSACICLICOS
PETROLERA	1267-3-PETRALEOBRUTO
PETROLERA	HIDROCARBUROSCICLICOS
PETROLERA	PETRSLEOBRUTO
PETROLERA	CRUDOEXPORTECNICAS
PETROLERA	CRUDODERIO
PETROLERA	CRUDOCA??OSUR
PETROLERA	CRUDOMORICHE
PETROLERA	EQUIPOPETROLERO
PETROLERA	1267-3-PETR?LEOBRUTO
PETROLERA	LEOBRUTO
PETROLERA	ALCOHOL
PETROLERA	COMBUS
PETROLERA	HIDROCARBUROSACICLICOS
MUEBLES Y ELECTRODOMESTICOS	MUEBLES

Actividad	Descripción
MUEBLES Y ELECTRO-DOMESTICOS	MUEBLESYENSERES
MUEBLES Y ELECTRO-DOMESTICOS	ESTIBAS
MUEBLES Y ELECTRO-DOMESTICOS	LOSDEMASMUEBLESYSUSPARTES
MUEBLES Y ELECTRO-DOMESTICOS	MOBILIARIODEOFICINA
MUEBLES Y ELECTRO-DOMESTICOS	MUEBLESDEOFICINA
MUEBLES Y ELECTRO-DOMESTICOS	LAMPARAS
MUEBLES Y ELECTRO-DOMESTICOS	MUEBLES Y ELECTRODOMESTICOSVARIOS
MUEBLES Y ELECTRO-DOMESTICOS	MUEBLESMODULARES
MUEBLES Y ELECTRO-DOMESTICOS	MUEBLESYENSERES
MUEBLES Y ELECTRO-DOMESTICOS	MUEBLESDEMADERAENCAJAPARAARMAR
MUEBLES Y ELECTRO-DOMESTICOS	TABLEROSDEFIBRADEMADERAUOTRAS MATERIAS-LE?OSAS,INCLU
MUEBLES Y ELECTRO-DOMESTICOS	MUEBLESENSERESCAJAS
MUEBLES Y ELECTRO-DOMESTICOS	ELECTRODOMESTICOS
ALIMENTOSYBEBIDAS	PRODUCTOSBAVARIA
ALIMENTOSYBEBIDAS	ALIMENTOREFRIGERADO
ALIMENTOSYBEBIDAS	ALIMENTOBALANCEADOS
ALIMENTOSYBEBIDAS	AZUCAR
ALIMENTOSYBEBIDAS	PRODUCTOSDEMOLINER?A,MALTA, AL-MID?NYF?CULA,INULINA,GL
ALIMENTOSYBEBIDAS	AZ?CARESYART?CULOSDECONFITER?A
ALIMENTOSYBEBIDAS	AGUA
ALIMENTOSYBEBIDAS	SAL(INCLUIDASLADEMESAYLADESNATURALIZADA) YCLORURODE
ALIMENTOSYBEBIDAS	HARINADETRIGO
ALIMENTOSYBEBIDAS	GASEOSASYJUGOS
ALIMENTOSYBEBIDAS	PRODUCTOSALIMENTICIOS
ALIMENTOSYBEBIDAS	SoyaTortaEstandar
ALIMENTOSYBEBIDAS	VIVERESYABARROTES
ALIMENTOSYBEBIDAS	CARNICOS
ALIMENTOSYBEBIDAS	AZUCARENBULTO
ALIMENTOSYBEBIDAS	HELADOS
ALIMENTOSYBEBIDAS	productosdicorp
ALIMENTOSYBEBIDAS	LECHELIQUIDAENBOLSAYDERIVADOS
ALIMENTOSYBEBIDAS	AZUCARDECA?ADEREMOLACHA
ALIMENTOSYBEBIDAS	DULCES
ALIMENTOSYBEBIDAS	PULPA
ALIMENTOSYBEBIDAS	ALIMEBTOS
ALIMENTOSYBEBIDAS	ALIMEBTOSPERECEDEROS
ALIMENTOSYBEBIDAS	PRODUCTOSLACTEOS

Actividad	Descripción
ALIMENTOSYBEBIDAS	TORTASYDEMASRESIDUOSSOLIDOS DELAEXTRACCION-DEGRASASO
ALIMENTOSYBEBIDAS	COMESTIBLES
ALIMENTOSYBEBIDAS	ALIMENTOSYBEBIDAS
ALIMENTOSYBEBIDAS	HARINADECEREALESEXCEPTODE TRIGOODEMORCA-JOOTRANQUIL
ALIMENTOSYBEBIDAS	ACEITEDESOYA
ALIMENTOSYBEBIDAS	PRODUCTOSDICORP
ALIMENTOSYBEBIDAS	GRASASYACEITES
ALIMENTOSYBEBIDAS	DestiladodeMaiz
ALIMENTOSYBEBIDAS	ARTICULOSDECONFITERIASINCACAO (INCLUIDOEL-CHOCOLATEBL
ALIMENTOSYBEBIDAS	AGUA,INCLUIDAELAGUAMINERALYLAGASIFICADA,AZUCARADA,E
ALIMENTOSYBEBIDAS	TORTADESOYA
ALIMENTOSYBEBIDAS	GASEOSA
ALIMENTOSYBEBIDAS	ACEITEDEPALMA
ALIMENTOSYBEBIDAS	JARABEDEMAIZGLUCOSA
ALIMENTOSYBEBIDAS	JARABEDEMAIZMALTOSA
ALIMENTOSYBEBIDAS	HARINA
ALIMENTOSYBEBIDAS	ALIME00
ALIMENTOSYBEBIDAS	ALIM.FASE1
ALIMENTOSYBEBIDAS	CAFEEXCELSO
ALIMENTOSYBEBIDAS	PANADERIAYPASTELERIA
ALIMENTOSYBEBIDAS	LEVADURAS
ALIMENTOSYBEBIDAS	HELADOSYPRODUCTOSSIMILARES INCLUSOCONCACAO
ALIMENTOSYBEBIDAS	LECHE
ALIMENTOSYBEBIDAS	LECHEYSUSDERIVADOS
ALIMENTOSYBEBIDAS	ACEITEDEPALMAYSUSFRACCIONES, INCLUSOREFINA-DO,PEROS
ALIMENTOSYBEBIDAS	ALIMENTOS
ALIMENTOSYBEBIDAS	ACEITECRUDODEPALMA
ALIMENTOSYBEBIDAS	MIELNATURAL
ALIMENTOSYBEBIDAS	HARINADETRIGOOMORCAJO/PASTAALIMENTICIA
ALIMENTOSYBEBIDAS	PASTASALIMENTICIAS
ALIMENTOSYBEBIDAS	PRODUCTOSDEPANADERIA
ALIMENTOSYBEBIDAS	PANADER??AYPASTELER??A
ALIMENTOSYBEBIDAS	PRODUDCTOSALIMENTICIOS
ALIMENTOSYBEBIDAS	PREPARACIONESALIMENTICIAS
ALIMENTOSYBEBIDAS	GRANOSEMPAQUETADOS
ALIMENTOSYBEBIDAS	ACEITE
ALIMENTOSYBEBIDAS	BEBIDAS,L?QUIDOSALCOH?LICOSYVINAGRE,
ALIMENTOSYBEBIDAS	PRODUCTOTERMINADO
ALIMENTOSYBEBIDAS	INSUMOS
AUTOMOTRIZ	VEHICULO
AUTOMOTRIZ	vehiculo
AUTOMOTRIZ	AUTOMOVIL
AUTOMOTRIZ	VOLQUETA
AUTOMOTRIZ	AUTOS
AUTOMOTRIZ	PARTESACCESORIOSDEVEHICULOS AUTOMOVILESDE-LASPARTIDAS
AUTOMOTRIZ	VEHICULO
AUTOMOTRIZ	REPUESTOS

Actividad	Descripción
AUTOMOTRIZ	BICICLETASYDEMASCICLOS (INCLUIDOSLOSTRICICLOSDEREPART
AUTOMOTRIZ	MOTOCICLETAS
AUTOMOTRIZ	VEHICULOSNUEVOS
AUTOMOTRIZ	LUBRICANTES,GRASAS
AUTOMOTRIZ	RETROEXCAVADORA
AUTOMOTRIZ	PARTESYACCESORIOSDEVEHICU
AUTOMOTRIZ	AUTOPARTES
AUTOMOTRIZ	MOTOS
AUTOMOTRIZ	AUTOPARTESRENAULT
AUTOMOTRIZ	vehiculo el añosenni
AUTOMOTRIZ	LUBRICANTES
AUTOMOTRIZ	ACEITES
AUTOMOTRIZ	LLANTAS
AUTOMOTRIZ	ACEITEPRODUCTOTERMINADO
AUTOMOTRIZ	MOTOSAKT
AUTOMOTRIZ	MOTOCICLETAS(INCLUSOCONPEDALES) YCICLOSCONMOTORAUXILI
AUTOMOTRIZ	CAMION
AUTOMOTRIZ	VEHICULOSNACIONALIZADOS
AUTOMOTRIZ	MOTO
AUTOMOTRIZ	VEHICULOSENNI??ERAMARCAKIA.-
CONTENEDORES	C40
CONTENEDORES	C20
CONTENEDORES	CONTEN
CONTENEDORES	CONTENEDORVACIO
CONTENEDORES	ISOTANK
CONTENEDORES	CONTENEDORESDE20”
CONTENEDORES	CONTENEDOSX40
CONTENEDORES	C-20
CONTENEDORES	CONTENDEVDE40”
CONTENEDORES	TANQUEDEACERO
CONTENEDORES	CONTENEDORDE40”
CONTENEDORES	CONTENEDOR
CONTENEDORES	CONTENEDORES(INCLUIDOSLOS CONTENEDORESCISTERNAYLOSCONT
CONTENEDORES	CONTENEDORVACIOENDEV
CONTENEDORES	CONTENEDORESVACIOS
CONTENEDORES	C-40
CONTENEDORES	VACIO
CONTENEDORES	VACIO0
ALIMENTOANIMALES	CONCENTRADO
ALIMENTOANIMALES	AlimentoBalanceadoParaAnimales
ALIMENTOANIMALES	ALIM.ENGORDE,CONCENTRADO
ALIMENTOANIMALES	ALIM.FINALIZADOR
ALIMENTOANIMALES	ALIMENTOPARAPOLLOS
ALIMENTOANIMALES	SALMINERALIZADAPARAGANADOS
ALIMENTOANIMALES	ALIM.POLLITO
ALIMENTOANIMALES	CONCENTRADOS
ALIMENTOANIMALES	RESINAS
ALIMENTOANIMALES	ALIM.PREPICO
ALIMENTOANIMALES	PRODUCTOSVETERINARIOS

Actividad	Descripción
ALIMENTOANIMALES	ABONOFERTILIZANTE
ALIMENTOANIMALES	ALIMENTOPARAANIMALES
ALIMENTOANIMALES	CONCENTRADOPARAANIMALES
ALIMENTOANIMALES	ALIMENTOCONCENTRADO
ALIMENTOANIMALES	ALIM.PREINICIADOR
ALIMENTOANIMALES	ALIM.ENGORDE
ALIMENTOANIMALES	ALIMENTOSCONCENTRADOS
COSMETICOSYASEO	ASEO
COSMETICOSYASEO	HIGIENE
COSMETICOSYASEO	TISSUE
COSMETICOSYASEO	ARTICULOSDEUSODOMESTICO,DEIGIENEODETOCADOR,YSUS
COSMETICOSYASEO	PULVERIZADORESDETOCADOR,SUS MONTURASYCABEZASDEMONTUR
COSMETICOSYASEO	MATERIAPRIMACOSMETICOS
COSMETICOSYASEO	JABON
COSMETICOSYASEO	PRODUCTOSTENSOACTIVOSUSADO COMOJABON,ENBARRAS,
COSMETICOSYASEO	PA?'ALESDESECHABLESNI?'OSYA
COSMETICOSYASEO	TOALLASDEPAPEL
COSMETICOSYASEO	productosdersa
COSMETICOSYASEO	PAPELHIGIENICO
COSMETICOSYASEO	DETERGENTE
COSMETICOSYASEO	DETERGENTES
COSMETICOSYASEO	COSMETICOS
COSMETICOSYASEO	PAPELHIGIENCO
COSMETICOSYASEO	PAPELHIGIENICO-SERVILLETA
COSMETICOSYASEO	T.HIGIENICA
COSMETICOSYASEO	PAPELDETIPOHIGIENICO
COSMETICOSYASEO	PERFUMESYAGUASDETOCADOR
COSMETICOSYASEO	PRODUCTOSDEBELLEZA
COSMETICOSYASEO	PRODUCTOSDEASEO
ENVASESYEMPAQUES	CARTON
ENVASESYEMPAQUES	CAJASDECARTON
ENVASESYEMPAQUES	CANASTAS
ENVASESYEMPAQUES	CAJAS
ENVASESYEMPAQUES	SACOS,BOLSAS
ENVASESYEMPAQUES	CUCURUCHOSYDEMASENVAESDEPAPEL
ENVASESYEMPAQUES	ENVASESPLASTICOS
ENVASESYEMPAQUES	CAJAS,SACOS,BOLSASDEPAPEL
ENVASESYEMPAQUES	CANASTILLASPLASTICAS
ENVASESYEMPAQUES	ENVASESDEMETAL
ENVASESYEMPAQUES	ENVASESVACIOS
ENVASESYEMPAQUES	ENVASEDEVIDRIO
ENVASESYEMPAQUES	MARQUETERIAYTARACEA
ENVASESYEMPAQUES	COFRES,CAJASYESTUCHESPARAJOYERI
ENVASESYEMPAQUES	ENVASEVACIO
ENVASESYEMPAQUES	CAJASSACOSBOLSASCUCURUCHOSY DEMASENVASESDEPAPELC
ENVASESYEMPAQUES	CAJASCARTON
ENVASESYEMPAQUES	ENV
ENVASESYEMPAQUES	CANASTILLASPLASTICASGRANDES
ENVASESYEMPAQUES	ENVASESDEHOJALATA

Actividad	Descripción
ENVASESYEMPAQUES	PRODUCTOSPLASTICOS
ENVASESYEMPAQUES	ARTICULOSDEPLASTICOS
ENVASESYEMPAQUES	BOLSAPLASTICA
ENVASESYEMPAQUES	PLASTICOS
ENVASESYEMPAQUES	PLASTICO
VARIOS	VARIAS
VARIOS	PAQUETEO
VARIOS	CAJAS
VARIOS	PAQUETESVARIOS
VARIOS	MISCELANCONTENPAQUETES
VARIOS	MISCELANEOSCONTENIDOSENPAQUETES(PAQUETEO)
VARIOS	MISCELANEOS
VARIOS	PAQUETEOVARIAS
VARIOS	PAQUETES
VARIOS	SOBRES
VARIOS	ProductosVarios
VIDRIOSYCERAMICAS	LASDEMASMANUFACTURASDECERAMICA
VIDRIOSYCERAMICAS	VIDRIO
VIDRIOSYCERAMICAS	CERAMICA
QUIMICOS	QUIMIC

Tabla 9. Diccionario de datos aplicado al campo descripción corta producto

Bibliografía

[1] Wilson Adarme Jaimes, *Desarrollo metodológico para la optimización de la cadena de suministro esbelta con m proveedores y n demandantes bajo condiciones de incertidumbre. Caso aplicado a empresas navieras colombianas*, Ph.D. thesis, 2011, pp. 1–88.

[2] B. Agard and A. Kusiak *, *Data-mining-based methodology for the design of product families*, International Journal of Production Research **42** (2004), no. 15, 2955–2969.

[3] Rakesh Agrawal, T Imielinski, and A Swami, *Mining association rules between sets of items in large databases*, ACM SIGMOD Record (1993), no. May, 1–10.

[4] Ozturk N. Aksoy, A., *Supplier selection and performance evaluation in just- in-time production environments*, Expert Systems with Applications **38** (2011), no. 5, 6351–6359.

[5] Fareed Akthar and Caroline Hahne, *RapidMiner 5: Operator Reference*, Rapid-I GmbH (2012).

[6] V Amado, *Expanding the use of pavement management data*, 2000 MTC Transportation Scholars Conference, Ames, ... (2000), 2–18.

[7] S.S. Anand, A.G. Büchner, and Financial Times Management, *Decision support using data mining*, Financial Times management briefings: Information technology, Financial Times Management, 1998.

[8] Rebecca Angeles, *Emerging Technologies: Supply-Chain Applications and Implementation Issues*, Information Systems Management (2005), 51–65.

[9] Jean-François Arvis, Daniel Saslavsky, Lauri Ojala, Ben Shepherd, Christina Busch, and Anasuya Raj, *LPI 2014 - Connecting to Compete, Trade Logistics in the Global Economy*, Worldbank (2014), 72.

[10] JF Arvis, MA Mustra, and J Panzer, *Connecting to compete: Trade logistics in the global economy*, World Bank. Washington, (2012).

[11] R.H. Ballou, *Logística: administración de la cadena de suministro*, Pearson educación, Pearson Educación, 2004.

[12] SK Barai, *Data mining applications in transportation engineering*, Transport **XVIII** (2003), no. April 2014, 37–41.

[13] D. Beeferman and A. Berger, *Agglomerative clustering of a search engine query log*, 2000, cited By 0, pp. 407–416.

[14] T. Berry and A. Ahmed, *The consequences of inter-firm supply chains for management accounting.*, Management Accounting: Magazine for Chartered Management Accountants **75** (1997), no. 10, 74.

[15] Fernando Berzal, Ignacio Blanco, DS\'anchez, and MA Vila, *Measuring the accuracy and interest of association rules: A new framework*, Intelligent Data Analysis **6** (2002), 221–235.

[16] By Erik Brynjolfsson and Andrew Mcafee, *The big data boom is the innovation story of our time*, (2011), 1–5.

[17] Cristian Bucur, *Implications and directions of development of web business intelligence systems for business community*, Economic Insights - Trends and Challenges **LXIV** (2012), no. 2, 96–109.

[18] CAG Camargo, *Evaluación de la utilidad de la minería de datos para la planeación de vías para el transporte de carga regional caso de estudio: Bolivia*, Revista Internacional Administración & Finanzas (2011), 61–73.

[19] Hsien Chao Chang-Yi Chen, Tien-Yin Chou, Ching-Yun Mu, Bing-Jean Lee, Magesh Chandramouli and Geographic, *Using Data Mining Techniques on Fleet Management System*, Geographic Information Systems Research Center, Feng Chia University, 1–11.

[20] P Chapman, J Clinton, and R Kerber, *Crisp-dm 1.0 step-by-step data mining guide*, (2000).

[21] Surajit Chaudhuri, Umeshwar Dayal, and Vivek Narasayya, *An overview of business intelligence technology*, Communications of the ACM **54** (2011), no. 8, 88.

[22] Z.H. Che, *A two-phase hybrid approach to supplier selection through cluster analysis with multiple dimensions*, International Journal of Innovative Computing Information and Control (2010), 4093–4111.

[23] Hsinchun Chen and Guoqing Chen, *Business intelligence and analytics : Research directions*, ACM Transactions on Management Information Systems **3** (2013), no. 4, 1–10.

[24] Hsinchun Chen and Veda C Storey, *Business intelligence and analytics: From big data to big impact*, MIS Quarterly **36** (2012), no. 4, 1165–1188.

[25] M CHEN and H WU, *An association-based clustering approach to order batching considering customer demand patterns*, Omega **33** (2005), no. 4, 333–343.

[26] Mu-Chen Chen, Cheng-Lung Huang, Kai-Ying Chen, and Hsiao-Pin Wu, *Aggregation of orders in distribution centers using data mining*, Expert Systems with Applications **28** (2005), no. 3, 453–460.

[27] Sunil Chopra and Peter Meindl, *Supply chain management. strategy, planning & operation*, Das Summa Summarum des Management, Springer, 2007, pp. 265–275.

[28] K Choy, *An intelligent supplier management tool for benchmarking suppliers in outsource manufacturing*, Expert Systems with Applications **22** (2002), no. 3, 213–224.

[29] King Lun Choy, W. B. Lee, Henry C. W. Lau, Dawei Lu, and Victor Lo, *Design of an intelligent supplier relationship management system for new product development*, Int. J. Computer Integrated Manufacturing (2004), 692–715.

[30] M. Christopher, *Supply chain managementthe industrial organisation perspective*, Logistics and Supply Chain Management Pitman Publishing (1992).

[31] Krzysztof J. Cios, *Data mining: a knowledge discovery approach*, 2007.

[32] Stacy Collett, *Why big data is a big deal*, (2011), 18–24.

[33] Vladimir Coric, *Data mining algorithms for traffic sampling, estimation and forecasting*, Ph.D. thesis, TEMPLE UNIVERSITY, 2014.

[34] Tom Costello and Beverly Prohaska, *2013 trends and strategies*, IT Professional (2013), no. February, 0–2.

[35] Terence Critchlow, *Big data ecosystems enable scientific discovery*, HPC Source (2011), no. November, 35–39.

[36] Universidad Oberta de Cataluña, *Curso Introducción al Business Intelligence y al Big Data*, 2016.

[37] Ministerio de Transporte de Colombia, *Transporte en cifras. Documento estadístico del sector transporte.*, Tech. report, 2010.

[38] Consejo Privado de Competitividad, *Informe nacional de competitividad 2014-2015*, Tech. report, 2014.

[39] Departamento Nacional de Planeación, *Conpes 3489 Política Nacional De Transporte Público Automotor De Carga*, 2007.

[40] ———, *Conpes 3527 política nacional de competitividad y productividad*, 2008.

[41] ———, *Conpes 3547 política nacional logística*, 2008.

[42] ———, *Plan nacional de desarrollo*, 2011.

[43] Departamento Nacional de Planeación, *Plan Nacional de Desarrollo 2014 - 2018*, 2014.

[44] Departamento Nacional de Planeación, *Encuesta Nacional de Logística 2015*, 2015.

[45] Ministerio de Transporte, *Registro nacional de despacho de transporte de carga rndc: Manual de usuario*, (2012), 1–105.

[46] Jeffrey Dean and Sanjay Ghemawat, *Mapreduce*, Communications of the ACM **53** (2010), no. 1, 72.

[47] W H DeLone and E R McLean, *Information systems success: The quest for the dependent variable*, Information Systems Research **3** (1992), no. 1, 60–95.

[48] Departamento Administrativo Nacional de Estadística - DANE, *Caracterización temática transporte*, 2011, pp. 1–77.

[49] Revista Dinero, *Competitividad del transporte de carga*, 2009, pp. 0–2.

[50] ———, *Absurdos costos del transporte*, 2015.

[51] L.M. Ellram, *Supply chain managementthe industrial organisation perspective*, International Journal of Physical Distribution and Logistics Management **21** (1991), no. 1, 13–22.

[52] Chwei-Jen Fan, Li-Chuan Wang, and Huan-Ming Chuang, *The applications of business intelligence to the improvement of supply chain management - a case of an electronic company*, Journal of Software **6** (2011), no. 11, 2173–2177.

[53] Usama Fayyad, Gregory Piatetsky-shapiro, and Padhraic Smyth, *From data mining to knowledge discovery in databases*, American Association for Artificial Intelligence (1996), 37–54.

[54] Georgina Febré and Gabriel Pérez Salas, *Sistemas inteligentes de transporte en la logística portuaria latinoamericana*, 2012.

[55] Luciano Ferreira and Denis Borenstein, *A fuzzy-bayesian model for supplier selection*, Expert Systems with Applications **39** (2012), no. 9, 7834–7844.

[56] R Fildes, K Nikolopoulos, S F Crone, and a a Syntetos, *Forecasting and operational research: a review*, Journal of the Operational Research Society **59** (2008), no. 9, 1150–1172.

[57] Key Findings, *Predicts 2013 : Information Innovation*, Gartner Group (2013), no. December 2012.

[58] Stefan Foell, Gerd Kortuem, Reza Rawassizadeh, Santi Phithakkitnukoon, Marco Veloso, and Carlos Bento, *Mining temporal patterns of transport behaviour for predicting future transport usage*, Proceedings of the 2013 ACM conference on Pervasive and ubiquitous computing adjunct publication - UbiComp '13 Adjunct (2013), 1239–1248.

[59] Eugenia G. (ed.), *Data mining in medical and biological research*, InTech, November 2008.

[60] Gartner, *Magic quadrant for business intelligence platforms*, (2012), 1–27.

[61] M. Ghazanfari, M. Jafari, and S. Rouhani, *A tool to evaluate the business intelligence of enterprise systems*, Scientia Iranica **18** (2011), no. 6, 1579–1590.

[62] Hector Gonzalez, Jiawei Han, and Xiaolei Li, *Mining compressed commodity workflows from massive rfid data sets*, Proceedings of the 15th ACM international conference on Information and knowledge management - CIKM '06 (New York, New York, USA), ACM Press, 2006, p. 162.

[63] Michel Goossens, Frank Mittlebach, and Alexander Samarin, *The LaTeX Companion*, aw, 1994.

[64] Daniela Grigori, Fabio Casati, Malu Castellanos, Umeshwar Dayal, Mehmet Sayal, and Ming-Chien Shan, *Business process intelligence*, Computers in Industry **53** (2004), no. 3, 321–343.

[65] a Gunasekaran and E.W.T Ngai, *Information systems in supply chain integration and management*, European Journal of Operational Research **159** (2004), no. 2, 269–295.

[66] S Ha and R Krishnan, *A hybrid approach to supplier selection for the maintenance of a competitive supply chain*, Expert Systems with Applications **34** (2008), no. 2, 1303–1311.

[67] P Haluzová, *Effective Data Mining for a Transportation Information System*, Acta Polytechnica **48** (2008).

[68] Jiawei Han, Hector Gonzalez, Xiaolei Li, and Diego Klabjan, *Warehousing and mining massive rfid data sets*, (2006), 1–18.

[69] Jiawei Han, Jian Pei, and Yiwen Yin, *Mining frequent patterns without candidate generation*, ACM SIGMOD Record (2000), 1–20.

[70] C M Harland, *Supply chain management : Relationships , chains and networks*, British Journal of Management **7** (1996), no. March, S63–S80.

[71] Petri Helo and Bulcsu Szekely, *Logistics information systems: An analysis of software solutions for supply chain co-ordination*, Industrial Management & Data Systems **105** (2005), 5–18.

[72] Gareth Herschel, Alexander Linden, and Lisa Kart, *Magic Quadrant for Advanced Analytics Platforms*, Gartner Group (2014), no. February.

[73] F.S. Hillier and G.J. Lieberman, *Introduction to operations research*, McGraw-Hill Education.

[74] Robinson P.J. Green Hinkle, C.L., *Vendor evaluation using cluster analysis*, Journal of Purchasing (1969), 49–58.

[75] C. Billington H.L. Lee, *Managing supply chain inventorypitfalls and opportunities*, Sloan Management Review **33** (1992), no. 3, 65–73.

[76] S.M. Ng H.L. Lee, *Introduction to the special issue on global supply chain management*, Production and Operations Management **6** (1997), no. 3, 191–192.

[77] D.W. Huang, B.T. Sherman, and R.A. Lempicki, *Systematic and integrative analysis of large gene lists using david bioinformatics resources*, Nature Protocols **4** (2009), no. 1, 44–57, cited By 0.

[78] SAS Institute Inc, *Sas — semma*, 2013.

[79] Instituto nacional de vías - INVIAS, *Consulta de las principales rutas viales y sus tarifas de peaje*, 2015.

[80] W.H. Ip, Min Huang, K.L. Yung, and Dingwei Wang, *Genetic algorithm solution for a risk-based partner selection problem in a virtual enterprise*, Computers & Operations Research **30** (2003), no. 2, 213–231.

[81] AK Jain, RPW Duin, and Jianchang Mao, *Statistical pattern recognition: A review*, Pattern Analysis and Machine ... **22** (2000), no. 1, 4–37.

[82] V. Jain, S. Wadhwa, and S.G. Deshmukh, *Supplier selection using fuzzy association rules mining approach.*, International Journal of Production Research **45** (2007), no. 6, 1323–1353.

[83] Vipul Jain, Lyes Benyoucef, and S.G. Deshmukh, *A new approach for evaluating agility in supply chains using fuzzy association rules mining*, Engineering Applications of Artificial Intelligence **21** (2008), no. 3, 367–385.

[84] Jeanne E Johnson, *Big data + big analytics = big oportunity*, Financial Executive (2012), 50 – 53.

[85] R.B. Handfield K.C. Tan, V.R. Kannan, *Supply chain managementsupplier performance and firm performance.*, International Journal of Purchasing and Material Management **34** (1998), no. 3, 2–9.

[86] Kdnuggets, *Poll What Analytics, Big Data, Data mining, Data Science software you used in the past 12 months*, 2013.

[87] ———, *Poll What Analytics, Data Mining, Data Science software/tools you used in the past 12 months for a real project Poll*, 2014.

[88] L.R. Kopczak, *Logistics partnership and supply chain restructuring: survey results from the us computer industry*, Production and Operations Management **6** (1997), no. 3, 226–247.

[89] R.J. Kuo, *A sales forecasting system based on fuzzy neural network with initial weights generated by genetic algorithm*, European Journal of Operational Research **129** (2001), no. 3, 496–517.

[90] Colprensa La República, *Colombia es el tercer país con los peajes más costosos de América Latina*, 2015.

[91] Neal Lathia, Jon Froehlich, and Licia Capra, *Mining Public Transport Usage for Personalised Intelligent Transport Systems*, 2010 IEEE International Conference on Data Mining, no. October 2009, IEEE, 2010, pp. 887–892.

[92] H.C.W. Lau, G.T.S. Ho, Y. Zhao, and N.S.H. Chung, *Development of a process mining system for supporting knowledge discovery in a supply chain network*, International Journal of Production Economics **122** (2009), no. 1, 176–187.

[93] Kenneth C. Laudon and Jane P. Laudon, *Management information systems*, twelfth ed ed., Prentice Hall, 2012.

[94] Ou-Yang C. Lee, C.C., *A neural networks approach for forecasting the supplier's bid prices in supplier selection negotiation process*, Expert Systems with Applications **2** (2009), 2961–2970.

[95] Guo-Dong Li, Daisuke Yamaguchi, and Masatake Nagai, *A grey-based rough decision-making approach to supplier selection*, The International Journal of Advanced Manufacturing Technology **36** (2007), no. 9-10, 1032–1040.

[96] Shu-hsien Liao, Ya-ning Chen, and Yu-yia Tseng, *Mining demand chain knowledge of life insurance market for new product development*, Expert Systems with Applications **36** (2009), no. 5, 9422–9437.

[97] Shu-Hsien Liao, Pei-Hui Chu, and Pei-Yuan Hsiao, *Data mining techniques and applications - a decade review from 2000 to 2011*, Expert Systems with Applications **39** (2012), no. 12, 11303–11311.

[98] Rong-ho Lin, Chun-ling Chuang, James J.H. Liou, and Guo-dong Wu, *An integrated method for finding key suppliers in scm*, Expert Systems with Applications **36** (2009), no. 3, 6461–6465.

[99] H. P. Luhn, *A business intelligence system*, IBM Journal (1958), 314–319.

[100] Serbanescu Luminita and Radulescu Magdalena, *Optimizing time in business with business intelligence solution*, Procedia - Social and Behavioral Sciences **62** (2012), 638–648.

[101] Ministerio de Transporte, *Ministerio de Transporte de Colombia*, 2015.

[102] Mintransporte, *Parque automotor de transporte de carga en colombia*, Ministerio de Transporte (2006), 109.

[103] María N Moreno and Vivian F López, *Uso de Técnicas no Supervisadas en la Construcción de Modelos de Clasificación en Ingeniería del Software*, Departamento de Informática y Automática. Universidad de Salamanca.

[104] Glenn J. Myatt, *Making sense of data: a practical guide to exploratory data analysis and data mining*, 2002.

[105] ———, *Making senser of data ii: a practical guide to data visualization, advanced data mining methods, and applications*, 2009.

[106] E.W.T. Ngai, L. Xiu, and D.C.K. Chau, *Application of data mining techniques in customer relationship management: A literature review and classification*, Expert Systems with Applications **36** (2009), no. 2 PART 2, 2592–2602, cited By 0.

[107] E.W.T. Ngai, Li Xiu, and D.C.K. Chau, *Application of data mining techniques in customer relationship management: A literature review and classification*, Expert Systems with Applications **36** (2009), no. 2, 2592–2602.

[108] Tho Manh Nguyen, Josef Schiefer, and A. Min Tjoa, *Sense & response service architecture (saresa): an approach towards a real-time business intelligence solution and its use for a fraud detection application*, DOLAP '05: Proceedings of the 8th ACM international workshop on Data warehousing and OLAP (New York, NY, USA), ACM Press, 2005, pp. 77–86.

[109] T.T.T. Nguyen and G. Armitage, *A survey of techniques for internet traffic classification using machine learning*, IEEE Communications Surveys and Tutorials **10** (2008), no. 4, 56–76, cited By 0.

[110] Marcos Paulo Valadares De Oliveira, Kevin McCormack, and Peter Trkman, *Business analytics in supply chains - the contingent effect of business process maturity*, Expert Systems with Applications **39** (2012), no. 5, 5488–5498.

[111] Myerson P., *Lean supply chain and logistics management*, McGraw-Hill Education, 2012.

[112] B Padmanabhan and A Tuzhilin, *On the use of optimization for data mining: Theoretical interactions and ecrm opportunities*, Management Science **49** (2003), no. 10, 1327–1343.

[113] By Gregory Piatetsky, *CRISP-DM , still the top methodology for analytics , data mining , or data science projects*, 2014.

[114] Selwyn Piramuthu, *Knowledge-based framework for automated dynamic supply chain configuration*, European Journal of Operational Research **165** (2005), no. 1, 219–230.

[115] Council of Supply Chain Management Professionals, *Cscmp supply chain management — council of supply chain management professionals*, 2013.

[116] Ricardo J. Rabelo, Alexandra a. Pereira-Klen, and Edmilson R. Klen, *Effective management of dynamic and multiple supply chains*, International Journal of Networking and Virtual Organisations **2** (2004), no. 3, 193.

[117] FA Rahman, MI Desa, and Antoni Wibowo, *A review of KDD-data mining framework and its application in logistics and transportation*, Networked Computing and ... (2011), 175–180.

[118] María F Rey, *Encuesta nacional logística*, (2008).

[119] A Ruiz-Rua and A Calatayud, *Mejores prácticas en logística internacional*, Banco Interamericano de Desarrollo (2012).

[120] B. Ryu, D.S. Kim, A.M. DeLuca, and R.M. Alani, *Comprehensive expression profiling of tumor cell lines identifies molecular signatures of melanoma progression*, PLoS ONE **2** (2007), no. 7, cited By 0.

[121] B.S. Sahay and Jayanthi Ranjan, *Real time business intelligence in supply chain analytics*, Information Management & Computer Security **16** (2008), no. 1, 28–48.

[122] G. Salton, *Automatic text processing*, Addison Wesley (1989).

[123] SAP, *Welcome — sap hana*.

[124] Linderman M.D. Sorenson J. Lee L. Nolan Schadt, E.E., *Computational solutions to large-scale data management and analysis*, NIH Public Access **11** (2011), no. 9, 647–657.

[125] Klaus Schwab, *The Global Competitiveness Report*, Tech. report, World Economic Forum, 2015.

[126] Farzad Shafiei and David Sundaram, *Multi-enterprise collaborative enterprise resource planning and decision support systems*, 37th Annual Hawaii International Conference on System Sciences, 2004. Proceedings of the, vol. 00, IEEE, 2004, p. 10 pp.

[127] Farzad Shafiei, David Sundaram, and Selwyn Piramuthu, *Multi-enterprise collaborative decision support system*, Expert Systems with Applications **39** (2012), no. 9, 7637–7651.

[128] Tipawan Silwattananusarn, *Data mining and its applications for knowledge management : A literature review from 2007 to 2012*, International Journal of Data Mining & Knowledge Management Process **2** (2012), no. 5, 13–24.

[129] Z. Song and A. Kusiak, *Optimising product configurations with a data-mining approach*, International Journal of Production Research **47** (2009), no. 7, 1733–1751.

[130] Samir K. Srivastava, *Green supply-chain management: A state-of-the-art literature review*, International Journal of Management Reviews **9** (2007), no. 1, 53–80.

[131] Sundar Swaminathan, *The effects of big data on the logistics industry*, Oracle (2012), 3–4.

[132] Andreas L Symeonidis, Dionisis D Kehagias, and Pericles A Mitkas, *Intelligent policy recommendations on enterprise resource planning by the use of agent technology and data mining techniques*, Expert Systems with Applications **25** (2003), no. 4, 589–602.

[133] Colin Tankard, *Big data security*, Network Security **2012** (2012), no. 7, 5–8.

[134] Frédéric Thiesse, Christian Floerkemeier, Mark Harrison, Florian Michahelles, and Christof Roduner, *Technology, standards, and real-world deployments of the epc network*, IEEE Internet Computing **13** (2009), no. 2, 36–43.

[135] Douglas J Thomas, Douglas J Thomas, Paul M Griffin, and Paul M Griffin, *Coordinated supply chain management*, European Journal of Operational Research **94** (1996), no. 96, 1–15.

[136] Sébastien Thomassey, *Sales forecasts in clothing industry: The key success factor of the supply chain management*, International Journal of Production Economics **128** (2010), no. 2, 470–483.

[137] Chitriki Thotappa and K Ravindranath, *Data mining aided proficient approach for optimal inventory control in supply chain management*, Proceedings of the World Congress on Engineering **I** (2010).

[138] Rui Tian, Zhaosheng Yang, and M Zhang, *Method of road traffic accidents causes analysis based on data mining*, Computational Intelligence and ... (2010), 4–7.

[139] EL TIEMPO, *Mintransporte impone millonaria sanción al runt*, Portafolio (2013).

[140] Peter Trkman, Kevin McCormack, Marcos Paulo Valadares de Oliveira, and Marcelo Bronzo Ladeira, *The impact of business analytics on supply chain performance*, Decision Support Systems **49** (2010), no. 3, 318–327.

[141] Tzu-Liang Tseng, *Performance evaluation for pull-type supply chains using an agent-based approach*, American Journal of Industrial and Business Management **03** (2013), no. 01, 91–100.

[142] Mihaela Filofteia Tutunea and Rozalia Veronica Rus, *Business intelligence solutions for sme's*, Procedia Economics and Finance **3** (2012), no. 12, 865–870.

[143] V Venkatesh, MG Morris, GB Davis, and FD Davis, *User acceptance of information technology: Toward a unified view*, MIS quarterly **46** (2003), no. 1, 75–81.

[144] Guohui Wang and T. S. Eugene Ng, *Big data, but are we ready?*, Trelles,Oswaldo,Pjotr, Prins (2010), 1–9.

[145] Barnes D. Rosenberg D. Luo X.X. Wu, C., *An analytic network process-mixed integer multi-objective programming model for partner selection in agile supply chains*, Production Planning and Control **20** (2009), no. 3, 254–275.

[146] Chong Wu and David Barnes, *A literature review of decision-making models and approaches for partner selection in agile supply chains*, Journal of Purchasing and Supply Management **17** (2011), no. 4, 256–274.

[147] Desheng Wu, *Supplier selection: A hybrid model using dea, decision tree and neural network*, Expert Systems with Applications **36** (2009), no. 5, 9105–9112.

[148] Jui-Yu Wu, *Computational intelligence-based intelligent business intelligence system: Concept and framework*, 2010 Second International Conference on Computer and Network Technology (2010), 334–338.

[149] Xindong Wu, Vipin Kumar, J. Ross Quinlan, Joydeep Ghosh, Qiang Yang, Hiroshi Motoda, Geoffrey J. McLachlan, Angus Ng, Bing Liu, Philip S. Yu, Zhi-Hua Zhou, Michael Steinbach, David J. Hand, and Dan Steinberg, *Top 10 algorithms in data mining*, Knowledge and Information Systems **14** (2007), no. 1, 1–37.

[150] Tito Yepes, Juan Mauricio Ramírez, and Leonardo Villar, *Infraestructura de Transporte en Colombia: ¿luz al final del túnel?*, 9° Congreso Nacional de la Infraestructura - Cámara Colombiana de la Infraestructura (2012), 61.

[151] Kai Zhao and Xin Yu, *A case based reasoning approach on supplier selection in petroleum enterprises*, Expert Systems with Applications **38** (2011), no. 6, 6839–6847.