



UNIVERSIDAD NACIONAL DE COLOMBIA

---

# **Aplicaciones de la Analítica y la Minería de Datos en la Gestión de Recursos Humanos**

**Natalia Moreno Rodríguez**

**Universidad Nacional de Colombia**

**Facultad de Minas, Departamento de Ingeniería de la Organización**

**Medellín, Colombia**

**2018**



# **Aplicaciones de la Analítica y la Minería de Datos en la Gestión de Recursos Humanos**

**Natalia Moreno Rodríguez**

*Ingeniera administradora, Especialista en Ingeniería Financiera*

Trabajo de investigación presentado como requisito parcial para optar al título de:  
**Magister en Ingeniería Administrativa**

Director

Juan David Velásquez Henao, MSc, PhD  
Profesor Titular

Universidad Nacional de Colombia  
Facultad de Minas, Departamento de Ingeniería de la Organización  
Medellín, Colombia  
2018



## Resumen

La analítica y los sistemas de información de Recursos Humanos les permiten a los profesionales de esta área tener mayor credibilidad profesional y rigor en la toma de decisiones. A pesar de esto, el uso de la analítica en las áreas de Recursos Humanos se ha desarrollado de manera más lenta que en otras funciones del negocio y, aunque en los últimos años la literatura alrededor del uso de la analítica en la Gestión de Recursos Humanos ha venido aumentando, ésta aún es considerada como una innovación en estado de difusión. El objetivo de esta investigación es identificar los diferentes usos o aplicaciones que la analítica y la minería de datos ha tenido en los diversos campos de la Gestión de Recursos Humanos, por medio de una revisión sistemática de literatura, para brindar herramientas a los profesionales de Recursos Humanos así como a los investigadores, para decidir sobre futuras inversiones e investigaciones adicionales que permitan que el uso de la analítica en Recursos Humanos continúe expandiéndose. Los resultados permiten concluir que la analítica y la minería de datos tiene una gran variedad de aplicaciones en la Gestión de Recursos Humanos permitiendo resolver problemas en áreas como la atracción, selección e incorporación, la evaluación del desempeño, el desarrollo y planes de sucesión, formación, análisis y descripción de puestos, retención y otras, en donde se han aplicado una gran diversidad de técnicas que permiten dar solución a estos problemas. Es posible concluir que existe un gran potencial de investigación en la aplicación de analítica y minería de datos para dar solución a los problemas y apoyar la toma de decisiones de los profesionales en esta área.

Palabras clave: analítica, minería de datos, Recursos Humanos, Gestión de Recursos Humanos (GRH), analítica de Recursos Humanos

## Abstract

Analytics and HR information Systems allow HR professionals to have greater professional credibility and rigorousness in their decision making. However, the use of analytics in Human Resources areas has developed more slowly than in other business functions and, although in the past few years the literature around the use of analytics in Human Resource Management has been on the rise, it is still considered to be an innovation in diffusion state. The objective of this research it to identify the different uses or applications that analytics and data mining have had on the various areas of Human Resource Management, through a systematic literature review, in order to provide tools to HR Professionals and researchers to decide on future investments and

research that will allow the use of analytics in Human Resources to continue growing. The results lead to conclude that analytics and data mining have a great variety of applications in Human Resource Management and can help solve problems in areas such as talent attraction, selection and incorporation, performance assessment, development and succession planning, training, job analysis and descriptions, retention and others, where diverse techniques have been applied in order to solve these problems. It is possible to conclude that there is great potential for future research in the application of analytics and data mining in order to solve the problems and support decision making of professionals in this area.

Keywords: analytics, data mining, Human Resources, Human Resource Management (HRM), HR analytics

# Contenido

<b>Resumen.....</b>	<b>V</b>
<b>Lista de Figuras.....</b>	<b>VII</b>
<b>Lista de Tablas .....</b>	<b>IX</b>
<b>Introducción .....</b>	<b>1</b>
<b>1. Generalidades .....</b>	<b>3</b>
1.1 Gestión de Recursos Humanos .....	3
1.2 Analítica y minería de datos .....	4
<b>2. Protocolo de Investigación.....</b>	<b>7</b>
<b>3. Resultados .....</b>	<b>9</b>
<b>4. Discusión .....</b>	<b>17</b>
4.1 ¿Qué aplicaciones se le han dado a la analítica y la minería de datos en la Gestión de Recursos Humanos y en qué áreas o campos? .....	17
4.2 ¿Cuáles son las técnicas o herramientas de analítica y de minería de datos más utilizadas en Recursos Humanos?.....	20
4.3 ¿Cómo ha evolucionado el uso de analítica en Recursos Humanos a través de los años? .....	23
4.4 ¿Cuáles son los principales resultados obtenidos en las investigaciones? .....	26
<b>5. Conclusiones .....</b>	<b>30</b>
<b>Bibliografía .....</b>	<b>32</b>

## Lista de Figuras

Figura 1: Cantidad de documentos publicados por año .....	9
Figura 2: Cantidad de citas, según el año de publicación .....	10
Figura 3: Histograma de índice h de los autores .....	13
Figura 4: Número de artículos según el país de procedencia .....	15
Figura 5: Histograma de Palabras Clave que se repiten más de una vez.....	16



## Lista de Tablas

Tabla 1: Citaciones por artículo .....	10
Tabla 2: Artículos con más de 10 citaciones .....	11
Tabla 3: Autores con el mayor índice H .....	12
Tabla 4: Número de Artículos según fuente de Publicación .....	14
Tabla 5: Número de casos encontrados por área de aplicación.....	18
Tabla 6: Número de casos en los que se aplica cada tipo de técnica de la analítica o minería de datos, según el área de Recursos Humanos .....	21
Tabla 7: Técnicas de minería de datos utilizadas y número de aplicaciones.....	22
Tabla 8: Número de modelos descritos según su tipo y el año .....	25
Tabla 9: Número de modelos descritos, según el tipo de técnica utilizada y el año.....	25

# Introducción

El área de Recursos Humanos se ha venido transformando desde hace algunos años, pasando de ser un área administrativa o de mantenimiento, a una función estratégica para el negocio cuyo propósito debe ser el de agregar valor para éste [1]. En la medida en que la tecnología se propaga en los procesos relacionados con la Gestión de Recursos Humanos, la analítica y los sistemas de información de Recursos Humanos, los cuales han evolucionado desde sistemas de transacciones simples hasta sistemas de soporte a la decisión, han sido parte de esta transformación y han permitido cambiar los servicios que Recursos Humanos brinda al interior de las organizaciones, así como su rol [2]. Aunque muchas decisiones requieren el uso del juicio, estas herramientas le permiten al área de Recursos Humanos tener mayor credibilidad profesional y rigor en la toma de decisiones. Por esto, Ulrich y Dulebohn [1] identifican la analítica de Recursos Humanos como una de las 4 áreas hacia las cuales se debe dirigir la inversión futura en las áreas de Recursos Humanos.

A pesar de esto, el uso de la analítica en las áreas de Recursos Humanos se ha desarrollado de manera más lenta que en otras funciones del negocio, como por ejemplo Mercadeo, donde se ha visto una revolución en la manera cómo las compañías buscan entender y brindar soluciones a sus consumidores, gracias a que la capacidad para recolectar, almacenar y analizar grandes volúmenes de datos ha cambiado la manera en la que se toman las decisiones en los negocios, aquellas que anteriormente se tomaban con base en el instinto o la experiencia [3]. De acuerdo con la investigación realizada por Marler y Bourdeau [4], el término “HR Analytics” aparece por primera vez en la literatura en el año 2003, más enfocado al contexto de la inteligencia de negocios que a la definición actual de Analytics y, a pesar de que en los últimos años la literatura alrededor del uso de la analítica en Gestión de Recursos Humanos ha venido aumentando, ésta aún es considerada como una innovación en estado de difusión y el porcentaje de compañías que declaran utilizar analítica en la definición o implementación de la estrategia de Recursos Humanos es sorprendentemente bajo.

Uno de los motivos por los que la analítica ha tenido una evolución más lenta en el campo de Recursos Humanos que en otros es debido a que los profesionales de esta área no tienen el conocimiento, las habilidades y la visión necesarios para hacerse las preguntas correctas sobre la data que tienen a la mano, esto sumado a que los profesionales de Recursos Humanos usualmente no están formados en el área de la analítica y los equipos de analítica usualmente no están

familiarizados con los Recursos Humanos [5]. Otros artículos resaltan la falta de habilidades de los profesionales de Recursos Humanos alrededor de la analítica como una de las principales barreras para avanzar en su incorporación [1] [6] [7] [8] [9]. Por otro lado, la mayoría de la literatura que aborda el tema de analítica en Recursos Humanos es de carácter promocional y brinda poca información que permita llevar las ideas a la práctica [10]; se enfoca en responder la pregunta ¿Cómo funciona la analítica en Recursos Humanos? y presenta un marco de referencia teórico [4]. Debido a que la implementación y difusión de la analítica en las áreas de Recursos Humanos depende en gran medida de los profesionales que lideran estas áreas, es relevante contar con referentes y estudios sobre las aplicaciones o usos que la analítica puede tener en las diferentes sub-áreas de Recursos Humanos.

El objetivo de esta investigación es identificar los diferentes usos o aplicaciones específicas que la analítica puede tener en los diversos campos de la Gestión de Recursos Humanos para brindar herramientas a los profesionales de Recursos Humanos así como a los investigadores, para decidir sobre futuras inversiones e investigaciones adicionales que permitan que el uso de la analítica en Recursos Humanos pase de la teoría a la práctica y continúe expandiéndose. Por lo tanto, esta investigación busca responder las siguientes preguntas, a través de un proceso de Revisión Sistemática de Literatura:

- 1- ¿Qué aplicaciones se le han dado a la analítica en el campo de la Gestión de Recursos Humanos y en qué áreas?
- 2- ¿Cuáles son las técnicas o herramientas de analítica y minería de datos más utilizadas en Recursos Humanos?
- 3- ¿Cómo ha evolucionado el uso de analítica en Recursos Humanos a través de los años?
- 4- ¿Cuáles son los principales resultados obtenidos?

# 1. Generalidades

## 1.1 Gestión de Recursos Humanos

El área de Recursos Humanos surgió en las organizaciones a principios del siglo XX con el objetivo de generar valor por medio de la gestión efectiva y la racionalización de las relaciones con los empleados, como resultado de las condiciones presentadas por la primera Guerra Mundial, incluyendo la escasez de mano de obra y el rápido incremento en los salarios [1]. Posterior a la guerra, el enfoque de Recursos Humanos fue interno, y buscaba manejar los retos que presentaban los sindicatos y las relaciones laborales, sin embargo, en los años 80 comenzó una transformación en la Gestión de Recursos Humanos, de un área administrativa hacia una función central del negocio que contribuye a la efectividad organizacional [1]. Los cambios en el entorno, incluyendo la globalización, cambios en la demografía, la búsqueda de la rentabilidad por medio del crecimiento, cambios en la tecnología, el capital humano y las organizaciones mismas han llevado a que, cada vez más, la Gestión de Recursos Humanos sea vista como una fuente de ventaja competitiva y cobre mayor importancia en las organizaciones [11].

Martha Alicia Alles en su libro *Dirección Estratégica de Recursos Humanos: Gestión por Competencias* [12], define un modelo para la Gestión de Recursos Humanos compuesto por las siguientes seis áreas o campos de trabajo:

- **Análisis y Descripción de Puestos:** su objetivo es recoger la información básica de los puestos de trabajo necesarios en una organización para definirlos correctamente, incluyendo las tareas a realizar, el contexto en el que se realizan, los requerimientos específicos y el tipo de personas que deben ocupar cada posición.
- **Atracción, Selección e Incorporación:** su objetivo es realizar una correcta selección de personas, asegurando que exista una correspondencia entre las expectativas del candidato y la empresa y buscando generar una relación fructífera entre ambos.
- **Desarrollo y Planes de Sucesión:** busca incrementar las capacidades intelectuales y conductuales de los empleados y de esta manera asegurar su crecimiento y avance de carrera, con el objetivo final de cuidar el capital intelectual de la compañía. Este campo está estrechamente relacionado con el de Formación y el de Evaluación de Desempeño.
- **Formación:** busca lograr que las personas realicen mejor su trabajo, preparándolas para desempeñarse con éxito en su puesto, en el presente y también a futuro.

- Evaluación del Desempeño: su objetivo es valorar el comportamiento del empleado en relación al trabajo y de esta manera poder tomar decisiones relacionadas con desarrollo, remuneración y formación.
- Remuneración y Beneficios: su objetivo es asegurar el pago equitativo dentro de la compañía, manteniendo su competitividad en el reclutamiento, contratación y en la retención de los empleados.

Los Sistemas de Información que se utilizan en la Gestión de Recursos Humanos almacenan información sobre los empleados en cada uno de estos procesos, incluyendo información demográfica, sobre su historia laboral, habilidades y competencias, educación, desempeño, la formación y desarrollo que ha tenido el empleado, casos disciplinarios, esquemas de participación, resultados de encuestas de actitud, entre otras [5], la cual puede ser utilizada para soportar la toma de decisiones.

De acuerdo con Ulrich y Dulebohn [1], hacia el futuro, Recursos Humanos debe estar enfocado en generar valor a la organización y para lograrlo, debe enfocar sus inversiones en:

- El departamento de Recursos Humanos, es decir, transformar la estructura del área
- Las prácticas de Recursos Humanos, buscando que estén alineadas con la estrategia y los grupos de interés externos, que sean integradas entre sí y que sean innovadoras.
- Los profesionales de Recursos Humanos, mejorando y actualizando las habilidades y competencias de los profesionales que trabajan en el área
- Analítica de Recursos Humanos, que brinde mayor rigor en la toma de decisiones y le permita al área tener mayor credibilidad profesional.

## 1.2 Analítica y minería de datos

La analítica es un método para la manipulación tecnológica de información que provee una visión sobre un asunto relevante y permite tomar decisiones con base en las evidencias [10]. Los datos analizados pueden ser estructurados, no estructurados, o una mezcla de ambos, es decir híbridos. Por lo tanto, la analítica permite comprender una situación a partir de una mezcla compleja de datos y encontrar valor oculto en éstos. Según Fitz-enz y Mattox [13], existen tres niveles de la analítica: la descriptiva, que describe relaciones entre datos y patrones históricos; la predictiva, que utiliza datos actuales e históricos para hacer predicciones del futuro con base en probabilidades y posibles resultados; y la prescriptiva, que va más allá de las predicciones y presenta alternativas de decisiones y sus impactos. La estadística es uno de los componentes

principales de la analítica pero ésta también implica el entendimiento del problema, sus interacciones y relaciones.

La minería de datos es un proceso de analítica que consiste en analizar grandes volúmenes de datos que están disponibles en distintas fuentes para transformarlos en información útil, es decir, es la búsqueda de información nueva, valiosa y no trivial, en grandes volúmenes de datos y es el resultado del esfuerzo conjunto entre las habilidades humanas para describir los problemas y las metas y las habilidades computacionales de búsqueda. El proceso de minería de datos cuenta de 5 pasos: 1) definir el problema, 2) recolectar los datos, 3) pre-procesar los datos, 4) estimar el modelo y 5) interpretar el modelo y sacar conclusiones. Al igual que la analítica, la minería de datos también se clasifica en descriptiva, cuando busca identificar patrones y describir la información para que pueda ser interpretada; o en predictiva, cuando busca predecir valores futuros o desconocidos de otras variables de interés [14] [15]. Las técnicas o métodos de minería de datos se clasifican en:

**Técnicas de clasificación:** permiten clasificar los datos en una o varias categorías previamente identificadas. Algunas de las técnicas de clasificación más utilizadas son [14] [15]:

- Aquellas basados en la teoría Bayesiana de decisión. El objetivo de esta teoría es minimizar la probabilidad de tomar una decisión equivocada.
- Aquellas basadas en reglas de clasificación, que utilizan reglas SI-ENTONCES para predecir la clase. En esta categoría se incluyen técnicas de árboles de decisión, algoritmos evolutivos y algoritmos genéticos.
- Redes neuronales, las cuales aprenden de la experiencia y errores del pasado y entrega un resultado con base en la comparación con un límite o umbral establecido.
- Máquinas de Soporte Vectorial (SVM), cuyo objetivo es encontrar un hiperplano óptimo en el espacio de los datos, con base en el principio de minimización del riesgo estructural (SRM) y la teoría de aprendizaje estadístico.
- K-nearest neighbor (KNN). Esta técnica busca identificar el valor óptimo de K que minimiza el error al calcular la distancia entre los datos de prueba y los de entrenamiento.
- Razonamiento basado en casos (CBR). Esta técnica almacena los datos como casos. Cuando existe un nuevo caso para clasificar, primero busca si ya existe un caso igual para determinar el resultado. Si no existe, busca aquellos que sean similares al nuevo caso.

**Técnicas de agrupamiento:** permiten identificar una serie de categorías en las que se pueden agrupar los datos [15]. Es similar a las técnicas de clasificación ya que agrupa los datos en categorías con la diferencia de que en este caso las categorías no están identificadas previamente. Algunas de las técnicas de agrupamiento son [14]:

- K-means. Es uno de los algoritmos no supervisados más simples y se utiliza para distinguir entre datos similares o no similares y generar una agrupación significativa.
- Fuzzy C-means. Esta técnica se basa en el principio de que cada dato puede pertenecer a más de un cluster y en cada uno tiene un valor diferente entre 0 y 1, de manera que la suma de todos es 1.

**Reglas de asociación:** Se utilizan para determinar la relación entre los elementos de un set de datos. Las reglas deben satisfacer un límite mínimo de soporte y un límite mínimo de confianza que son definidos por el usuario o por expertos en el área. Algunas técnicas de reglas de asociación son:

- El Algoritmo a priori. Es una técnica de reglas de asociación que se utiliza para analizar ítems frecuentes con base el conocimiento previo de las propiedades de éstos.
- Reglas de asociación multidimensionales. Se utilizan para analizar atributos que tienen más de una dimensión al mismo tiempo y las reglas de asociación cuantitativas, que se refieren a un tipo especial de reglas en la forma de  $X \rightarrow Y$  en la que al menos una de las variables debe ser numérica.
- Descubrimiento de secuencias. Se utiliza para identificar patrones en los eventos que están organizados dentro de un set de datos.

**Técnicas de regresión:** permite identificar una función de predicción para predecir el valor de variable [15]. Existen varios métodos de regresión, incluyendo análisis de series de tiempo, regresión no paramétrica, regresión robusta, regresión ridge, regresión no lineal, aprendizaje profundo o machine learning [14].

Otras técnicas de minería de datos incluyen [15]:

- Resumen (summarization): permite encontrar una descripción compacta de una serie de datos
- Modelos de dependencia: permite identificar un modelo que describe las dependencias entre variables
- Detección de cambios y desviaciones: permite identificar los cambios más significativos en una serie de datos.
- Algoritmos genéticos, que combinan métodos de evolución computacionales con algoritmos de optimización.
- Teoría de Rough Set (RS). Se utiliza para encontrar relaciones estructurales entre datos imprecisos o con ruido en el espacio de clasificación.
- Aquellas con enfoque Fuzzy. La teoría Fuzzy set permite trabajar con datos vagos o inexactos.

## 2. Protocolo de Investigación

En esta sección se describe el protocolo de investigación o los pasos metodológicos para lograr responder a las preguntas de investigación.

El método de investigación que se utilizará para responder a las preguntas de investigación presentadas anteriormente es el mapeo sistemático de literatura. Un mapeo sistemático de literatura es una metodología para identificar, evaluar e interpretar la literatura disponible sobre un tema en particular utilizando estudios primarios ya existentes y publicados [16]. A diferencia de una revisión de literatura simple, una revisión sistemática debe ser rigurosa y justa y se realiza siguiendo una estrategia de búsqueda definida previamente que incluye una etapa de planeación, una de ejecución y una de reporte [16]. Una revisión sistemática de literatura permite sintetizar la evidencia sobre un tema, identificar vacíos en la investigación actual y sugerir áreas para investigaciones futuras, así como proveer un marco de referencia para éstas [16].

A continuación se describe el protocolo de investigación:

- Base de datos bibliográfica: Scopus.
- Tipo de documentos: artículos publicados en revistas indexadas.
- Período analizado: toda la información disponible hasta agosto de 2017.
- Cadena de Búsqueda: Con el objetivo de obtener la mayor cantidad de artículos relacionados con el uso de la analítica en Recursos Humanos, se definieron palabras claves relacionadas con estas dos áreas de estudio (la analítica y Recursos Humanos). Debido a que el área de Recursos Humanos es amplia y tiene muchas sub-áreas y procesos relacionados, la cadena de búsqueda utilizada incluyó palabras clave identificadas en artículos revisados en la preparación inicial de la investigación.

```
((TITLE-ABS-KEY ( "analytics" ) OR TITLE-ABS-KEY ( "data analytics" ) OR TITLE-ABS-KEY ( "data science" ) OR TITLE-ABS-KEY ( "data mining" ) OR TITLE-ABS-KEY ( "big data" )) AND ( TITLE-ABS-KEY ( "employee recruit*" ) OR TITLE-ABS-KEY ( "talent staff*" ) OR TITLE-ABS-KEY ( "employee selection" ) OR TITLE-ABS-KEY ( "hiring" ) OR TITLE-ABS-KEY ( "Talent attract*" ) OR TITLE-ABS-KEY ( "human resources" ) OR TITLE-ABS-KEY ( "human resource management" ) OR TITLE-ABS-KEY ( "talent manag*" ) OR TITLE-ABS-KEY ( "career manag*" ) OR TITLE-ABS-KEY ( "Succession plan*" ) OR TITLE-ABS-KEY ( "career plan*" ) OR TITLE-ABS-KEY ( "Employee Assess*" ) OR TITLE-ABS-KEY ( "performance manag*" ) OR TITLE-ABS-KEY ( "Employee Develop*" ) OR TITLE-ABS-KEY ( "employee reten*" ) OR TITLE-ABS-KEY ( "turnover prediction" ) OR TITLE-ABS-KEY ( "employee engag*" ) OR TITLE-ABS-KEY ( "Talent Plan*" ) OR TITLE-ABS-KEY ( "WORKFORCE PLAN*" ) OR ( title AND abs KEY ( "compensation" AND "employee" ) ) ) ) )
```



```
AND (LIMIT-TO (DOCTYPE , "ar")) AND (LIMIT-TO (SRCTYPE , "j")) AND (EXCLUDE (SUBJAREA , "MEDI") OR EXCLUDE (SUBJAREA , "ENER") OR EXCLUDE (SUBJAREA , "ENVI") OR EXCLUDE (SUBJAREA , "AGRI") OR EXCLUDE (SUBJAREA , "HEAL") OR EXCLUDE (SUBJAREA , "MATE") OR EXCLUDE (SUBJAREA , "EART") OR EXCLUDE (SUBJAREA , "PHYS") OR EXCLUDE (SUBJAREA , "BIOC") OR EXCLUDE (SUBJAREA , "CENG") OR EXCLUDE (SUBJAREA , "CHEM") OR EXCLUDE (SUBJAREA , "NEUR") OR EXCLUDE (SUBJAREA , "NURS"))
```

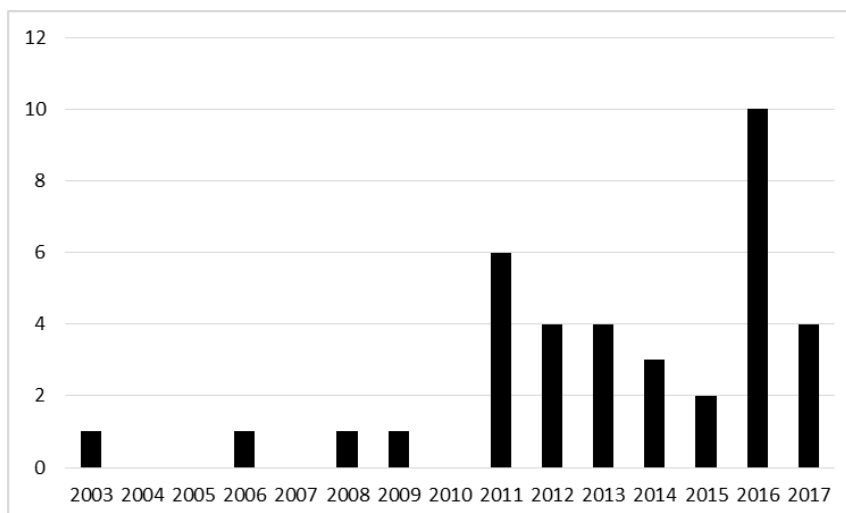
- Criterios de inclusión y exclusión: Luego de la recuperación automática de la cadena de búsqueda descrita anteriormente, se realizó una revisión manual del título y el resumen de los artículos. Se incluyeron artículos con base en los siguientes criterios:
  - (1) Se da solución a un problema específico de una de las áreas relacionadas con la gestión del Talento Humano en un negocio.
  - (2) Hace referencia al uso de herramientas de analítica para la solución del problema.
- Áreas temáticas: Se excluyeron las áreas de Medicina, Enfermería, Ciencia del Medio Ambiente, Agricultura y Ciencias biológicas, Salud, Matemáticas, Energía, Física, Ingeniería Química, Neurología, Ciencia de los Materiales, Ciencia de la Tierra y los Planetas, Química, Neurociencia, Bioquímica, Genética y Biología Molecular.
- Información recolectada: Los siguientes datos fueron extraídos de cada artículo seleccionado para realizar el análisis de los resultados:
  - Autores.
  - Nombre de la revista.
  - Año de publicación.
  - Citaciones al artículo.
  - Índice H de los autores.
  - País de procedencia
  - Palabras clave
  - Descripción del problema
  - Áreas de la Gestión del Talento Humano a las que se refiere el problema abordado
  - Técnicas de analítica y/o minería de datos utilizadas en la solución del problema
  - Tipo de modelo utilizado

### 3. Resultados

Como resultado de la búsqueda utilizando la cadena descrita anteriormente se encontraron 200 documentos. Después de realizar la revisión manual y aplicar los criterios de inclusión y exclusión se seleccionaron 37 artículos para responder las preguntas de investigación, los cuales se encuentren incluidos en la parte final de las referencias bibliográficas [17 – 52]. Esto implica que la cadena de búsqueda solo tiene una precisión de 37/200, lo que es bastante bajo e implica que un estudio posterior a este trabajo, estaría basado fundamentalmente en la revisión por parte del experto más que en la obtención automática de los documentos, lo que dificulta enormemente la tarea de actualización del estudio.

La Figura 1 muestra la cantidad de artículos escritos por año. El primer artículo encontrado fue del año 2003 y posteriormente se mantuvo en niveles muy bajos hasta el año 2011. En este año hubo un incremento significativo con respecto a los años anteriores, pero no se sostuvo para los años siguientes, disminuyendo en el año 2012 y manteniéndose estable hasta el 2015. Posteriormente, el año 2016 registró la mayor cantidad de artículos, mostrando un incremento significativo, pero disminuyó de nuevo en el año 2017 a las cantidades que se venían generando en los años anteriores.

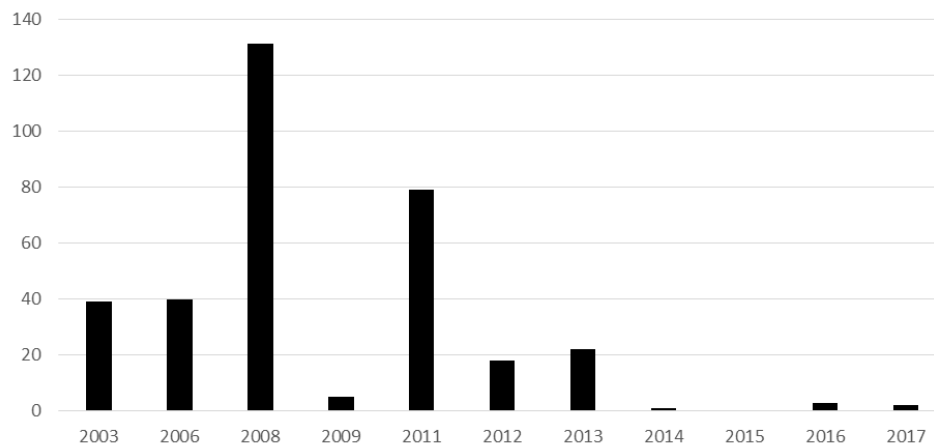
**Figura 1: Cantidad de documentos publicados por año**



Estos resultados reflejan la lenta evolución que ha tenido el uso de la analítica en los campos de la Gestión de Recursos Humanos, como se mencionó en la introducción de esta investigación, a pesar del gran potencial que existe en esta área.

Por otro lado, la Figura 2 muestra el número de citaciones de los artículos, según su año de publicación.

**Figura 2: Cantidad de citaciones, según el año de publicación**



Se puede concluir que los primeros artículos publicados que abordan las aplicaciones de la analítica y la minería de datos en las áreas de Recursos Humanos, comprendidos entre el 2003 y el 2008, son los de mayor relevancia para futuros estudios, especialmente aquellos escritos en el año 2008, en donde sólo se identificaron 2 artículos publicados pero los cuales han tenido la mayor cantidad de citaciones.

El 43% de los artículos no han tenido ninguna citación y tan sólo el 22% tienen más de 10 citaciones, como se muestra en la Tabla 1, lo cual indica que los investigadores usan como referencia los mismos artículos base, que, como se evidenció en la Figura 2, son aquellos artículos escritos inicialmente.

**Tabla 1: Citaciones por artículo**

Citaciones	Número de Artículos	Porcentaje del total
Ninguna	16	43%
Entre 1 y 10	13	35%
Más de 10	8	22%

En la Tabla 2 se listan los artículos que han tenido más de 10 citas, incluyendo la información sobre sus autores. El artículo que tiene la mayor la cantidad de citas es del año 2008, seguido por los artículos publicados en los años 2003 y 2006, lo cual coincide con lo observado previamente en la Figura 2. Esto permite concluir que estos artículos son los de mayor relevancia para las investigaciones que se han desarrollado posteriormente. El año 2011, que fue el año con mayor número de citas después del 2008, tiene tres artículos incluidos en la Tabla 2.

**Tabla 2: Artículos con más de 10 citas**

Título	Año	Autor (índice H)	Índice H del autor	Citaciones
Data mining to improve personnel selection and enhance human capital: A case study in high-technology industry	2008	Chien, C.F. Chen, L.F	23 9	131
Integrating fuzzy data mining and fuzzy artificial neural networks for discovering implicit knowledge	2006	Huang, M.J. Tsou, Y.L. Lee, S.C.	7 1 2	40
Developing the profiles of truck drivers for their successful recruitment and retention: A data mining approach	2003	Min, H. Emam, A.	28 3	39
Formal model for assigning human resources to teams in software projects	2011	André, M. Baldoquín, M.G. Acuña, S.T.	1 3 9	37
Data mining techniques for better decisions in human resource management systems	2008	Ranjan, J. Goyal, D.P. Ahson, S.I.	9 15 8	31
Employee churn prediction	2011	Saradhi, V.V. Palshikar, G.K.	2 7	17
Visualizing Informal Professional Development Networks: Building a Case for Learning Analytics in the Workplace	2013	de, Laat, M. Schreurs, B.	13 4	14
Talent and analytics: new approaches, higher ROI	2011	Harris, J.G. Craig, E. Light, D.A.	10 2 4	13
An integrated e-recruitment system for automated personality mining and applicant ranking	2012	Faliagka, E. Tsakalidis, A. Tzimas, G.	4 18 9	11

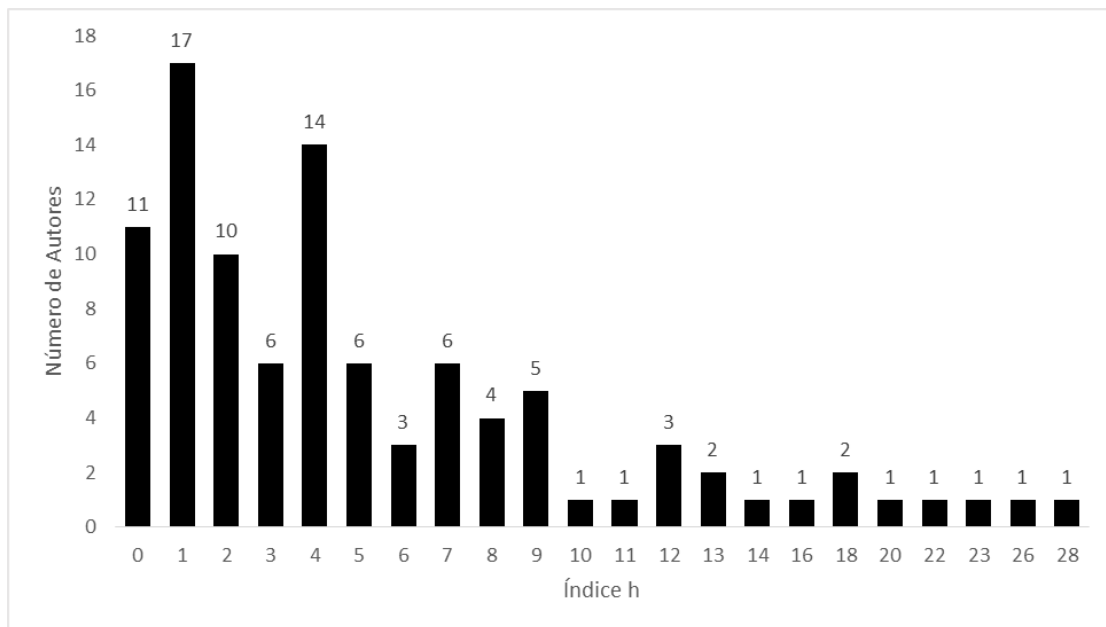
Ninguno de los autores se repite dentro de la Tabla 2 y sólo cuatro de estos autores se encuentran dentro de los diez autores con el mayor índice H, como se observa en la Tabla 3: Min, H. (índice H: 28), Chien, C.F. (índice H: 23), Tsakalidis, A. (índice H: 18) y Goyal, D.P. (índice H: 15). Lo

anterior también es un indicador de la baja dedicación de estos autores a los temas bajo estudio en esta investigación, como se mencionó anteriormente. Es decir, no hay una continuidad en la investigación en esta área por parte de los autores; esto va en contra de la maduración del campo, ya que se espera que la publicación de mayor calidad se dé después de que el autor ha ganado experiencia en el campo, Por otra parte, y a pesar de que hay autores con índices H altos y una gran cantidad de citas, éstas no corresponden con el tema central de la presente investigación.

**Tabla 3: Autores con el mayor índice H**

<b>Autor</b>	<b>Cantidad de citas en general</b>	<b>Índice H del autor</b>
Min, H.	3423	28
Angelis, L.	1803	26
Chien C.F.	2280	23
Colomo-Palacios, R.	324	22
Datta, A.	1922	20
García-Peñalvo, F.J.	1447	18
Tsakalidis, A.	1118	18
Fan, C.Y.	875	16
Nembhard, D.A.	686	14
De Laat, M.	806	13
Hoppe, H.U.	630	13

Los 37 artículos seleccionados fueron escritos por un total de 98 autores. El 98% de los autores sólo tienen 1 artículo que hace parte de los resultados de esta búsqueda y sólo dos autores (Chien, C.F. y Chen L.F.) cuentan con 2 publicaciones que están incluidas en los documentos seleccionados para la presente investigación. Adicionalmente, en la Figura 3 se observa que la mayor concentración de autores está en los índices H entre cero y uno, seguido de índices H entre 2 y 7. Lo anterior indica que el área de estudio seleccionada no tiene una dedicación por parte de los autores y adicionalmente éstos tienen una baja actividad académica o investigativa y, por el contrario, son muy pocos los autores que han escrito sobre el área sujeto de investigación, que tienen una relevancia más significativa a nivel científico.

**Figura 3: Histograma de índice h de los autores**

La Tabla 4 muestra los nombres de las revistas donde se encuentran artículos que hacen referencia al uso de la analítica en las áreas de Recursos Humanos. Se observa que hay gran dispersión en las fuentes que abordan este tema. La revista que tiene la mayor cantidad de artículos es “Expert Systems With Applications”, con un total de 5 artículos. Las demás publicaciones solo tiene un artículo.

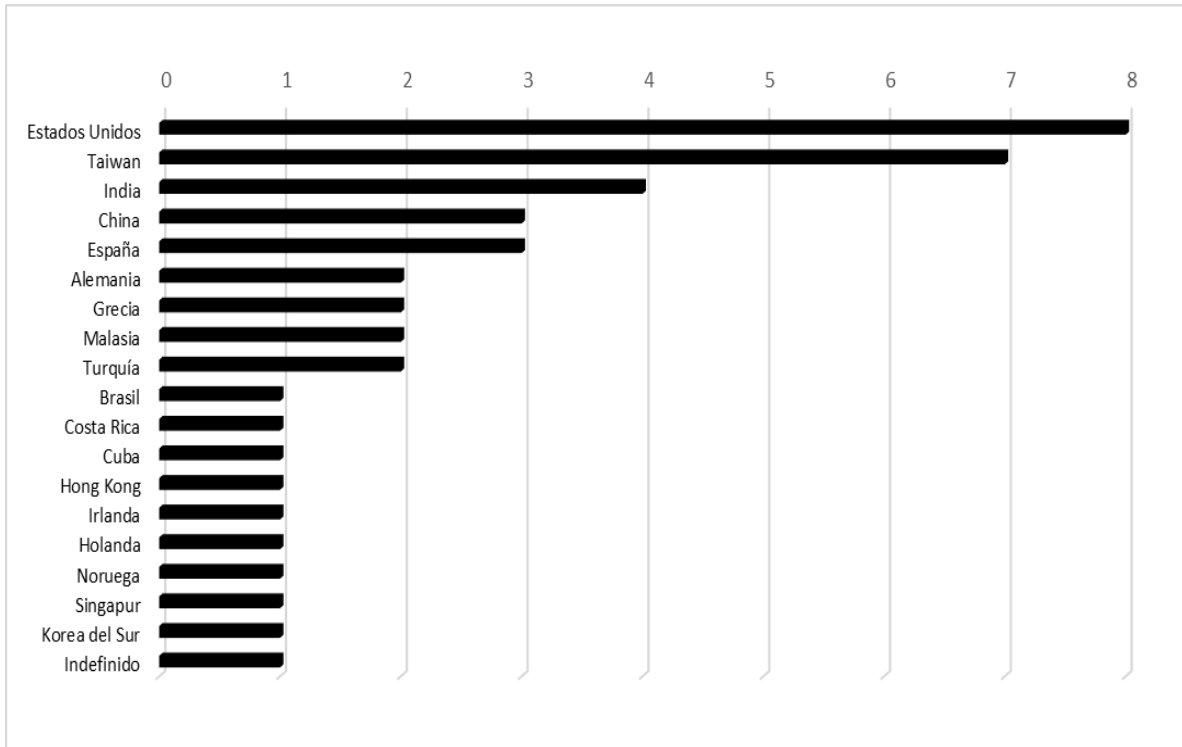
**Tabla 4: Número de Artículos según fuente de Publicación**

<b>Nombre de la Revista</b>	<b>Artículos</b>
Expert Systems with Applications	5
American Behavioral Scientist	1
Computers and Industrial Engineering	1
Education and Information Technologies	1
Flexible Services and Manufacturing Journal	1
Human Factors and Ergonomics In Manufacturing	1
Human Resource Development Review	1
Human Resource Management International Digest	1
Indian Journal of Science and Technology	1
Information and Software Technology	1
INFORMS Journal on Computing	1
International Journal of Advanced Manufacturing Technology	1
International Journal of Industrial Engineering : Theory Applications and Practice	1
International Journal of Physical Distribution and Logistics Management	1
International Journal of Production Research	1
International Journal of Reasoning-based Intelligent Systems	1
International Journal of Sports Science and Coaching	1
Internet Research	1
Journal of Business Strategy	1
Journal of Computers	1
Journal of Leadership Studies	1
Journal of Network and Computer Applications	1
Journal of Networks	1
Journal of Quality	1
Journal of the Association for Information Science and Technology	1
Journal of Theoretical and Applied Information Technology	1
Knowledge Management and E-Learning	1
Knowledge-Based Systems	1
Management Research Review	1
Procesamiento de Lenguaje Natural	1
Science of Computer Programming	1
Turkish Journal of Electrical Engineering and Computer Sciences	1
World Academy of Science, Engineering and Technology	1

Al analizar los países de los cuales provienen los artículos, se observa en la Figura 4 que tres países lideran las investigaciones en el área de analítica aplicada en la Gestión de Recursos Humanos. Estados Unidos es el país que produce la mayor cantidad de artículos, con un total de 8. Le siguen Taiwan e India, con un total de 7 y 4 respectivamente. Hay una ausencia de trabajos

originados en América Latina, y esto podría explicar, al menos en parte, la ausencia de experiencia en el uso de la analítica tanto a nivel empresarial como investigativo.

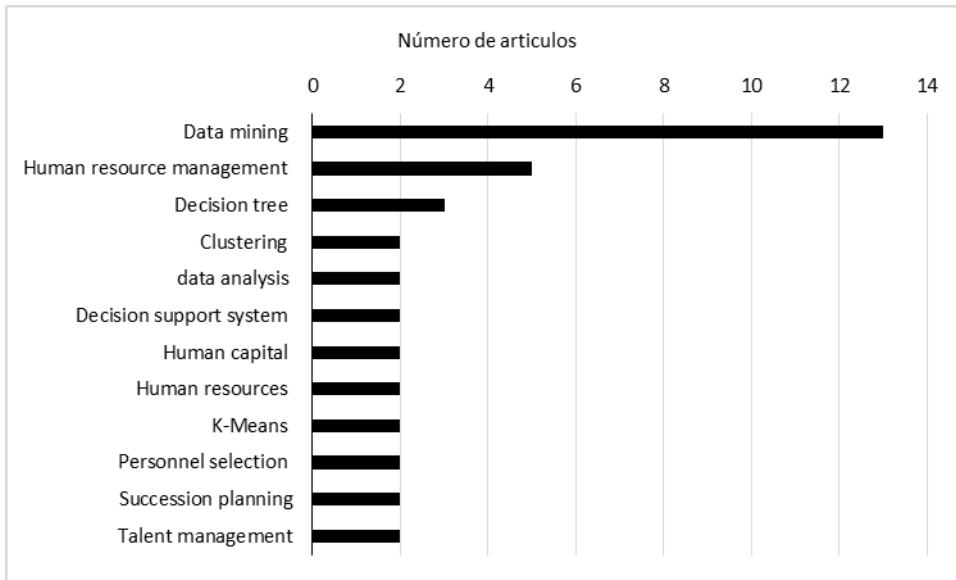
**Figura 4: Número de artículos según el país de procedencia**



Finalmente, se encontraron un total de 151 palabras clave. Esta gran dispersión en las palabras clave usadas en los artículos muestra el amplio rango de áreas que se consideran al investigar la analítica en la Gestión de Recursos Humanos. El 92% de las palabras clave se encuentran presentes sólo en un artículo. En la Figura 5 se presenta el histograma de palabras clave que se repiten más de una vez. La palabra clave más utilizada es data mining, que se encuentra en 13 de los 37 artículos, lo cual permite pensar que la minería de datos es una de las herramientas de analítica más utilizadas en Recursos Humanos. El problema radica en que en la minería de datos se usan técnicas provenientes tanto de la estadística como de la inteligencia artificial, por lo que el término por sí mismo resulta demasiado amplio como palabra clave que ayude a la recuperación de un artículo en una base de datos bibliográfica. Este supuesto será revisado en detalle más adelante en esta investigación.



**Figura 5: Histograma de Palabras Clave que se repiten más de una vez**

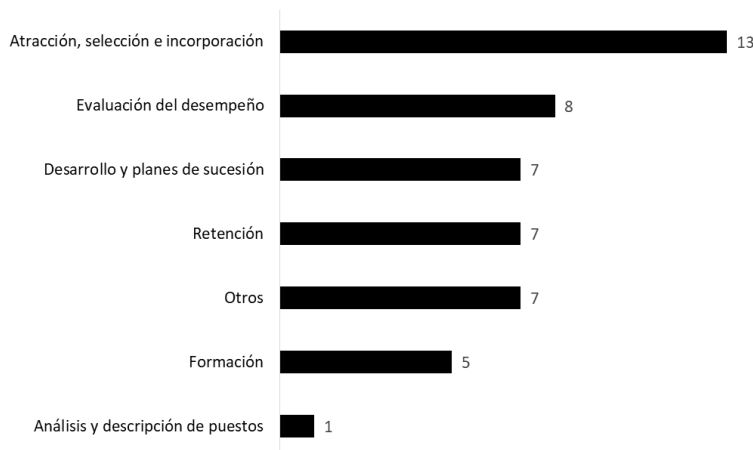


## **4. Discusión**

A continuación se responden las preguntas de investigación planteadas al inicio de este documento. Algunos de los artículos describen casos de aplicación de la analítica que impactan varios procesos en diferentes campos de Recursos Humanos y adicionalmente algunos artículos describen diferentes casos de aplicación [17] [18] [19] [20] [21] [22]. En estos casos, los resultados que se presentan a continuación consideran cada una de las aplicaciones de manera independiente. Por otro lado, algunos de los artículos abordan aplicaciones de la analítica y la minería de datos en Recursos Humanos, sin embargo, no describen el modelo o la técnica utilizada. Lo anterior da como resultado un total de 29 modelos que tienen 48 aplicaciones diferentes identificadas en los 37 artículos.

### **4.1 ¿Qué aplicaciones se le han dado a la analítica y la minería de datos en la Gestión de Recursos Humanos y en qué áreas o campos?**

Para efectos de esta investigación se utilizaron los siguientes campos o áreas dentro de la Gestión de Recursos Humanos, propuestos por Martha Alicia Allice: 1- Análisis y Descripción de Puestos; 2- Atracción, Selección e Incorporación; 3- Desarrollo y Planes de Sucesión (el cual incluye el aprovisionamiento interno o asignación de empleados); 4- Formación; 5- Evaluación del Desempeño; 6- Remuneración y Beneficios [12]. Adicionalmente, se incluyó un área adicional denominada “Retención”, debido a la relevancia de este campo en los resultados encontrados. Y se definió un categoría adicional de “Otros” para incluir otras aplicaciones que no están contenidas en los campos inicialmente descritos. Utilizando esta clasificación, los resultados de la investigación se presentan en la Tabla 5.

**Tabla 5: Número de casos encontrados por área de aplicación**

Como se observa, en la literatura analizada no se encontró evidencia de aplicaciones de la analítica o de minería de datos en el área de Remuneración y Beneficios. La principal aplicación que ha tenido la analítica en la Gestión de Recursos Humanos ha sido en el área de Atracción, Selección e Incorporación de Empleados, donde se ha utilizado con los siguientes propósitos: identificar los factores o características más relevantes a tener en cuenta en la búsqueda de candidatos que aseguren su éxito o su buen desempeño [18] [22] [23]; desarrollar modelos de selección que se adapten a los cambios en las estrategias de selección [24]; identificar los candidatos que mejor se ajustan a los requisitos de un perfil o cargo o a la cultura organizacional, ya sea en la etapa inicial del filtro o en la etapa de la selección [17] [18] [21] [25] [26], facilitar la lectura de hojas de vida extrayendo de éstas la información requerida [27], predecir comportamiento en el trabajo de los candidatos a una vacante, incluyendo predicciones de desempeño y/o de retención [20] [28]; y evaluar los proveedores o fuentes de talento [20].

El campo de la Evaluación del Desempeño es el segundo que ha tenido mayor uso de analítica para dar solución a los problemas que se presentan en éste. En este campo, las aplicaciones que ha tenido la analítica tienen principalmente tres objetivos: mejorar el modelo de evaluación de desempeño al identificar los criterios más relevantes para predecir el desempeño y que impactan directamente en éste [19] [29] [30] [31]; mejorar la calidad en las evaluaciones de desempeño, al reducir su subjetividad y/o predecir el desempeño futuro de los empleados [32] [33]; y evaluar las competencias de los empleados con respecto a un rol o actividad específicos [17] [21].

En el área de Desarrollo y Planes de Sucesión, la analítica ha sido utilizada principalmente en la asignación de empleados internos a roles, actividades específicas o equipos de trabajo, identificando aquellos que mejor se ajustan a los requisitos de éstos [18] [34] [35] [36] [37]. También ha sido utilizada para identificar las brechas en competencias y conocimientos (necesidades de desarrollo) de los empleados con respecto a roles específicos con el fin de diseñar

planes de desarrollo ajustados y/o planes de carrera [20]; y para identificar las competencias de liderazgo más relevantes que empujan los resultados del negocio, con el fin de enfocar el desarrollo de los líderes en estas competencias [38].

En el campo de Retención de Empleados, la analítica ha sido utilizada para predecir el abandono de los empleados, identificando aquellos con mayor riesgo de abandono y/o los motivos por los cuales los empleados deciden abandonar la compañía, lo cual permite implementar estrategias y acciones para evitarlo [10] [18] [20] [21] [39] [40]. También se ha utilizado para medir aquellos empleados más valiosos, hacia los cuales deben enfocarse estrategias de retención [22] [40].

En el área de Formación, la analítica ha sido utilizada buscando optimizar los recursos invertidos en ésta, por medio estrategias como: aumentar la eficiencia de los Sistemas de Gestión del Aprendizaje (o Learning Management Systems, LMS) al proveer a cada empleado entrenamientos que se ajusten a su perfil de aprendizaje [41]; explorar la calidad de los sistemas de aprendizaje e identificar indicadores críticos para su evaluación [42]; evaluar la efectividad de los entrenamientos (a través del desempeño y comportamientos de los estudiantes) [43]; permitir la identificación y mejor aprovechamiento de las redes informales de aprendizaje [44] y mejorar la planeación de los programas de formación [45].

En el área de Análisis y Descripción de Puestos sólo se identificó una aplicación de la analítica, la cual tiene como objetivo la creación y mantenimiento automático de la ontología de competencias [46].

Finalmente, en la categoría Otros se agruparon aquellos artículos que no están relacionados con las áreas ya descritas, incluyendo: tres aplicaciones en la gestión de la colaboración, una de las cuales busca identificar los patrones de colaboración que mejoran la eficiencia de los equipos de trabajo [47]; y otras dos pretenden identificar las relaciones y comportamientos colaborativos entre personas, cargas de trabajo y autonomía de quienes que hacen parte de un flujo de trabajo o un proyecto [48] [49]. Dos investigaciones están relacionadas con la planeación o asignación del personal: por un lado, una de éstas tiene como objetivo generar una serie de reglas para la asignación de trabajadores a determinadas máquinas o tareas que permitan optimizar los recursos (cubrir las necesidades de producción y al mismo tiempo brindar el entrenamiento requerido para generar flexibilidad en la fuerza de trabajo) [50]; y la otra busca predecir las necesidades de personal futuras con base en posibles escenarios, identificando número el número de personas requeridas, el tipo (interno o externo) y las habilidades requeridas [20]. Otro de los artículos tiene como objetivo identificar los factores más relevantes que se deben considerar en cierto tipo de compañías para la implementación de sistemas de sugerencias, en comparación con los sistemas de sugerencias exitosos en otras compañías [51]. Finalmente, uno de los artículos analizados se refiere a la aplicación de técnicas de minería de datos en la Gestión de Recursos Humanos de

manera general para analizar la información contenida en los sistemas de Recursos Humanos, sin referirse a un área en particular [52].

## **4.2 ¿Cuáles son las técnicas o herramientas de analítica y de minería de datos más utilizadas en Recursos Humanos?**

La minería de datos es el área de la analítica que ha tenido mayor aplicación en la Gestión de Recursos Humanos. En los documentos analizados, el 59% de los modelos descritos utilizan técnicas de minería de datos para dar solución al problema que se planea y el 20% utilizan otras técnicas de analítica. El 22% de las aplicaciones encontradas no describen el tipo de técnica utilizada.

La minería de datos se utiliza en los artículos para dar solución a problemas relacionados principalmente con Clasificación (en el 36% de las aplicaciones) Reglas de Asociación (en el 27%), Agrupamiento (en el 23%), Selección de características (en el 8%) y Regresión (2%), ya sea de manera individual o una combinación de éstos. Las técnicas de Agrupamiento son aquellas que se utilizan con mayor frecuencia en combinación con otras (en el 24% de las aplicaciones), principalmente con aquellas de Clasificación o Reglas de Asociación. Adicionalmente, uno de los artículos combina minería de datos con la Analítica de Redes Sociales. .

Otros campos diferentes a la minería de datos, han sido aplicados en la Gestión de Recursos Humanos, pero en menor medida, incluyendo Analítica o Minería de Texto [18] [26] [27] [51], Analítica de Redes sociales [44] [46] [48], Analítica visual [49] e Inteligencia Colectiva [34].

La Tabla 6 muestra el número de casos en los que se aplica cada tipo de técnica, según el área de Recursos Humanos, ya sea de manera individual o en combinación con otras.

**Tabla 6: Número de casos en los que se aplica cada tipo de técnica de la analítica o minería de datos, según el área de Recursos Humanos**

Técnica \ Área de la GRH		Atracción, selección e incorporación	Evaluación del desempeño	Desarrollo y planes de sucesión	Otros	Retention	Formación	Análisis y descripción de puestos	Total	
Técnicas Utilizadas	Minería de Datos	Clasificación	5	3	2	-	4	3	-	17
		Reglas de Asociación	4	3	1	3	1		1	13
		Agrupamiento	2	2	1	1	-	4	1	11
		Selección de Características	1	2	-	-	1	-	-	4
		Regresión	-	-	-	1	-	-	-	1
	Otras técnicas de Analítica	Análítica de Texto	2	-	-	1	1	-	-	4
		Análítica de Redes Sociales	-	-	-	1	-	1	1	3
		Análítica Visual	-	-		1	-	-		1
		Inteligencia Colectiva	-	-	1	-	-	-	-	1
	Aplicaciones en que se combinan varias técnicas		3	3	3	2	1	4	1	17
Técnica no especificada		3	2	3	1	1	-	-	10	
Total Aplicaciones identificadas		13	8	7	7	7	5	1	48	

Se puede observar que en todas las áreas de Recursos Humanos han sido aplicadas diversas técnicas de minería de datos o de analítica para resolver los problemas inherentes a éstas, es decir, ninguna técnica se utiliza de forma exclusiva en una de las áreas o viceversa. Las técnicas de minería de datos son las más utilizadas en todas las áreas. Otras técnicas de analítica han sido utilizadas en menor medida, principalmente en campos identificados como “Otros”, en el que se han aplicado la mayor cantidad de técnicas de analítica diferentes a la minería de datos, incluyendo Analítica de Redes Sociales, de Texto y Visual.

La Tabla 7 muestra las diferentes técnicas de minería de datos descritas en los artículos, para aquellos casos en los que se describen de manera detallada, que fueron aplicadas ya sea en modelos simples o híbridos. En los casos en los cuales la investigación incluyó una etapa de evaluación de varios modelos y selección del mejor, sólo se incluye aquella técnica utilizada en el modelo seleccionado. Se observa que hay una gran variedad de técnicas aplicadas. K-Means, Redes Neuronales y Árboles de Decisión son las más utilizadas, considerando las aplicaciones en modelos simples e híbridos, seguido por las técnicas de Algoritmo a priori y Máquinas de Soporte Vectorial. Las demás técnicas sólo utilizan en una de las aplicaciones, y en su mayoría se utilizan en modelos híbridos, excepto K-nearest neighbor (en una versión mejorada), Análisis de Clases Latentes, Fuzzy C-means y Gradient Boosted Machine, que fueron aplicadas en modelos simples.

**Tabla 7: Técnicas de minería de datos utilizadas y número de aplicaciones**

Técnica	Tipo de Problema	Cantidad de veces utilizada en modelos simples	Cantidad de veces utilizada en modelos híbridos	Número total de veces aplicada
K-means	Agrupamiento	1	5	6
<b>Redes Neuronales</b>		<b>2</b>	<b>4</b>	<b>6</b>
Redes Neuronales de propagación hacia atrás	Clasificación / Agrupamiento	-	4	4
Redes Neuronales Fuzzy	Reglas de Asociación	1	-	1
Redes Neuronales Combinadas	Reglas de Asociación	1	-	1
Árboles de Decisión	Clasificación / Reglas de Asociación	3	3	6
Algoritmo a priori	Reglas de Asociación	2	2	4
Máquinas de Soporte Vectorial	Clasificación	1	1	2
K-nearest neighbor (mejorado)	Clasificación	1	-	1
Análisis de clases latentes	Reglas de Asociación	1	-	1
Fuzzy C-means	Agrupamiento	1	-	1
Gradient Boosted Machine	Clasificación	1	-	1
Self-organizing map (SOM)	Agrupamiento	-	2	2
Regression Operating Characteristic (ROC)	Clasificación	-	1	1
RST (rough set theory)	Selección de Características	-	1	1
Minería de Secuencias	Reglas de Asociación	-	1	1
Regresión Cox	Regresión	-	1	1
Algoritmo Scott-Knott (SK)	Agrupamiento	-	1	1
Método de Varianza Mínima de Ward	Agrupamiento	-	1	1
Algoritmo GRASP	Clasificación	-	1	1
Algoritmo de Escalada Simple	Clasificación	-	1	1
Algoritmo de recocido simulado	Clasificación	-	1	1
Búsqueda Tabu	Clasificación	-	1	1

### Otras Técnicas de analítica

En las otras aplicaciones de analítica, se identificaron las siguientes técnicas específicas:

- En la Analítica de Redes Sociales se utilizaron modelos de descubrimiento, medidas de centralidad y medidas de prestigio.

- Al aplicar Analítica de Texto se utilizaron Procesos de Jerarquía Analítica (o proceso jerárquico analítico), Algoritmo Jaro Winkler, Análisis de Semántica Latente, Pellet Reasoner y SMS (Semantic Matching Step)
- En el caso en que se aplicó Inteligencia Colectiva se utilizaron técnicas como Coeficiente de Agrupamiento, Crawling y Crowd Sourcing

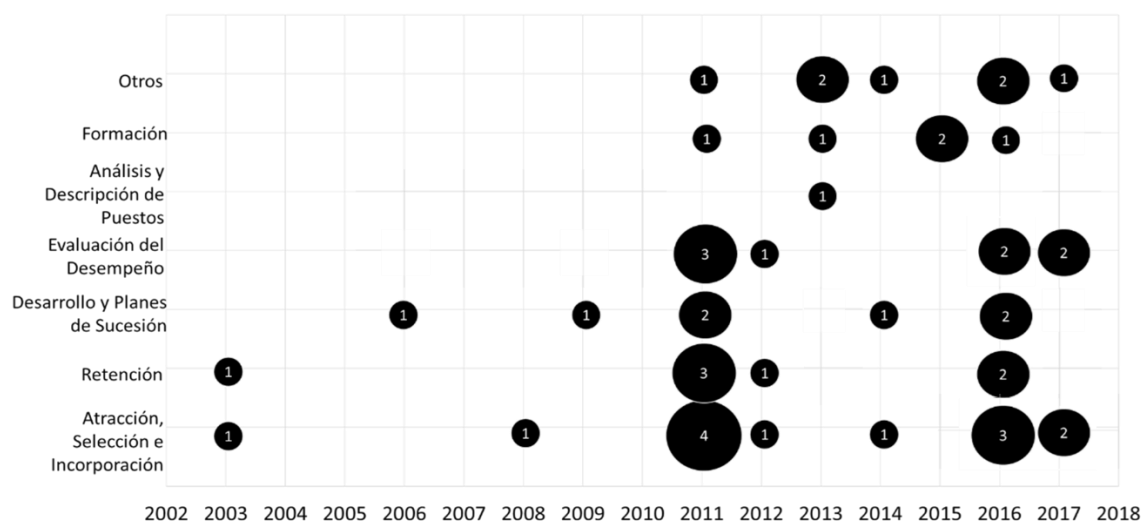
En conclusión, existe una gran cantidad de técnicas utilizadas al aplicar analítica y minería de datos en los problemas de la Gestión de Recursos Humanos. En general, las que han tenido mayor aplicación han sido técnicas de Clasificación, Agrupamiento y Reglas de Asociación, en función del problema específico que se quiere resolver.

### 4.3 ¿Cómo ha evolucionado el uso de analítica en Recursos Humanos a través de los años?

Para responder esta pregunta se analizó la evolución que ha tenido la analítica y la minería de datos a través de los años según el campo de aplicación y las técnicas utilizadas.

La Figura 6 muestra el número de aplicaciones descritas en los artículos publicados, según el área de Recursos Humanos que se aborda y el año de publicación.

**Figura 6: Número de aplicaciones descritas según el área de Recursos Humanos y el año de publicación**



Se observa que las investigaciones publicadas durante los primeros 6 años que abordan las aplicaciones de la analítica y minería de datos en las áreas de Recursos Humanos están enfocadas



en problemas de Atracción, Selección e Incorporación, de Retención, así como de Desarrollo y Planes de Sucesión. Estas áreas son el foco de estudio hasta el año 2009 y siguen siendo estudiadas en los años posteriores hasta el 2017. Sólo hasta el año 2011, en el cual se registró un aumento en la literatura, se encontraron aplicaciones en el área de Formación, Evaluación del Desempeño y en Otras áreas, y en el 2013 se registró una nueva aplicación en el área de Análisis y descripción de puestos. Luego de este año, todas las áreas de Recursos Humanos se han mantenido vigentes y siguen siendo incluidos en las investigaciones de manera consistente.

La Tabla 8 y la Tabla 9 muestran, para cada año, el número de modelos que se describen en los artículos según su tipo (simple o híbrido) y según el tipo de técnica aplicada respectivamente. Los modelos clasificados como “híbrido” son aquellos que se construyen con la combinación de varios algoritmos o técnicas y los clasificados como “individual” son aquellos que sólo incluyen una técnica específica.

Como se observa en las Tablas 8 y 9, en los años iniciales (entre el 2003 y el 2009) los artículos que describen aplicaciones de analítica y minería de datos en la Gestión de Recursos Humanos registran modelos de minería de datos, en los cuales se resuelven problemas de Clasificación y Reglas de Asociación de manera individual (no híbridos). A partir del año 2011, se comenzaron a registrar modelos híbridos en las aplicaciones, inicialmente combinando principalmente técnicas de Clasificación, Agrupamiento y Reglas de Asociación y posteriormente con otras técnicas como Selección de Características, Regresión y analítica de Redes Sociales. A partir de este año se han aplicado los modelos híbridos en mayor medida que aquellos en los cuales se utiliza una sola técnica. A partir del año 2012 se encontraron registros de aplicaciones que utilizan técnicas de analítica diferentes a la minería de datos, incluyendo Analítica de Redes Sociales, Analítica de Texto, Analítica Visual e Inteligencia Colectiva. Aunque estas técnicas han sido aplicadas en menor medida que la minería de datos, sus aplicaciones han ido aumentando a través de los años desde el 2012 hasta el 2016, incluyendo también casos en los cuales se aplican éstas junto con otras técnicas de minería de datos. A pesar de la investigación de nuevas técnicas de analítica en las áreas de Recursos Humanos, las investigaciones que aplican la minería de datos siguen ocupando la mayor parte de la literatura, siendo los más importantes aquellos relacionados con Clasificación, Agrupamiento y Reglas de Asociación.

**Tabla 8: Número de modelos descritos según su tipo y el año**

Tipo de Modelo	Año											Total
	2003	2006	2008	2009	2011	2012	2013	2014	2015	2016	2017	
Simple	1	1	1		2	1	1	2		4	1	14
Híbrido					3	2	2	1	2	4	2	16
No Especificado				1	6		1			2	1	11
<b>Total</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>11</b>	<b>3</b>	<b>4</b>	<b>3</b>	<b>2</b>	<b>10</b>	<b>4</b>	<b>41</b>

**Tabla 9: Número de modelos descritos, según el tipo de técnica utilizada y el año**

Tipo de Técnica	2003	2006	2008	2009	2011	2012	2013	2014	2015	2016	2017	Total
<b>Minería de Datos</b>	<b>1</b>	<b>1</b>	<b>1</b>		<b>5</b>	<b>2</b>	<b>2</b>	<b>1</b>	<b>2</b>	<b>5</b>	<b>3</b>	<b>23</b>
Clasificación	1				1					3		5
Reglas de Asociación		1	1		1		1	1			1	6
Clasificación + Reglas de Asociación + Selección de Características					1							1
Clasificación + Agrupamiento					2				2		1	5
Agrupamiento						2				1		3
Agrupamiento + Reglas de Asociación							1					1
Agrupamiento + Selección de Características										1		1
Reglas de Asociación + Regresión											1	1
<b>Otras</b>						<b>1</b>	<b>1</b>	<b>2</b>		<b>3</b>		<b>7</b>
Analítica de Texto						1		1		1		3
Analítica de Redes Sociales							1			1		2
Inteligencia Colectiva								1				1
Analítica Visual										1		1
<b>Minería de Datos y Otras</b>							<b>1</b>			<b>1</b>		<b>2</b>
Agrupamiento + Reglas de Asociación + Analítica de Redes Sociales							1					1
Agrupamiento + Clasificación + Analítica de Texto										1		1
<b>No Especificada</b>				<b>1</b>	<b>6</b>					<b>1</b>	<b>1</b>	<b>9</b>

Al analizar la evolución que ha tenido la aplicación de técnicas específicas a través de los años se evidencia que Árboles de decisión fue la primera técnica utilizada y, aunque fue aplicada de nuevo en años posteriores, se ha explorado una gran variedad de técnicas de minería de datos y no existe una que haya sido utilizada de manera consistente a través de los años, sino que por el contrario,

se incorporan nuevas técnicas a las investigaciones. Lo mismo ocurre con otras técnicas de analítica, las cuales fueron incorporadas a partir del 2012 y a partir de este año, se han incorporado nuevas técnicas no utilizadas previamente.

#### **4.4 ¿Cuáles son los principales resultados obtenidos en las investigaciones?**

Dado que hay una gran variabilidad en las aplicaciones descritas en los artículos (como se evidenció en la Pregunta #1), y a que también hay una alta variabilidad en las técnicas de analítica o minería de datos que se utilizan para resolverlos (como se evidenció en la Pregunta #2), los resultados obtenidos en cada artículo son específicos para éste y no se pueden generalizar los resultados en términos de las aplicaciones o de las técnicas utilizadas. Sin embargo, es posible presentar algunas conclusiones sobre la utilidad de la analítica o la minería de datos en la Gestión de Recursos Humanos.

Primero, en todos los casos analizados que incluyen el desarrollo de un modelo para resolver el problema que se aborda, los resultados son satisfactorios según los parámetros definidos por los investigadores, y en todos los casos se obtiene información valiosa para soportar la toma de decisiones en la Gestión de Recursos Humanos en cada uno de los campos descritos en la Pregunta número 1 [10] [21] [22] [23] [24] [25] [28] [29] [30] [31] [33] [37] [39] [40] [41] [42] [43] [45] [47] [50] [51] [52]. Los niveles de precisión o confianza de los modelos desarrollados varían y van desde el 60% al 98% (para aquellos artículos en los cuales se especifica). Sin embargo, a excepción de uno de los modelos que generó reglas con un nivel de confianza del 60% [25], todos los demás tienen resultados superiores al 80% en sus niveles de confianza. Se puede concluir entonces que las aplicaciones de analítica y minería de datos en la Gestión de Recursos Humanos permiten tomar mejores decisiones que los métodos tradicionales, los cuales se basan principalmente en el criterio de expertos. Los resultados encontrados en los artículos evidencian que las compañías pueden aumentar su eficiencia en los procesos, al automatizar o facilitar las diferentes etapas que los confirman, así como tomar mejores decisiones, al contar con información relevante que de otra manera no estaría disponible. También se puede concluir que los modelos desarrollados pueden ser aplicados en otros campos o contextos sólo si se cuenta con los datos necesarios para implementar el modelo.

Entre los resultados obtenidos en los artículos se resalta el desarrollo de algunas herramientas de software que pueden ser utilizadas por los profesionales del área de Recursos Humanos para dar solución a los problemas que se les presentan. Estas herramientas tienen las siguientes funcionalidades: extraer información sobre características, educación y experiencia de las hojas

de vida no estructuradas que están almacenadas en las bases de datos y presentarla de forma organizada [27], predecir con precisión el candidato más apropiado para ocupar una vacante, así como identificar los vacíos en competencias de empleados internos [17], identificar las tendencias en los perfiles profesionales que se encuentran en el mercado para elaborar perfiles de cargo más ajustados, lo cual facilita los procesos de reclutamiento y selección, así como el desarrollo de las competencias que se están volviendo relevantes en el mercado [46], automatizar los procesos de pre-selección de candidatos, realizando un ranking de éstos [26], extraer información relevante sobre los mensajes recogidos en encuestas de satisfacción en el trabajo y entrevistas de retiro para identificar los motivos de abandono y poder evitarlos [18], identificar y visualizar redes informales de aprendizaje del día a día en organizaciones y permitirle a los empleados conectarse con otras personas, según sus necesidades e intereses de aprendizaje en el día a día [44], identificar las contribuciones realizadas por los miembros de los equipos en proyectos de desarrollo de Software e identificar patrones de colaboración [49], brindar recomendaciones para la conformación de equipos de trabajo, según el área de especialidad requerida para éstos y otras preferencias del usuario [34] e identificar las reglas para la conformación de equipos de trabajo, específicamente para el desarrollo de Software [35].

Adicionalmente, hay un grupo de artículos que incluyen la comparación entre dos o más modelos y se presenta aquel que genera mejores resultados [10] [21] [24] [29] [33] [39] [40] [42] [43].

En estos casos los principales resultados obtenidos en la comparación de los modelos son:

- El modelo de “Gradient Boosted Machine” produjo los mejores resultados para predecir los empleados con mayor probabilidad de abandonar su empleo, con una precisión general del 81% en sus predicciones, en comparación con otras técnicas como árboles de decisión, regresión logística y k-means [17].
- Los árboles de decisión C4.5 fueron los de mayor precisión comparados con los métodos Random Tree, REP Tree y CART en la generación de reglas de selección para soportar la identificación de candidatos que mejor se ajustan a un perfil, con una precisión del 88,7%. Sin embargo, la precisión se aumentó por encima del 90% al construir un modelo combinado adaptable, que permite incorporar las reglas adaptables generadas con los otros tres algoritmos [24].
- En comparación con siete algoritmos tradicionales KNN (K-Nearest Neighbor), el algoritmo KNN mejorado que se desarrolló tuvo un mejor desempeño, con mayor precisión y menor varianza. Este algoritmo mostró un máximo error de predicción inferior al 2% cuando la distancia ponderada entre el individuo que se predice y la muestra es por debajo de 0.9. Cuando la distancia está entre 0.9 y 2, el error sube a 5,31%, pero cuando ésta sube por encima de 2, el máximo error es inestable y produce resultados inaceptables [33].

- En la identificación de los parámetros más relevantes para predecir el desempeño y que impactan directamente en éste, el algoritmo K-means continuo generó mejores resultados que los algoritmos K-means y “Pairwise agglomerative”, al medir parámetros como tasa de error, entropía y pureza [29].
- Los árboles de decisión C5.0 tuvieron mejor precisión de clasificación que los métodos CHAID y CART, con una tasa de precisión del 89,4% en la identificación de los indicadores críticos para evaluar los entrenamientos [42].
- Los modelos de Redes Neuronales de propagación hacia atrás BPNN (tradicional y adaptado), obtuvieron una mayor precisión comparado con los árboles de decisión y la regresión logística, tanto en la predicción de los comportamientos de aprendizaje de los estudiantes, así como en la predicción del desempeño de éstos. El modelo de BPNN adaptado tuvo una mayor precisión que el tradicional (98,28% vs. 95,15%) en la predicción de los comportamientos de aprendizaje de los estudiantes, y el modelo BPNN tradicional tuvo mayor precisión que el adaptado (95,17% vs. 91,04%) en la predicción del desempeño de los estudiantes [43].
- En la predicción de la rotación de profesionales de tecnología, un modelo híbrido (SOM + BPNN) obtuvo una precisión del 92%, superior a aquella obtenida con los modelos K-Means (63,5%) y redes neuronales BPNN (87,2%) [39].
- Con un coeficiente de confianza del 95%, se concluye que el algoritmo híbrido RST-SVM-DT, que combina algoritmos de RST (Rough Set Theory), SVM (Support Vector Machines) y Árboles de decisión tiene mayor precisión en la extracción de información para la toma de decisiones relacionadas con selección, predicción del desempeño y retención, que cada uno de los algoritmos aplicados de manera independiente o combinados entre sí. Comparado con el algoritmo SVM, la precisión es levemente mayor, aunque no es significativo estadísticamente, por lo cual ambos modelos son comparables [21].
- Para la predicción de la rotación de empleados, se compararon los modelos SVM, Random Forest y Naïve Bayes. Aunque los tres modelos son comparables en su precisión total, el modelo SVM tuvo resultados superiores a los otros dos modelos en el reconocimiento de verdaderos positivos, que son aquellos empleados que han renunciado [40].

Finalmente, hay un grupo de artículos incluidos en el análisis, que corresponde al 14% del total, que hacen mención a otras potenciales aplicaciones de la analítica y la minería de datos en la Gestión de Recursos Humanos, sin embargo, no presentan el desarrollo de un modelo específico ni resultados concretos para soportar las hipótesis que plantean [19] [20] [32] [36] [38]. Por lo tanto, existe oportunidad de realizar investigaciones adicionales para desarrollar los modelos prácticos más adecuados a utilizar en la solución de los problemas identificados en cada caso. También hay oportunidad de realizar investigaciones futuras enfocadas en mejorar los modelos

---

ya desarrollados, especialmente en aquellos casos que presentan limitaciones, ya sea a través de la identificación de una técnica que permita mejorar los niveles de confianza o incorporando variables o información adicional que también genere mejores resultados; o en desarrollar nuevos modelos para resolver problemáticas que aún no han sido abordadas, incluyendo por ejemplo problemas de Remuneración y Beneficios, los cuales no fueron incluidos en las investigaciones analizadas.

## 5. Conclusiones

En relación a cada una de las preguntas de investigación planteadas, se pueden realizar las siguientes conclusiones:

### **P1: ¿Qué aplicaciones se le han dado a la analítica y la minería de datos en la Gestión de Recursos Humanos y en qué áreas o campos?**

Como resultados del análisis sistemático de literatura se evidenció que la analítica y la minería de datos han sido aplicadas en todas las áreas de la Gestión de Recursos Humanos a excepción del área de Remuneración y Beneficios, en donde existe potencial de realizar investigaciones futuras. El área que ha tenido mayor aplicación a través de los años ha sido la de Atracción, Reclutamiento y Selección de empleados, seguido por el área de Evaluación del Desempeño. Las aplicaciones de la analítica y la minería de datos en todas las áreas son muy variadas y responden a diferentes problemáticas que se presentan en cada una.

### **P2: ¿Cuáles son las técnicas o herramientas de analítica y de minería de datos más utilizadas en Recursos Humanos?**

La minería de datos es el área de la analítica que ha tenido mayor aplicación en la Gestión de Recursos Humanos. Se utilizan principalmente modelos de Clasificación, Agrupamiento y Reglas de Asociación, tanto en modelos híbridos como simples. Al analizar las áreas de la Gestión de Recursos Humanos, es posible identificar diferentes técnicas y modelos aplicados en todas las áreas, es decir, hay una gran variedad de técnicas utilizadas en los casos analizados y no existe una tendencia a utilizar cierto tipo de modelo para cada problema. K-means, Redes Neuronales, Árboles de decisión y Algoritmo a priori fueron las que más se aplicaron en forma repetitiva.

### **P3: ¿Cómo ha evolucionado el uso de analítica en Recursos Humanos a través de los años?**

Las áreas de Atracción, Selección e Incorporación, Retención de empleados y Desarrollo y Planes de Sucesión, han sido el foco de estudio desde las investigaciones iniciales y aún continúan vigentes. Con el tiempo han surgido problemas en nuevas áreas que se han buscado resolver a través de la analítica y la minería de datos, incluyendo problemas en el área de Evaluación de

Desempeño, Formación y otros. Ha habido dos momentos importantes desde el 2003 en los cuales se ha visto un interés particular por estudiar las aplicaciones de la analítica y la minería de datos en la Gestión de Recursos Humanos. El primero fue en el 2011, año en el cual se registró un aumento importante en la literatura y en el cual se incorporaron nuevas áreas y nuevas técnicas que no habían sido estudiadas previamente. Posteriormente en el año 2016 se registró de nuevo un incremento en la literatura. Aunque en este año no se incursionó en áreas nuevas, sí se incluyeron nuevas técnicas, especialmente técnicas de analítica diferentes a la minería de datos que no habían sido estudiadas previamente, incluyendo Analítica de Redes Sociales, Analítica de Texto, Analítica Visual e Inteligencia Colectiva. Sin embargo, estas nuevas técnicas no han sido aplicadas posteriormente en otras áreas o campos.

**P4: 4.4 ¿Cuáles son los principales resultados obtenidos en las investigaciones?**

A pesar de que los resultados prácticos de cada una de las investigaciones son diferentes, se puede concluir de manera general que las aplicaciones de analítica y minería de datos en la Gestión de Recursos Humanos permiten aumentar la eficiencia en sus procesos, al automatizar o facilitar las diferentes etapas de éstos, así como tomar mejores decisiones que los métodos tradicionales al contar con información relevante que de otra manera no estaría disponible. También se puede concluir que los modelos desarrollados pueden ser aplicados en otros campos o contextos siempre y cuando se cuente con los datos necesarios para implementar el modelo.



## Bibliografia

- [1] D. Ulrich and J.H. Dulebohn, J. H. "Are we there yet? What's next for HR?", *HR Mgmt. Rev.*, vol. 25, no. 2, pp. 188-204, 2015
- [2] J. H. Dulebohn and R. D. Johnson, "Human resource metrics and decision support: A classification framework", *HR Mgmt. Rev.*, vol. 23, no. 1, pp. 71-83, 2013
- [3] C. Russell and N. Bennett, "Big data and talent management: Using hard data to make the soft stuff easy" *Bus. Hrzn.*, vol 58, no. 3, pp. 237-242, 2015.
- [4] J. H. Marler and J. W. Boudreau, "An evidence-based review of HR Analytics", *The Intl. J. of HR Mgmt.*, vol. 28, no. 1, pp. 3-26, 2017
- [5] D. Angrave, A. Charlwood, I. Kirkpatrick, M. Lawrence and M. Stuart, "HR and analytics: why HR is set to fail the big data challenge", *HR Mgmt. J.*, vol. 26, no. 1, pp. 1-11, 2016
- [6] Katherine Jones, "Conquering HR Analytics: Do You Need a Rocket Scientist or a Crystal Ball?", *Workforce Solutions Review*, July 2014
- [7] Bersin, Josh, Karen O'Leonard, and Wendy Wang-Audia, "High-impact talent analytics: Building a world-class HR measurement and analytics function" *Bersin by Deloitte*, 2013
- [8] Bennett, C., and L. Collins, "HR and People Analytics: Stuck in neutral", *Global Human Capital Trends 2015*, 2015
- [9] Chartered Institute of Personnel and Development. "Talent analytics and big data—The challenge for HR", 2013
- [10] King, K.G, "Data Analytics in Human Resources: A Case Study and Critical Review", *HR. Develop. Rev.*, vol. 15, no. 4, pp. 487-495, December 2016
- [11] Krishnan Sandeep, K., and Manjari Singh, "Strategic Human Resource Management: Three-Stage Process and Influencing Organisational Factors" *Indian Institute of Management Ahmedabad*, Research and Publication Department, 2004

- [12] M. A. Alles. "Dirección estratégica de recursos humanos: gestión por competencias", vol. 1, Ediciones Granica SA, 2006
- [13] Jac Fitz-enz and John Mattox, "Predictive Analytics for Human Resources", Hoboken, NJ: John Wiley, 2014.
- [14] Pruengkarn, Ratchakoon, Kok Wai Wong, and Chun Che Fung, "A review of data mining techniques and applications." *Journal of Advanced Computational Intelligence and Intelligent Informatics*, vol. 21, no. 1, pp. 31-48, January 2017
- [15] Mehmed Kantardzic, "Data-Mining Concepts, "Data Mining: Concepts, Models, Methods, and Algorithms, Second Edition, 2011, pp. 1-25
- [16] B. Kitchenham. "Procedures for performing systematic reviews", Keele, UK, Keele University ST5 5BG, 2004
- [17] M. Bohlouli, N. Mittas, G. Kakarontzas, T. Theodosiou, L. Angelis and M. Fathi, "Competence assessment as an expert system for human resource management: A mathematical approach", *Expert Syst. with Appl.*, vol. 70, pp. 83-102, March 2017
- [18] J.V. Román, J.C.G. Cristóbal and J.A.G. Vázquez, "TALENT + Tecnologías avanzadas para la Gestión del Talento", *Procesamiento de Lenguaje Natural*, vol. 57, pp. 159-162, September 2016
- [19] K.F. Shen, "The analytics of critical talent management", *Ppl. & Strat.*, vol. 34, no. 2 pp. 50-57, 2011
- [20] J.G. Harris, E. Craig and D.A. Light, "Talent and analytics: new approaches, higher ROI", *J. of Bus. Strat.*, vol. 32, no. 6, pp. 4-13, October 2011
- [21] L.F. Chen and C.F. Chien "Manufacturing intelligence for class prediction and rule generation to support human capital decisions for high-tech industries", *Flex. Serv. Manuf. J.*, vol. 23, no. 3, pp. 263-289, September 2011
- [22] H. Min and A. Emam, "Developing the profiles of truck drivers for their successful recruitment and retention: A data mining approach", *Intl. J. of Phys. Distrib. Log. Mgmt.*, vol. 33, no. 2, pp. 149-162, March 2003

- [23] D.A. Pierce, J.E. Johnson, B.D. Krohn and L.W. Judge "Who should we hire?: Examining coaching succession in NCAA Division I women's basketball", *Intl. J. of Sports Sci. & Coach.*, vol. 12, no. 2, pp. 151-161, 2017
- [24] M.A. Shehu and F. Saeed, "An adaptive personnel selection model for recruitment using domain-driven data mining", *J. of Theoret. Appl. Inf. Technol.*, vol. 91, no. 1, pp. 117-130, September 2016
- [25] E. Ferneda, H.A. Do Prado, A.G. Cancian Sobrinho and R. Balaniuk, "Using data mining in the selection process of high performance managers", *Intl. J. of Reas.-based Intel. Syst.*, vol. 6, no. 1-2, pp. 40-48, January 2014
- [26] E. Faliagka, A. Tsakalidis and G. Tzimas, "An integrated e-recruitment system for automated personality mining and applicant ranking", *Internet Res.*, vol. 22, no. 5, pp. 551-568, October 2012
- [27] D. Çelik, "Towards a semantic-based information extraction system for matching résumés to job openings", *Turk. J. of Electrical Eng. Comput. Sci.*, vol. 24, no. 1, pp. 141-159, 2016
- [28] C. Chien and L.F. Chen, "Data mining to improve personnel selection and enhance human capital: A case study in high-technology industry", *Expert Syst. with Appl.*, vol. 34, no. 1, pp. 280-290, January 2008
- [29] P. Bafna, S. Pillai and D. Pramod, "Quantifying performance appraisal parameters: A forward feature selection approach", *Indian J. of Sci. Technol.*, vol. 9, no. 21, 2016
- [30] A. Aktepe and S. Ersoz, "A quantitative performance evaluation model based on a job satisfaction-performance matrix and application in a manufacturing company", *Intl. J. of Ind. Eng.: Theory Appl. Pract.*, vol. 19, no. 6, pp. 264-277, 2012
- [31] Z. Danping and D. Jin, "The data mining of the human resources data warehouse in university based on association rule", *J. of Comput.*, vol. 6, no. 1, pp. 139-146, 2011
- [32] A. Sharma and T. Sharma, "HR analytics and performance appraisal system: A conceptual framework for employee performance improvement", *Mgmt. Res. Rev.*, vol. 40, no. 6, pp. 684-697, 2017

- [33] N. Li, H. Kong, Y. Ma, G. Gong and W. Huai, "Human performance modeling for manufacturing based on an improved KNN algorithm", *Intl. J. of Adv. Manuf. Technol.*, vol. 84, no. 1-4, pp. 473-483, April 2016
- [34] A. Datta, J.T.T. Yong and S. Braghin, "The zen of multidisciplinary team recommendation" *J. of the Ass. for Inf. Sci. Technol.*, vol. 65, no. 12, pp. 2518-2533, December 2014
- [35] M. André, M.G. Baldoquín and S.T. Acuña, "Formal model for assigning human resources to teams in software projects", *Inf. Softw. Technol.*, vol. 53, no. 3, pp. 259-275, March 2011
- [36] H. Jantan, A.R. Hamdan and Z.A. Othman, "Knowledge discovery techniques for talent forecasting in human resource application" *World Acad. of Sci., Eng. Technol.*, vol. 38, pp. 803-811, February 2009
- [37] M.J. Huang, Y.L. Tsou and S.C. Lee, "Integrating fuzzy data mining and fuzzy artificial neural networks for discovering implicit knowledge", *Know. Based Syst.*, vol. 19, no. 6, pp. 396-403, October 2006
- [38] L. Bassi and D. McMurrer, "Four Lessons Learned in How to Use Human Resource Analytics to Improve the Effectiveness of Leadership Development", *J. of Leaders. Stud.*, vol. 10, no. 2, pp. 39-43, 2016
- [39] C.Y. Fan, P.S. Fan, T.Y. Chan and S.H. Chang, "Using hybrid data mining and machine learning clustering analysis to predict the turnover rate for technology professionals", *Expert Syst. with Appl.*, vol. 39, no. 10, pp. 8844-8851, August 2012
- [40] V.V. Saradhi and G.K. Palshikar, "Employee churn prediction", *Expert Syst. with Appl.*, vol. 38, no. 3, pp. 1999-2006, March 2011
- [41] A.S. Sabitha, D. Mehrotra and A. Bansal, "Delivery of learning knowledge objects using fuzzy clustering", *Edu. Inf. Technol.*, vol. 21, no. 5, pp. 1329-1349, September 2016
- [42] Y. Jou, Y. Wu and W. Lin, "Applying decision tree and neural network to raise the performance of human training quality", *J. of Qlty.*, vol. 22, no. 5, pp. 383-403, December 2015
- [43] J. Wang, Y.I. Lin, Y. S.Y. Hou, "A data mining approach for training evaluation in simulation-based training", *Comput. Ind. Eng.*, vol. 80, no. 1, pp. 171-180, 2015

- [44] M. de Laat and B. Schreurs, "Visualizing Informal Professional Development Networks: Building a Case for Learning Analytics in the Workplace", *Americ. Behav. Sci.*, vol. 57, no. 10, pp. 1421-1438, October 2013
- [45] W.T. Lin, S.J. Wang, Y.C. Wu, and T.C. Ye, "An empirical analysis on auto corporation training program planning by data mining techniques", *Expert Syst. with Appl.*, vol. 38, no. 5, pp. 5841-5850, May 2011
- [46] N. Malzahn, S. Ziebarth and H.U. Hoppe, "Semi-automatic creation and exploitation of competence ontologies for trend aware profiling, matching and planning", *Know. Mgmt. E-Learn.*, vol. 5, no. 1, pp. 84-103, 2013
- [47] S. Fan, X. Li and J. L. Zhao, "Collaboration process pattern approach to improving teamwork performance: A data mining-based methodology", *J. on Comput.*, vol. 29, no. 3, pp. 438-456, 2017
- [48] Park, M., Ahn, H. and Kim, K.P., "Workflow-supported social networks: Discovery, analyses, and system", *J. of Netw. Comput. Appl.*, vol. 75, pp. 355-373, November 2016
- [49] A. González-Torres, F.J. García-Peñalvo, R. Therón-Sánchez and R. Colomo-Palacios, "Knowledge discovery in software teams by means of evolutionary visual software analytics", *Sci. of Comput. Program.*, vol. 121, pp. 55-74, June 2016
- [50] S. Kim and D.A. Nembhard, "Rule mining for scheduling cross training with a heterogeneous workforce", *Intl. J. of Product. Res.*, vol. 51, no. 8, pp. 2281-2300, 2013
- [51] P. Marksberry, J. Church, and M. Schmidt, "The employee suggestion system: A new approach using latent semantic analysis", *Hum. Fact. Ergon. In Manuf.*, vol. 24, no. 1, pp. 29-39, January 2014
- [52] L. He and J. Qi, "Enterprise human resources information mining based on improved apriori algorithm", *JNW*, vol. 8, no. 5, pp. 1138-1145, 2013