



UNIVERSIDAD NACIONAL DE COLOMBIA

Modelo de aprendizaje para estructurar los datos de las hojas de vida de maquinaria amarilla

Lina Fernanda Prado Gamba

Universidad Nacional de Colombia
Facultad de Ciencias, Departamento Matemáticas
Bogotá, Colombia
2022

Modelo de aprendizaje para estructurar los datos de las hojas de vida de maquinaria amarilla

Lina Fernanda Prado Gamba

Tesis presentada como requisito para optar al título de:
Magíster en Matemática Aplicada

Director:

Ph.D., en Ingeniería - Sistemas y Computación. Francisco Albeiro Gómez Jaramillo

Línea de Investigación:

Aprendizaje de máquina y matemática aplicada

Grupo de Investigación:

Computational Modeling of Biological Systems Research Group - COMBIOS

Universidad Nacional de Colombia
Facultad de Ciencias, Departamento Matemáticas
Bogotá, Colombia
2022

A mi madre.

Agradecimientos

En primer lugar quiero agradecer a mis padres y al profesor Francisco Albeiro Gómez cuya orientación y apoyo permitieron llevar a cabo este proyecto.

También quiero agradecer al Ing. Wilfredo Campos y a la empresa McAllister SAS por facilitarme los datos y brindarme una guía sobre las necesidades de la industria.

Por último, quiero agradecer al Ing. Felipe Torres quien me ayudó con la evaluación del etiquetado, con el entendimiento de la terminología y el funcionamiento de los equipos.

Índice general

Agradecimientos	vii
Resumen	xi
Lista de figuras	xiii
Lista de tablas	1
1. Introducción	2
1.1. Objetivo	3
1.2. Justificación	3
2. Estado del arte	5
2.1. Estrategias y actividades de mantenimiento	6
2.1.1. Criterios de optimización	8
2.1.2. Key performance indicators	8
2.2. Minería de texto	9
2.3. Natural language processing	10
2.3.1. Preprocesamiento	10
2.3.2. Representación	11
2.4. NLP en el área de mantenimiento	15
3. Materiales y métodos	20
3.1. Datos	21
3.1.1. Hojas de vida	21
3.2. Etiquetado	25
3.3. Preprocesamiento y vectorización	36
3.3.1. Escalamiento	37
3.4. Estructuración automática	37
3.4.1. Métricas de evaluación	37
3.4.2. Prueba de permutación	45
3.5. KPIs	45
4. Resultados	48
4.1. Tipología	51

4.2. Falla encontrada	52
4.3. Estado final	54
4.4. Sistema	55
4.5. Prueba de permutación	57
4.6. KPIs	57
5. Conclusiones y recomendaciones	60
A. Análisis inicial de datos	62
B. Variación de Representación	65
C. Reemplazos	67
Bibliografía	69

Resumen

Modelo de aprendizaje para estructurar los datos de las hojas de vida de maquinaria amarilla

La falta de información relevante para la toma de decisiones es uno de los grandes problemas a los que se enfrentan los departamentos de mantenimiento en las empresas. Esta tesis explora un método automático para aportar a la solución de este problema mediante la extracción de información relevante de los registros históricos de las actividades de mantenimiento realizadas en los equipos. Dada la naturaleza de los datos, texto no estructurado con lenguaje técnico, se plantea la implementación de diferentes representaciones (*bag of words*, *term frequency-inverse document frequency*, *Fasttext* y *Doc2vec*) para alimentar los modelos de aprendizaje de máquina que realizan la estructuración de información importante contenida en los documentos. En la búsqueda del modelo con mejor rendimiento se compararon modelos de *support vector machine*, *random forest*, *gaussian naive bayes* y *gradient boosting trees*. Estos modelos se aplicaron a datos provenientes de un negocio de venta y renta de maquinaria amarilla; se consideraron 15 montacargas de 3 modelos diferentes y 4 variables independientes en las cuales se extrae información: tipología, falla encontrada, estado final y sistema. Los modelos con mejor rendimiento alcanzaron un *f1-score macro* **0,86**, **0,8**, **0,81** y **0,68** con 3 *support vector machine* y un *gradient boosting trees*. Se concluye que para obtener mejores resultados el paso a seguir es aumentar la base de datos y expandir el campo de aplicación.

Palabras claves: Mantenimiento, Registros de mantenimiento, Extracción de información, Aprendizaje de máquina y Procesamiento de lenguaje natural.

Abstract

Machine learning model to structure yellow machinery logs

The lack of relevant information for decision-making is one of the major problems that maintenance departments face. In this thesis, an automatic method is explored to contribute to the solution of this problem by extracting relevant information from the records that are kept of the maintenance activities carried out on the equipment. Given the nature of the data, unstructured text with technical language, the implementation of different representations (*bag of words*, *term frequency-inverse document frequency*, *Fasttext* and *Doc2vec*) is proposed for the machine learning models that carry out the structuring of relevant information contained in the documents. In the search for the best performing model, *support vector machine*, *random forest*, *gaussian naive bayes* and *gradient boosting trees* models were compared. The models were applied to data from a business of sale and rental of yellow machinery; 15 forklifts of 3 different models and four independent variables in which information is extracted were considered: typology, fault found, final state and system. The best performing models achieved *f1-score macro* **0,86**, **0,8**, **0,81** y **0,68** with 3 *support vector*

machines and one gradient boosting trees. It is concluded that to obtain better results the next step is to increase the database and expand the field of application.

Keywords: Maintenance, Maintenance logs, Information extraction, Machine learning and Natural language processing.

Lista de Figuras

3-1. Proceso de minería de texto. Inicia con las DTRs las cuales se etiquetan y se preprocesan. Después se vectorizan y se utilizan para construir modelos de aprendizaje de máquina supervisado. Posteriormente se afinan los parámetros, se elige el modelo con mejor rendimiento, se aplica en los datos y finalmente se calculan los KPIs.	20
3-2. Ejemplo de las hojas de vida en la que se almacenan los registros	21
3-3. Flujo de trabajo y los documentos adquiridos en cada uno de los pasos. Inicia con una avería en un equipo, se realiza una solicitud inicial, luego se agenda un trabajo y se genera una OT. Luego se realiza la reparación y se registran las actividades realizadas en la hoja de vida.	24
3-4. Flujo de etiquetado compuesto por tres fases. En la primera el anotador asigna las etiquetas a las DTRs, luego el evaluador de calidad (QA) evalúa los resultados y en caso de encontrar incongruencias o errores se comunica con el anotador. Después de realizar las correcciones necesarias decide si aprobar el trabajo o realizar más ajustes.	26
3-5. Esquema de etiquetado. Incluye cuatro etiquetas con sus correspondientes clases.	26
3-6. Matriz de confusión binaria donde se evidencia el rendimiento de un modelo, contiene el conteo de las instancias positivas y negativas correctamente etiquetadas (VP y VN). Adicionalmente incluye el cálculo de los registros positivos y negativos etiquetadas erróneamente (FP y FN)	38
4-1. Distribución de las clases de etiquetado para tipología, falla encontrada, estado final y sistema. Se evidencia las clases dominantes: correctivo (tipología), verdadero (estado final) y chasis (sistema).	48
4-2. Partición de entrenamiento y de evaluación. Distribución de las clases para cada una de las variables en la partición 70 a 30.	49
4-3. Matriz de confusión del modelo seleccionado para tipología sobre los datos de evaluación.	52
4-4. Distribución de los puntajes se SVM al variar las representaciones. Se evidencia una distribución similar entre métricas al variar las vectorizaciones.	53
4-5. Matriz de confusión del modelo seleccionado para falla encontrada sobre el conjunto de evaluación.	54

4-6. Matriz de confusión del modelo seleccionado para estado final sobre el conjunto de evaluación.	55
4-7. Matriz de confusión del modelo seleccionado para sistema sobre el conjunto de evaluación.	56
4-8. Evolución del ensamblaje seleccionado para sistema al ir agregando estimadores. 56	
4-9. Completitud de las DTRs con respecto a la falla encontrada y estado final. El KPI denotado como C que cuantifica la completitud de las DTRs con respecto a falla encontrada y estado final.	57
4-10. Tipología de las DTRs. Combinación de los KPIs CT_{clase} para las clases de la variable tipología. Adicionalmente se complementa con el conteo de actividades correctivas por trimestre.	58
4-11. Caracterización de correctivos. Combinación de los KPIs denotados como CC_{clase} para todas las clases de la variable sistema.	58
4-12. Promedio de días de indisponibilidad. El KPI denotado como TI, complementado con la suma total días de indisponibilidad.	59
A-1. Comparación de la distribución de las palabras por documento de los textos originales, los LPP y los APP.	62
A-2. Nubes de palabras en las que el tamaño de las palabras es proporcional a su frecuencia de ocurrencia.	63

Lista de Tablas

3-1.	Estructura de las hojas de vida de los equipos compuesta por cuatro secciones: información general, servicios prestados, trabajos realizados y repuestos. Contiene el nombre de las variables y sus respectivas descripciones.	23
3-2.	Definiciones de los KPI. Incluye nombre, objetivo bajo el cual se plantea, la descripción, el tipo de variable y notas adicionales.	46
4-1.	Valores alcanzados con la representación y el modelo de línea base. Puntajes obtenidos por SVM sobre la representación BoW de los textos APP y LPP. La métrica de entrenamiento corresponde a G-mean para tipología y estado final, a accuracy para falla encontrada y F1-score macro para sistema. . .	50
4-2.	Valores de las métricas alcanzados por el modelo de línea base en los DRT LPP y APP. Considerando el puntaje de entrenamiento más alto (resaltado en amarillo), se seleccionaron las siguientes representaciones, representación 8 (APP) (tipología), representación 10 (LPP) (falla encontrada y estado final) y representación 1 (APP) (sistema).	51
4-3.	Resultados de GNB, RF, SVM y GBT para tipología representación 8 (APP). El G-mean más alto (0,84) se alcanza con SVM con parámetros ajustados. .	51
4-4.	Resultados de GNB, RF, SVM y GBT para falla encontrada representación 10 (LPP). El accuracy más alto (0,8) se alcanza con SVM con parámetros ajustados.	53
4-5.	Resultados de GNB, RF, SVM y GBT para estado final representación 10 (LPP). El G-mean más alto (0,81) se alcanza con SVM con parámetros ajustados.	54
4-6.	Resultados de GNB, RF, SVM y GBT para sistema representación 1 (APP). El F1-score más alto (0.72) se alcanza con GBT con parámetros ajustados. .	55
B-1.	Modelos utilizados para generar las representaciones junto con los parámetros seleccionados.	66
C-1.	Lista de reemplazos realizados en el texto para la lematización y la corrección ortográfica.	68

1. Introducción

Los recientes avances en la adquisición y procesamiento de datos han abierto las puertas para grandes cambios en la industria. Junto con estos avances ha surgido la necesidad de herramientas que permitan el aprovechamiento de los datos. Al ser diseñadas para suplir las exigencias particulares del negocio al que van dirigidas pueden llegar a tener un gran impacto en las empresas. En este trabajo se propone crear un modelo matemático, que genere valor agregado a un negocio de renta y venta de maquinaria amarilla. Este tipo de equipos corresponden a maquinaria de grandes dimensiones y complejidad [de Academias de la Lengua Española y Real Academia Española, 2021], incluye equipos de construcción, de excavación y montacargas, se conocen como maquinaria amarilla debido a que usualmente llevan ese color.

Una tarea fundamental en las empresas de renta de maquinaria es la estimación de la cuota mensual que se cobra a los arrendatarios. Esta cuota abarca los costos base, los cuales corresponden a los gastos conocidos que seguramente tendrán lugar en el plazo establecido del contrato, por ejemplo, mantenimientos preventivos programados y seguros. También cubre los costos con relación a las posibles contingencias que se presenten, por ejemplo, mantenimientos correctivos. El costo de estas contingencias se cuantifica por medio de cálculos aproximados y se agregan al presupuesto. Usualmente, estos cálculos se basan en conocimiento empírico de la maquinaria, por lo cual cabe la posibilidad que los desfases entre el valor estimado y los gastos reales, resulten en pérdidas económicas u obstáculos para mejorar la competitividad de estas empresas en el mercado.

En caso de que la empresa decida modificar la cuota mensual, debido a que los costos base son fijos, es necesario recurrir a los costos relacionados con contingencias. Para lograrlo se requiere hacer cambios en las estrategias de mantenimiento, ya sea disminuyendo la cantidad de actividades o reduciendo la incertidumbre que rodea a las contingencias.

Con el objetivo de reducir el presupuesto de la cuota correspondiente a mantenimiento artículos recientes proponen recurrir a estrategias de mantenimiento predictivas. Para desarrollar este tipo de estrategias o explorar otras alternativas que permitan modificar la cuota es indispensable considerar la información relevante disponible; una fuente de datos importante y poco usada son los registros de mantenimiento [Brundage et al., 2021].

En general, las empresas cuentan con un sistema de gestión y en este almacenan los registros de mantenimiento en las hojas de vida de la maquinaria. Las hojas de vida de la maquinaria son documentos en los que se describen, en lenguaje natural, las intervenciones realizadas en los equipos. Como el personal de mantenimiento registra los aspectos de la visita que llaman

su atención y que en su opinión son relevantes los registros pueden contener cualquier palabra o número [Bokinsky et al., 2013], es decir, que no son estructurados [Stenström et al., 2015]. Sin embargo, las hojas de vida también incluyen datos estructurados como: modelo, fecha de adquisición, entre otros. En la literatura, se ha demostrado que los campos en los que los operarios describen los procesos de mantenimiento contiene información relevante, aunque sea difícil su procesamiento [Stenström et al., 2015].

El problema de no considerar la parte no estructurada de los registros en la toma de decisiones se puede ver a través de una analogía médica: suponga que, para diagnosticar quejas de indigestión y dificultad para respirar, un médico ignorara notas anteriores, estudios o historial de salud y sólo usara pruebas de diagnóstico. Al estudiar las pruebas de diagnóstico, sin un contexto adicional proporcionado por los registros, este médico podría perder conexiones más allá de los problemas estomacales, como por ejemplo un ataque cardíaco y podría tomar decisiones que perjudicarían a su paciente [Brundage et al., 2021]. Lo mismo podría pasar con una máquina amarilla, por no considerar los registros, las decisiones tomadas podrían no ser las indicadas, lo cual implicaría gastos innecesarios o el deterioro del equipo. Este problema ocasiona un vacío en los procesos de mantenimiento en las empresas del país y en el análisis de lenguaje técnico en idioma español, vacío que se podría llenar con este trabajo.

1.1. Objetivo

Construir un modelo de aprendizaje supervisado para estructurar los registros de las descripciones de las intervenciones realizadas en la maquinaria de tal manera que el modelo generado permita soportar la toma de decisiones en el departamento de mantenimiento. Se define el proceso de estructuración de los registros como la asignación de las siguientes características: tipología, reconocimiento de fallas, estado final y sistema. En el capítulo de materiales y métodos se describirán cada uno de estos atributos en detalle.

1.2. Justificación

Teniendo en cuenta que la falta de información relevante es uno de los principales problemas para la toma de decisiones en los procesos de mantenimiento [Parida, 2006] y que la estructuración de los datos, permitiría una extracción eficiente de información valiosa, se considera que este trabajo será de gran interés para las empresas de renta y en general para el mantenimiento industrial.

Se espera que la información extraída permita el análisis de la completitud de la descripción, la confiabilidad de los equipos, facilite la caracterización de las fallas, la evaluación de la efectividad de las actividades y cumplimiento de los planes de mantenimiento.

Mediante la estructuración automática de los registros sería posible soportar la toma de decisiones con respecto a la contratación del servicio de maquinaria en contextos similares y

mejorar la tarifa de mantenimiento para ser más competitivos en el mercado o reducir posibles pérdidas económicas. Además, esta información puede reflejar la necesidad de capacitaciones adicionales para los operarios.

En el presente proyecto se emplea *natural language processing* (NLP) para transformar el texto no estructurado en una representación adecuada para el análisis o para implementar modelos de aprendizaje de máquina orientados a mejorar la gestión del mantenimiento.

Debido a que hay diferencias estructurales entre el lenguaje usado en los registros de mantenimiento y el lenguaje usual para el cual se han desarrollado la mayoría de herramientas de NLP, al procesar los registros relacionados con mantenimiento, se presentan diferentes retos. Estos retos incluyen palabras importantes frecuentemente abreviadas o mal escritas, palabras concatenadas, vocabulario especializado propio del área de mantenimiento, palabras en otro idioma (inglés) o anglicismos, inconsistencias en las abreviaciones y en el uso de acrónimos [Bokinsky et al., 2013].

Adicionalmente, según la revisión de la literatura, aunque existen trabajos en los que procesan los registros de mantenimiento, estos no se han desarrollado en idioma español y hasta donde se tiene conocimiento no se han implementado en el país. Por lo cual, con el desarrollo de este trabajo se espera contribuir en la eliminación de este vacío, dando la posibilidad a las empresas de incorporar información para la toma de decisiones.

Con el fin de ilustrar el uso de la información estructurada, además del problema de estructuración, en el presente documento se generan *key performance indicators* (KPIs), mediante de los cuales sería posible soportar la toma de decisiones en el área de mantenimiento de la empresa que brinda los datos.

El resto de la tesis se estructura de la siguiente manera: en el siguiente capítulo se define terminología clave para el desarrollo de la propuesta junto con sus antecedentes, en el tercer capítulo se describen los datos y el proceso implementado para cumplir el objetivo planteado. En el cuarto se exponen los resultados obtenidos por el etiquetado, el modelado y la aplicación de los KPIs. En el quinto se relatan las conclusiones y se dan recomendaciones para trabajos futuros.

2. Estado del arte

Las empresas dedicadas a la renta y venta de maquinaria amarilla pertenecen a una línea de negocio en la que se ofrece el servicio de diferentes máquinas como: montacargas, plataformas para trabajo en altura, *telehandlers*, entre otros, a cambio de una cuota mensual. Estas empresas operan sobre contratos de tres a cinco años, en los cuales la compañía que renta la maquinaria se encarga de los costos que conlleva la titularidad del equipo [Mcallister, 2021]. La renta de maquinaria es una práctica altamente usada en la industria y ofrece beneficios económicos, financieros, operativos y administrativos. No afecta el nivel de endeudamiento, ni se registra la maquinaria como activo, depreciación o pasivo, brinda un alivio fiscal, administrativo y contable. Facilita la renovación tecnológica, ofrece una gran variedad de opciones de maquinaria con cero horas de trabajo, a la cual se le hace seguimiento en tiempo real y no es necesario preocuparse por el personal técnico. Adicionalmente elimina costos inciertos [Mcallister, 2021].

Los equipos rentados se desgastan debido al uso y a la exposición a los factores ambientales. Con el tiempo empiezan a presentar fallas, las cuales pueden comprometer la seguridad del personal, la calidad o capacidad de realizar trabajos, entre otros. Hace algunas décadas, la gestión de los bienes se manejaba realizando actividades en los equipos después de que se presentaba la falla, con el tiempo las empresas se han aprendido que la gestión efectiva de mantenimiento aumenta la eficiencia y la confiabilidad de los equipos, lo cual a su vez permite evitar los tiempos de indisponibilidad (tiempo muerto) no planificados y producción de mala calidad [de Jonge and Scarf, 2020].

Con el propósito de optimizar la gestión de mantenimiento, definida como “todas las actividades de la gestión que determinan los objetivos de mantenimiento, las estrategias y las responsabilidades, y las realizan por medio de la planificación del mantenimiento, control y supervisión del mantenimiento, mejora de los métodos en la organización incluyendo los aspectos económicos” [Único organismo de normalización en España, 2021], se han desarrollado diversos modelos matemáticos. Como lo mencionan en [de Jonge and Scarf, 2020] el desarrollo de estos modelos suele ser desafiante debido que deben ser interpretados, tienen que estar acompañados de un sistema de soporte de decisiones y un conjunto de datos.

Para la administración del área de mantenimiento se manejan diferentes estrategias y tipos de actividades. Dependiendo de la estrategia de mantenimiento establecida se eligen y se programan las actividades.

Las estrategias de mantenimiento de la empresa son establecidas por el área de mantenimiento. Esta área es la encargada de garantizar el buen funcionamiento y prolongar la vida

útil de los bienes; según la asociación española de normalización el mantenimiento engloba “las acciones técnicas, administrativas y de gestión durante el ciclo de vida de un elemento, destinado a conservarlo o a devolverlo a un estado en el cual pueda desarrollar la función requerida” [Único organismo de normalización en España, 2021]. Es un área muy importante; las acciones de mantenimiento no deberían ser vistas como un costo, sino como una inversión que puede asociarse al crecimiento futuro de los ingresos de la empresa [Prasertrunguang and Hadikusumo, 2007].

2.1. Estrategias y actividades de mantenimiento

Las estrategias de gestión de mantenimiento se pueden agrupar en tres categorías según la manera en que se manejan las fallas y los cuidados de los equipos [Ding and Kamaruddin, 2015]. Se tienen estrategias correctivas, preventivas y predictivas [Carvalho et al., 2019] y se detallan a continuación:

1. Las estrategias de mantenimiento correctivas son aquellas en las que las acciones se llevan a cabo después de que se presenta la falla. El problema de éstas es que la maquinaria se detiene y los tiempos muertos son largos y costosos, puede causar grandes daños al sistema, los operarios o al ambiente [de Jonge and Scarf, 2020] [Ding and Kamaruddin, 2015].

Los tiempos de la avería incluyen el tiempo de agendamiento, el tiempo que se demora el técnico en llegar y reparar la máquina, y en caso de que se necesiten repuestos, se requiere adicionar el tiempo de importación o de envío de las piezas. Por otro lado, debido a que se espera hasta que las fallas sean visibles, en algunos casos, esto implica que éstas sean de mayor magnitud siendo la aglomeración de fallas previas tratables [Çınar et al., 2020].

2. Las preventivas son las que se llevan a cabo en periodos de tiempo predeterminados, o después de cierta cantidad de horas de uso de la maquinaria. Se utiliza para reducir y hasta prevenir las fallas. Aunque es la estrategia más usada, disminuye los tiempos muertos y los gastos relacionados con la estrategia anterior, tiene la desventaja de que incluyen actividades innecesarias que implican gastos evitables [Çınar et al., 2020] [Ding and Kamaruddin, 2015].
3. Las predictivas son aquellas en las que se agendan las actividades de mantenimiento a partir de la predicción de las fallas de los equipos, aunque para implementarlas es necesario contar con una gran cantidad de datos y con personal capacitado para el análisis de datos. Permite desarrollar herramientas para la optimización de inventarios, de agendamientos y disminuir tiempos muertos, actividades de mantenimiento, costos asociados, alargar la vida útil y mejorar la calidad del servicio [Çınar et al., 2020].

Por otro lado, las actividades de mantenimiento se distribuyen sobre las siguientes opciones [Único organismo de normalización en España, 2021]:

- Inspección: actividad en la que revisa el equipo para evaluar su estado, realizando mediciones, observaciones, ensayos o calibraciones.
- Monitoreo: acción que se realiza manual o automáticamente en la que se observa el estado actual de un elemento. Se distinguen de las inspecciones dado a que éstas se realizan para analizar el cambio de parámetros en un periodo de tiempo.
- Ensayo de conformidad: pruebas que se realizan con el objetivo de evidenciar si una propiedad o característica del equipo es conforme a las especificaciones establecidas.
- Verificación de la función: actividad realizada posteriormente a las acciones de mantenimiento para confirmar que el equipo puede realizar la función requerida.
- Mantenimiento de rutina: eventos regulares, repetitivos o sencillos en los que no es necesario solicitar autorizaciones ni herramientas. Incluye limpiezas y lubricaciones, entre otras.
- Revisión: grupo de exámenes y acciones realizadas para mantener la seguridad y la disponibilidad del equipo. Abarca situaciones en las que sea necesario el desmonte parcial o completo de los componentes del equipo.
- Reconstrucción: actividad realizada después de un desmonte, incluye modificaciones y/o mejoras.
- Reparación: acción física realizada para recuperar la función requerida dado la falla de un elemento. En caso de que el elemento no sea reparable, las acciones de mantenimiento correspondientes siempre son reemplazos.
- Reparación temporal: similar a una reparación con la diferencia de que en este caso la función requerida se restablece por un periodo de tiempo limitado hasta que se realiza la reparación.
- Diagnóstico de averías: tarea que se lleva a cabo con el objetivo de reconocer y caracterizar las averías identificando localización y causa de la falla. La localización se refiere a la búsqueda del nivel apropiado para realizar la intervención requerida.
- Mejora: acciones en las que espera aumentar la seguridad de funcionamiento, sin comprometer la función del equipo.

A grandes rasgos las actividades de mantenimiento se distinguen entre preventivas y correctivas, que en concordancia con las estrategias de mantenimiento se relacionan con el momento

en que se realizan. Las preventivas se ejecutan en los elementos a consecuencia de parámetros previamente establecidos (como tiempo de trabajo) o motivadas por la condición, y las correctivas son provocadas por la presencia de una avería.

Entre las actividades preventivas se distinguen subcategorías como *time-based* o *condition-based* que incluye las acciones pronósticas y las predictivas, las cuales dependen de la información del estado de los equipos. A su vez, se puede diferenciar por la manera de monitorear el estado del equipo, puede tenerse un seguimiento continuo o uno realizado mediante inspecciones (agendadas con anterioridad o de manera dinámica).

2.1.1. Criterios de optimización

Los criterios de optimización del departamento de mantenimiento dependen de las estrategias de las empresas. Algunos ejemplos son la minimización de la tasa de costos, la maximización de la disponibilidad a largo plazo [de Jonge and Scarf, 2020], la maximización de la confiabilidad del equipo (tiempo que puede durar operando el equipo sin presentar una falla) y la disponibilidad (cuantificada como la probabilidad de que pueda operar en cierto grado de tolerancia en un momento determinado) [Nakagawa, 2006].

Para evaluar la confiabilidad del equipo se puede recurrir a indicadores como el tiempo promedio entre fallas, disponibilidad y número esperado de fallas.

Para el modelamiento del estado de las piezas se pueden considerar diferentes conjuntos de opciones, como por ejemplo conjuntos de dos estados (funcional o no funcional), con tres estados (bueno, defectuoso o no funcional) o una elección entre opciones continuas de estados. Para el presente trabajo, se considera el primer conjunto de estados que se define como la pareja de estado funcional y no funcional.

Por otro lado, en el área de mantenimiento se han desarrollado modelos para cuantificar los efectos de las políticas de mantenimiento de la empresa que requieren grandes cantidades de datos que usualmente no están disponibles. Además, debido a carencias de evidencia empírica no se suelen emplear en la industria [Lundgren et al., 2018].

El problema relacionado con la disponibilidad de los datos también afecta la utilización de métricas de rendimiento del departamento, las cuales se utilizan para identificar las falencias y resaltar los aspectos a mejorar. Son medidas que se utilizan para valorar, revisar las políticas y técnicas de mantenimiento, para evaluar la asignación de recursos y la seguridad de los equipos [Kumar et al., 2013].

2.1.2. Key performance indicators

Los key performance indicators (KPIs) son métricas que se utilizan para soportar las decisiones, se construyen a partir de las necesidades y los datos disponibles. Son indicadores alineados con las metas de la empresa y se utilizan para evaluar y comunicar el desempeño del área.

Debido a que se utilizan como soporte para la toma de decisiones suelen ser calculados mediante fórmulas intuitivas altamente interpretables. Para su construcción existen diferentes metodologías, en particular se puede empezar por los objetivos y estrategias de la empresa, seguido por la identificación de los factores claves de éxito, con la planeación de los indicadores, la determinación del umbral de aceptación, el diseño de la fórmula, la asignación de recursos, la medición y el ajuste. Y finaliza con la formalización, la implementación y las mejoras continuas [Beltrán, 2000].

La implementación de estas métricas le permite a los tomadores de decisiones saber si se está realizando lo correcto correctamente. El mantenimiento de equipos de gran inversión, que alguna vez se pensó que era un mal necesario, ahora se considera clave para mejorar la rentabilidad de una operación, creando valor adicional al brindar mejores y más innovadores servicios al cliente [Kumar et al., 2013].

Comúnmente las empresas de renta de maquinaria operan bajo una estrategia de mantenimientos preventivos considerando la recomendación del fabricante de los equipos, con mantenimientos adicionales pactados en los contratos. A pesar de que la estrategia es preventiva, el área de mantenimiento continúa prestando servicios correctivos en caso de fallas inesperadas de la maquinaria.

Hay diferentes herramientas para analizar los datos producidos en las actividades de mantenimiento. En particular se puede utilizar la minería de texto.

2.2. Minería de texto

La minería de texto es un conjunto de técnicas que se utilizan para procesar y explorar grandes cantidades de datos. Estas herramientas han estado presentes desde mediados de la década de 1980, donde se realizaban los trabajos de manera manual. Con el paso del tiempo y gracias a los avances tecnológicos el área se ha desarrollado en gran medida, actualmente incluye técnicas que abarcan la recuperación de la información, aprendizaje automático, *natural language processing* (NLP), entre otros [Ghosh et al., 2012].

La minería de texto es el descubrimiento de nueva información, que antes no se conocía, mediante la extracción automática de información de recursos escritos [Pelham and Hockley, 2017]. Este tipo de aproximación se ha utilizado en finanzas, biología molecular, entre otras [Zhang et al., 2015].

Se estima que aproximadamente el 80% de la información se almacena por medio de texto, por lo cual la minería de texto tiene un gran potencial para generar valor agregado a las empresas [Ghosh et al., 2012].

2.3. Natural language processing

Es un campo que busca que los computadores entiendan el lenguaje natural como lo hacen las personas [Szűcs and Ballagi, 2020] [Ghosh et al., 2012].

Corresponde a un conjunto de técnicas computacionales que han sido aplicadas a múltiples áreas del conocimiento, entre las cuales encontramos cuidado de la salud, publicidad y redes sociales, entre otros [Szűcs and Ballagi, 2020]. Las técnicas de NLP se han utilizado en diferentes problemas, incluyendo, traducciones automáticas, reconocimiento de enfermedades, reconocimiento de discurso, generación de texto y análisis de sentimientos, entre otros [Hirschberg and Manning, 2015][Luque et al., 2019].

Una rama de NLP es la extracción de información, ésta ha sido desarrollada para aislar información de fuentes de texto no estructurado, no obstante la mayoría de sus aplicaciones son para idioma inglés. Los avances en el procesamiento de lenguaje técnico taquigráfico¹ se encuentran principalmente en el análisis de registros médicos [Bokinsky et al., 2013]. Sin embargo, se han llevado a cabo pocas investigaciones sobre la extracción de información en el dominio de registros de mantenimiento [McKenzie et al., 2010][Brundage et al., 2021].

2.3.1. Preprocesamiento

Debido a la naturaleza de los datos, textos en lenguaje natural no estructurados, es necesario realizar un conjunto de tareas que faciliten una representación apropiada. Entre las tareas más usuales en el preprocesamiento se encuentran las siguientes [Brundage et al., 2021][Bortolini and Forcada, 2020]:

Estandarización de formato

La estandarización de formato es una tarea que agrupa acciones como dejar todo el texto en minúsculas, retirar caracteres especiales (por ejemplo #@) y marcas diacríticas [Bortolini and Forcada, 2020]. De esta manera palabras iguales pero con acentos diferentes se identifican con la misma cadena de caracteres con el formato homogeneizado.

Eliminación de palabras de parada y puntuación

Mediante esta tarea se remueven las palabras frecuentes que no aportan información para caracterizar los registros, éstas se denotan como palabras de parada y se eliminan los signos de puntuación, los cuales, en vez de enriquecer la representación pueden generar ruido innecesario [Brundage et al., 2021].

Las palabras de parada corresponden a palabras altamente usadas en el idioma trabajado, usualmente incluye conjunciones, artículos y pronombres.

¹Se define en el diccionario de la lengua española como la “técnica de escribir tan deprisa como se habla, por medio de ciertos signos y abreviaturas”.

Lematización

Es un proceso por medio del cual se extrae la forma canónica (lema) de las formas flexionadas, es decir, variaciones de los lemas por medio de alteraciones que reflejan las distintas funciones de la palabra dentro de la oración y su relación con su contexto (por ejemplo conjugaciones). La lematización debe considerar la intención con la que se utiliza la palabra y asignar un representante por convención. Se puede seguir las normas que utilizan los diccionarios para su elección: singular para sustantivos, masculino singular para adjetivos e infinitivo para verbos.

Corrección ortográfica

Hace referencia a reemplazar las palabras mal escritas, en particular corregir aquellas en las que se omiten o se emplean caracteres incorrectos.

2.3.2. Representación

Posterior al preprocesamiento, en NLP generalmente se recurren a herramientas de representación que brindan una alternativa numérica del texto. A continuación se describen algunas de éstas.

Bag of Words

Bag of Words (BoW) es un método clásico para la representación numérica del texto, es una herramienta que se utiliza regularmente en aplicaciones de NLP y en particular en aplicaciones relacionadas con clasificación de texto. Permite generar una versión numérica en un vector de longitud fija [Le and Mikolov, 2014].

Inicialmente se construyen *tokens* particionando el texto en cada uno de los espacios entre palabras. Luego, se estructura un vector cuya longitud es igual a la cantidad de tokens presentes en todo el conjunto de datos. Cada una de las entradas debe corresponder a un token de manera fija y para la representación del texto las entradas deben almacenar la frecuencia del token en el texto correspondiente. En otras palabras, este método consiste en transformar textos en vectores que almacenan la frecuencia de las palabras. Se construye un diccionario con todas las palabras utilizadas emparejadas con su frecuencia de ocurrencia en el texto que se quiere representar. Las BoW se pueden ver como vectores, en los que cada entrada corresponde a la frecuencia de una palabra del diccionario. Para la representación de múltiples documentos se realiza la representación en vectores que mantienen el mismo orden de correspondencia entre entradas y palabras, y se construyen matrices donde cada fila corresponde a un documento.

Siguiendo el anterior proceso es posible representar el conjunto de datos como una matriz donde cada fila corresponde a una descripción de un trabajo realizado y cada columna esta asociada con un token.

Esta técnica no es muy eficiente si se aplica al texto original, debido a que procesa las conjugaciones y errores ortográficos como diferentes palabras, adicionalmente le puede dar un gran peso a palabras que no son relevantes para la representación del texto (como por ejemplo el artículo “el”) debido que estas suelen ser usadas continuamente en los escritos [Usuga Cadavid et al., 2020]. Con el objetivo de mitigar este tipo de complicaciones es necesario llevar a cabo actividades de preprocesamiento del texto, tales como lematización, corrección ortográfica, remoción de palabras de parada, puntuación y caracteres especiales. La representación lograda mediante este método se caracteriza porque en general los textos contienen un subconjunto pequeño de los tokens y la representación obtenida son vectores dispersos, en los cuales la mayoría de entradas corresponden a ceros.

Aunque este método es intuitivo, fácil de interpretar y de analizar, y obtiene buenos resultados en algunas aplicaciones [Le and Mikolov, 2014], sus mayores desventajas son: las palabras pierden su ordenamiento al no considerar la posición relativa, ignora la semántica y los vectores resultantes suelen tener una gran dimensión [Le and Mikolov, 2014]. Además, requiere un preprocesamiento extenso del texto y cuando se implementa en datos que incluyan tokens que no se encontraban presentes en el corpus inicial, es necesario volver a representar los datos.

Term frequency-inverse document frequency

Term frequency-inverse document frequency (TF-IDF) es una variante de BoW en la que las entradas de los vectores están asociadas con la relevancia de las palabras en los textos. La relevancia de la palabra t en el documento d se cuantifica como $\text{tf-idf}(t, d)$ y es la frecuencia de los términos multiplicada por el inverso de la frecuencia de los documentos.

$$\text{tf-idf}(t, d) = \text{tf}(t, d) * \text{idf}(d), \quad (2-1)$$

donde,

$$\text{idf}(d) = \log \frac{1 + n}{1 + \text{df}(t)} + 1, \quad (2-2)$$

siendo n la cantidad de documentos en todo el conjunto y $\text{df}(t)$ la cantidad de documentos que contienen la palabra t . De esta manera las entradas con mayor peso corresponden a los tokens que aparecen más veces en el documento y menos veces en los otros registros, es decir que diferencian al texto del resto de documentos del conjunto de datos. Adicionalmente, se divide cada vector por su norma.

Fasttext

Es una herramienta de representación no contextual desarrollada por *Facebook Research* [Bojanowski et al., 2017]. Su objetivo es ayudar a construir soluciones escalables tanto para representación como para clasificación de texto. No requiere preprocesamiento del texto

y se puede aplicar a datos no incluidos en el conjunto de entrenamiento ya que maneja representaciones de subcadenas de las palabras. Esta vectorización se caracteriza por ser un método que implementa una red neuronal recurrente con 650 unidades de *Long short-term memory* (LSTM), es rápido y es capaz de retener información del lenguaje que le permite a los computadores procesar características lingüísticas como la similitud entre palabras y analogías [Bojanowski et al., 2017].

En términos de la vectorización esta herramienta ofrece modelos no supervisados, que capturan la morfología de la palabra mediante las subcadenas de caracteres que conforman la palabra, dependiendo de la longitud establecida (n) estas subcadenas se conocen como n -gramas.

En este método se busca que la representación de cada uno de los n -gramas que componen la palabra objetivo alimenten un modelo que prediga correctamente las palabras que aparecen en su contexto. Formalmente, considerando el corpus de tamaño T , formado por la serie de palabras w_1, \dots, w_T , se entrena un modelo para minimizar la *negative log-likelihood* relacionada con la presencia de la palabra w_c en el contexto de w_t .

$$\sum_{t=1}^T \left[\sum_{c \in C_t} \log(1 + e^{-s(w_t, w_c)}) + \sum_{n \in N_{t,c}} \log(1 + e^{s(w_t, n)}) \right], \quad (2-3)$$

optimizando esta fórmula con un gradiente descendiente estocástico [Bojanowski et al., 2017]. Teniendo en cuenta que C_t es el conjunto de índices que corresponden a las palabras que se encuentran alrededor del término correspondiente a w_t en el conjunto de entrenamiento y $N_{t,c}$ es un conjunto de muestras del vocabulario que no se encuentran representadas en C_t . Adicionalmente, se recurre a una función de puntaje que mapea la pareja w_t, C_t a \mathbf{R} , definida como el producto escalar de las representaciones de los n -gramas contenidos en la palabra correspondiente a w_t y a la representación v_{w_c} .

$$s(w_t, w_c) = \sum_{g \in G_{w_t}} z_g^\top v_{w_c}, \quad (2-4)$$

La extracción de los n -gramas se realiza añadiendo un símbolo especial al principio y al final de la palabra, luego se coleccionan las cadenas de caracteres de tamaño n , por último se le adiciona a la colección la palabra completa.

En caso de que sea necesario representar una palabra que no estaba presente en el conjunto de entrenamiento, se le asigna el promedio de los vectores de los n -gramas que la conforman. A su vez, la vectorización de los textos se obtiene promediando las representaciones normalizadas de los términos utilizados.

Fasttext ofrece dos tipos de modelos los que utilizan *skipgrams* y los que implementan *continuous bag of words* (CBOW). El primero corresponde al mencionado anteriormente y el

segundo funciona de manera similar pero se utilizan las palabras del contexto para predecir la palabra objetivo.

En [Bojanowski et al., 2017] propone utilizar los n -gramas con longitud mayor igual que 3 y menor igual que 6 para el idioma inglés y alemán, pero el parámetro se puede ajustar según las necesidades del usuario. Adicionalmente, esta librería ofrece modelos preentrenados para 294 idiomas, incluido el español, entrenados en textos de wikipedia.

Doc2Vec

Herramienta desarrollada por *Google Inc.* [Le and Mikolov, 2014] que por medio de un algoritmo no supervisado, basado en redes neuronales, se representan los documentos con vectores densos. Ofrece dos modelos los *Distributed Memory version of Paragraph Vector (PV-DM)* y *Distributed Bag of Words version of Paragraph Vector (PV-DBOW)*.

La construcción de PV-DM inicia con la creación de dos matrices D y W , en las columnas de la primera matriz se almacenan las representaciones únicas de los textos y en las de la segunda matriz la vectorización de cada una de las palabras. Luego se combinan promediando los vectores para predecir la siguiente palabra.

Aunque la elección de la representación inicial es aleatoria, debido a que el modelo se entrena para predecir la siguiente palabra, al final la representación es capaz de capturar información sintáctica [Le and Mikolov, 2014].

Esta vectorización funciona mediante un modelo que establece la representación con el objetivo de que contribuyan en la predicción de la siguiente palabra basándose en un muestreo del contexto (ventana de longitud fija que se desliza sobre el conjunto de entrenamiento). El objetivo del entrenamiento, es maximizar el promedio de la probabilidad logarítmica

$$\frac{1}{T} \sum_{t=k}^{T-k} \log p(w_t | w_{t-k}, \dots, w_{t+k}), \quad (2-5)$$

donde la probabilidad es estimada con una función *softmax*,

$$p(w_t | w_{t-k}, \dots, w_{t+k}) = \frac{e^{y_{w_t}}}{\sum_i e^{y_i}}, \quad (2-6)$$

y w_1, \dots, w_T es la secuencia de palabras de entrenamiento.

Cada y_i es una probabilidad no normalizada de la palabra i que se busca predecir. Depende tanto de los parámetros b , U y del promedio de los vectores extraídos de D y de W y se calcula como:

$$y = b + Uh(w_{t-k}, \dots, w_{t+k}, d; W, D). \quad (2-7)$$

El modelo se entrena utilizando gradiente descendente estocástico y en cada paso del algoritmo muestrea un contexto, el cual incluye palabras y un vector de párrafo, el cual es compartido por todos los contextos generados que pertenecen al mismo párrafo, pero variante entre párrafos [Le and Mikolov, 2014].

Para tratar con textos nuevos, se agrega una etapa de inferencia en la que se mantienen los parámetros de la función softmax y W , se agregan nuevas columnas a D , y se vuelve a aplicar el gradiente en descenso.

Es importante resaltar que en esta primera alternativa, el usuario puede elegir concatenar representaciones del contexto en vez de promediar.

PV-DBOW tiene una estructura similar a PV-DM, pero en esta alternativa se ignora la posición relativa de las palabras, y en cada iteración del gradiente en descenso se elige el vector de párrafo para predecir las palabras en una ventana de contexto.

La principal ventaja de este algoritmo es la protección de la semántica en los textos, es decir que resguarda la cercanía entre palabras con relación de significado. También, maneja dimensiones menores de representación, lo cual reduce el nivel de dificultad al generalizar [Le and Mikolov, 2014].

Los vectores son entrenados usando gradiente descendente y *backpropagation*, estos pueden llegar a representar textos de cualquier longitud (frases, párrafos o documentos).

En la siguiente sección se describen algunos de los trabajos relacionados con NLP en el área de mantenimiento desarrollados en la literatura.

2.4. NLP en el área de mantenimiento

Los registros de mantenimiento son descripciones que los operarios generan al realizar trabajos en los equipos, estos registros usualmente se encuentran en campos de texto libre que contienen lenguaje técnico. Extraer información relevante de las descripciones de los procesos de mantenimiento no es una tarea sencilla, debido a los grandes volúmenes de información, al vocabulario especializado y taquigráfico. La estructuración manual con expertos consume mucho tiempo, es una actividad tediosa, costosa y extenuante [López, 2019] [Marzec et al., 2014] [Pelham and Hockley, 2017].

Se han realizado diferentes trabajos relacionados con los históricos de las actividades de mantenimiento. A continuación se describirán algunos de estos, empezando por aquellos en los se procesan los registros para cumplir un objetivo específico y terminando con los que adecuan herramientas de procesamiento para mejorar su rendimiento en esa clase de textos. En [Devaney et al., 2005] utilizan los datos de una empresa manufacturera de estados unidos y por medio de ontologías junto con actividades de preprocesamiento (corrección ortográfica, *stemming* y eliminación de palabras de parada) construyen aglomerados en los que se extraen un conjunto canónico de componentes y subsistemas físicos de la máquina y sus fallas asociadas. Sin embargo, esta aproximación no permite la identificación de las características específicas solicitadas.

Butters y Ciravegna en [Butters and Ciravegna, 2008] trabajan en el reconocimiento de términos para extraer los componentes nombrados en los registros de mantenimiento utilizando métricas de cadenas. Dos años después en [Butters and Ciravegna, 2010] desarrollan una metodología para mejorar la búsqueda y posibilitar la recuperación de documentos en dominios técnicos mediante el uso de ontologías y una herramienta de identificación de texto basada en un modelo de *Name Entity Recognition* (NER); pero no consideran extraer la información que caracteriza a los registros.

Marzec et al., en [Marzec et al., 2014] hacen una revisión de la literatura con el objetivo de determinar si los métodos de extracción de información existentes son lo suficientemente precisos para ser utilizados en la clasificación de datos de mantenimiento. Al llegar a la conclusión de que los métodos no alcanzan el nivel de precisión esperado se centran en el problema de clasificar las descripciones como correctivas o preventivas para poder analizar la frecuencia de las fallas. Utilizan *Singular Value Decomposition* para disminuir la dimensión del problema y aprendizaje supervisado. Esta estrategia es aplicada en datos generados por una flotilla de buses y se encuentran en idioma polaco. Aunque esta aproximación es similar al problema abordado en esta tesis, dicho acercamiento no extrae toda la información que se espera. Además presenta otros desafíos por estar en otro idioma.

En el mismo año Zhao et al., en [Zhao et al., 2014] hacen una clasificación de fallas mediante modelamiento de tópicos no supervisado. El modelo es evaluado con datos del ferrocarril de alta velocidad Wuhan-Guangzhou, que están en idioma chino. Sin embargo, esta aproximación presenta el mismo inconveniente que la anterior, el idioma.

Un año después en [Stenström et al., 2015] muestran como el uso de herramientas básicas de NLP en los registros de mantenimiento pueden enriquecer el análisis del desempeño de los equipos. Evalúan la frecuencia de los términos de mantenimiento de un tramo de ferrocarril sueco. Aunque el análisis de frecuencias da una idea general de la información contenida en los registros, por falta de contexto se podrían malinterpretar los resultados.

En 2017 Arif et al., en [Arif-Uz-Zaman et al., 2017] se centran en la construcción de un modelo para reconocer la confiabilidad de los equipos. Usan minería de texto para extraer información precisa sobre los tiempos de falla mediante el procesamiento de órdenes de trabajo y datos de tiempo de inactividad. Se aplica a datos de compañías australianas, una de electricidad y la otra de procesamiento de azúcar. Sin embargo, no siempre se cuenta con datos de tiempo de inactividad.

Pelahm et al., en [Pelham and Hockley, 2017] hacen la comparación del rendimiento de diferentes modelos de clasificación en la tarea de identificar actividades de mantenimiento en las que no se encontraron fallas.

En [Yang et al., 2017] mediante un trabajo manual se procesan 56 órdenes de trabajo del mantenimiento de equipos de calefacción, ventilación y aire acondicionado. A partir de los datos calculan el porcentaje de falla, la probabilidad del modo de falla, tiempo medio entre fallas y severidad.

El siguiente año en [Nagasaka et al., 2018] buscan optimizar el reemplazo de componentes

usando una combinación de herramientas de aprendizaje de máquina, minería de texto y ontologías. Este estudio se aplica a datos de elevadores. Además de que se emplea en datos de otro idioma (inglés), utilizan ontologías las cuales necesitan un conocimiento profundo del área.

Gunay et al. en [Gunay et al., 2019] recurren a técnicas de minería de texto para extraer reglas de asociación. Encuentran patrones de fallas a nivel de componentes/equipos y las fallas más usuales. Para esto utilizan *association rule-mining* y estudian las concurrencias de los términos con relación al soporte (probabilidad conjunta), la confianza (probabilidad condicional) y el *lift* (proporción entre probabilidad conjunta y la multiplicación de las probabilidades). Adicionalmente usan *agglomerative hierarchical clustering* con el objetivo de aislar los registros de mayor interés con respecto a las fallas. Aunque extraen patrones, éstos no necesariamente coinciden con la información que se solicita.

En trabajos más recientes como [Wang et al., 2020] procesan 96 MB de datos correspondientes a las especificaciones, registros y tiquetes de una compañía de red eléctrica en China. En este trabajo proponen un marco que resume la minería de texto corto (200 palabras) de operación y mantenimiento de equipos eléctricos para analizar las decisiones para el tratamiento de los fallos y la clasificación de los grados de defecto en tres categorías: general, importante y urgente. Utilizan herramientas de preprocesamiento (segmentación de palabras, etiquetado POS y eliminación de palabras de parada), realizan una limpieza de los datos (reconocimiento de errores e inconsistencias), emplean representaciones estructuradas (TF-IDF y Word2vec) y una no estructurada (estructura de grafos). Además utilizan modelos de aprendizaje de máquina (support vector machine, *convolutional neural networks*) y métodos basados en reglas.

El mismo año en [Bortolini and Forcada, 2020] trabajan sobre 6830 solicitudes de mantenimiento derivadas de 46 edificios entre oficinas, edificios académicos y laboratorios durante dos años y medio para extraer el tipo de falla y la severidad de las solicitudes de mantenimiento de los usuarios finales. Realizan un preprocesamiento (eliminación de puntuación, marcas diacríticas, estandarización de formato y *stemming*) y utilizan un cuestionario a expertos sobre palabras claves para clasificar por tipo y severidad, adicionalmente aplican una corrección manual para ajustar el clasificador.

En [Yang et al., 2020] a partir de 628 textos cortos con menos de 8 palabras correspondientes a registros de mantenimiento de cucharones de 26 excavadoras, definen agrupaciones para identificar grados de degradación de los equipos. Plantean un modelo estocástico de degradación multiestado y optimización de costos usando herramientas como Word2Vec, *latent Semantic Analysis*, convolutional neural network, *K-means clustering* y simulaciones de Monte Carlo.

A partir del 2020 se han empezado a implementar herramientas más complejas de NLP. Por ejemplo en [Szűcs and Ballagi, 2020] utilizan un modelo de *named entity recognition* (NER) para identificar la localización de la falla en los equipos. Debido a que se desarrolla en idioma inglés, presenta diferentes desafíos a los de esta propuesta.

En [Bouabdallaoui et al., 2020] se construye un clasificador de registros de mantenimiento de instalaciones para encontrar la naturaleza de la falla con el objetivo de automatizar la gestión de las órdenes de trabajo utilizando técnicas de aprendizaje profundo (convolutional neural network y *Long short-term memory*) en datos en francés. Sin embargo, para este tipo de acercamientos es necesario contar con una gran cantidad de datos.

Usuga et al., en [Usuga Cadavid et al., 2020] comparan diferentes vectorizaciones de los registros de mantenimiento con el objetivo de identificar criticidad y duración de los problemas con los equipos. En este trabajo optan por un acercamiento mediante diferentes clasificadores de aprendizaje de máquina y lo aplica en datos en idioma francés. Al igual que el caso anterior es necesario contar con una gran cantidad de datos, además hay que tener en cuenta las diferencias en el idioma.

Por otro lado, la literatura también reporta la adecuación de modelos para el procesamiento de los registros de mantenimiento. Desde 2010, McKenzie et al., [McKenzie et al., 2010] crean herramientas de extracción de la información adecuadas para ser usadas en textos de mantenimiento, construyen algoritmos para etiquetado *part of speech* (PoS), *entity chunking* y *relation extraction*. Este estudio es desarrollado por *Condition-Based Maintenance Center* y se implementa en registros de mantenimiento de helicópteros, que se encuentran en inglés. Tres años después en [Bokinsky et al., 2013] modifican herramientas de NLP para mejorar la extracción de información, adecúan el etiquetador PoS y el entity chunking de la librería NLTK. Aunque estas herramientas serían útiles para el procesamiento de los datos, no serían suficientes para caracterizarlos.

En [Akhbardeh et al., 2020a] desarrollan una nueva herramienta para facilitar y fomentar la investigación en el área de mantenimiento, MaintNet es una librería colaborativa de código libre para procesamiento de lenguaje técnico y de dominio específico. La librería cuenta con conjuntos de datos de aviación, automotriz e instalaciones. Implementa recursos lingüísticos como lista de abreviaciones, listas de información morfosintáctica y bancos de términos. También generan paquetes de python para el preprocesamiento del lenguaje técnico como: eliminación de palabras de parada, lematizadores, etiquetado PoS, métodos de agrupamiento y más. En [Akhbardeh et al., 2020b] describen y evalúan los componentes de la librería. Por el momento la librería está desarrollada solamente para idioma inglés. Sin embargo, no tienen datos de mantenimiento de maquinaria amarilla y además se pueden presentar inconvenientes por el idioma.

En [Brundage et al., 2021], reconocen la necesidad del desarrollo de herramientas de NLP dirigidas al procesamiento de lenguaje técnico. Mencionan las ventajas y desventajas del uso de los algoritmos usuales de NLP al ser implementadas en lenguaje técnico. Resaltan que debido a que se carece de un esfuerzo similar al de otros dominios con respecto a los conjuntos de datos anotados y a compartir los lineamientos de entrenamiento de los modelos, en mantenimiento no se ha experimentado el mismo renacimiento impulsado por la NLP.

Aunque se han contemplado para el futuro, en la literatura no se identifican herramientas o investigaciones dirigidas al procesamiento de los registros de mantenimiento en idioma

español.

A nivel nacional no se han encontrado estudios en los que se analicen los registros de mantenimientos. Es importante aclarar que si se han llevado a cabo trabajos relacionados con la implementación de estrategias de mantenimiento predictivos, pero en éstos se consideran únicamente variables correspondientes a la condición en tiempo real de los equipos, es decir, información sensorial de las vibraciones, temperaturas, entre otras características [Márquez Vásquez et al., 2011] [Ubaque Castillo et al., 2019] [Olarte et al., 2010] [Sánchez Gómez et al., 2017]. Adicionalmente hay empresas en el país que aún no manejan sistemas electrónicos para la gestión de mantenimiento y en general las empresas nacionales registran las actividades pero en vez de analizar los registros los guardan por cuestiones de trazabilidad y de facturación. Infortunadamente, en la mayoría de los casos estos registros no se consideran una fuente importante para la toma de decisiones.

3. Materiales y métodos

Para alcanzar el objetivo planteado, es necesario recolectar bases de datos relacionadas con mantenimiento y procesarlas para estructurar los datos de interés, utilizando una metodología que garantice la reproducibilidad de los resultados. En la Figura 3-1 se ilustra el proceso de minería de texto que se emplea en este trabajo. Se inicia con las descripciones de los trabajos realizados (DTRs), se le adicionan etiquetas, luego se genera una representación del texto mediante el uso de herramientas NLP, dicha vectorización se estructura con modelos de aprendizaje de máquina y, por último, los datos estructurados se utilizan para el cálculo de un conjunto de indicadores claves del procedimiento de mantenimiento.

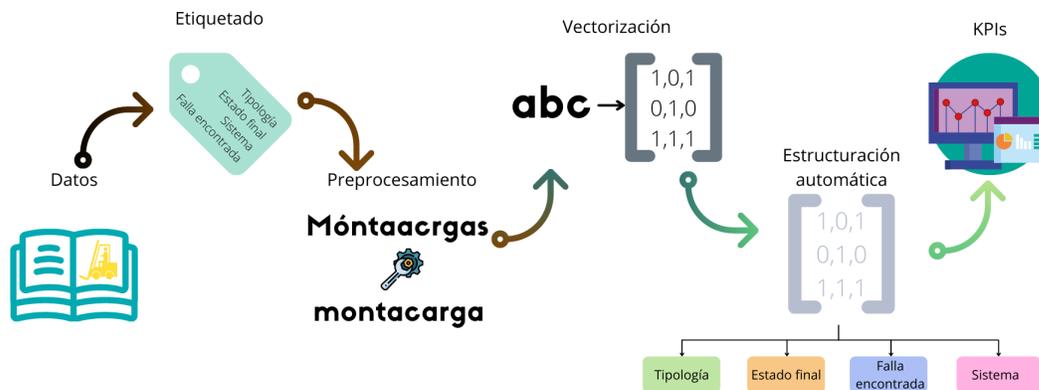


Figura 3-1.: Proceso de minería de texto. Inicia con las DTRs las cuales se etiquetan y se preprocesan. Después se vectorizan y se utilizan para construir modelos de aprendizaje de máquina supervisado. Posteriormente se afinan los parámetros, se elige el modelo con mejor rendimiento, se aplica en los datos y finalmente se calculan los KPIs.

3.1. Datos

El flujo expuesto comienza con la obtención de las DTRs (descripción de los trabajos realizados). Esta tarea se realiza mediante la colaboración de una empresa con una amplia trayectoria en el mercado de renta y venta de maquinaria amarilla en Colombia, quienes facilitaron las hojas de vida de los equipos donde se registran las actividades de mantenimiento.

3.1.1. Hojas de vida

Las hojas de vida de los equipos son documentos en los que se anotan los registros de mantenimiento (Figura 3-2). Éstas usualmente se recolectan y almacenan en el *enterprise resource planning* o en software de gestión de mantenimiento de las empresas, estos son programas que utilizan las organizaciones para automatizar y administrar los procesos. En el caso de la empresa que suministró los datos utiliza un programa de sistema de administración de mantenimiento dirigido a empresas de prestación de servicios, en el que se guarda la información de operación en línea y se almacenan los registros para hacerle seguimiento a las máquinas.

Descripción: [MONTACARGAS ELECTRICO LINDE R20HD]		Horómetro:				
Serial:		Ciudad:				
Localización:		Contacto:				
Teléfono:		Sucursal:				
Cliente:		Garantía hasta:				
Rec. Instalación:		Fecha de Fabricación:				
Vida Útil:						
Otros datos						
MONTACARGAS ELECTRICO PASILLO ANGOSTO						
SERVICIOS PRESTADOS						
Fecha	OT	Tipo Servicio	Solicitud	Lista de Chequeo	Contrato	Estado
25/05/2017	OTR-15176		ALG-671			Cerrada
01/06/2017	OTT-17664	PREVENTIVOS	ALG-671			Cerrada
01/08/2017	OTT-18018	PREVENTIVOS	ALG-671			Cerrada
01/10/2017	OTT-18211	PREVENTIVOS	--			Cerrada
01/10/2017	OTT-18480	PREVENTIVOS	--			Cerrada
01/12/2017	OTT-18670	PREVENTIVOS	--			Cerrada
01/12/2017	OTT-18927	PREVENTIVOS	--			Cerrada
01/02/2018	OTT-19307	PREVENTIVOS	--			Cerrada
01/03/2018	OTT-19549	PREVENTIVOS	--			Cerrada
01/04/2018	OTT-19694	PREVENTIVOS	--			Cerrada
14/12/2020	OTT-20000	PREVENTIVOS				Cerrada
TRABAJOS REALIZADOS						
14/12/2020 08:00:00 a.m.			12743			
Mantenimiento preventivo Mensual R20-01 Hr: 12743						
Se realiza mantenimiento preventivo del equipo, se sopletea, se realiza limpieza general del equipo, se inspecciona sistema hidráulico y sistema eléctrico, se realiza función a los mandos hidráulicos y no se evidencian defectos en la parte externa de las mangueras, tampoco se evidencian fugas de aceite, se realiza lubricación de rodamientos de mástil y carraje, se realiza lubricación de cadenas, se inspecciona nivel de aceite hidráulico y líquido de frenos, se inspecciona estado de las ruedas, el equipo queda en buen estado de funcionamiento y aseo. Además se realiza limpieza general e hidratación a las 2 baterías, se realiza limpieza general y lubricación a la cama de baterías y se realiza limpieza general del cargador. El equipo queda a satisfacción del cliente.						
Referencia	Nombre		Cantidad			
680008016	LUBRICANTE 5-56		1			
681158007	LIMPIADOR ELECTRONICO S-220 8onz		1			
PT1163ACUADE	Agua desmineralizada por litros		16			
PT1163ACUADE	Agua desmineralizada por litros		16			
680008016	LUBRICANTE 5-56		1			
681158007	LIMPIADOR ELECTRONICO S-220 8onz		2			

Figura 3-2.: Ejemplo de las hojas de vida en la que se almacenan los registros

En este programa, en el momento en que se entrega el equipo al cliente, se genera una hoja de vida, en esta se anota la información general del contrato bajo el cual se entrega, los servicios prestados, los trabajos realizados y repuestos. A mayor detalle, las hojas de vida incluye los datos descritos en la Tabla 3-1.

Datos	
Variable	Descripción
Información general	
	Nombre del equipo
Descripción	Modelo del equipo
	Serial del equipo
Serial	Código de identificación de los equipos
Teléfono	Número telefónico del cliente
Cliente	Nombre de la empresa que contrata el servicio
Fecha de instalación	Día, mes y año en el que se instaló el equipo en las instalaciones del contratante
Vida útil	Intervalo de tiempo que termina en el momento en que la frecuencia de fallos del equipo se hace inaceptable o la máquina se considera irreparable
Horómetro	Última lectura de la cantidad de horas trabajadas, junto con la fecha en la que realizó la lectura
Ciudad	Ubicación en la que se instaló el equipo
Contacto	Persona en la empresa contratante que hace de enlace entre ambas compañías
Sucursal	Establecimiento propiedad del contratante en la que se encuentra la máquina
Garantía hasta	Fecha límite para solicitar la garantía
Fecha de fabricación	Día, mes y año en el que se instaló el equipo en las instalaciones del contratante
Servicios prestados	
Fecha	Día, mes y año en la que se prestó el servicio
OT	Código asociado a la orden de trabajo, documento en el que se autoriza y se describe el trabajo solicitado
Tipo de servicio	Clasificación del servicio solicitado (preventivo, correctivo o diagnóstico)
Solicitud	Código relacionado con el documento generado para registrar la solicitud relacionada
Lista de chequeo	Código relacionado con la lista de actividades
Contrato	Código relacionado con el contrato bajo el cual se da la prestación del equipo
Estado	Condición en la que se encuentra el servicio solicitado (cerrada, repuestos pedidos, aprobada o en ejecución)

Trabajos realizados	
Fecha	Día, mes, año y hora en que se realizó el trabajo en cuestión
Horómetro	Lectura de las horas acumuladas de trabajo registradas al realizar el trabajo
Técnico	Persona que lleva a cabo el trabajo
Descripción	Texto corto que incluye la explicación de la actividad llevada a cabo
Repuestos	
Referencia	Código que identifica la pieza
Nombre	Nombre del repuesto utilizado
Cantidad	Cantidad de repuestos usados

Tabla 3-1.: Estructura de las hojas de vida de los equipos compuesta por cuatro secciones: información general, servicios prestados, trabajos realizados y repuestos. Contiene el nombre de las variables y sus respectivas descripciones.

Las variables presentes en las hojas de vida se pueden separar en dos grupos, aquellas que se ingresan en la creación del documento y las que corresponden a entradas que se van agregando a medida que pasa el tiempo y se realizan trabajos sobre el equipo. El segundo grupo tiene información temporal del estado del equipo y contiene lo relacionado con órdenes de trabajo (OT), las solicitudes y los trabajos realizados.

Estos datos se recolectan según el flujo de trabajo que se registra en la Figura 3-3. Se inicia con la avería del equipo que se reporta por medio de una solicitud, se realiza una orden de trabajo con las especificaciones de agendamiento y descripción de las actividades solicitadas, y termina con la descripción de los trabajos realizados. Cabe resaltar que el flujo de trabajo puede evitar el paso de la solicitud en caso de que sea un trabajo preventivo.

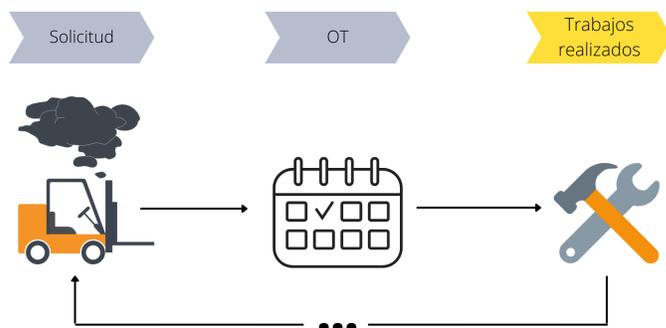


Figura 3-3.: Flujo de trabajo y los documentos adquiridos en cada uno de los pasos. Inicia con una avería en un equipo, se realiza una solicitud inicial, luego se agenda un trabajo y se genera una OT. Luego se realiza la reparación y se registran las actividades realizadas en la hoja de vida.

Las descripciones de los trabajos realizados corresponden a la fuente de datos que permiten extraer las características necesarias para cumplir los objetivos propuestos. Estas se encuentran en lenguaje natural en campos de texto libre, es decir que pueden contener cualquier cadena de caracteres, en general comprende la explicación de las actividades realizadas, los hallazgos y evaluaciones de la situación de los equipos. Son relatos cuyo contenido varía de acuerdo a lo que llama la atención del técnico y al tipo de actividad, por lo cual no tienen una estructura definida ni una longitud fija.

El análisis de los registros se puede plantear a diferentes niveles, pueden ser evaluados a nivel de la empresa, a nivel de contrato, de sucursal donde están instalados o de equipo. Por otro lado, se pueden agrupar por el tipo de equipo al que pertenecen o el modelo. Para el desarrollo de la tesis se cuenta con las hojas de vida de los montacargas que se encuentran bajo un contrato determinado, estos datos describen las tareas de mantenimiento realizadas en 15 equipos de 3 modelos diferentes.

La elección del contrato bajo el cual se encuentran rentadas las máquinas fue definido en conjunto con personal del departamento de mantenimiento de la empresa de renta de maquinaria, al igual que la opción de los montacargas como primer acercamiento al problema, basado en el hecho de que los montacargas son los equipos con mayor cantidad de registros. Se decidió acotar los datos a un tipo específico de maquinaria, para limitar el tipo de actividades y el conjunto de piezas que se emplean en su mantenimiento.

El subconjunto de los registros de mantenimiento seleccionados para este trabajo contiene 1112 descripciones de intervenciones realizadas, donde cada una contiene entre 1 y 409 palabras. En el anexo A se encuentran más detalles sobre los datos.

El hecho de que los datos se generen en una empresa y no se hayan recolectados con fines investigativos ni académicos implica que no se tiene un *gold standard* que permita juzgar los

resultados, lo que a su vez es necesario para crear modelos supervisados para encontrar las variables objetivo.

Para la creación del conjunto gold standard se recurre a la anotación manual por parte de expertos siguiendo el protocolo de etiquetado descrito a continuación.

3.2. Etiquetado

El proceso de etiquetado consiste en la asignación de anotaciones a los datos, que identifican elementos o características de interés. Se define como el proceso de añadir una o más etiquetas informativas y relevantes, para construir un contexto a partir del cual es posible entrenar un modelo de aprendizaje [Amazon Web Services, 2021]. Puesto que el objetivo del trabajo de estructuración de los registros de mantenimiento es extraer información relevante por medio de herramientas de aprendizaje de máquina supervisado, es necesario contar con datos debidamente etiquetados. Obtener un etiquetado apropiado es esencial para el futuro desempeño de los modelos. A continuación se detallan las instrucciones que se utilizaron para hacer el etiquetado de las DTRs. Esta propuesta busca estandarizar el proceso de etiquetado y obtener resultados de calidad minimizando el riesgo de introducción de sesgos.

Flujo de etiquetado

El proceso de etiquetado se lleva a cabo siguiendo el flujo de la Figura 3-4. Se inicia con el anotador, quien asigna las etiquetas siguiendo las instrucciones de la siguiente sección, luego pasa a manos de un evaluador para que asegure la calidad del etiquetado, centrándose en aquellos registros en los que el anotador mostró dudas. En caso necesario el evaluador será el encargado de solucionar las inquietudes de etiquetado y de aprobar el trabajo o realizar más ajustes.



Figura 3-4.: Flujo de etiquetado compuesto por tres fases. En la primera el anotador asigna las etiquetas a las DTRs, luego el evaluador de calidad (QA) evalúa los resultados y en caso de encontrar incongruencias o errores se comunica con el anotador. Después de realizar las correcciones necesarias decide si aprobar el trabajo o realizar más ajustes.

Esquema

La información que se desea extraer está relacionada con diferentes aspectos del departamento de mantenimiento. Para lograr este objetivo se realiza el etiquetado de distintas variables. Se le asignan etiquetas a los textos teniendo en cuenta el esquema de la Figura 3-5.

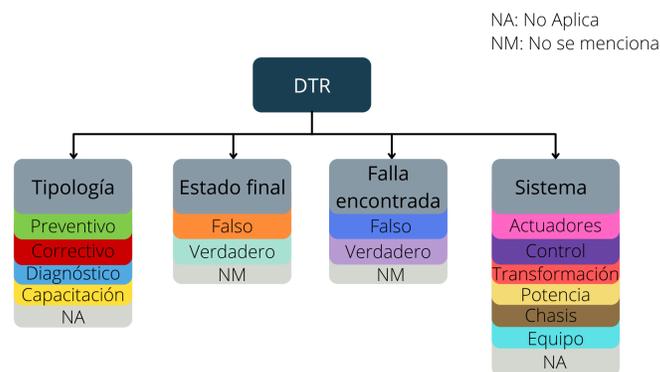


Figura 3-5.: Esquema de etiquetado. Incluye cuatro etiquetas con sus correspondientes clases.

Instrucciones de etiquetado

A continuación se da una definición de la información que se va a extraer mediante este procedimiento¹ y se ponen a disposición un conjunto de ejemplos para usarlos como referencia.

Tipología

Variable que captura la intención con la que fue realizado el trabajo de mantenimiento. Ésta puede asumir cuatro valores diferentes preventivo, correctivo, diagnóstico, capacitación o no aplica (NA).

Preventivos: corresponden a las DTRs que relatan trabajos que se realizaron con la finalidad de mantener el estado del equipo (por ejemplo mantenimiento de rutina y revisiones). Textos que hacen referencia a las actividades programadas con anterioridad que se realizan según las horas de uso del equipo o lo pactado en el contrato. Contiene las actividades que se encuentran descritas en la guía de mantenimiento preventivo de los manuales de cada uno de los modelos o son actividades solicitadas por la fábrica (por ejemplo campañas).

Ej 1. mantenimiento preventivo 500 mensual modelo horometro se realiza mantenimiento preventivo del equipo , se sopletea , se realiza limpieza general del equipo , se inspecciona sistema hidraulico y sistema electrico , se realiza funcion a los mandos hidraulicos y no se evidencian defectos en la parte externa de las mangueras , tampoco se evidencian fugas de aceite , se realiza lubricacion de rodamientos de mastil y carraje , se realiza lubricacion de cadenas , se inspecciona nivel de aceite hidraulico y liquido de frenos , se inspecciona estado de las ruedas , el equipo queda en buen estado de funcionamiento y aseo . ademas se realiza limpieza general e hidratacion a las 2 baterias , se realiza limpieza general y lubricacion a la cama de baterias y se realiza limpieza general del cargador . el equipo queda a satisfaccion del cliente .

Ej. 2 cambio de cofres de baterias : consiste en el retiro de todas las baterias de las instalaciones de nestle para realizar el cambio de los cofres , esta actividad se hace con la ayuda de un tecnico de la empresa multiservicios cargar , y por medio de una montacarga modelo y eslinga realizamos en levantamiento de las baterias de su respetiva cama , para ubicarlas en estibas y poder llevarlas al muelle para su despacho en un camion , el cambio de cofre es realizado por la empresa tronex , se hace el retiro cada 4 horas en promedio y en medida que vamos retirando ellos nos hacen el ingreso de las baterias con cofre nuevo . por cada montacarga son 2 baterias . 14 montacargas 28 baterias nota : ya se habia hecho el cambio de 2 cofres de montacarga modelo , esta vez solo se cambiaron 16 cofres .

¹Es importante resaltar que para el proceso de etiquetado se estandariza los textos en formato con codificación utf-8, se reemplazaron las lecturas de horómetro por **horometro**, las menciones de los nombres de los modelos por **modelo** y los meses por **mes**. Para el etiquetado se seleccionaron los textos únicos, por lo cual se extrajo un subconjunto de 485 registros.

Ej. 3. hidratacion de baterias .

Ej. 4. campanna garantia linde se realiza el desmontaje del conjunto assy esamble y se instala uno totalmente nuevo como indica el fabricante en su campanna .

Correctivo: se relacionan con actividades en las que se corrigió una falla en la máquina. Descripciones en las que se mencionan reconstrucciones, reparaciones, reparaciones temporales, mejoras, ajustes y/o reemplazo de partes, con el objetivo de regresar el equipo a un estado de buen funcionamiento.

Ej. 1. instalacion 4ta via hidraulico auxiliar para side shift se desmontan horquillas del equipo se desenergiza el equipo desconectandolo de la bateria , se desmonta el piso y tapas para poder retirar el bloque valvular . se sopletea el equipo , se marcan conectores y mangueras del bloque se desconectan conectores y se sueltan mangueras . se suelta y se desmonta el bloque valvular se retiran racores del antiguo bloque y se montan en el nuevo. se instala el nuevo bloque valvular de 4 vias se conectan mangueras y conectores se energiza el equipo y se realizan pruebas , el equipo opera normalmente se suelta cadena del carro portahorquillas y se desmonta se retiran rodamientos del carro portahorquillas viejo y se le instalan al nuevo se monta el nuevo carro portahorquillas y se tensiona la cadena se instalan soportes , polea y accesorios para las mangueras de la nueva funcion hidraulica se montan mangueras (hace falta una manguera) se completa nivel de aceite hidraulico se prueba el funcionamiento del side shift ok

Ej. 2. arreglo de fuga en trasmision . horometro se continua con el desarme del equipo , se retira trasmision del equipo .

Ej. 3. modelo : cambio de rueda de traccion por una nueva : consiste en tramitar los permisos de trabajo . con la e25 nos ubicamos en la parte trasera de la montacarga modelo y con las unnas levantamos para dejarla bloqueada sobre madera . con herramienta manual , rache extension de 1 / 2 y copa 19 retiramos 7 pernos y retiramos la rueda e instalamos la nueva y torqueamos los pernos . levantamos de nuevo con la montacarga e25 , retiramos la madera y entregamos la maquina operativa .

Diagnóstico: corresponde a los registros que describen trabajos en los que se evalúa y se analiza el estado del equipo. Textos en los que describen actividades de inspección, monitorización, ensayos de conformidad, verificación de la función y/o diagnósticos de averías justificadas por la presencia de fallas o irregularidades en el funcionamiento de los equipos.

Ej. 1. equipo modelo bloqueado . horometro se continuan las inspecciones del equipo , se inspecciona sistema electrico de la carretilla , se inspeccionan fusibles y reles , se sueltan conectores y se aplica limpia contactos para descartar mal contacto en los conectores , se comprueba continuidad en el cableado para descartar que estos esten abiertos , el equipo continua bloqueado .

Ej. 2. revision y diagnostico . se realiza desmontaje de elementos pasadores , y demas para desmontar el sistema de la elevacion principal .

se detecta que no tiene bujes y tiene demasiado juego en el sistema de la elevacion principal . se procede a buscar todos los numeros de parte para corregir el fallo del movimiento anormal en la elevacion principal .

Capacitación: se le designa a las DTRs vinculadas con acciones que se realizan para enriquecer los conocimientos y habilidades de las personas que pertenecen a la empresa que contrata el servicio de renta (por ejemplo manejo y cuidado de los equipos).

Ej. 1. se presta ayuda al sennor germanico en induccion de operacion de montacargas .

No aplica (NA): se le otorga a los registros que relatan trabajos realizados en los que no se ejecutó ninguna actividad de mantenimiento ni capacitación sobre el equipo.

Ej. 1. cambio de manguera 7 mts modelo horometro se inicia diligenciamiento de permiso , se va a enfermeria para toma de presion , se autoriza permiso muy tarde y no se puede realizar la actividad .

Ej. 2. prueba

Ej. 3. no hubo operacion en bodega , no se podia ingresar .

Nota: En caso de presentarse actividades de más de un tipo, el orden de prioridad definido es el siguiente: correctivo, preventivo, diagnóstico y capacitación.

mantenimiento preventivo mensual modelo se realiza mantenimiento preventivo del equipo , se sopletea , se realiza limpieza general del equipo , se inspecciona sistema hidraulico y sistema electrico , se realiza funcion a los mandos hidraulicos y no se evidencian defectos en la parte externa de las mangueras , tampoco se evidencian fugas de aceite , se inspeccionan escobillas , se realiza lubricacion de rodamientos de mastil , se realiza lubricacion de cadenas , se inspecciona nivel de aceite hidraulico , se inspecciona estado de las ruedas , el equipo queda en buen estado de funcionamiento y aseo . ademas se realiza limpieza general e hidratacion a las 2 baterias , se realiza limpieza general y lubricacion a la cama de baterias y se realiza limpieza general del cargador . el equipo queda a satisfaccion del cliente . ademas se realiza cambio de motor hidraulico. (etiqueta asignada: Correctivo)

Falla encontrada

Variable que indica que la persona que realiza el trabajo identifica o no las averías en el equipo más allá de los síntomas que presenta. Se relaciona con las actividades de evaluación de la condición del equipo, inspecciones, monitoreos y/o ensayos de conformidad, que resultaron en el hallazgo de una falla o avería en el equipo. Falla encontrada puede asumir tres valores: verdadero, falso y no se menciona (NM).

Falso: Descripciones en las que se menciona explícitamente que no se encontraron fallas o en las que no se menciona explícitamente que se encontraron fallas, y tampoco se llevaron a cabo actividades correctivas en los equipos.

Ej. 1. se continua con la inspeccion del equipo modelo para identificar alguna posible falla pero no se logra encontrar nada .

Ej. 2. calibracion de luz de los 3 mts . se realiza calibracion de luz de los 3 mts a 15 equipos montacargas .

Verdadero: Textos en los que se menciona explícitamente que se encontró una falla en el equipo o se menciona que falla se encontró.

Ej. 1. arreglo maguera de extencion modelo . maquina llega con fuga en racor de la manguera , se retira la manguera , se saca para realizar cambio de el racor , se instala nuevamente la manguera , se realiza limpieza de el equipo , se realizan pruebas al equipo quedando este en buen estado de funcionamiento . se hace entrega del equipo .

Ej. 2. armado de equipo modelo horometro se diligencia permiso de trabajo , se inicia armado del equipo , al momento de estar terminando el armado nos enteramos que hubo una falla en la altura de las mordazas y no se puede terminar la labor , se vuelve a desmontar el eje de elevacion y se retira para pedir garantia del trabajo .

No se menciona (NM): Textos en los que no se menciona explícitamente si se encontró fallas en el equipo, pero se describen la realización de actividades correctivas.

Ej. 1. acrilico del equipo en mal estado . horometro se retira el acrilico , se desengrasa , se realiza arrello del acrilico , se brilla , se instala nuevamente el acrilico , el equipo queda en buen estado y listo para su operacion .

Estado Final

Variable que modela la información relacionada con la disponibilidad del equipo, hace referencia al estado funcional, teniendo en cuenta sólo dos estados: funcional y no funcional. Estado final puede asumir tres valores: verdadero, falso y no se menciona (NM).

Falso: Descripciones en las que se hace referencia a que el equipo no queda funcional al terminar el servicio. Mencionan que quedan problemas en la máquina que pueden impedir su funcionamiento y no se repara el problema. Adicionalmente, textos que implican que se sigue trabajando en el equipo o se describe que no era capaz de realizar sus funciones y no se hace referencia a un cambio de un estado.

Ej. 1. se realiza inspeccion al sistema electrico del equipo que presenta corto , se inspecciona estado de las bobinas de los contactores , se evidencia que dos contactores no estan funcionando , se identifica que las bobinas no se estan energizando .

Ej. 2. cambio eje de elevacion modelo horometro se diligencia permiso de trabajo , se sennaliza area de trabajo , se inicia desarme de equipo .

Ej. 3. desmonte eje de elevacion modelo horometro se continua desarme del equipo , se realiza desmonte del eje de elevacion que se encontraba fracturado , se lleva eje de elevacion a taller industrial para su reparacion .

Verdadero: Descripciones en las que se hace referencia a que el equipo queda funcional o no hace referencia a problemas en el equipo y solo realizan actividades preventivas. También incluye aquellos textos en los que no se realizan actividades de mantenimiento o mencionan

fallas que afectan directamente el estado funcional del equipo pero también realizan las actividades correctivas correspondientes.

Ej. 1. mantenimiento preventivo 500 mensual modelo horometro se realiza mantenimiento preventivo del equipo , se sopletea , se realiza limpieza general del equipo , se inspecciona sistema hidraulico y sistema electrico , se realiza funcion a los mandos hidraulicos y no se evidencian defectos en la parte externa de las mangueras , tampoco se evidencian fugas de aceite , se realiza lubricacion de rodamientos de mastil y carraje , se realiza lubricacion de cadenas , se inspecciona nivel de aceite hidraulico y liquido de frenos , se inspecciona estado de las ruedas , el equipo queda en buen estado de funcionamiento y aseo . ademas se realiza limpieza general e hidratacion a las 2 baterias , se realiza limpieza general y lubricacion a la cama de baterias y se realiza limpieza general del cargador . el equipo queda a satisfaccion del cliente .

Ej. 2. hidratacion y limpieza de baterias para montacarga modelo baterias : 2 agua : 2 galones

Ej. 3. modelo se realizo mantenimiento preventivo de la monta carga modelo , el cual comprendio una serie de tareas tales como : limpieza general del equipo con la sopladora electrica para quitar el polvo y otros residuos como papel , plastico etc . engrase general del equipo : cadenas del mastil en buen estado se lubrico el mastil , los rodillos del mastil , las cadenas de elevacion y las poleas de las cadenas , los canales por donde se desliza el mastil (extraer y retraer) , el porta horquillas , las almohadillas de corredera de inclinacion y desplazamiento lateral . (se utilizo wd40 (cantidad 2) y grasa chevron moly grease ep2) . el control dinamico del mastil (dmc) , funciona correctamente . limpieza y nivelacion de agua a las dos baterias del equipo , se necesito agua desionizada (cantidad : 4 galones * 1 bateria) y limpiador electronico (cantidad : 2 sprays * 1 bateria) . los enchufes y cables de la bateria se encuentran en buen estado . se hizo revision de las guias para el carro de bateria y el mecanismo correspondiente . el sistema de enclavamiento de la bateria funciona correctamente . los pernos de la rueda motriz estan apretados de forma segura . revision del nivel de liquido de freno (full) . las ruedas del equipo estan en buen estado , no hay presencia de objetos extrannos . el sistema hidraulico : se reviso el nivel de aceite , nose encontraron fugas en las mangueras y tanque . el sistema electrico los cables y conexiones se encuentran en buen estado , se realizo prueba a todas las funciones del equipo tales como ; los controles para las funciones de elevacion del mastil (reach) , traccion , direccion . la cabina , apoya brazos , pantalla y resolucio de la camara , pedales , silla , boton de parado de emergencia , teclado , pito , y luz (bue spot) , (funciones y accesorios en estado optimo) . se hizo limpieza general del chasis del monta carga , del acrilico y de la cabina del operador . se hizo la limpieza de errores a traves del software pathfinder (traccion , elevacion , direccion) y se programo la advertencia de mantenimiento para el proximo mes de mes o 500 horas .

Ej. 4. mantenimiento preventivo mensual modelo 12748 se realiza mantenimiento preventivo del equipo , se sopletea , se realiza limpieza general del equipo , se inspecciona sistema hidraulico y sistema electrico , se realiza funcion a los mandos hidraulicos y no se evidencian defectos en la parte externa de las mangueras , tampoco se evidencian fugas de aceite , se inspeccionan escobillas , se realiza lubricacion de rodamientos de mastil , se realiza lubricacion de cadenas , se inspecciona nivel de aceite hidraulico , se inspecciona estado de las ruedas , el equipo queda en buen estado de funcionamiento y aseo . ademas se realiza limpieza general e hidratacion a las 2 baterias , se realiza limpieza general y lubricacion a la cama de baterias y se realiza limpieza general del cargador . el equipo queda a satisfaccion del cliente .

No se menciona (NM): Textos en los que no se hace referencia al estado final del equipo o se mencionan fallas que no comprometen el funcionamiento del equipo directamente.

Ej. 1. Apesar de que se suponga que el equipo se encuentra operativo, no se puede asegurar que la falla no comprometa el estado funcional. montacarga modelo : se corrige el problema que presentaba la unna izquierda del montacarga que hacia que se atravesara la mercancia debido a que la empaquetadura del cilindro que acciona la unna estaba desgastada . el problema se corrigio desmontando la unna , luego el cilindro y por ultimo cambiado la empaquetadura por unos oring nuevos . cantidad : 2

Ej. 2. cambio de rueda motriz modelo horometro se realiza cambio de rueda motris que se encontraba en mal estado por desgaste , se instala rueda nueva .

Sistema

Variable que hace referencia a la información de las piezas en las que se realizaron reparaciones o reemplazos. Es importante notar que las reparaciones que se tiene en cuenta son un subconjunto de todas las posibles, dicho subconjunto incluye modificaciones, reemplazo, montajes (armado y desarmado) y calibraciones de las piezas. Sin embargo, no incluye el uso de consumibles ni ajustes.

Con el fin de que la variable sistema se adapte a diferentes equipos se crearon clases entre las cuales se reparten las piezas de las máquinas teniendo en cuenta su funcionalidad. De acuerdo a lo anterior, la variable toma el valor de alguna de las clases, si la reparación o reemplazo descrito se realiza en una pieza que pertenece a dicha etiqueta. Se definen las siguientes clases: actuador, control, transformación, potencia, chasis, equipo y no aplica (NA).

Actuador: Descripciones en las que lleva a cabo trabajos en partes que pertenecen al sistema actuador. Se entiende que una parte pertenece al sistema actuador si la pieza transforma la energía hidráulica, neumática o eléctrica en la activación de un proceso con la finalidad de generar un efecto sobre un proceso automatizado.

Ej. 1. se realiza cambio de baliza del equipo

Ej. 2. montacarga modelo : se corrige el problema que presentaba la unna izquierda del montacarga que hacia que se atravesara la mercancia debido a que la empaquetadura del cilindro que acciona la unna estaba desgastada . el problema se corrigio desmontando la unna , luego el cilindro y por ultimo cambiado la empaquetadura por unos oring nuevos . cantidad : 2

Ej. 3. arreglo patin de side shift . horometro se retira sistema de patin de side shift , se manda a fabricar lamina de sujecion nueva , se repara patin con soldadura , se instalan laminas de nylon para un mejor deslizamiento del side shift , se instala sistema de patin , se realizan pruebas al equipo , el equipo queda en buen estado de funcionamiento .

Control: Descripciones en las que lleva a cabo reparaciones en partes que pertenecen al sistema de control. Se entiende que una parte pertenece al sistema de control si la pieza comunica al usuario con la máquina o comunica las necesidades a los actuadores.

Ej. 1. mantenimiento preventivo mensual de equipo modelo horometro se diligencia permiso de trabajo , se sennaliza area de trabajo , se destapa el equipo , se sopletea , se realiza limpieza general , se realiza inspeccion visual de los sistemas hidraulico , electrico y mecanico , se realiza limpieza y ajuste de conectores electrico , se realiza lubricacion del equipo , se arma nuevamente el equipo , se conecta el computador , se hace reseteo de errores , se realiza calibracion de sensores , se realizan pruebas al equipo quedando este en buen estado de funcionamiento . adicionalmente se realiza limpieza e hidratacion a las 2 baterias , se realiza limpieza y lubricacion de la cama de baterias y se realiza limpieza general al cargado .

Ej. 2. programacion del preselector de altura se calibra encoder de altura por medio del pathfinder se realiza la calibracion en 3 zonas distintas en sus diferentes niveles de estanterias con ayuda de un montacarguista (bodega nueva . antigua bodega y mezzanine) . se realizan las pruebas de funcionamiento del preselector . ok

Transformación: Descripciones en las que lleva a cabo reparaciones en partes que pertenecen al sistema de transformación. Se entiende que una parte pertenece al sistema de transformación si la funcionalidad de la pieza es transformar el tipo de energía: hidráulica, neumática, mecánica o eléctrica para uso del equipo.

Ej. 1. arreglo de fuga en trasmision modelo horometro se diligencia permiso de trabajo , se sennaliza area de trabajo , se inicia desmonte de trasmision .

Potencia: Descripciones en las que lleva a cabo reparaciones en partes que pertenecen al sistema de potencia. Se entiende que una parte pertenece al sistema de potencia si la funcionalidad de la pieza es brindarle la energía necesaria al equipo para su funcionamiento.

Ej. 2. bateria se encuentra en mal estado . se inspecciona la bateria , se identifica que esta no carga , se desconectan conexiones de la bateria , se realiza limpieza y desulfatacion , se inspeccionan postes , se inspeccionan conexiones , se encuentran dos conexiones rotas , se realiza cambio de las conexiones , una de las conexiones es suministrada por multiservicios , se realizan pruebas a la bateria quedando esta en buen estado de funcionamiento .

Ej. 1. cambio de cofres de baterias : consiste en el retiro de todas las baterias de las instalaciones de nestle para realizar el cambio de los cofres , esta actividad se hace con la ayuda de un tecnico de la empresa multiservicios cargar , y por medio de una montacarga modelo y eslinga realizamos en levantamiento de las baterias de su respetiva cama , para ubicarlas en estibas y poder llevarlas al muelle para su despacho en un camion , el cambio de cofre es realizado por la empresa tronex , se hace el retiro cada 4 horas en promedio y en medida que vamos retirando ellos nos hacen el ingreso de las baterias con cofre nuevo . por cada montacarga son 2 baterias . 14 montacargas 28 baterias nota : ya se habia hecho el cambio de 2 cofres de montacarga modelo , esta vez solo se cambiaron 16 cofres .

Chasis: Descripciones en las que lleva a cabo reparaciones en partes que pertenecen al sistema de chasis. Se entiende que una parte pertenece al sistema de chasis si no pertenece a ninguno de los sistemas anteriores.

Ej. 1. cambio de silla modelo horometro se diligencia permiso de trabajo , se sennaliza area de trabajo , se retira la silla en mal estado , se instala silla reparada , el equipo queda en buen estado de funcionamiento .

Ej. 2. rueda motriz equipo modelo gastada . horometro se diligencia permiso de trabajo , se bloquea el equipo sobre tacos , se sennaliza area de trabajo , se retira rueda mala , se instala rueda motriz nueva , se retiran tacos del equipo , se retira sennalizacion del equipo , se realizan pruebas al equipo quedando este en buen estado de funcionamiento .

Equipo: Descripciones en las que lleva a cabo reparaciones en partes que pertenecen a más de uno de los sistemas mencionados. O no se menciona en que parte se realiza el trabajo.

Ej. 1. instalacion 4ta via hidraulico auxiliar para side shift se desmontan horquillas del equipo se desenergiza el equipo desconectandolo de la bateria , se desmonta el piso y tapas para poder retirar el bloque valvular . se sopletea el equipo , se marcan conectores y mangueras del bloque se desconectan conectores y se sueltan mangueras . se suelta y se desmonta el bloque valvular se retiran racores del antiguo bloque y se montan en el nuevo se instala el nuevo bloque valvular de 4 vias se conectan mangueras y conectores se energiza el equipo y se realizan pruebas , el equipo opera normalmente se suelta cadena del carro portahorquillas y se desmonta se retiran rodamientos del carro portahorquillas viejo y se le instalan al nuevo se monta el nuevo carro portahorquillas y se tensiona la cadena se instalan soportes , polea y accesorios para las mangueras de la nueva funcion hidraulica se montan mangueras (hace falta una manguera) se completa nivel de aceite hidraulico se prueba el funcionamiento del side shift ok

Ej. 2. cambio de empaques del equipo modelo

Ej. 3. ingresa equipo con bloqueo hidraulico , se traslada equipo hasta el taller , se realiza inspeccion y posterior desbloqueo del equipo , se hace entrega del equipo sin novedad .

No aplica (NA): Descripciones en las que no lleva a cabo reparaciones o no se mencionan explícitamente.

Ej. 1. mantenimiento preventivo 500 mensual modelo horometro se realiza mantenimiento preventivo del equipo , se sopletea , se realiza limpieza general del equipo , se inspecciona sistema hidraulico y sistema electrico , se realiza funcion a los mandos hidraulicos y no se evidencian defectos en la parte externa de las mangueras , tampoco se evidencian fugas de aceite , se realiza lubricacion de rodamientos de mastil y carraje , se realiza lubricacion de cadenas , se inspecciona nivel de aceite hidraulico y liquido de frenos , se inspecciona estado de las ruedas , el equipo queda en buen estado de funcionamiento y aseo . ademas se realiza limpieza general e hidratacion a las 2 baterias , se realiza limpieza general y lubricacion a la cama de baterias y se realiza limpieza general del cargador . el equipo queda a satisfaccion del cliente .

Ej. 2. modelo se realizo mantenimiento preventivo de la monta carga modelo , el cual comprendio una serie de tareas tales como : limpieza general del equipo con la sopladora electrica para quitar el polvo y otros residuos como papel , plastico etc . engrase general del equipo : cadenas del mastil en buen estado se lubrico el mastil , los rodillos del mastil , las cadenas de elevacion y las poleas de las cadenas , los canales por donde se desliza el mastil (extraer y retraer) , el porta horquillas , las almohadillas de corredera de inclinacion y desplazamiento lateral . se utilizo wd40 (cantidad 2) y grasa chevron moly grease ep2) . el control dinamico del mastil (dmc) , funciona correctamente . limpieza y nivelacion de agua a las dos baterias del equipo , se necesito agua desionizada (cantidad : 4 galones * 1 bateria) y limpiador electronico (cantidad : 2 sprays * 1 bateria) . los enchufes y cables de la bateria se encuentran en buen estado . se hizo revision de las guias para el carro de bateria y el mecanismo correspondiente . el sistema de enclavamiento de la bateria funciona correctamente . los pernos de la rueda motriz estan apretados de forma segura . revision del nivel de liquido de freno (full) . las ruedas del equipo estan en buen estado , no hay presencia de objetos extrannos . el sistema hidraulico : se reviso el nivel de aceite , nose encontraron fugas en las mangueras y tanque . el sistema electrico los cables y conexiones se encuentran en buen estado , se realizo prueba a todas las funciones del equipo tales como ; los controles para las funciones de elevacion del mastil (reach) , traccion , direccion . la cabina , apoya brazos , pantalla y resolucio de la camara , pedales , silla , boton de parado de emergencia , teclado , pito , y luz (bue spot) , (funciones y accesorios en estado optimo) . se hizo limpieza general del chasis del monta carga , del acrilico y de la cabina del operador . se hizo la limpieza de errores a traves del software pathfinder (traccion , elevacion , direccion) y se programo la advertencia de mantenimiento para el proximo mes de mes o 500 horas .

La base de datos construida para alimentar el modelo está constituída por las DTRs junto con las cuatro variables solicitadas: tipología, falla encontrada, estado final y sistema. Dado que los datos son textos no estructurados, no se pueden ingresar directamente a los modelos, por lo tanto el siguiente paso es codificarlos en términos que un computador los pueda entender, para esto se recurre al NLP.

3.3. Preprocesamiento y vectorización

El primer paso al aplicar NLP es realizar preprocesamiento de los textos, en el presente trabajo se definen tres tipos de texto según la cantidad de herramientas de preprocesamiento que se hayan aplicado. Si no se les aplica ninguna herramienta, se considera que los textos son originales, mientras que si se les aplica la estandarización del formato se considera un texto ligeramente preprocesado (LPP). Si se adiciona la lematización, la corrección ortográfica y eliminación de palabras de parada y puntuación, se tienen textos altamente preprocesados (APP).

La lematización se realiza por medio de un proceso semi-automático, en el cual se seleccionan manualmente los lemas bajo la consideración de los propuestos por la herramienta de lematización preentrenada de stanza². Ésto con el objetivo de tener en cuenta el lenguaje técnico.

La corrección ortográfica se hizo bajo el supuesto de que los errores cometidos corresponden a eventos ocasionales, por lo cual las palabras mal escritas se podrían identificar por una baja frecuencia de presencia en los registros. El proceso de corrección se hizo de manera manual, priorizando las palabras menos frecuentes y considerando para el reemplazo las otras palabras usadas cuya distancia *levenshtein* es menor a 3. En el anexo C se encuentran las sustituciones realizadas.

La distancia de levenshtein entre dos cadenas de caracteres a y b se define de forma recursiva como:

$$\text{lev}(a, b) = \begin{cases} 0, & \text{si } |a| = 0 \text{ o } |b| = 0, \\ \text{lev}(\text{tail}(a), \text{tail}(b)), & \text{si } a[0] = b[0], \\ 1 + \text{mín} \left\{ \text{lev}(a, \text{tail}(b)), \text{lev}(b, \text{tail}(a)), \text{lev}(\text{tail}(a), \text{tail}(b)) \right\}, & \text{caso contrario.} \end{cases} \quad (3-1)$$

donde $|a|$ longitud de la palabra, $a[n]$ indica el n -ésimo caracter y $\text{tail}(a)$ hace referencia a la cadena resultante al retirar el primer elemento.

Mediante las herramientas de NLP es posible asociar a los textos vectores numéricos que abstraen diferentes propiedades del lenguaje. Dependiendo de la representación elegida y del preprocesamiento que se ejecute, puede que los modelos encuentren diferentes patrones aunque los textos y las etiquetas sean los mismos. Para el presente proyecto se exploraron cuatro tipos de representación: bag of words (BoW), term frequency-inverse document frequency (TF-IDF), Fasttext y Doc2Vec las cuales capturan información codificada en el lenguaje, una descripción de estas representaciones se encuentra en la sección 2.3.2. Después de representar el texto se realiza el escalamiento del vector resultante.

²Conjunto de herramientas desarrolladas para el análisis lingüístico. Contiene algoritmos para lematizar, para etiquetado PoS y reconocimiento de entidades, entre otros. Hay modelos preentrenados que soportan 66 idiomas incluyendo el español [Qi et al., 2020].

3.3.1. Escalamiento

El escalamiento de los atributos es una práctica altamente recomendada para algunos algoritmos debido a que consideran la distancia entre observaciones o porque se entrenan mediante métodos de gradiente en descenso, y estos varían dependiendo de las dimensiones de los atributos de entrada.

Hay diferentes métodos de escalamiento entre los cuales resalta la estandarización, también conocida como *Z-score normalization*, en la cual se escalan los atributos con el objetivo de que tengan media cero y desviación estándar uno. Cada uno de los valores x de cada una de las entradas de la vectorización es escalada mediante la fórmula

$$z = \frac{x - \mu}{\sigma}, \quad (3-2)$$

donde μ y σ son la media y la desviación estándar de la entrada correspondiente.

Al terminar el preprocesamiento y el escalamiento se obtiene una representación numérica de las DTRs. Se selecciona BoW como la representación para el modelo de línea base.

3.4. Estructuración automática

Previo a la elección de un modelo, en el desarrollo de un proyecto es esencial determinar el criterio bajo el cual se evalúa la eficiencia de la solución planteada. Se han desarrollado indicadores de rendimiento para evaluar y comparar varios modelos o para afinar los parámetros. Para la elección de la métrica apropiada es necesario considerar diferentes aspectos, entre los cuales se encuentra el objetivo del proyecto (predecir varias clases), las métricas sugeridas en la literatura y características particulares del problema (imbalance de clases).

Debido a que cada clasificador se enfrenta a diferentes desafíos, las métricas apropiadas para evaluar los parámetros y comparar modelos pueden variar. A continuación se definen las métricas consideradas para el problema de estructuración de los registros de mantenimiento.

3.4.1. Métricas de evaluación

La definición de las métricas utilizadas parten de una herramienta conocida como matriz de confusión.

Matriz de confusión

La matriz de confusión es una tabla que evidencia la relación entre las predicciones que genera el modelo y las etiquetas reales. Esta tabla se puede utilizar para generar una gráfica que evidencia el rendimiento del modelo, donde las columnas están relacionadas con las etiquetas asignadas por el modelo y las filas corresponden a las etiquetas originales, de tal manera que cada celda corresponde a las ocurrencias de clasificación.

La matriz de confusión tiene la misma cantidad de filas y de columnas, estos equivalen al número de clases [Kohavi, 1998]. Adicionalmente, cumple la propiedad de que al sumar las columnas en cada fila se obtiene el número original de la etiqueta correspondiente, de manera análoga la suma de las filas equivale al número total de predicciones de la categoría correspondiente a dicha columna.

Debido a que las correspondencia de las columnas y de las filas sigue el mismo orden, en la diagonal de la matriz se encuentra la cantidad de registros correctamente clasificados para cada una de las clases. Considerando que algunas de las versiones para clasificadores multiclase de las métricas consideradas son extensiones de la versión binaria, en la Figura 3-6 se ilustra la versión correspondiente de la matriz de confusión. Los verdaderos positivos (VP) y verdaderos negativos (VN) componen la frecuencia de las etiquetas correctamente. De manera similar los falsos positivos (FP) y falsos negativos (FN) constituyen las predichas incorrectamente.

Matriz de confusión

		Clase A	Clase B
Real	Clase A	VP	FP
	Clase B	FN	VN
		Predicción	

Figura 3-6.: Matriz de confusión binaria donde se evidencia el rendimiento de un modelo, contiene el conteo de las instancias positivas y negativas correctamente etiquetadas (VP y VN). Adicionalmente incluye el cálculo de los registros positivos y negativos etiquetadas erróneamente (FP y FN)

Accuracy

Esta métrica se define en términos de la matriz de confusión como la proporción entre las predicciones correctas y el total de predicciones, el caso binario se define como

$$\text{accuracy} = \frac{\text{VP} + \text{VN}}{\text{VP} + \text{VN} + \text{FP} + \text{FN}}, \quad (3-3)$$

y es equivalente a la probabilidad de que el modelo asigne la etiqueta correcta.

Para los modelos multiclase se sigue el mismo principio. En caso de que se presente desbalance de clases, aquella con mayor representación en los datos tendrá una mayor influencia

sobre los resultados, por lo cual, cabe la posibilidad de que la métrica alcance valores altos pero el modelo no sea capaz de etiquetar correctamente a ningún representante de la clase minoritaria. Este índice de rendimiento es bastante intuitivo y fácil de usar [Grandini et al., 2020], usualmente es la métrica predilecta en la literatura y en problemas similares al que se trabaja en el presente proyecto [Marzec et al., 2014] [Pelham and Hockley, 2017] [Bokinsky et al., 2013] [Akhbardeh et al., 2020b] [Bouabdallaoui et al., 2020] y es la métrica más usada para evaluar clasificadores [Bouabdallaoui et al., 2020].

Por otro lado, se han desarrollado variantes en las que se considera la frecuencia de las clases en el conjunto de datos pero éstas no se van considerar en el presente proyecto.

Otra métrica altamente usada en diferentes aplicaciones y en literatura relacionada es la *F1-score*.

F1-score

Corresponde a la media armónica entre los indicadores de *precision* y *recall*. Precision es la proporción de predicciones que fueron etiquetadas como positivas y realmente eran positivas, refleja la confianza que tiene el modelo en asignarle a un individuo la etiqueta de positivo [Grandini et al., 2020], y la recall corresponde a la probabilidad de clasificar correctamente la clase positiva. Debido a que la media armónica toma valores cercanos a uno cuando precision y recall encuentran un equilibrio la F1-score refleja el balance entre ambas cantidades. Esta medida se caracteriza por dar la misma importancia a las clases sin importar su participación. La ecuación:

$$\begin{aligned} \text{precision} &= \frac{VP}{VP + FP}, & \text{recall} &= \frac{VP}{VP + FN}, \\ \text{F1-score} &= \frac{2 * \text{precision} * \text{recall}}{\text{precision} + \text{recall}}, \end{aligned} \tag{3-4}$$

corresponde a la manera en que se calcula la versión binaria de la métrica mencionada.

En el caso de los modelos multiclase, se puede generalizar este indicador de diferentes maneras, entre las cuales encontramos la F1-score macro.

Esta variante se calcula de diferentes maneras. En este estudio se sigue una variante en la que se calcula el indicador para cada una de las clases y posteriormente se evalúa el promedio de dichas cantidades, para evitar clasificadores muy sesgados o producir puntuaciones engañosamente altas.

Un valor alto de este indicador refleja un buen rendimiento general a todas las clases, mientras que un valor pequeño advierte una predicción pobre general [Grandini et al., 2020].

Esta variante se calcula similar a la fórmula 3-4, pero se construye a partir de la *precision macro* y del *recall macro*. Estas se definen como el promedio de las medidas pertinentes evaluadas en las diferentes clases. Por lo tanto, en esta medida todas las clases tienen la misma influencia a pesar de que haya algún desbalance en la distribución de las etiquetas [Grandini et al., 2020].

Por otro lado, entre las métricas usadas en casos de imbalance de clases, también encontramos la *G-mean*.

G-Mean

También conocido como el puntaje de la media geométrica, como su nombre lo indica, se calcula como la media geométrica del recall de cada una de las n clases. En la literatura se ha explorado esta medida para capturar el rendimiento en casos en los que se trabaja con imbalances de datos puesto que le da igual importancia a la precisión de clasificación de todas las clases sin importar su representación en el conjunto.

Para ser usado en un modelo con n clases la G-mean se calcula por medio de la fórmula presente en la ecuación 3-5.

$$\text{G-mean} = \sqrt[n]{\prod_{i=1}^n \text{recall}_i} \quad (3-5)$$

Teniendo en cuenta los diferentes niveles de desbalance, se eligieron las siguientes métricas para la afinación de parámetros: Para tipología y estado final se utiliza G-mean, para falla encontrada se selecciona accuracy y para sistema F1-score.

La asignación de G-mean se hizo basándose en que esta métrica favorece de manera fuerte a los modelos que clasifican correctamente las clases menos representadas.

Lo anterior es provechoso para la variable tipología debido a que la identificación de las actividades en las que no se realizan acciones correctivas ni preventivas sobre los equipos puede ser utilizada para la creación de indicadores de eficiencia del área y el costo de clasificar incorrectamente la clase mayoritaria se podría corregir con el uso de las otras etiquetas. En el caso de estado final el comportamiento mencionado también es beneficioso puesto que la clase minoritaria (falso) contiene mayor información para la generación de indicadores de eficiencia y calidad de servicio.

Por otro lado, la elección de la métrica accuracy para afinar los parámetros de falla encontrada se realiza teniendo en cuenta que esta etiqueta tiene un desbalance de clases muy leve y esta métrica es fácil de interpretar.

Por último, se seleccionó F1-score macro para sistema puesto que es métrica que favorece de manera leve a los modelos que clasifican bien las clases minoritarias. Esto es provechoso para esta variable, debido a que se espera que el modelo asigne las clases minoritarias de manera correcta sin descuidar las clases mayoritarias.

Adicionalmente, se toma la decisión de usar F1-score macro para evidenciar los resultados finales obtenidos por los modelos. Con el objetivo de que sea posible comparar el rendimiento con futuros trabajos que se realicen en el área.

Se establecen los criterios de éxito para cada una de las variables considerando las cantidad de etiquetas. Para tipología, falla encontrada y estado final se espera que se alcance un

F1-score superior o igual a 0,8 y para sistema mayor o igual a 0,65.

Para cada una de las variables objetivo, se particiona el conjunto de datos en una proporción de 70 a 30, de manera tal que cada partición contiene las clases en aproximadamente la misma proporción que en el conjunto original. El grupo más grande se destina al entrenamiento y el otro a la evaluación. Luego se continúa con la elección de los modelos para la extracción de cada una de las características solicitadas para la estructuración de los registros.

En la literatura existen múltiples herramientas que permiten solucionar problemas de clasificación como los presentes en este trabajo. Cada uno de estos tienen ventajas y desventajas. Debido a la naturaleza de los datos y los resultados alcanzados en trabajos similares, se consideraron los siguientes modelos: support vector machine (SVM), random forest (RF), gradient boosting trees (GBT) y gaussian naive bayes (GNB).

Support vector machine

Es un algoritmo que hace parte de los métodos de kernel, los cuales eran de las principales herramientas para realizar aprendizaje de máquina antes del aprendizaje profundo, y aún mantienen importancia [Mahesh, 2020] en el medio debido a sus características beneficiosas como su fuerte fundamentación matemática, adicionalmente es bastante usado debido a que tiene un buen desempeño cuando se cuenta con pocos datos como en el presente problema. Los métodos de kernel nos permiten partir del espacio original del problema y hacer un mapeo en un espacio de atributos de mayor dimensión de manera implícita (con el truco del kernel), en el que es más sencilla la separación de las clases.

SVM es el método de kernel más utilizado y sirve para clasificar datos, su principal objetivo es encontrar un hiperplano que separe de la mejor manera los datos correspondientes a distintas clases. Separar de la mejor manera hace referencia a elegir el hiperplano que maximice el margen, es decir, sea el hiperplano que se encuentre a mayor distancia de los puntos más cercanos de diferentes clases [Alpaydin, 2020]. Se elige este criterio esperando que más adelante el modelo generalice bien con nuevos datos. Por lo anterior, la elección del hiperplano reside en los datos más cercanos entre diferentes clases, a los cuales se les asigna el nombre de vectores de soporte.

Este modelo cuenta con un parámetro de regularización para suavizar el margen y evitar *overfitting*, este parámetro se conoce como C y media el balance entre la reducción del error de entrenamiento y las violaciones al margen (complejidad del modelo). Si C es grande se le da más importancia al error de clasificación, pero si es pequeño se prioriza la simplicidad del margen, por lo cual al disminuir C aumenta el bias (*underfitting*) y se reduce el *variance* (*overfitting*) [Alpaydin, 2020]. Es posible ajustar el modelo para conjuntos con desbalance de clases, agregando pesos inversamente proporcionales a su distribución para ajustar C , para así relajar un poco más la frontera para las clases menos representadas en el conjunto de entrenamiento.

A mayor detalle, este algoritmo tiene como objetivo hallar los α que minimicen la fórmula

presente en la ecuación 3-6, teniendo en cuenta que m es la cantidad de datos de entrenamiento y los x_i corresponden al dato i . Y $K(x_i, x_j)$ es el kernel [Bishop, 2014],

$$\frac{1}{2} \sum_{i,j=1}^m \alpha_i \alpha_j y_i y_j K(x_i, x_j) - \sum_{i=1}^m \alpha_i \quad (3-6)$$

$$\sum_{i=1}^m \alpha_i y_i = 0, \quad C \geq \alpha_i \geq 0, i = 1, \dots, m.$$

Hay diferentes opciones de elección del kernel, entre las cuales encontramos el lineal

$$K_{\text{lineal}}(x_i, x_j) = \langle x_i, x_j \rangle, \quad (3-7)$$

y el *polynomial and radial basis function* (RBF),

$$K_{\text{rbf}}(x_i, x_j) = e^{-\gamma \|x_i - x_j\|^2}. \quad (3-8)$$

Debido a que en este trabajo se consideran más de dos clases para todas las variables, recurrimos a la forma de la función de decisión uno contra uno (ovo), es decir que, si hay n clases, se enfrentan $\frac{n(n-1)}{2}$ decisiones, se comparan todas entre sí y se elige aquella con mayor cantidad de votos.

En la literatura han mostrado buenos resultados en problemas en los que se tienen bastantes atributos pero no tantos registros y son bastantes versátiles gracias al truco del kernel. Un ejemplo de su uso se encuentra en [Usuga Cadavid et al., 2020] [Butters and Ciravegna, 2010] [Arif-Uz-Zaman et al., 2017]. En [Arif-Uz-Zaman et al., 2017] recomiendan usar el kernel RBF puesto que permite un equilibrio entre la complejidad de los límites de decisión y el nivel de generalización. Otro modelo empleado en la literatura es gaussian naive bayes.

Gaussian naive bayes

Es uno de los algoritmos más eficientes y efectivos empleados en aprendizaje de máquina y minería de datos [Zhang, 2004]. Parte del supuesto de que las clases son independientes y aproxima la probabilidad conjunta de los términos usando el teorema de Bayes.

La elección de la clase y correspondiente se hace mediante la ecuación

$$\arg \max_c p(y) \prod_{i=1}^n p(x_i|y), \quad (3-9)$$

en la variante gaussiana, la probabilidad se estima según la fórmula

$$p(x_i|y) = \frac{1}{2\pi\theta_y^2} e^{-\frac{(x_i - \mu_y)^2}{2\theta_y^2}}, \quad (3-10)$$

cuyos parámetros se calculan usando *maximum likelihood estimation*.

Se ha empleado en diferentes clasificadores de texto como los filtros de correos, clasificaciones de tópicos y clasificación de opiniones. En [Nagasaka et al., 2018] [Pelham and Hockley, 2017] [Arif-Uz-Zaman et al., 2017] optan por este tipo de acercamiento.

Por otro lado, también existen métodos de ensamblado de modelos, en los que se acoplan estimadores con el objetivo de mejorar las predicciones que realizan los algoritmos solos al reducir el overfitting. Hay dos familias de ensamblado: métodos de promediado y los de *boosting*.

Los métodos de boosting son ensambles que se construyen al ordenar secuencialmente los algoritmos. Mientras que los métodos de promediado, como su nombre lo indica, funcionan bajo el principio de promediar las predicciones de los algoritmos que componen el ensamblado. Entre los primeros se encuentran los GBT y entre los segundos RF.

Random forest

RF es un modelo conformado por una gran cantidad de árboles de decisión individuales que operan como un conjunto. Cada modelo individual se entrena en una muestra con reemplazo de vectores aleatorios independientes idénticamente distribuidos utilizando optimización aleatoria de nodos [Breiman, 2001].

Cada uno de los árboles que componen el ensamblaje se construyen agregando nodos teniendo en cuenta una medida para evaluar la calidad de la partición. Entre las métricas de elección, para n clases, se encuentra la medida de impureza *gini* y *entropy*

$$\text{gini} = \sum_{i=1}^n p(i)(1 - p(i)), \tag{3-11}$$

$$\text{entropy} = - \sum_{i=1}^n p(i) \log(p(i)).$$

Uno de los beneficios de usar el ensamblaje en vez de un árbol de decisión es que, debido a que la clase asignada por RF corresponde a aquella que tenga la probabilidad promedio más alta, se supera la limitación de overfitting característica de los árboles de decisión. En [Usuga Cadavid et al., 2020] eligen este modelo y en [Marzec et al., 2014] optan por el uso de árboles de regresión.

A pesar de sus ventajas también tiene desventajas, entre las cuales está que requiere bastante poder computacional y tiempo de acuerdo a la cantidad de árboles que contenga el bosque. Adicionalmente, debido al ensamble la importancia de las variables de entrada es difícil de determinar.

Otra manera de construir un ensamblaje a partir de árboles de decisión es enlazarlos de manera secuencial, como se hace en gradient boosting trees.

Gradient boosting trees

Este ensamblaje se construye con el objetivo de que cada modelo que se agrega vaya corrigiendo los errores de los anteriores y la predicción se realiza adicionando los resultados obtenidos. Formalmente, las fórmulas en la siguiente ecuación muestran la manera en que se construyen las predicciones cada vez que se agrega un árbol al ensamble.

$$\begin{aligned}\hat{y}_i^{(t)} &= \hat{y}_i^{(t-1)} + f_t(x_i), \\ \text{obj}^{(t)} &= \sum_{i=1}^n [g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i)] + \Omega(f_t), \\ g_i &= \delta_{\hat{y}_i^{(t-1)}} l(y_i, \hat{y}_i^{(t-1)}), \\ h_i &= \delta_{\hat{y}_i^{(t-1)}}^2 l(y_i, \hat{y}_i^{(t-1)}),\end{aligned}\tag{3-12}$$

donde $\hat{y}_i^{(t)}$ denota el valor predicho, l es la función de pérdida (en este caso softmax), $\Omega(f_t)$ es el término de regularización y $f_t(x_i)$ corresponde a la función de decisión en el paso t . Escribiendo la función de decisión en términos del vector de puntajes de las hojas (w) y definiendo el término de regularización como:

$$\Omega(f) = \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T w_j^2,\tag{3-13}$$

se obtiene la siguiente función objetivo para evaluar la estructura (q) del árbol:

$$\begin{aligned}\text{obj}^* &= -\frac{1}{2} \sum_{j=1}^T \frac{G_j^2}{H_j + \lambda} + \gamma T, \\ G_j &= \sum_{i \in I_j} g_i, \\ H_j &= \sum_{i \in I_j} h_i,\end{aligned}\tag{3-14}$$

donde $I_j = \{i | q(x_i) = j\}$, w es el vector de los puntajes de las hojas y T es el número de hojas. Este modelo es bastante usado, muy eficiente y alcanza altos puntajes en las competencias. Como modelo base se selecciona un SVM y se valora si cumple los criterios de éxito. En caso negativo se exploran más opciones de representación, y si aún no se cumplen los criterios se exploran entre SVM, RF, GNB y GBT con afinación de parámetros. Cuando se cumpla con los objetivos del proyecto, bajo el criterio de las métricas de evaluación, o se cumpla con los plazos de tiempo establecidos para completar el proyecto, se selecciona el modelo más eficiente.

Después de las elecciones de los algoritmos para cada una de las variables, se hace un análisis y una prueba de permutación de etiquetas para identificar las fortalezas y debilidades del modelo elegido.

3.4.2. Prueba de permutación

Es una prueba que se implementa para evaluar el rendimiento de los modelos. Incluye una evaluación en la que se analiza si el modelo captura la verdadera estructura de los datos, puesto que un clasificador puede tener una medida de error bajo pero la data como tal puede que no tenga estructura [Ojala and Garriga, 2010]. Su objetivo es determinar si el modelo se aprovecha de la dependencia de los atributos al clasificar.

Se calcula un p-valor según la ecuación

$$\text{p-valor} = \frac{|D' \in \hat{D} : e(f, D') \leq e(f, D)| + 1}{k + 1}, \quad (3-15)$$

donde D es la matriz original de atributos y etiquetas, f es la función de decisión del algoritmo, $e(f, D)$ es la función con la que se calcula el error de entrenamiento y \hat{D} es un conjunto de k versiones alteradas de los datos originales muestreados aleatoriamente de una distribución nula, obtenida por medio de permutaciones en las etiquetas. La hipótesis nula es que los atributos y la variable objetivo son independientes.

Este p-valor refleja si los datos originales contiene una estructura y si el modelo es capaz de capturar cierta relación entre los atributos y la variable objetivo. Es la probabilidad de que se obtengan mejores resultados que con datos al azar.

Si el p-valor es menor a cierto umbral ($\alpha = 0,05$) se concluye que el error inicial es significativamente pequeño, además implica el rechazo de la hipótesis nula, es decir que la variable objetivo depende de los atributos en el conjunto de datos original.

Se compone el modelo final con los algoritmos seleccionados para cada una de las variables y se aplica en los datos con el objetivo de obtener los KPIs para mostrar cómo sería su uso en la industria.

3.5. KPIs

La información obtenida al aplicar los modelos seleccionados se utiliza para calcular el conjunto de KPIs presentes en la Tabla 3-2.

		KPI			
Nombre	Objetivo	Descripción	Tipo	Frecuencia de cálculo	Nota

TI	Tiempo de indisponibilidad	Días en los que el equipo estuvo detenido por averías	Numérico discreto	3 meses	Niveles: serial nota: es posible calcularlo como el promedio de los días detenido
$C_{variable}$	Complejidad de descripciones	Porcentaje de las DTRs en las que la variable toma el valor NM	Numérico continuo [0,1]	6 meses	Nivel: técnico Variables: falla encontrada y estado final
CT_{clase}	Proporción de las actividades correspondientes a la clase indicada	Porcentaje de las DTRs en las que la variable toma el valor <i>clase</i> de los posibles valores de tipología	Numérico continuo [0,1]	3 meses	Niveles: serial, modelo y contrato Nota: También se puede calcular con cantidades totales
CC_{clase}	Caracterizar de los equipos	Porcentaje de las DTRs en las que las averías están relacionadas con la <i>clase</i> de los posibles valores de sistema	Numérico continuo [0,1]	6 meses	Niveles: serial, modelo y contrato Nota: También se puede calcular con cantidades totales

Tabla 3-2.: Definiciones de los KPI. Incluye nombre, objetivo bajo el cual se plantea, la descripción, el tipo de variable y notas adicionales.

El desarrollo se lleva a cabo mediante *python* usando diferentes librerías, entre las cuales destacan *sklearn*, *xgboost* para la construcción de los modelos, *stanza gensim*, Fasttext para la representación de los textos y *matplotlib* para las visualizaciones.

4. Resultados

En la búsqueda de cumplir los objetivos planteados se construyó un modelo que integra cuatro modelos independientes correspondientes a cada una de las variables objetivo (tipología, falla encontrada, estado final y sistema).

Siguiendo el método propuesto, se etiquetaron los datos y se obtuvieron las distribuciones de clases correspondientes a la Figura 4-1. Se hace evidente que para todas las variables, con excepción de falla encontrada, había una clase altamente predominante, en tipología era la clase **correctivo** (71,75 %), en estado final era **verdadero** (68,45 %) y era sistema es **chasis** (43,71 %).

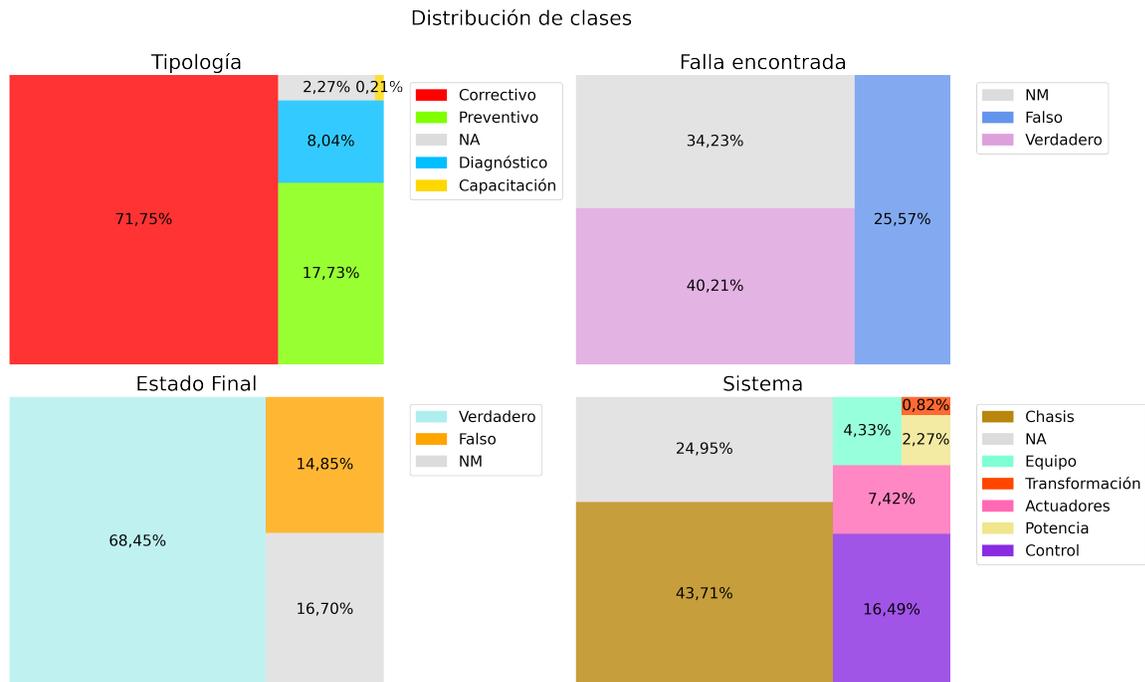


Figura 4-1.: Distribución de las clases de etiquetado para tipología, falla encontrada, estado final y sistema. Se evidencia las clases dominantes: correctivo (tipología), verdadero (estado final) y chasis (sistema).

En tipología las clases **capacitación** (1), **NA** (11) y **diagnóstico** (39) no tenían suficientes representantes en el conjunto de datos para que un modelo aprendiera a identificarlas, así que se tomó la decisión de agruparlas en una nueva clase sin actividades-**SA** en la que pertenecen

las DTRs en los que no se realizan actividades de mantenimiento que afecten o mantengan el estado del equipo.

De manera similar en sistema, las categorías **transformación** (4), **potencia**(11) y **equipo** (21), no alcanzaban a estar ni la mitad de lo esperado en caso de que el conjunto de datos fuera balanceado, es decir, tenían menos de 69 DTRs correspondientes. Por esta razón fue necesario agruparlas en una nueva clase denominada **otros**.

Hay dos maneras de trabajar con datos desbalanceados: modificando la distribución de las clases o mediante parámetros en los modelos. Es posible modificar las proporciones de las clases de diferentes maneras, ya sea agregando más registros de las clases menos representadas (creando datos de manera artificial o solicitando más datos a la empresa) o disminuyendo los representantes de la clase dominante. Teniendo en cuenta que no se contaba con muchos datos para el desarrollo del proyecto, se optó por la utilización de parámetros que le den mayor importancia a las clases minoritarias o que les de mayor margen de error (como C para SVM).

Continuando con la construcción de un modelo de línea base, se partitionaron los datos en un conjunto de entrenamiento y uno de evaluación, asegurándose que ambos conjuntos mantuvieran la misma proporción de representantes de cada clase (Figura 4-2). Se tomó BoW para representar las DTRs y un SVM como modelo para las cuatro variables.

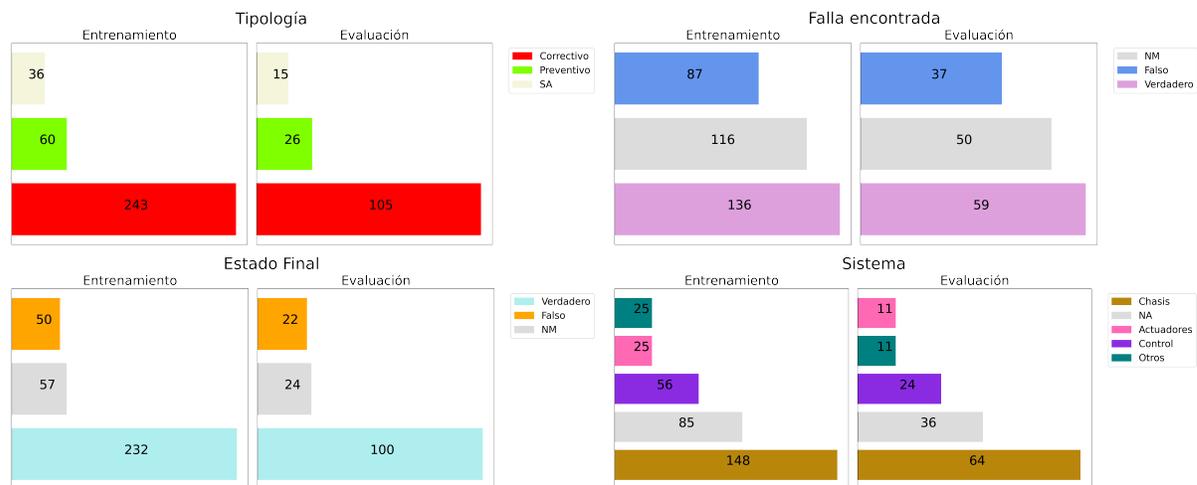


Figura 4-2.: Partición de entrenamiento y de evaluación. Distribución de las clases para cada una de las variables en la partición 70 a 30.

Al aplicar el modelo BoW sobre los textos se realizaron los tipos de preprocesamientos mencionados en el capítulo anterior, ligeramente preprocesados (LPP) y altamente preprocesados (APP).

Luego se aplicó un SVM con los parámetros predeterminados en la librería usada. La Tabla 4-1 reporta los resultados obtenidos, donde la métrica de entrenamiento es el promedio calculado en la validación cruzada (5-fold estratificados). Corresponde a G-mean para tipología

y estado final, accuracy para falla encontrada y F1-score para sistema.

	APP		LPP			
	Métrica de entrenamiento	F1-score	Métrica de entrenamiento	F1-score		
Tipología		0,12	0,59		0,04	0,49
Falla encontrada		0,76	0,73		0,68	0,73
Estado final		0,45	0,73		0,14	0,48
Sistema		0,41	0,5		0,35	0,36

Tabla 4-1.: Valores alcanzados con la representación y el modelo de línea base. Puntajes obtenidos por SVM sobre la representación BoW de los textos APP y LPP. La métrica de entrenamiento corresponde a G-mean para tipología y estado final, a **accuracy** para falla encontrada y **F1-score macro** para sistema.

Puesto que ninguno de los modelos alcanzó a tener un F1-score macro superior al criterio de éxito, se continuó evaluando el mismo modelo pero variando las representaciones empleadas. En particular, se evaluaron 25 representaciones diferentes (anexo B), aplicadas a las DTRs APP y LPP. Como lo muestra la Tabla 4-2.

En el caso de tipología la representación que alcanzó el mejor puntaje en la métrica de entrenamiento fue la 8 (APP), para falla encontrada y estado final fue la representación 10 (LPP) y para sistema fue la 1 (APP). Con respecto a los mejores puntajes conseguidos por el modelo de línea base, tipología obtuvo un G-mean mayor por 0,52 y su F1-score mejoró por 0,19, falla encontrada superó el accuracy por 0,02 mientras que el F1-score se mantuvo igual, los puntajes de estado final aumentaron 0,19 (G-mean) y 0,12 y los de sistema incrementaron 0,11 y 0,07.

	Tipología									
	APP					LPP				
	Representación_0	Representación_1	Representación_2	Representación_3	Representación_4	Representación_0	Representación_1	Representación_2	Representación_3	Representación_4
G-mean_cv	0,12	0,29	0,43	0,41	0,04	0,04	0,19	0,41	0,32	0,04
F1-score macro	0,59	0,67	0,68	0,68	0,58	0,49	0,59	0,67	0,67	0,59
	Representación_5	Representación_6	Representación_7	Representación_8	Representación_9	Representación_5	Representación_6	Representación_7	Representación_8	Representación_9
G-mean_cv	0,04	0,04	0,03	0,64	0,61	0,04	0,04	0,04	0,59	0,61
F1-score macro	0,58	0,58	0,54	0,78	0,75	0,6	0,58	0,58	0,74	0,74
	Representación_10	Representación_11	Representación_12	Representación_13	Representación_14	Representación_10	Representación_11	Representación_12	Representación_13	Representación_14
G-mean_cv	0,64	0,55	0,52	0,55	0,46	0,52	0,35	0,35	0,44	0,23
F1-score macro	0,75	0,72	0,72	0,72	0,67	0,74	0,7	0,67	0,67	0,59
	Representación_15	Representación_16	Representación_17	Representación_18	Representación_19	Representación_15	Representación_16	Representación_17	Representación_18	Representación_19
G-mean_cv	0,04	0,04	0,63	0,04	0,61	0,04	0,04	0,61	0,04	0,6
F1-score macro	0,48	0,55	0,74	0,53	0,74	0,48	0,57	0,72	0,57	0,69
	Representación_20	Representación_21	Representación_22	Representación_23	Representación_24	Representación_20	Representación_21	Representación_22	Representación_23	Representación_24
G-mean_cv	0,04	0,02	0,38	0,02	0,37	0,04	0,03	0,37	0,03	0,37
F1-score macro	0,53	0,37	0,5	0,39	0,6	0,54	0,35	0,62	0,35	0,51
	Falla encontrada									
	APP					LPP				
	Representación_0	Representación_1	Representación_2	Representación_3	Representación_4	Representación_0	Representación_1	Representación_2	Representación_3	Representación_4
Accuracy_cv	0,76	0,74	0,68	0,68	0,63	0,68	0,74	0,74	0,72	0,67
F1-score macro	0,73	0,76	0,72	0,69	0,55	0,73	0,76	0,72	0,67	0,58
	Representación_5	Representación_6	Representación_7	Representación_8	Representación_9	Representación_5	Representación_6	Representación_7	Representación_8	Representación_9
Accuracy_cv	0,64	0,63	0,59	0,76	0,76	0,68	0,67	0,67	0,77	0,77
F1-score macro	0,62	0,62	0,51	0,75	0,74	0,62	0,61	0,55	0,75	0,73
	Representación_10	Representación_11	Representación_12	Representación_13	Representación_14	Representación_10	Representación_11	Representación_12	Representación_13	Representación_14
Accuracy_cv	0,77	0,72	0,72	0,74	0,76	0,78	0,71	0,7	0,71	0,76
F1-score macro	0,77	0,71	0,72	0,73	0,75	0,73	0,72	0,71	0,71	0,7
	Representación_15	Representación_16	Representación_17	Representación_18	Representación_19	Representación_15	Representación_16	Representación_17	Representación_18	Representación_19
Accuracy_cv	0,59	0,63	0,75	0,62	0,74	0,63	0,66	0,77	0,66	0,76
F1-score macro	0,57	0,61	0,71	0,61	0,71	0,59	0,65	0,74	0,63	0,73
	Representación_20	Representación_21	Representación_22	Representación_23	Representación_24	Representación_20	Representación_21	Representación_22	Representación_23	Representación_24
Accuracy_cv	0,62	0,47	0,7	0,52	0,68	0,63	0,45	0,67	0,46	0,69
F1-score macro	0,61	0,42	0,68	0,51	0,68	0,61	0,45	0,7	0,48	0,66

Estado final										
	APP					LPP				
	Representación_0	Representación_1	Representación_2	Representación_3	Representación_4	Representación_0	Representación_1	Representación_2	Representación_3	Representación_4
G-mean_cv	0.45	0.41	0.56	0.55	0	0.14	0.43	0.59	0.58	0.39
F1-score macro	0.73	0.61	0.79	0.8	0.27	0.48	0.55	0.77	0.78	0.52
	Representación_5	Representación_6	Representación_7	Representación_8	Representación_9	Representación_5	Representación_6	Representación_7	Representación_8	Representación_9
G-mean_cv	0.1	0.01	0	0.6	0.58	0.41	0.43	0.11	0.63	0.63
F1-score macro	0.5	0.4	0.27	0.83	0.82	0.55	0.55	0.4	0.86	0.87
	Representación_10	Representación_11	Representación_12	Representación_13	Representación_14	Representación_10	Representación_11	Representación_12	Representación_13	Representación_14
G-mean_cv	0.6	0.6	0.57	0.6	0.44	0.64	0.61	0.61	0.59	0.43
F1-score macro	0.82	0.75	0.74	0.76	0.76	0.85	0.77	0.78	0.77	0.79
	Representación_15	Representación_16	Representación_17	Representación_18	Representación_19	Representación_15	Representación_16	Representación_17	Representación_18	Representación_19
G-mean_cv	0	0	0.28	0	0.2	0	0.06	0.48	0.05	0.49
F1-score macro	0.27	0.27	0.6	0.27	0.56	0.27	0.42	0.7	0.4	0.74
	Representación_20	Representación_21	Representación_22	Representación_23	Representación_24	Representación_20	Representación_21	Representación_22	Representación_23	Representación_24
G-mean_cv	0.01	0	0.02	0.01	0.12	0.02	0	0.02	0	0.08
F1-score macro	0.35	0.27	0.46	0.27	0.55	0.51	0.27	0.42	0.27	0.44
Sistema										
	APP					LPP				
	Representación_0	Representación_1	Representación_2	Representación_3	Representación_4	Representación_0	Representación_1	Representación_2	Representación_3	Representación_4
F1-score macro_cv	0.41	0.52	0.45	0.44	0.3	0.35	0.48	0.4	0.4	0.28
F1-score macro	0.5	0.57	0.42	0.42	0.28	0.36	0.58	0.48	0.45	0.26
	Representación_5	Representación_6	Representación_7	Representación_8	Representación_9	Representación_5	Representación_6	Representación_7	Representación_8	Representación_9
F1-score macro_cv	0.32	0.31	0.33	0.51	0.49	0.32	0.3	0.28	0.44	0.44
F1-score macro	0.31	0.3	0.26	0.5	0.52	0.32	0.28	0.25	0.49	0.49
	Representación_10	Representación_11	Representación_12	Representación_13	Representación_14	Representación_10	Representación_11	Representación_12	Representación_13	Representación_14
F1-score macro_cv	0.5	0.49	0.48	0.48	0.5	0.43	0.42	0.41	0.42	0.44
F1-score macro	0.58	0.44	0.41	0.41	0.51	0.49	0.48	0.46	0.46	0.45
	Representación_15	Representación_16	Representación_17	Representación_18	Representación_19	Representación_15	Representación_16	Representación_17	Representación_18	Representación_19
F1-score macro_cv	0.24	0.31	0.47	0.33	0.48	0.25	0.36	0.42	0.35	0.41
F1-score macro	0.25	0.29	0.58	0.3	0.59	0.25	0.34	0.56	0.32	0.53
	Representación_20	Representación_21	Representación_22	Representación_23	Representación_24	Representación_20	Representación_21	Representación_22	Representación_23	Representación_24
F1-score macro_cv	0.28	0.23	0.42	0.26	0.42	0.26	0.17	0.37	0.16	0.37
F1-score macro	0.25	0.38	0.48	0.36	0.48	0.27	0.16	0.45	0.16	0.42

Tabla 4-2.: Valores de las métricas alcanzados por el modelo de línea base en los DRT LPP y APP. Considerando el puntaje de entrenamiento más alto (resaltado en amarillo), se seleccionaron las siguientes representaciones, representación 8 (APP) (tipología), representación 10 (LPP) (falla encontrada y estado final) y representación 1 (APP) (sistema).

A partir de este punto se evaluaron los resultados obtenidos organizados por variables.

4.1. Tipología

Variando la representación base, no se alcanzó el criterio de éxito, por lo cual, se evaluaron los modelos GNB, RF, SVM y GBT en la representación 8 (APP), se afinaron los parámetros correspondientes y se obtuvieron los resultados reportados en la Tabla 4-3.

	Tipología	
	G-mean	F1-score macro
SVM	0,84	0,86
GNB	0,76	0,79
RF	0,82	0,8
GBT	0,82	0,82

Tabla 4-3.: Resultados de GNB, RF, SVM y GBT para tipología representación 8 (APP). El G-mean más alto (0,84) se alcanza con SVM con parámetros ajustados.

El modelo que mejores resultados obtuvo fue SVM con los siguientes parámetros: C igual a 100, los pesos de clase balanceados, RBF kernel y gamma 0, 1, junto con una estandarización

con media pero sin desviación estándar. Este modelo fue seleccionado puesto que cumple con lo esperado y su aplicación en el conjunto de evaluación se evidencia en la Figura 4-3.



Figura 4-3.: Matriz de confusión del modelo seleccionado para tipología sobre los datos de evaluación.

Se analizaron las DTRs de evaluación en los que el modelo cometió errores de clasificación y se llegó a la conclusión de que es posible que confunda las clases debido a que en una DTR se pueden describir tanto acciones preventivas como correctivas y usualmente siempre realizan una de SA. Adicionalmente, la mayoría del texto puede referirse a la clase que predice, pero debido a unas pocas palabras la etiqueta no concuerda. A continuación se presentan DTRs que ejemplifican la afirmación anterior, en la cual la mayoría del texto corresponde a actividades preventivas, pero en la última frase menciona el cambio de un motor, por lo cual la tipología toma el valor de correctivo.

Ej. mantenimiento preventivo mensual modelo se realiza mantenimiento preventivo del equipo , se sopletea , se realiza limpieza general del equipo , se inspecciona sistema hidraulico y sistema electrico , se realiza funcion a los mandos hidraulicos y no se evidencian defectos en la parte externa de las mangueras , tampoco se evidencian fugas de aceite , se inspeccionan escobillas , se realiza lubricacion de rodamientos de mastil , se realiza lubricacion de cadenas , se inspecciona nivel de aceite hidraulico , se inspecciona estado de las ruedas , el equipo queda en buen estado de funcionamiento y aseo . ademas se realiza limpieza general e hidratacion a las 2 baterias , se realiza limpieza general y lubricacion a la cama de baterias y se realiza limpieza general del cargador . el equipo queda a satisfaccion del cliente . ademas se realiza cambio de motor hidraulico (etiqueta real: correctivo, etiqueta predicha: preventivo)

4.2. Falla encontrada

En el caso de esta variable tampoco se cumplió el criterio de éxito con el modelo de línea base, así que se evaluaron los otros modelos en la representación 10 (LPP) y se alcanzaron los puntajes reportados en la Tabla 4-4. El modelo que mejor rendimiento alcanzó fue un

SVM con C igual a 10, kernel RBF, pesos de clases balanceados y con estandarización, pero aún no obtuvo un puntaje suficiente para ser seleccionado.

	Falla encontrada	
	Accuracy	F1-score macro
SVM	0,8	0,79
GNB	0,68	0,66
RF	0,76	0,66
GBT	0,78	0,73

Tabla 4-4.: Resultados de GNB, RF, SVM y GBT para falla encontrada representación 10 (LPP). El accuracy más alto (0,8) se alcanza con SVM con parámetros ajustados.

Por lo tanto se tomó el SVM como modelo, se variaron las representaciones y se afinaron los parámetros. La distribución del accuracy promedio de la validación cruzada y del F1-score macro del conjunto de evaluación se ilustran en la Figura 4-4

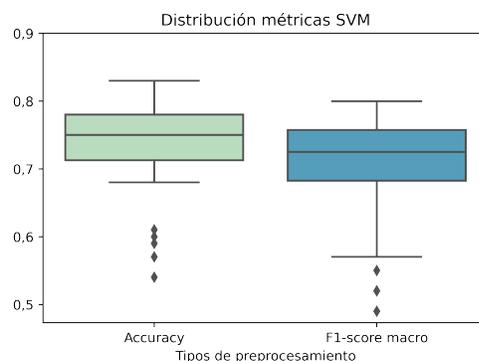


Figura 4-4.: Distribución de los puntajes se SVM al variar las representaciones. Se evidencia una distribución similar entre métricas al variar las vectorizaciones.

La representación en la que se alcanzó la mejor accuracy 0,83 fue la representación 1 (LPP) la cual obtuvo un F1-score macro de 0,78. Entre los modelos explorados, en la representación 1 (APP) el SVM con parámetros afinados de C igual a 100, pesos de clase balanceados, γ 0,1, kernel RBF y estandarización sin considerar la desviación estándar, alcanzó un puntaje de accuracy igual a 0,79 y F1-score macro de 0,8. Debido a que se igualó el criterio de éxito al sacrificar solamente 0,04 accuracy se tomó la decisión de elegir este modelo para esta variable.

En la Figura 4-5 se ilustra el comportamiento del modelo seleccionado sobre el conjunto de evaluación.

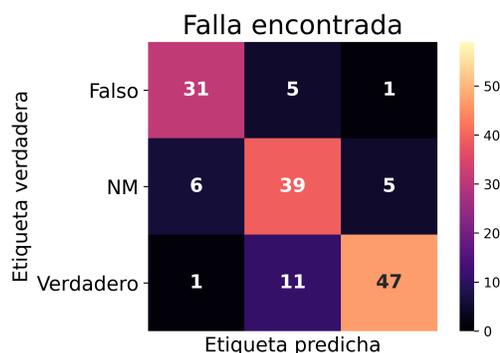


Figura 4-5.: Matriz de confusión del modelo seleccionado para falla encontrada sobre el conjunto de evaluación.

Al revisar los errores cometidos por el clasificador se encontró que esta confundiendo la descripción de los síntomas con la mención de la falla.

Ej.cambio de ruedas de apoyo modelo horometro se diligencia permiso de trabajo , se realiza sennalizacion de area , se bloque equipo sobre tacos , se retiran ruedas de carga en mal estado , se instalan ruedas de carga nuevas , se realizan pruebas al equipo quedando este en buen estado de funcionamiento . adicional se realiza cambio de cableado de timon con el equipo modelo . (etiqueta real: nm, etiqueta predicha: verdadero)

4.3. Estado final

En el caso de la variable estado final, como se ilustra en la Tabla 4-5, el modelo seleccionado correspondió a un SVM con parámetro de regularización igual a 10, peso de clases balanceado, kernel RBF, gamma 0,01 y estandarización. Este logró el mejor puntaje entre los modelos evaluados en la representación 10 (LPP).

	Estado Final	
	G-mean	F1-score macro
SVM	0,81	0,81
GNB	0,74	0,54
RF	0,78	0,6
GBT	0,77	0,65

Tabla 4-5.: Resultados de GNB, RF, SVM y GBT para estado final representación 10 (LPP). El G-mean más alto (0,81) se alcanza con SVM con parámetros ajustados.

En la Figura 4-6 se evidencia los resultados de la clasificación del conjunto de evaluación realizada por el modelo elegido.



Figura 4-6.: Matriz de confusión del modelo seleccionado para estado final sobre el conjunto de evaluación.

Se analizaron los casos en los que el modelo comete errores de clasificación, pero no se evidenció una razón consistente con el aspecto temático que justifique las confusiones. Es posible que como la mayoría de las DTRs corresponden a la descripción de las actividades, y sólo una parte ínfima hace referencia al estado final (en especial cuando se escribe explícitamente) el texto restante puede generar ruido que afecta la clasificación.

Ej. instalacion de estator en equipo modelo horometro se instala estator en equipo modelo , se cambia un cable de poder que se encontraba defectuoso , el equipo funciona bien . (etiqueta real: verdadero, etiqueta predicha: falso)

4.4. Sistema

En el caso de sistema los resultados de los modelos se encuentran en la Tabla 4-6. El modelo seleccionado fue GTB con los siguientes parámetros: fracción de las columnas para ser submuestreados igual a 0,82, *subsample* 0,66, tasa de aprendizaje 0,23, γ 2, 03, profundidad máxima de 6, *min child weight* 1, α igual a 0, λ igual a 0,12, 41 estimadores y estandarización sin desviación estándar.

	Sistema	
	F1-score macro	F1-score macro
SVM	0,67	0,66
GNB	0,37	0,54
RF	0,58	0,6
GBT	0,72	0,69

Tabla 4-6.: Resultados de GNB, RF, SVM y GBT para sistema representación 1 (APP). El F1-score más alto (0.72) se alcanza con GBT con parámetros ajustados.

A mayor detalle, la Figura 4-7 revela la clasificación obtenida por el modelo en los datos de evaluación.

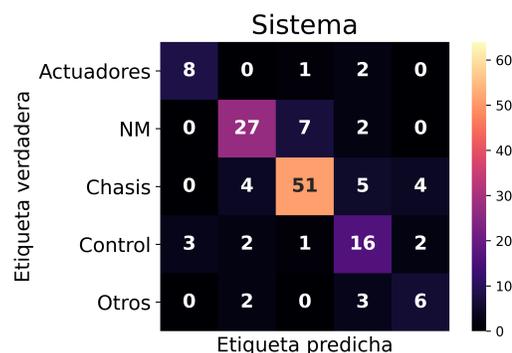


Figura 4-7.: Matriz de confusión del modelo seleccionado para sistema sobre el conjunto de evaluación.

Adicionalmente la Figura 4-8 evidencia la evolución del F1-score macro a medida que se agregaron árboles al ensamblaje.

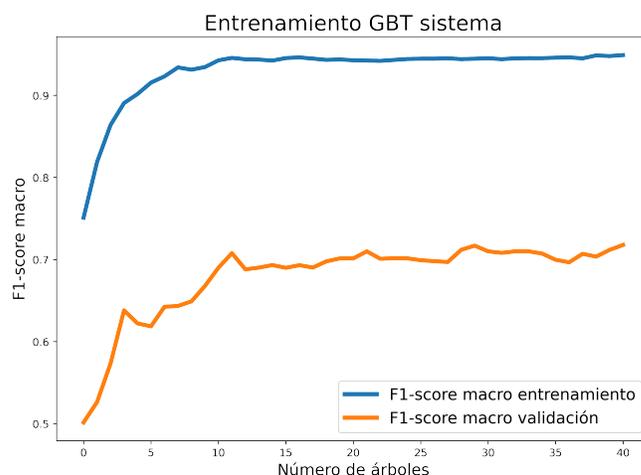


Figura 4-8.: Evolución del ensamblaje seleccionado para sistema al ir agregando estimadores.

En general, al analizar los registros de prueba en los que se cometieron errores de clasificación no se evidenció ningún patrón, hay casos particulares como el siguiente ejemplo, en el cual la confusión de etiqueta se puede deber a que la DTR menciona la actividad que se iba a realizar en el sistema que predice pero se aclara que no fue posible llevarla a cabo.

Ej. cambio de manguera 7 mts modelo horometro se inicia diligenciamiento de permiso , se va a enfermeria para toma de presion , se autoriza permiso muy tarde y no se puede realizar la actividad . (etiqueta real: NA, etiqueta predicha: chasis)

Los modelos seleccionados se construyeron a partir de las siguientes representaciones 1, 8 (APP) y 10 (LPP), es decir sobre la representación TI-IDF del texto altamente procesado y Fasttext con skipgrams. Esto indica que las actividades de preprocesamiento ejecutadas

para las DTRs (APP) sí ayudaron a mejorar los resultados obtenidos en tres de los cuatro modelos bajo estudio.

4.5. Prueba de permutación

Con respecto a la prueba de permutación los p-valores obtenidos por los modelos y las representaciones seleccionadas correspondieron al mejor valor posible (0.0099), a excepción de sistema (0.02). Al ser menores que el umbral establecido ($\alpha = 0,05$) se concluyó la negación de la hipótesis nula de que los atributos y las variables objetivo son independientes. Además, permitió afirmar que los modelos capturan una estructura de los datos.

4.6. KPIs

El modelo final corresponde al enlace de aquellos seleccionados para cada una de las variables. Gracias a los resultados obtenidos por éste, es posible calcular aproximaciones a diferentes KPIs que sirvan de apoyo para la toma de decisiones en el área de mantenimiento, a continuación se muestran algunos de ellos calculados con las predicciones generadas. Es importante resaltar que entre mejor sea el desempeño de los modelos, más cercano estarían los KPIs calculados al valor real.

En la Figura 4-9 se ilustra los índices $C_{falla.encontrada}$ y $C_{estado.final}$ para todos los trimestres a partir 2017-05-01 y hasta el 2021-04-01 para el técnico 1. Estos hacen referencia al porcentaje de presencia de NM en la variable falla encontrada y estado final en las DTRs ingresados por el técnico en cuestión y en el respectivo semestre. Se relaciona con que tan completos están los registros y podrían ser usados para evidenciar la necesidad de capacitaciones de los técnicos de mantenimiento.

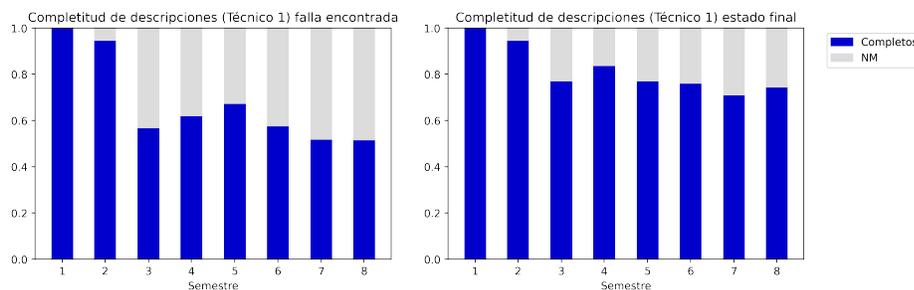


Figura 4-9.: Complejidad de las DTRs con respecto a la falla encontrada y estado final. El KPI denotado como C que cuantifica la completitud de las DTRs con respecto a la falla encontrada y estado final.

Otro indicador que resulta útil son las proporciones de tipología de los trabajos realizados. Por ejemplo, en la Figura 4-10 se muestran los valores correspondientes a los trimestres entre

2017-05-01 y hasta el 2021-04-01 del equipo 1. Por el comportamiento de las proporciones y su progresión en el tiempo es posible evidenciar irregularidades con los equipos. Adicionalmente, se puede analizar el conteo de correctivos en el trimestre, valor que se puede asociar con la cantidad de fallas.

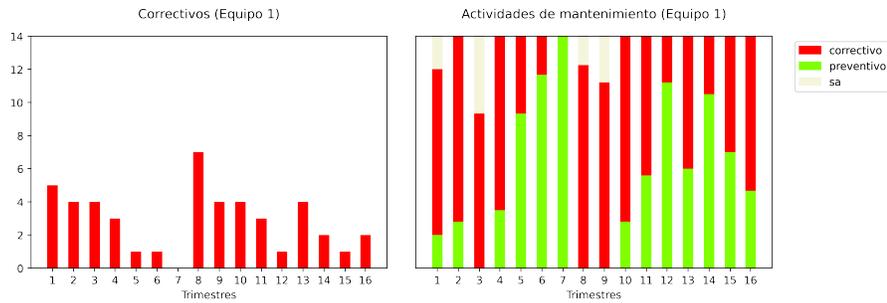


Figura 4-10.: Tipología de las DTRs. Combinación de los KPIs CT_{clase} para las clase de la variable tipología. Adicionalmente se complementa con el conteo de actividades correctivas por trimestre.

Con respecto a la caracterización de los correctivos, como se ilustra en la Figura 4-11, es posible diferenciarlos por el sistema en el que se realizaron los trabajos. Con información adicional, sería posible aproximar los costos relacionados con las fallas.

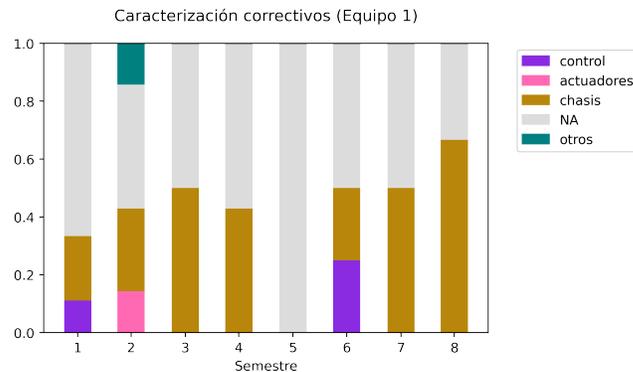


Figura 4-11.: Caracterización de correctivos. Combinación de los KPIs denotados como CC_{clase} para todas las clases de la variable sistema.

Por último, la Figura 4-12 corresponde a la cantidad total y promedio de los días que el equipo estuvo detenido debido a una falla. Valores altos implican que en el trimestre correspondiente no se presta un buen servicio lo cual, si se ignora puede implicar el desgaste en la relación con el cliente.

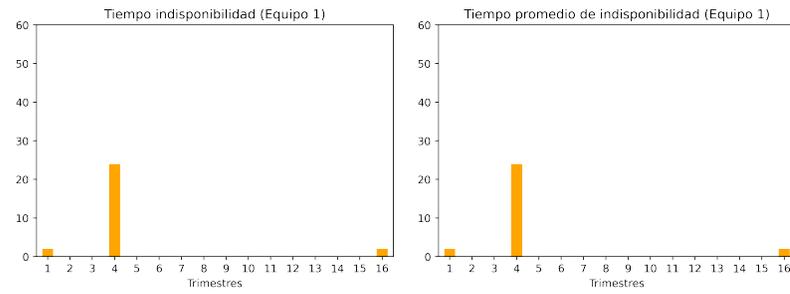


Figura 4-12.: Promedio de días de indisponibilidad. El KPI denotado como TI, complementado con la suma total días de indisponibilidad.

5. Conclusiones y recomendaciones

En la presente tesis fue posible crear un modelo mediante el cual es posible asignar a las DTRs su tipología, la falla encontrada, el estado final y el sistema. El modelo consta de 3 SVM y un GBT que alcanzan un F1-score macro de 0,86, 0,8, 0,81 y 0,69 de manera respectiva.

Los resultados de la prueba de permutación de los modelos individuales indican que las etiquetas correspondientes a cada una de las variables capturan una estructura de los datos y evidencian que las clases no son independientes de los componentes de las representaciones usadas. Adicionalmente, los resultados fueron utilizados para la estimación de algunos KPIs del área de mantenimiento ejemplificando el uso del modelo.

Hasta donde se tiene conocimiento es el primer trabajo realizado en el país con respecto a la estructuración de los registros de mantenimiento, que a diferencia de las aproximaciones del estado del arte en problemas similares desarrolladas en Colombia, no depende de la instalación de sensores en la maquinaria para monitorear el estado en tiempo real de los equipos, la cual suele ser una alternativa costosa de implementar y de mantener. Esto permite que el modelo resultante se pueda aplicar a las empresas que no han recurrido a la alternativa del monitoreo en tiempo real. Por otro lado, en caso de que si cuentan con los sensores, este trabajo serviría para complementar los datos obtenidos y brindar aún más información relevante para la toma de decisiones.

Adicionalmente, se considera que es un avance para la estructuración de registros con lenguaje técnico en idioma español, puesto que se genera una base de datos anotada que puede ser utilizada para entrenar otros modelos que pueden ser adecuados para circunstancias similares.

Los indicadores generados, de aplicarse en la industria, podrían ser utilizados para facilitar la toma de decisiones y la comunicación entre departamentos de las empresas. Se espera que al utilizar la información estructurada y los KPIs, las decisiones tomadas en el departamento se vean reflejadas en aumentos en las ganancias generadas por la renta de los equipos, en la mejora de la relación con el cliente y/o la adquisición de los datos.

Por otro lado, se recuerda que los KPIs calculados son una aproximación a los valores reales, cuya relación con la realidad se estrecha a medida que los modelos tengan un mejor rendimiento. Al ser aplicados en la industria, es importante que el tomador de decisiones relacione los índices con las observaciones del personal del área.

Adicionalmente, se aclara que debido a que el modelo se construye a partir de los datos generados por el personal de mantenimiento, la fiabilidad de los resultados obtenidos depende

en gran medida de la buena fe con que son recolectados.

Considerando que el modelo construido se entrenó con un conjunto de datos de un solo contrato y un tipo de maquinaria específico, para trabajos futuros se recomienda aumentar los datos adicionando las DTRs de montacargas rentados bajo diferentes contratos para enriquecer la base anotada (utilizando el protocolo establecido) con el fin mejorar los puntajes de las métricas, teniendo en cuenta que la diversidad de ejemplos de entrenamiento también crecería y las vectorizaciones serían más robustas. Se sugiere explorar la extensión del modelo a DTRs de diferentes tipos de equipos y probar otras representaciones cuando se cuente con un conjunto de datos más extenso. Y se aconseja agregar un paso adicional en el que se filtren las DTRs con omisiones de información o con datos contradictorios.

A. Análisis inicial de datos

La distribución de la cantidad palabras por DTR se observa en la Figura A-1. Originalmente se cuenta con 1112 registros con un mínimo de 1 palabra, un máximo de 409 y un promedio de 95,53. Al aplicar las primeras tareas de preprocesamiento se obtienen textos levemente preprocesados (LPP), que resulta en 485 DTRs únicas, con un máximo de 489 palabras por registro (la cantidad aumentó, puesto que los signos de puntuación se incluyen en la cuenta) y un promedio de 68.56 palabras por registro. Al llevar a cabo más tareas de preprocesamientos, se obtuvo un registro altamente preprocesado (APP) con un promedio de 33,50 palabras por documento.

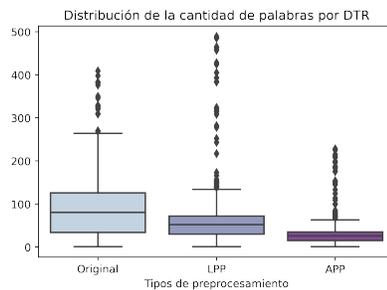


Figura A-1.: Comparación de la distribución de las palabras por documento de los textos originales, los LPP y los APP.

Como se evidencia en la Figura A-2 de 2699 palabras únicas iniciales, las más frecuentes son **de** (8442), **se** (8417), **y** (4645) y **realiza** (3817). En los textos LPP hay 147 palabras únicas y las más frecuentes son **se** (2696), **de** (2394), **,** (2357) y **.** (1157). Por último, las DTRs APP tienen 905 términos singulares, los más usados son **equipo** (1153), **realizar** (772), **rueda** (349) y **quedar** (316).

Las palabras más utilizadas son coherentes con el nivel de preprocesamiento. En los dos primeros casos las palabras más usadas corresponden a pronombres, conjunciones, signos de puntuación y preposiciones, que no dan mucha información mientras que en los textos APP las expresiones más usadas se relacionan un poco más con el área de estudio.

Como se evidencia en las figuras anteriores las actividades de preprocesamiento ayudan a que la cantidad de palabras por documento tenga menos variación. Además, reducen de manera sustancial la cantidad de palabras únicas empleadas, bajo el primer grupo de acciones de preprocesamiento se genera una reducción del 54,57% y las siguientes aplicadas tienen aún un mayor impacto con una reducción de 61,43% del vocabulario restante.

B. Variación de Representación

Se emplearon cuatro tipos de representaciones: BoW, TF-IDF, Fasttext y Doc2vec. En la Tabla **B-1** se listan los modelos junto con los parámetros que se emplearon para construir cada una de las 25 representaciones.

	Representación_0	Representación_1	Representación_2
Modelo	BoW	TF-IDF	Fasttext, skipgram, minn:3, maxn: 6, epoch: 20, dim:200
	Representación_3	Representación_4	Representación_5
Modelo	Fasttext, skipgram, minn: 3, maxn: 8, epoch :20, dim:300	Fasttext, skipgram	Fasttext, cbow, minn:3, maxn: 6, epoch:20, dim:200
	Representación_6	Representación_7	Representación_8
Modelo	Fasttext, cbow, minn: 3, maxn: 8, epoch :20, dim:300	Fasttext, cbow	Fasttext, skipgram, minn:3, maxn: 6, epoch: 50, dim:200
	Representación_9	Representación_10	Representación_11
Modelo	Fasttext, skipgram, minn: 3, maxn: 8, epoch :50, dim:300	Fasttext, skipgram, epoch:50	Fasttext, cbow, minn:3, maxn: 6, epoch:50, dim:200
	Representación_12	Representación_13	Representación_14
Modelo	Fasttext, cbow, minn: 3, maxn: 8, epoch:50, dim:300	Fasttext, cbow, epoch:50	Fasttext preentrena- do
	Representación_15	Representación_16	Representación_17
Modelo	Doc2vec, seed:42, dm:0	Doc2vec, vec- tor_size:200, min_count:2, epochs:20, seed:42, dm:0, dm_concat:1	Doc2vec, vec- tor_size:200, min_count:2, epochs:50, seed=42, dm=0
	Representación_18	Representación_19	Representación_20

Modelo	Doc2vec, vector_size:300, min_count:2, epochs:20, seed:42, dm:0, dm_concat:1	vec-	Doc2vec, vector_size:300, min_count:2, epochs:50, seed:42, dm:0	vec-	Doc2vec, dm:1, seed:42,
	Representación_21		Representación_22		Representación_23
Modelo	Doc2vec, vector_size:200, min_count:2, epochs:20, seed:42, dm:1, dm_concat:1	vec-	Doc2vec, vector_size:200, min_count:2, epochs:50, seed=42, dm=1	vec-	Doc2vec, vector_size:300, min_count:2, epochs:20, seed:42, dm:1, dm_concat:1
	Representación_24				
Modelo	Doc2vec, vector_size:300, min_count:2, epochs:50, seed:42, dm:1	vec-			

Tabla B-1.: Modelos utilizados para generar las representaciones junto con los parámetros seleccionados.

C. Reemplazos

En la Tabla C-1 se documentan los reemplazos realizados en la lematización y corrección ortográfica realizada.

Bibliografía

- [Akhbardeh et al., 2020a] Akhbardeh, F., Desell, T., and Zampieri, M. (2020a). Maintnet: A collaborative open-source library for predictive maintenance language resources. *arXiv preprint arXiv:2005.12443*.
- [Akhbardeh et al., 2020b] Akhbardeh, F., Desell, T., and Zampieri, M. (2020b). Nlp tools for predictive maintenance records in maintnet. In *Proceedings of the 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 10th International Joint Conference on Natural Language Processing: System Demonstrations*, pages 26–32.
- [Alpaydin, 2020] Alpaydin, E. (2020). *Introduction to machine learning*. MIT press.
- [Amazon Web Services, 2021] Amazon Web Services, I. (2021). What is data labeling? [en línea]. url<https://aws.amazon.com/es/sagemaker/data-labeling/what-is-data-labeling/>[5 de Diciembre].
- [Arif-Uz-Zaman et al., 2017] Arif-Uz-Zaman, K., Cholette, M. E., Ma, L., and Karim, A. (2017). Extracting failure time data from industrial maintenance records using text mining. *Advanced Engineering Informatics*, 33:388–396.
- [Beltrán, 2000] Beltrán, J. M. J. (2000). Indicadores de gestion una herramienta para lograr la competitividad.
- [Bishop, 2014] Bishop, C. (2014). Bishop-pattern recognition and machine learning-springer 2006. *Antimicrob. Agents Chemother*, pages 03728–14.
- [Bojanowski et al., 2017] Bojanowski, P., Grave, E., Joulin, A., and Mikolov, T. (2017). Enriching word vectors with subword information. *Transactions of the Association for Computational Linguistics*, 5:135–146.
- [Bokinsky et al., 2013] Bokinsky, H., McKenzie, A., Bayoumi, A., McCaslin, R., Patterson, A., Matthews, M., Schmidley, J., and Eisner, L. (2013). Application of natural language processing techniques to marine v-22 maintenance data for populating a cbm-oriented database. In *AHS Airworthiness, CBM, and HUMS Specialists' Meeting, Huntsville, AL*.

- [Bortolini and Forcada, 2020] Bortolini, R. and Forcada, N. (2020). Analysis of building maintenance requests using a text mining approach: Building services evaluation. *Building Research & Information*, 48(2):207–217.
- [Bouabdallaoui et al., 2020] Bouabdallaoui, Y., Lafhaj, Z., Yim, P., Ducoulombier, L., and Bennadji, B. (2020). Natural language processing model for managing maintenance requests in buildings. *Buildings*, 10(9):160.
- [Breiman, 2001] Breiman, L. (2001). Random forests. *Machine learning*, 45(1):5–32.
- [Brundage et al., 2021] Brundage, M. P., Sexton, T., Hodkiewicz, M., Dima, A., and Lukens, S. (2021). Technical language processing: Unlocking maintenance knowledge. *Manufacturing Letters*, 27:42–46.
- [Butters and Ciravegna, 2008] Butters, J. and Ciravegna, F. (2008). Using similarity metrics for terminology recognition. In *LREC*.
- [Butters and Ciravegna, 2010] Butters, J. and Ciravegna, F. (2010). Authoring technical documents for effective retrieval. In *International Conference on Knowledge Engineering and Knowledge Management*, pages 287–300. Springer.
- [Carvalho et al., 2019] Carvalho, T. P., Soares, F. A., Vita, R., Francisco, R. d. P., Basto, J. P., and Alcalá, S. G. (2019). A systematic literature review of machine learning methods applied to predictive maintenance. *Computers & Industrial Engineering*, 137:106024.
- [Çınar et al., 2020] Çınar, Z. M., Abdussalam Nuhu, A., Zeeshan, Q., Korhan, O., Asmael, M., and Safaei, B. (2020). Machine learning in predictive maintenance towards sustainable smart manufacturing in industry 4.0. *Sustainability*, 12(19):8211.
- [de Academias de la Lengua Española y Real Academia Española, 2021] de Academias de la Lengua Española y Real Academia Española, A. (2021). Diccionario de la lengua española - edición del tricentenario. [versión 23.5 en línea]. <https://dle.rae.es/> [5 de Diciembre].
- [de Jonge and Scarf, 2020] de Jonge, B. and Scarf, P. A. (2020). A review on maintenance optimization. *European journal of operational research*, 285(3):805–824.
- [Devaney et al., 2005] Devaney, M., Ram, A., Qiu, H., and Lee, J. (2005). Preventing failures by mining maintenance logs with case-based reasoning. In *Proceedings of the 59th meeting of the society for machinery failure prevention technology (MFPT-59)*.
- [Ding and Kamaruddin, 2015] Ding, S.-H. and Kamaruddin, S. (2015). Maintenance policy optimization—literature review and directions. *The international journal of advanced manufacturing technology*, 76(5):1263–1283.

- [Ghosh et al., 2012] Ghosh, S., Roy, S., and Bandyopadhyay, S. K. (2012). A tutorial review on text mining algorithms. *International Journal of Advanced Research in Computer and Communication Engineering*, 1(4):7.
- [Grandini et al., 2020] Grandini, M., Bagli, E., and Visani, G. (2020). Metrics for multi-class classification: an overview. *arXiv preprint arXiv:2008.05756*.
- [Gunay et al., 2019] Gunay, H. B., Shen, W., and Yang, C. (2019). Text-mining building maintenance work orders for component fault frequency. *Building Research & Information*, 47(5):518–533.
- [Hirschberg and Manning, 2015] Hirschberg, J. and Manning, C. D. (2015). Advances in natural language processing. *Science*, 349(6245):261–266.
- [Kohavi, 1998] Kohavi, R. (1998). Glossary of terms. *Special issue on applications of machine learning and the knowledge discovery process*, 30(271):127–132.
- [Kumar et al., 2013] Kumar, U., Galar, D., Parida, A., Stenström, C., and Berges, L. (2013). Maintenance performance metrics: a state-of-the-art review. *Journal of Quality in Maintenance Engineering*.
- [Le and Mikolov, 2014] Le, Q. and Mikolov, T. (2014). Distributed representations of sentences and documents. In *International conference on machine learning*, pages 1188–1196. PMLR.
- [Lundgren et al., 2018] Lundgren, C., Skoogh, A., and Bokrantz, J. (2018). Quantifying the effects of maintenance—a literature review of maintenance models. *Procedia CIRP*, 72:1305–1310.
- [Luque et al., 2019] Luque, C., Luna, J. M., Luque, M., and Ventura, S. (2019). An advanced review on text mining in medicine. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 9(3):e1302.
- [López, 2019] López (2019). Your guide to natural language processing (nlp). [en línea]. <https://towardsdatascience.com/your-guide-to-natural-language-processing-nlp-48ea2511f6e1> [5 de Diciembre].
- [Mahesh, 2020] Mahesh, B. (2020). Machine learning algorithms-a review. *International Journal of Science and Research (IJSR)*. [Internet], 9:381–386.
- [Márquez Vásquez et al., 2011] Márquez Vásquez, D. et al. (2011). Plan de negocios de una empresa que brinda servicios de mantenimiento predictivo en colombia. B.S. thesis, Uniandes.

- [Marzec et al., 2014] Marzec, M., Uhl, T., and Michalak, D. (2014). Verification of text mining techniques accuracy when dealing with urban buses maintenance data. *Diagnostyka*, 15.
- [Mcallister, 2021] Mcallister (2021). Mcallister. [en línea]. <https://mcallister.com.co> [5 de Diciembre].
- [McKenzie et al., 2010] McKenzie, A., Matthews, M., Goodman, N., and Bayoumi, A. (2010). Information extraction from helicopter maintenance records as a springboard for the future of maintenance text analysis. In *International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems*, pages 590–600. Springer.
- [Nagasaka et al., 2018] Nagasaka, M., Sato, M., and Kinoshita, E. (2018). Integrated analysis system for elevator optimization maintenance using ontology processing and text mining. In *Safety and Reliability—Safe Societies in a Changing World*, pages 3093–3098. CRC Press.
- [Nakagawa, 2006] Nakagawa, T. (2006). *Maintenance theory of reliability*. Springer Science & Business Media.
- [Ojala and Garriga, 2010] Ojala, M. and Garriga, G. C. (2010). Permutation tests for studying classifier performance. *Journal of Machine Learning Research*, 11(6).
- [Olarte et al., 2010] Olarte, W., Botero, M., and CAÑON, B. A. (2010). Análisis de vibraciones: una herramienta clave en el mantenimiento predictivo. *Scientia et Technica*, 16(45):219–222.
- [Parida, 2006] Parida, A. (2006). Maintenance performance measurement system: Application of ict and e-maintenance concepts. *International journal of COMADEM*, 9(4):30.
- [Pelham and Hockley, 2017] Pelham, J. G. and Hockley, C. (2017). Analysis of short form maintenance records for nff using nlp, phrase matching, and bayesian learning. *Procedia CIRP*, 59:257–262.
- [Prasertrungruang and Hadikusumo, 2007] Prasertrungruang, T. and Hadikusumo, B. (2007). Heavy equipment management practices and problems in thai highway contractors. *Engineering, construction and Architectural management*.
- [Qi et al., 2020] Qi, P., Zhang, Y., Zhang, Y., Bolton, J., and Manning, C. D. (2020). Stanza: A python natural language processing toolkit for many human languages. *arXiv preprint arXiv:2003.07082*.
- [Sánchez Gómez et al., 2017] Sánchez Gómez, A. M. et al. (2017). Técnicas de mantenimiento predictivo: metodología de aplicación en las organizaciones.

- [Stenström et al., 2015] Stenström, C., Al-Jumaili, M., and Parida, A. (2015). Natural language processing of maintenance records data. *International Journal of COMADEM*, 18(2):33–37.
- [Szűcs and Ballagi, 2020] Szűcs, B. and Ballagi, (2020). Artificial intelligence in maintenance: The industrial application of natural language processing.
- [Ubaque Castillo et al., 2019] Ubaque Castillo, Y. M., Aguirre Zabala, E. L., et al. (2019). Estructuración del programa de mantenimiento predictivo por condición para los equipos del área de producción en la empresa kellogg de colombia sa.
- [Usuga Cadavid et al., 2020] Usuga Cadavid, J. P., Grabot, B., Lamouri, S., Pellerin, R., and Fortin, A. (2020). Valuing free-form text data from maintenance logs through transfer learning with camembert. *Enterprise Information Systems*, pages 1–29.
- [Wang et al., 2020] Wang, H., Liu, Z., Xu, Y., Wei, X., and Wang, L. (2020). Short text mining framework with specific design for operation and maintenance of power equipment. *CSEE Journal of Power and Energy Systems*.
- [Yang et al., 2017] Yang, C., Chen, Q., Shen, W., and Gunay, B. (2017). Toward failure mode and effect analysis for heating, ventilation and air-conditioning. In *2017 IEEE 21st International Conference on Computer Supported Cooperative Work in Design (CSCWD)*, pages 408–413. IEEE.
- [Yang et al., 2020] Yang, Z., Baraldi, P., and Zio, E. (2020). A novel method for maintenance record clustering and its application to a case study of maintenance optimization. *Reliability Engineering & System Safety*, 203:107103.
- [Zhang, 2004] Zhang, H. (2004). The optimality of naive bayes. *AA*, 1(2):3.
- [Zhang et al., 2015] Zhang, Y., Chen, M., and Liu, L. (2015). A review on text mining. In *2015 6th IEEE International Conference on Software Engineering and Service Science (ICSESS)*, pages 681–685. IEEE.
- [Zhao et al., 2014] Zhao, Y., Xu, T.-h., and Hai-feng, W. (2014). Text mining based fault diagnosis of vehicle on-board equipment for high speed railway. In *17th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, pages 900–905. IEEE.
- [Único organismo de normalización en españa, 2021] Único organismo de normalización en españa (2021). Une-en 13306:2011. [en línea]. www.une.org[5 de Diciembre].