



**OSEA: Optimal system for employee analytics.**

Elaborado por:

Mariana Ruiz Rubiano

Fabian Andrés Suarez Molano

Jorge Mario Navas Mazuera

Sergio Esteban Molano Rojas

Universidad EAN

Escuela de Formación en Investigación

Proyecto de Investigación de Pregrado

Bogotá

02/04/2023

## Tabla de contenido

### Contenido

1.	Resumen Ejecutivo.....	3
2.	Introducción .....	4
3.	Objetivos.....	5
4.	Definición del problema .....	6
5.	Justificación .....	7
6.	Análisis de requerimientos .....	8
7.	Marco de Referencias .....	10
8.	Análisis de restricciones .....	19
9.	Metodología para la selección y desarrollo de la solución.....	21
10.	Análisis de costos.....	34
11.	Conclusiones .....	36
12.	Referencias.....	38

## 1. Resumen Ejecutivo

Una problemática que las empresas vienen presentando hoy en día y que es muy significativa internamente como compañía en desarrollo es la fuga de talentos. La retención de estos talentos es de igual importancia que la selección de personal para crear y mantener un equipo enfocado y que siga aportando en los objetivos de la empresa. Así como se han desarrollado estrategias empresariales enfocadas en esta selección de personal hoy en día se crearon estrategias para investigar y aplicar metodologías con el fin de predecir la fuga de talento.

El siguiente proyecto tiene como finalidad investigar y descubrir las metodologías que sean más eficientes y seguras para de este modo procesar la información obtenidas con estas en un algoritmo que permita pronosticar y preparar a las empresas de la fuga de talentos.

Lograr generar este algoritmo permitiría a las empresas retener, adaptarse y anticiparse a las vacantes libres que pueden presentarse cuando un personal altamente capacitado desea irse.

## 2. Introducción

Desde el inicio de nuestra vida académica y universitaria empezamos a desarrollar, conocer, practicar y mejorar nuestra habilidades y destrezas en línea con la carrera que hemos decidido estudiar, posteriormente utilizamos estas habilidades y destrezas para aportar valor como profesionales, técnicos o tecnólogos en una compañía.

Todas las personas tenemos habilidades, cualidades y capacidades diferentes, y es por esto que las empresas destinan gran parte de su capital y enfoquen su atención en los procesos de selección de personal a fines de designar en cada área a profesionales competentes que se acomoden a las empresas, los valores corporativos y los propósitos empresariales de manera óptima y confiable. Una vez se recluta a la persona o personas idóneas para cada área las empresas inician un proceso de capacitación, ambientación y proyección de estos nuevos perfiles en donde una vez más se invierten recursos económicos y de talento humano para cumplir con los objetivos dentro del proceso de vinculación de personal.

Sin embargo, estas inversiones pueden perderse debido a la rotación de personal y fuga de talentos. Con respecto a Colombia el 97% de las compañías consideran importante ofrecer beneficios adicionales a sus empleados como auxilios de estudio, convenios académicos, de bienestar y entretenimiento, seguros de vida, entre otros. Pero aun con estos beneficios los niveles de rotación han llegado al 74% predominando el personal entre los 20 años y los 29 años.

Hoy en día gracias a la tecnología y todas sus áreas como la revolución 4.0, la big data, la IA y el análisis de datos, podemos crear herramientas que mediante su implementación nos permitan mitigar los fenómenos de rotación y fuga de talentos. Y es este nuestro planteamiento de investigación en el desarrollo de este proyecto.

### 3. Objetivos

#### **Objetivo General**

Proponer, diseñar e implementar un algoritmo predictivo para disminuir y evitar la rotación y fuga de talentos de las empresas que contratan los servicios de GSH, a partir de la analítica de los datos otorgados por los empleados tomados de las encuestas proporcionadas por GSH.

#### **Objetivos Específicos**

Investigar diferentes metodologías para recopilación de datos para empresas que realicen encuestas de satisfacción con sus colaboradores.

Investigar sistemas de almacenamiento seguros y cómodos para procesar la información recolectada y que pueda ser aplicada en el algoritmo.

Presentar una estructura de uso de la metodología y almacenamiento de datos para ser aplicada en el algoritmo.

Procesar los datos obtenidos y almacenados de los empleados para la gestión de talento.

Realizar pruebas y validaciones del algoritmo OSEA.

#### 4. Definición del problema

Una de las problemáticas más recurrentes en el departamento de recursos humanos en una empresa, son los fenómenos de rotación y fuga de talentos, sin importar al sector al que pertenece la empresa. Este fenómeno causa graves consecuencias como pérdidas económicas, baja productividad e incumplimiento de los objetivos empresariales. Sin embargo, estos fenómenos pueden predecirse y evitarse, para esto la empresa debe conocer las causas que lo producen tales como falta de liderazgo, mal clima laboral, falta de motivación en los empleados, pérdida de los sentimientos de pertenencia, impasibilidad de progreso e incluso mala comunicación en el interior de la empresa.

Cuando hablamos de la rotación de personal y fuga de talentos en una empresa nos referimos a dos fenómenos que se producen cuando profesionales cualificados abandonan la empresa en la que trabajan para iniciar una nueva vida laboral en otra compañía. Muchas veces esas empresas pertenecen a la competencia y se encuentran en la misma ciudad que la empresa de la cual el profesional se ha ido. Pero ¿Por qué se produce esta situación? ¿Que motiva a un profesional a salir de una empresa? ¿En qué momento el profesional toma la decisión de salir de la empresa? Si en la empresa no existe un clima agradable, la oportunidad de seguir desarrollándose a nivel personal y profesional, no confía el talento en la compañía, no es casualidad que alguno de los asociados decida marcharse del equipo y dejar de trabajar en esa compañía.

Para las empresas y con el fin de mitigar estos fenómenos es muy importante conocer el talento de su personal al momento de atraerlo, vincularlo y por supuesto retenerlo. El objetivo principal es conseguir que los profesionales se sientan como un miembro valioso para la compañía y que se pueda prevenir de manera

oportuna el abandono de estos talentos porque siempre este fenómeno produce consecuencias negativas realmente graves para las compañías.

## **5. Justificación**

Antes de hablar de un diseño de un algoritmo confiable para evitar y predecir el fenómeno de rotación y fuga de personal es importante hablar de que es este fenómeno de fuga y rotación de personal y él porque es tan dañino para las organizaciones.

Cuando se habla de fuga o rotación de talento se hace referencia al fenómeno que se genera cuando personas cualificadas abandonan la empresa en la que trabajan, para iniciar una nueva vida laboral en otra empresa. En ocasiones esta se debe a múltiples factores como oportunidades en empresas extranjeras, mejores oportunidades o una vida laboral más agradable y de mejor calidad.

En muchas ocasiones esta fuga de talento lo que genera es una amenaza al desarrollo de la compañía dado que en ciertas ocasiones esta puede llevar a pérdidas económicas de hasta un 20% del salario anual del trabajador fugado, según estudios del Center for American Progress, para darnos una idea local en Colombia se encontró que cerca del 22% de las empresas requieren entre 4 y 6 meses para cubrir vacantes claves y cerca del 12% pueden tardar más de 6 meses en ocuparse y en adición, el 57.8% de las empresas del país tienen problemas para llenar estas vacantes que se generan por fuga de talentos y para la Mintic en el 2025 Colombia presentaría un déficit de profesionales. También un estudio realizado por Acrip revelo que el 44% de las empresas consultadas había tenido un 10% de su personal operativo en el año anterior y que para el mes de diciembre la rotación de personal es la mayor con un 24.1%. Y aunque son muchos los factores que influyen a que un colaborador decida dejar su puesto en la organización es notorio que esto acompañado de la incertidumbre que genero la pandemia del COVID-19 en los años anteriores genero más

factores de los que se evidencian comúnmente dado que después de esta los empleados dejaron de preocuparse por el factor económico y se centró en el factor de comodidad y bienestar para el mismo.

Dado un pequeño contexto también es necesario saber que ciertas de las circunstancias por las que el personal decide fugarse son muy claras en algunas ocasiones como las siguientes: una cultura organizacional muy agresiva, un mal clima laboral, poco engagement entre los equipos o las áreas, mala gestión del equipo de talento humano (que en ciertas ocasiones llega a ser menos humano de lo que se espera), malas relaciones laborales entre los compañeros y el jefe directo o las pocas oportunidades de crecimiento y el poco bienestar para el empleado.

Teniendo como contexto ciertas causas del porqué de la fuga de personal en las compañías lo que se quiere exponer en este proyecto un algoritmo el cual apoyado desde datos y análisis enviados por la empresa se pueda generar ciertas variables que logren predecir y mitigar la fuga del talento humano y la rotación de personal en la empresa GSH y así ellos poder entender como estas variables puede ingresarse en un algoritmo confiable para mitigar estas pérdidas que se generan en los diferentes aliados y así maximizar los resultados de las mismas.

## **6. Análisis de requerimientos**

Para diseñar un algoritmo confiable para que organizaciones puedan describir, explicar y predecir el fenómeno de rotación y fuga de talentos es necesario de manera amplia y general identificar, recopilar, almacenar, comprender y clasificar la información pertinente a las variables que pueden influir en la rotación o intención de renuncia de los colaboradores de las empresas.

Es claro que el área de recursos humanos no genera siempre variables cuantitativas y los responsables de gestión humana generan un formulario para recopilar la información pertinente a las clases a evaluar o relacionar a los fenómenos de interés. Después de la recolección la información se debe preparar ya que la data en crudo puede traer errores e información que no puede ser aplicada. En este caso la data se pasa a hojas de Excel para revisar y modificar los tipos de datos recolectados para darles el formato y la clasificación necesarias.

Esta información es importante poder almacenarla, consultarla y mantener su integridad, por lo que se sugiere usar una DB base de datos no relacional NoSQL debido a que la información cruda puede variar a futuro su estructura y con el fin de hacerla escalable sin incurrir en demoras o caída del servicio por actualización de la estructura. Una estructura de datos Columnar (Column-Family Store) se acomoda al formato Big Data, es rápida para consulta y fácilmente escalable.

Una vez asegurada la información en la DB es importante seleccionar las herramientas necesarias para escoger, probar y medir un algoritmo confiable y eficaz que nos permita predecir y explicar el fenómeno de rotación y fuga de talentos dentro de las empresas. Posibles herramientas pueden ser Tensor Flow de Google, PyTorch o WEKA (Weikato for knowledge Analysis) entre otras. Recomendable WEKA debido a que su set de herramientas incluye preparación de datos, algoritmos de clasificación, regresión y agrupación, así como herramientas de visualización y asociación de reglas.

Con las herramientas definidas es importante pasar a seleccionar los tipos de algoritmo que pueden implantarse para la predicción del fenómeno de rotación y fuga de talentos en las organizaciones, existen varios tipos de algoritmos pero los algoritmos de clasificación como Regresión logística, Naive Bayes, k-

Nearest Neighbors, Decision Tree y Support Vector Machines, pueden ser los más adecuados para analizar el sentimiento o las ganas de cambiar de trabajo de los colaboradores de las empresas.

El paso siguiente sería seleccionar algunas de estas técnicas, alimentarlas con la data suministrada y correr varias iteraciones llevando el registro de cada experimento para medir la precisión, exhaustividad, sensibilidad y puntuación F1. Al evaluar estas métricas es posible definir cuál de las técnicas usadas puede ser la más confiable y eficaz para predecir el fenómeno de rotación y fuga de talento en las empresas.

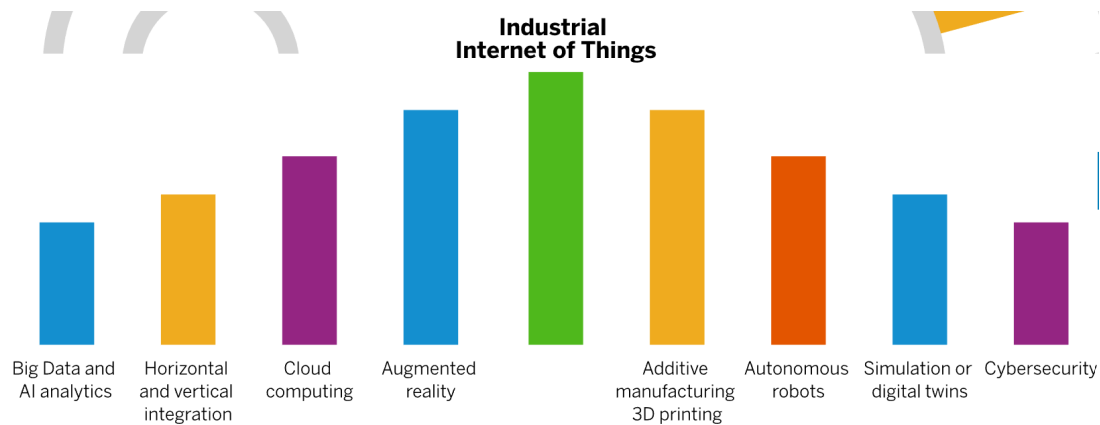
## **7. Marco de Referencias**

### **Industria 4.0**

Desde que ocurrió la primera revolución industrial desde el siglo 18, esta se centró en las necesidades de producción y cadenas de suministro. Con el paso del tiempo ha presentado constantemente diferentes cambios a raíz de las nuevas tecnologías que se han venido desarrollando con el fin de ser más eficientes, productivos y predictivos en los procesos que se manejan dentro de una empresa. La llegada del internet a finales de los 70s transformo por completo el modelo de trabajo de las diferentes compañías dando como inicio a desarrollos en tecnología sensitiva, diseño en sistemas de ingeniería, la promesa de la inteligencia artificial y a lo que se llamó como Industria 4.0. Bopaya (2023) describe la premisa fundamental de la I4.0 que la data y la información son parte integral de la materia y el flujo de procesos, en toda la cadena de valor.

Las tecnologías de la I4.0 se basan en nueve pilares tecnológicos.

**Imagen 1. Pilares de la I4.0**



Tomado de SAP <https://www.sap.com/latinamerica/insights/what-is-industry-4-0.html>

### **Ingeniería de datos**

El big data no ha sido definido totalmente debido a la amplia frontera que tiene dentro de las ciencias de la computación. López y Zarza (2017) definen el big data como la aplicación en un enfoque científico-práctico para resolver problemas de datos que necesitan ser atacados eficientemente. En otras palabras, el big data es el almacenamiento de un gran volumen de datos y de mayor complejidad que contienen mayor variedad, presentan volúmenes crecientes y a una velocidad superior.

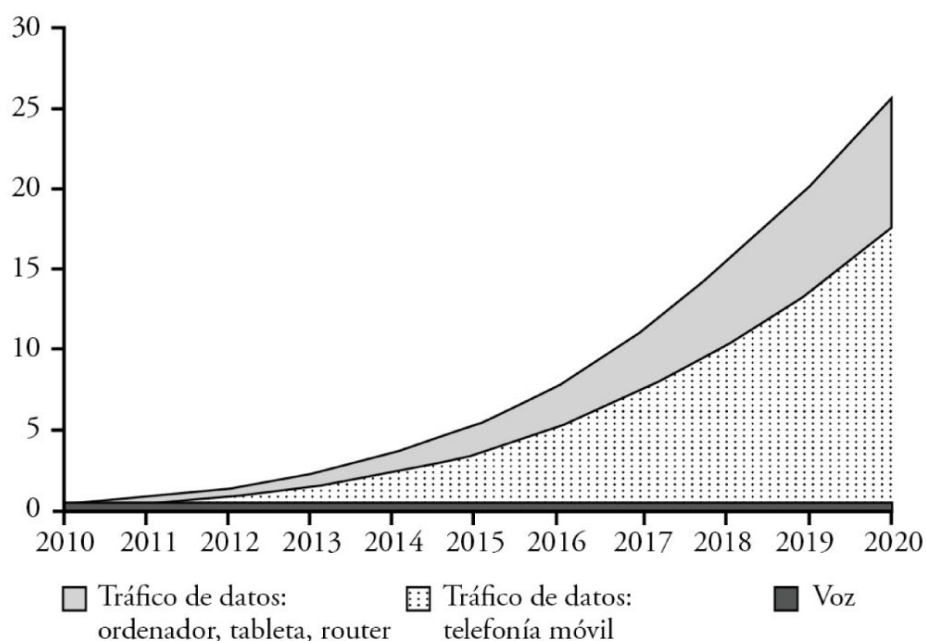
Estos atributos se pueden definir de la siguiente manera:

**Volumen:** Como su nombre lo indica es la gran cantidad de datos que se deben almacenar para tomar en cuenta en las aplicaciones a las que le vamos a dar uso y para aquellos problemas que puedan ser abordados de forma competente y resueltos de forma correcta.

**Velocidad:** Con la necesidad de tener respuestas y resultados en ventanas de corto tiempo el big data pretende manejar toda la información almacenada de forma eficiente. En este manejo de información se

derivan dos sistemas, el buffering el cual almacena primero los datos para luego ser procesados y la segunda que mantiene un flujo de procesamiento constante a medida que se almacenan los datos.

Variedad: Tradicionalmente el manejo de información y recopilación de datos se daba de manera estructurada. El big data refleja una diversidad en el formato y en el origen de los datos. Actualmente, el 90% de los datos no son estructurados lo que fundamenta la necesidad del big data.



### Software de recopilación

El software de recopilación de datos es un sistema computarizado que se utiliza para la recopilación y el almacenamiento de datos cualitativos y cuantitativos en forma digital. Algunos de los beneficios de su implementación es que eliminan el uso de las encuestas de papel que tienden a ser complejas en su proceso de aplicación, análisis de datos e informes. Estas herramientas digitales permiten rápidamente exportar la información de manera que obtener resultados es mucho más simple y además versátil para exponer los

análisis. Actualmente existen en el mercado múltiples herramientas para la recopilación y gestión de datos, pero hay una en particular que ha llamado nuestra atención llamada MAGPI.

MAGPI es una aplicación de registro de datos móviles que permite a cada uno de los usuarios de acuerdo con sus necesidades diseñar y crear formularios digitales de manera muy rápida e incluso con una conexión inestable de internet. MAGPI tiene varias funciones tales como encuestas móviles, estampado de GPS, fotos y actualizaciones automáticas. Esta herramienta también permite alimentar los datos del usuario casi en cualquier sistema desde la web.

### **Síndrome de Burnout**

El síndrome de burnout, también conocido como agotamiento laboral o síndrome del trabajador quemado, es un estado de estrés crónico y agotamiento emocional que resulta de la exposición prolongada a situaciones laborales estresantes. Se caracteriza por una combinación de síntomas físicos, emocionales y cognitivos que afectan negativamente el desempeño laboral y la calidad de vida de una persona.

El síndrome de burnout se manifiesta mediante un agotamiento emocional, donde la persona experimenta una sensación constante de falta de energía, fatiga y desgaste emocional. También se observa una actitud negativa y distante hacia el trabajo, los compañeros y la organización, lo que se conoce como despersonalización. Además, se puede presentar una disminución de la realización personal, donde la persona experimenta una sensación de ineficacia y baja autoestima en relación con su trabajo.

Este síndrome suele estar relacionado con profesiones que implican un alto nivel de demandas emocionales, como los servicios de atención médica, la educación, los servicios sociales y otras ocupaciones en las que el contacto humano y la responsabilidad emocional son frecuentes. Sin embargo, también puede afectar a personas en otros campos laborales.

El síndrome de burnout no solo afecta al individuo, sino que también tiene consecuencias negativas para las organizaciones, como el aumento del ausentismo, la disminución de la productividad y el deterioro del clima laboral. Es fundamental tomar medidas preventivas y de intervención para abordar el síndrome de burnout, como promover un equilibrio saludable entre el trabajo y la vida personal, fomentar el apoyo social y brindar recursos y herramientas para manejar el estrés laboral.

### **La encuesta**

Una de las herramientas más utilizadas al momento de recolectar información como procedimiento de investigación es la encuesta, esta nos permite elaborar y obtener datos de modo rápido y eficaz. La encuesta permite aplicaciones masivas en cuestión de segundos usando técnicas de muestreo adaptadas a las necesidades del encuestador y los propósitos de esta. El interés particular del investigador no es un sujeto en específico, sino la población o grupo al que pertenece y es por esto que es muy importante aplicar una técnica de muestreo apropiada.

Otra gran cualidad de las encuestas es que permiten recoger de modo estandarizado la información lo que facilita realizar comparaciones intergrupales ya que las instrucciones para todas las personas a quien se les aplicara la encuesta y las preguntas son las mismas.

Las encuestas pueden ser una herramienta valiosa en la gestión de la fuga de personal. Al realizar encuestas, se pueden obtener datos y opiniones de los empleados que se van y de los que se quedan, lo que puede ayudar a identificar las causas subyacentes de la fuga de personal y a desarrollar estrategias para retener a los empleados.

Las encuestas pueden abordar una variedad de temas relacionados con la fuga de personal, como la satisfacción laboral, la compensación y los beneficios, la cultura organizacional, el liderazgo y la oportunidad de crecimiento y desarrollo profesional. Las preguntas pueden ser abiertas o cerradas, y se pueden utilizar escalas de Likert para medir el grado de acuerdo o desacuerdo de los empleados con ciertas afirmaciones.

Es importante asegurarse de que las encuestas sean anónimas y confidenciales para que los empleados se sientan cómodos compartiendo sus opiniones y preocupaciones. Además, es crucial tomar medidas concretas basadas en los resultados de la encuesta para mejorar las áreas problemáticas identificadas por los empleados.

En resumen, las encuestas pueden ser una herramienta efectiva para ayudar a las organizaciones a entender las causas de la fuga de personal y a desarrollar estrategias para retener a los empleados valiosos.

### **Cuestionario de Maslach Burnout Inventory**

El cuestionario de Maslach de 1986 es el instrumento más utilizado en diferentes organizaciones alrededor del mundo. Esta escala tiene una fiabilidad del 90%. Para medir el desgaste emocional de las personas en el trabajo, este instrumento cuenta con 22 ítems en forma de afirmaciones donde enfatiza tres elementos importantes: la fatiga o cansancio emocional, la despersonalización en el trabajo, la falta de eficiencia o decadencia en el rendimiento.

La fatiga o agotamiento emocional: Esta variable se refiere a la sensación de agotamiento físico y emocional que experimenta un empleado debido a su trabajo. Se puede medir mediante preguntas relacionadas con el nivel de energía, cansancio y desgaste emocional que siente el empleado

Despersonalización: Se trata de la falta de conexión emocional y la actitud distante o negativa hacia los demás, como clientes, compañeros de trabajo o la organización en general. Puede medirse mediante preguntas que evalúen la percepción del empleado sobre su relación con los demás.

Realización personal, decadencia en el rendimiento: Esta variable se refiere al sentido de logros y competencias que un empleado siente en su trabajo. Puede medirse mediante preguntas relacionadas con el nivel de satisfacción y la sensación de crecimiento personal y profesional

### **DDG: Recopilación de datos digital**

Es un proceso de recopilación de datos electrónicamente mediante el uso de tecnología existente. Las principales razones por las que una organización recurre a la DDG es por motivo de calidad de datos. La recopilación de datos digitales garantiza que se apliquen estándares de recopilación estrictos, lo cual es una característica muy importante para controlar la calidad de los datos.

What is Digital Data Gathering? Learn how it can help you. (s/f). Translate.Goog. Recuperado el 2 de abril de 2023, de [https://www-adssomalia-com.translate.goog/what-is-digital-data-gathering/?\\_x\\_tr\\_sch=http&\\_x\\_tr\\_sl=en&\\_x\\_tr\\_tl=es&\\_x\\_tr\\_hl=es&\\_x\\_tr\\_pto=rq](https://www-adssomalia-com.translate.goog/what-is-digital-data-gathering/?_x_tr_sch=http&_x_tr_sl=en&_x_tr_tl=es&_x_tr_hl=es&_x_tr_pto=rq)

### **Banco de información**

Un banco de información es un sistema informático donde están almacenados cierta cantidad de datos, estos datos se encuentran organizados de manera específica de acuerdo con los requerimientos establecidos con el fin de hacer más fácil y rápido el acceso a esta información. A lo largo de los años siempre se ha requerido de un banco de datos o un banco de información para poder almacenar y manipular información

importante y de interés para las compañías, es por esto que existen las bibliotecas, archiveros, hemerotecas, entre otras.

La tecnología ha traído un gran avance en esta área, ya que nos ha permitido reemplazar el papel dándonos la facilidad de tener a nuestro alcance de manera oportuna y confiable cualquier tipo de información. Existen varios tipos de bancos de información:

- **Relacionales:** Son lo más comunes, hablamos de datos estructurados cuando son bases de datos muy amplias que deben ser organizadas rigurosamente. Decimos relacional porque tiene que ver con el hecho de que se utiliza un sistema que localiza los datos relacionados a la búsqueda ya que la información se encuentra estructurada y planificada.
- **No relacionales:** Contrario a la primera este modelo de bases de datos no exige una formalidad o información estructurada.
- **La nube:** Es el modelo de bases de datos de moda, que consiste en un sistema de bases de datos almacenado en la nube. Este modelo permite almacenar información sin ocupar necesariamente espacios dispositivos tecnológicos. De esta manera es posible alojar información de manera muy segura y con mucha comodidad.
- **Jerárquicas:** Es un banco de datos que organiza aplicaciones de alto rendimiento según la jerarquía.

### **Data Mining:**

El proceso de clasificar gran cantidad de datos para identificar patrones y relaciones que pueden ayudar a resolver problemas mediante el análisis de datos.

### **Decision Tree:**

Los algoritmos de árboles de decisión son útiles para problemas de clasificación y regresión, y pueden manejar tanto datos numéricos como categóricos. Algunas de las ventajas de los árboles de decisión

incluyen que son fáciles de entender e interpretar, y pueden ser visualizados gráficamente. También pueden manejar datos faltantes y ser utilizados para la selección de características

### **Support Vector Machines:**

Al igual que Decision Tree son un tipo de algoritmo de aprendizaje automático supervisado utilizado para clasificación y regresión. Funcionan encontrando una línea que pueda separar mejor los datos en diferentes clases, mientras se maximiza el margen entre la línea y los puntos de datos más cercanos de cada clase. Las SVM son efectivas en espacios de alta, pero pueden ser intensivas computacionalmente y sensibles a la elección de los hiperparámetros.

### **Machine Learning:**

Es una rama de la inteligencia artificial que se enfoca en el desarrollo de algoritmos y modelos estadísticos que permiten a una computadora aprender a realizar una tarea específica sin ser programada explícitamente para hacerlo. En lugar de seguir un conjunto de instrucciones fijas, un modelo de aprendizaje automático es capaz de mejorar su capacidad para realizar una tarea en función de la experiencia previa y de la retroalimentación proporcionada por los datos.

### **Tensor Flow**

Es una biblioteca de código abierto para la programación de redes neuronales y el aprendizaje profundo (deep learning). Fue desarrollada por el equipo de investigación de inteligencia artificial de Google y se utiliza ampliamente en aplicaciones de aprendizaje automático y análisis de datos.

## **Weikato for knowledge Analysis WEKA**

Es una herramienta de software libre para data Mining y Machine Learning. Proporciona una interfaz gráfica de usuario, y entre sus herramientas permite a los usuarios trabajar con conjuntos de datos, preprocesar los datos, aplicar técnicas de aprendizaje automático y visualizar los resultados. Cuenta con varios algoritmos, incluyendo árboles de decisión, redes neuronales, regresión logística, clasificación de vectores de soporte y agrupamiento, entre otros. Además, incluye herramientas para la evaluación de modelos, la selección de características y la optimización de parámetros.

## **8. Análisis de restricciones**

### **Restricciones ambientales.**

El uso de tecnologías para realizar algoritmos de predicción puede tener varios impactos ambientales, dependiendo de los recursos tecnológicos utilizados, los patrones de consumo de energía y la gestión de residuos electrónicos.

Consumo de energía, el procesamiento y almacenamiento de grandes conjuntos de datos requeridos para algoritmos de predicción y aprendizaje automático requiere una gran cantidad de energía. En los centros de datos, la energía utilizada para la refrigeración es también un factor importante. El exceso de energía utilizada contribuye a la emisión de gases de efecto invernadero.

Una investigación realizada por el laboratorio AI Now de la Universidad de Nueva York, en colaboración con otras organizaciones, aborda el impacto ambiental de la IA. Los investigadores destacan

que los centros de datos consumen grandes cantidades de energía y generan una cantidad significativa de emisiones de gases de efecto invernadero (GEI) y calor residual.

Otra investigación realizada por Code Green Networks y publicada en el Informe de Impacto Ambiental de Greenpeace, sostiene que la IA y la nube se están convirtiendo en un importante contribuyente a las emisiones de GEI, y que es fundamental la adopción de prácticas de energía renovable y eficiente para reducir su impacto.

Un estudio de la consultora PwC analiza el impacto de las tecnologías digitales, incluyendo la IA, en la reducción de las emisiones de GEI en diferentes sectores, destacando que, si se aplican de manera efectiva, estas tecnologías pueden ayudar a reducir las emisiones en un 15% para 2030.

La adopción de tecnologías para la realización de algoritmos de predicción es una parte importante del desarrollo tecnológico y, por lo tanto, tiene un impacto ambiental negativo. Sin embargo, la inversión en tecnologías limpias y la gestión adecuada de residuos electrónicos pueden ayudar a reducir estos impactos. Hay diversas investigaciones que abordan el impacto ambiental de la IA y que muestran la necesidad de adoptar prácticas de energía renovable y eficiente para reducir sus impactos.

### **Restricciones legales.**

En Colombia, la normativa de referencia en cuanto a la protección de datos personales es la Ley 1581 de 2012, la cual establece las disposiciones generales para la protección de datos personales y el derecho de habeas data. Además, la Resolución 01283 de 2014 adopta una metodología para la evaluación y gestión de riesgos asociados al tratamiento de datos personales.

En cuanto a las encuestas de satisfacción laboral, se deben tener en cuenta las disposiciones legales relacionadas con la protección de datos personales y la privacidad de los trabajadores. Es necesario obtener el consentimiento informado de los participantes y asegurarse de que los datos recopilados se utilicen únicamente para los fines previstos en la encuesta y que se mantengan en confidencialidad.

Es importante tener en cuenta que, en caso de que los resultados de la encuesta contengan información personal sensible, se deben aplicar medidas de seguridad adicionales para proteger dicha información.

Las normas colombianas para la seguridad de la información y la protección de datos personales son aplicables a las encuestas de satisfacción laboral en las empresas, y es importante garantizar el cumplimiento de las disposiciones legales para evitar sanciones y proteger la privacidad de los trabajadores.

## **9. Metodología para la selección y desarrollo de la solución**

Para la selección de variables y su grado de importancia que vamos a ingresar en el algoritmo se tomo como fuente el cuestionario de Maslach; encuesta realizada en la plataforma de LinkedIn; estudios realizados por Mckinsey & Company y datos revelados por Hays y KPMG, empresas consultoras de reclutamiento y selección.

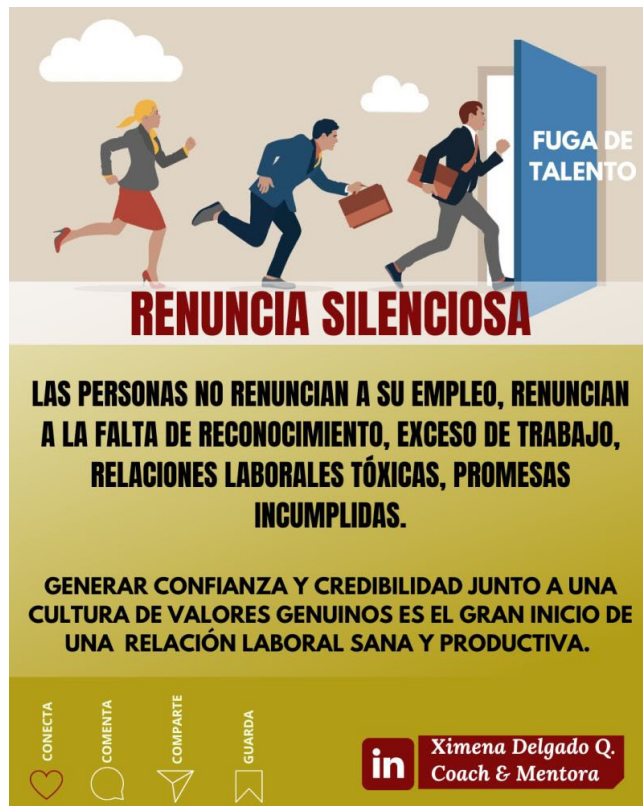
Los estudios de Mckinsey & Company indican que en Colombia las dos principales razones por las que hay deserción laboral se deben a falta de desarrollo profesional de carrera y compensación. Estos resultados son muy similares a estudios que se hicieron en otros países y se tienen en cuenta otras razones como flexibilidad en el trabajo, lideres indiferentes y rendimiento laboral insostenible.

Las encuestas realizadas por LinkedIn y las causas indicadas por Hays y KPMG coinciden que los factores mas influyentes para que una persona deje la empresa son insatisfacción con la cultura organizacional, falta de retos, pocas compensaciones por el trabajo realizado, y la falta de reconocimiento en las metas logradas. El 45% de las personas que renuncian lo hacen porque no ven oportunidades de crecimiento.

En la siguiente tabla enumeramos las variables con mayor importancia que serán ingresados en el algoritmo para realizar una simulación y encontrar una tendencia en la renuncia con base a los datos ofrecidos de las encuestas por parte de la empresa GSH.

<b>VARIABLE</b>
Satisfacción laboral
Posibilidad de crecimiento
Reconocimiento de trabajo
Agotamiento emocional
Factores sociodemográficos

Imagen 3. Renuncia silenciosa



Nota. Los 4 factores que fomentan la fuga de talento, 2022, <https://www.linkedin.com/news/story/los-4-factores-que-fomentan-la-fuga-de-talento-6026154/>

### **Los mejores softwares para la gestión de bases de datos en empresas.**

Los softwares de bases de datos se utilizan para almacenar y administrar datos en formatos estructurados, es por esto que se les denomina bases de datos. En general estos proporcionan una interfaz gráfica que permite a los usuarios crear, editar y administrar campos de datos y registros en forma tabular y organizada. Para cualquier compañía contar con una herramienta como estas les permite poder acceder a cualquier información que requieran en formatos sin procesar o basados en informes acompañados de graficas. Además, el administrador, un cliente o un trabajador podrá acceder a esta información de manera rápida maximizando la efectividad y la productividad.

En el siguiente cuadro realizaremos la comparación de algunos de los más populares softwares para la gestión de bases de datos.

<b>SOFTWARE</b>	<b>¿QUE ES?</b>	<b>VENTAJAS</b>	<b>DESVENTAJAS</b>
<b>Microsoft SQL server</b>	<i>Es el gestor de datos de Microsoft más eficiente del mercado.</i>	<ul style="list-style-type: none"> <li>-Una versión gratuita con una integración grafica llena de beneficios para los usuarios que les permitirá ahorrar tiempo.</li> <li>-Instalación desde 0.</li> <li>-Posibilidad de crear múltiples diseños y tabas.</li> <li>-Se puede integrar con programas similares creando consultas avanzadas usando la función Developer Network.</li> </ul>	<i>-Utiliza mucho espacio en la memoria RAM para su instalación y utilización.</i>
<b>MySQL</b>	<i>Es una herramienta de alta velocidad para el procesamiento de datos. Es una herramienta muy confiable ya que ofrece soporte técnico e implementa aplicaciones complejas que están orientadas a negocios</i>	<ul style="list-style-type: none"> <li>-En su nueva versión trae mejores opciones para recuperar datos.</li> <li>-Fácil de comprender e implementar.</li> <li>-No se necesitan conocimientos en programación.</li> <li>-Integración con los motores web apache</li> </ul>	<ul style="list-style-type: none"> <li>-No es tan rápida como otros administradores de bases de datos.</li> <li>-No es el más amigable con los programas que se utilizan actualmente.</li> <li>-Cuando se debe modificar la estructura de base de datos pueden existir ligeros fallos.</li> </ul>
<b>Amazon RDS</b>	<i>De las mejores herramientas para la administración de bases de datos desarrollada por Amazon. Esta herramienta brinda una conexión muy segura y que puede realizar copia automática de seguridad a través de una función incorporada en el software, además se</i>	<ul style="list-style-type: none"> <li>-El usuario puede procesar cargas de trabajo pesadas en una sola base de datos.</li> <li>-Solamente se paga por los recursos utilizados.</li> <li>-Se puede acceder a MySQL, Oracle y Microsoft SQL.</li> </ul>	<ul style="list-style-type: none"> <li>-No es apto para principiantes ya que el sistema es complejo para quienes no tiene experiencia previa.</li> <li>-Es una plataforma genérica que sirve para todo tipo de negocios pero que no está especializada en entornos regulados.</li> </ul>

	<i>puede modificar fácilmente el tamaño de las actividades de la base de datos.</i>		
<b>ORACLE RDBMS</b>	<i>Es el producto más destacado de la compañía. ORACLE nos ofrece un sistema de gestión de bases de datos relacionales de varios modelos diseñado principalmente para la computación grid empresarial y el almacenamiento de dato. Es una de las mejores opciones para las empresas que buscan herramientas confiables y rentables.</i>	<ul style="list-style-type: none"> <li><i>-Estándares flexibles.</i></li> <li><i>-Escala bajo demanda</i></li> <li><i>-Mejores costos.</i></li> <li><i>-Seguridad y soporte garantizados.</i></li> </ul>	<ul style="list-style-type: none"> <li><i>-Su plan de precios no es factible para pequeñas empresas.</i></li> <li><i>-El diseño parece un poco desfavorable ya que otras herramientas brindan una mejor interfaz.</i></li> <li><i>-Consume muchos recursos.</i></li> </ul>

### **Herramientas de software para la prueba y diseño de algoritmos en aprendizaje automático**

#### **(Machine Learning)**

Es importante utilizar un software de herramientas de aprendizaje automático para diseñar, probar y analizar algoritmos de predicción debido a que estos sistemas pueden procesar grandes cantidades de datos complejos y encontrar patrones que serían difíciles o imposibles de detectar manualmente. Además, estos softwares ofrecen una variedad de algoritmos de aprendizaje automático pre-entrenados que se pueden utilizar para una amplia gama de aplicaciones de predicción, desde la detección de fraude hasta la recomendación de productos. Al utilizar estas herramientas, los desarrolladores pueden acelerar el proceso de construcción de modelos y aumentar la precisión y eficiencia de sus sistemas de predicción.

En el siguiente cuadro comparamos algunas de las herramientas más populares en el uso de aprendizaje automático.

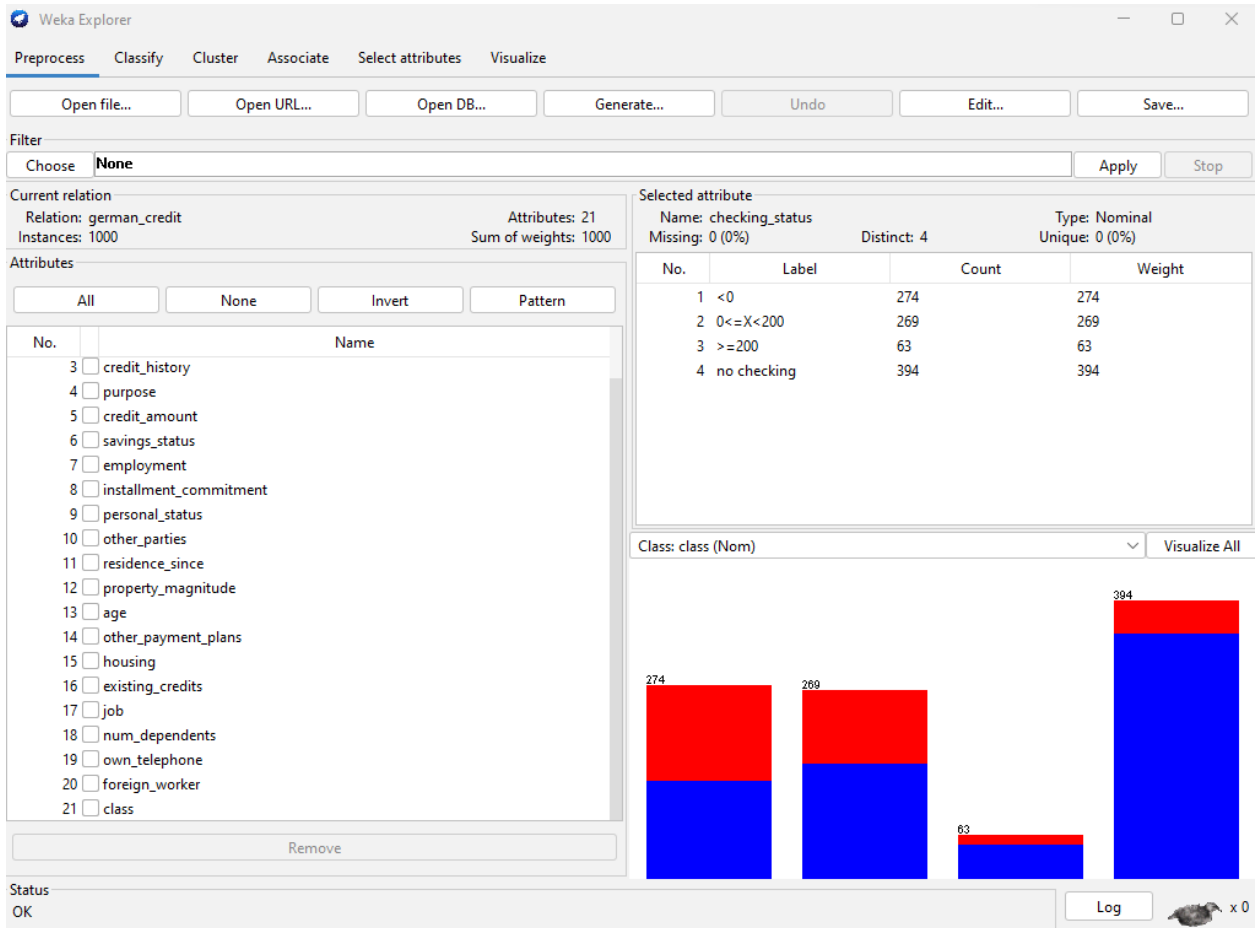
<i>Software</i>	<i>Plataforma</i>	<i>Lenguaje de programación</i>	<i>Algoritmos o características</i>
<i>Scikit Learn</i>	<i>Linux, Mac OS, Windows</i>	<i>Python, Cython, C, C++</i>	<i>Clasificación, Regresión, Agrupación, Selección de modelo de Preprocesamiento y Reducción de dimensionalidad.</i>
<i>PyTorch</i>	<i>Linux, Mac OS, Windows</i>	<i>Python, C++, CUDA</i>	<i>Autograd Module, Optim Module y nn Module</i>
<i>TensorFlow</i>	<i>Linux, Mac OS, Windows</i>	<i>Python, C++, CUDA</i>	<i>API de Librería para programación de flujo de datos</i>
<i>Weka</i>	<i>Linux, Mac OS, Windows</i>	<i>Java</i>	<i>Preparación de datos, Clasificación, Regresión, Agrupación, Visualización y Reglas de asociación de minado.</i>
<i>Apache Mahout</i>	<i>Local y servicios en la nube</i>	<i>Java, Scala</i>	<i>Preprocesamiento, Regresión, Agrupación, Recomendadores y Algebra lineal distribuida</i>
<i>Shogun</i>	<i>Linux, Mac OS, Windows, Unix</i>	<i>C++</i>	<i>Regresión, Clasificación, Agrupación, Máquinas de vectores de soporte, Reducción de dimensionalidad y aprendizaje en línea entre otras.</i>

Es recomendable utilizar una base de datos en la nube como Amazon RDS, siempre y cuando el usuario tenga la experiencia necesaria y una base de datos relacional con estructura similar a MySQL, para integrar el flujo de datos con las herramientas de software de machine learning de



Weka. La visualización de los datos es fundamental en el proceso de predicción, y el módulo de visualización de Weka permite examinar los datos y comparar diferentes modelos computacionales y algoritmos aplicados al mismo conjunto de datos. Esto ayuda a seleccionar el modelo que mejor se adapte a las necesidades de predicción. En conclusión, la combinación de una base de datos en la nube como Amazon RDS y una herramienta de software de machine learning como Weka, es una opción adecuada para mejorar los procesos de predicción y análisis de datos en el campo de la ciencia de datos.

**Preparando los datos:** En el uso de cualquier suite de machine learning o data mining es importante primero preparar los datos. Para el uso de los datos en WEKA es necesario convertir nuestra data de excel en formato CVS separado en comas (es muy importante que los datos no lleven comas porque generará errores al momento de importar), ya que el software no reconoce el formato xlsx, una vez se condicionan los datos en atributos y clases pueden ser visualizados en la interfaz grafica y pueden ser editados los datos.



Se pueden usar filtros para probar que atributos o clases se pueden eliminar o filtrar de los datos para lograr una mejor predicción o clasificación. En la ventana de visualización se pueden ver los gráficos de la data set y nos ayuda a ver gráficamente la clasificación que hace el algoritmo seleccionado. Es importante entender que estamos entrenando un algoritmo con datos para esto, pero no es necesario que sea 100% efectivo con los datos de entrenamiento ya que si bien los datos que se alimentaran a futuro tienen la misma procedencia la clasificación del algoritmo debe ser un poco general para que la predicción no se vea afectada en cada vez que se corre la clasificación. En el caso de tener solo un set de datos, es aconsejable usar 2/3 del set como datos de entrenamiento y 1/3 de estos como datos de prueba ya que si usamos los mismos datos de entrenamiento y prueba

no vamos a tener resultados de clasificación o predicción representativos de la aplicación del algoritmo en un set de datos nuevo.

Para realizar pruebas y la clasificación de datos contamos con la data “Sondeo Nacional 2023” donde se tiene el registro de la encuesta de 919 trabajadores que nos dan la información necesaria para comenzar con la asignación de variables y clases para tratar de crear reglas que puedan predecir o explicar de riesgo de rotación en cada colaborador que resuelva la encuesta.

Encuestado	Nivel de Riesgo Organizacional	Nivel de Riesgo Engagement	Nivel de Riesgo Reconocimiento de trabajo	Nivel de Riesgo por Crecimiento	Nivel de Riesgo Total
Encuestado #890	1.0	2.0	1.0	1.0	Bajo
Encuestado #891	1.0	1.0	1.0	1.0	Bajo
Encuestado #892	1.0	1.0	1.0	1.0	Bajo
Encuestado #893	2.0	2.0	2.0	1.0	Medio
Encuestado #894	1.0	1.0	1.0	1.0	Bajo
Encuestado #895	2.0	2.0	2.0	2.0	Medio
Encuestado #896	1.0	1.0	1.0	2.0	Bajo
Encuestado #897	1.0	1.0	2.0	1.0	Bajo
Encuestado #898	2.0	2.0	1.0	1.0	Bajo
Encuestado #899	2.0	2.0	2.0	2.0	Medio
Encuestado #900	2.0	1.0	2.0	2.0	Medio
Encuestado #901	3.0	2.0	2.0	2.0	Medio
Encuestado #902	3.0	2.0	2.0	1.0	Medio
Encuestado #903	2.0	3.0	2.0	1.0	Medio
Encuestado #904	1.0	1.0	1.0	2.0	Bajo
Encuestado #905	1.0	2.0	1.0	2.0	Bajo
Encuestado #906	1.0	2.0	2.0	1.0	Bajo
Encuestado #907	1.0	1.0	1.0	1.0	Bajo
Encuestado #908	3.0	1.0	1.0	1.0	Bajo
Encuestado #909	2.0	2.0	1.0	1.0	Bajo
Encuestado #910	2.0	3.0	2.0	2.0	Medio
Encuestado #911	1.0	2.0	1.0	2.0	Bajo
Encuestado #912	1.0	1.0	2.0	1.0	Bajo
Encuestado #913	1.0	1.0	1.0	1.0	Bajo
Encuestado #914	1.0	2.0	2.0	2.0	Medio
Encuestado #915	1.0	1.0	1.0	1.0	Bajo
Encuestado #916	2.0	2.0	2.0	2.0	Medio
Encuestado #917	2.0	2.0	2.0	3.0	Medio
Encuestado #918	1.0	1.0	1.0	1.0	Bajo
Encuestado #919	2.0	2.0	2.0	2.0	Medio

Clasificamos toda la información en nivel de riesgo organizacional, nivel de riesgo por engagement, nivel de riesgo por reconocimiento laboral y nivel de riesgo por oportunidad de crecimiento, la regla se clasifico según el promedio de los resultados de cada uno de los instrumentos y nos da valores de riesgo bajo, medio y alto.

Al emplear esta data curada y organizada en diferentes algoritmos tratamos de clasificar cuantos de estos colaboradores se verían en riesgo bajo, medio y alto de salir de sus cargos actuales.

Los algoritmos de clasificación usados para entrenar y probar el método de predicción son suministrados por el software de herramientas de machine learning WEKA de la universidad de Weikato de Nueva Zelanda, para las comprobaciones de clasificación se eligieron 5 algoritmos para obtener la información de cuales se pueden ajustar a la tarea de manera eficaz.

Para entrenar y evaluar cada algoritmo se optó por usar un porcentaje del 66% como data aleatoria de entrenamiento y el 34% restante como la data a clasificar o predecir el nivel de riesgo de rotación de los colaboradores del sondeo.

### ZeroR

Es una técnica básica que sirve como punto de referencia para evaluar el rendimiento de otros clasificadores. El clasificador ZeroR no tiene en cuenta ninguna característica de entrada y predice la etiqueta de clase más frecuente en los datos de entrenamiento para todas las instancias. Usamos el clasificador ZeroR como punto de referencia para comparar el rendimiento de otros clasificadores y determinar si ofrecen alguna mejora significativa en comparación con este enfoque básico.

Se obtuvieron los siguientes resultados:

Clasificador	ZeroR
Tiempo empleado en segundos	0
Clasificaciones correctas	196
Clasificaciones Incorrectas	116
% Clasificaciones correctas	62.82%
% Clasificaciones Incorrectas	37.18%
Total de instancias	312

Matriz de confusión			
Clasificación de Riesgo	a	b	c
a = Medio	0	102	0
b = Bajo	0	196	0
c = Alto	0	14	0

## DecisionStump

Es un algoritmo simple que constituye un solo nivel, Utiliza una característica para hacer predicciones basadas en un valor umbral. Si el valor de la característica es menor o igual al umbral, asigna una etiqueta de clase, y si es mayor, asigna otra etiqueta. Para características numéricas, busca el umbral que minimice el error de clasificación. Para características nominales, prueba cada valor como umbral. Si hay valores faltantes, predice la etiqueta de clase más común. Es un modelo simple y útil como punto de referencia para otros clasificadores.

Se obtuvieron los siguientes resultados:

Clasificador	DecisionStump
Tiempo empleado en segundos	0
Clasificaciones correctas	298
Clasificaciones Incorrectas	14
% Clasificaciones correctas	95.51%
% Clasificaciones Incorrectas	4.49%
Total de instancias	312

Matriz de confusión			
Clasificación de Riesgo	a	b	c
a = Medio	102	0	0
b = Bajo	0	196	0
c = Alto	14	0	0

## Decisión Table

utiliza una tabla de decisiones para realizar predicciones. Se basa en reglas de decisión definidas en función de los valores de las características de entrada. Cada regla de decisión consiste en una combinación específica de valores de características y la etiqueta de clase asociada. Durante la clasificación, el clasificador busca en la tabla de decisiones la regla que se ajuste a los valores de características de la instancia y predice la etiqueta de clase correspondiente. Si no se encuentra una regla coincidente, se utiliza una regla predeterminada. El clasificador DecisionTable es útil cuando los datos presentan reglas de decisión claras y explícitas, y puede ser interpretado fácilmente.

Se obtuvieron los siguientes resultados:

Clasificador	DecisionTable
Tiempo empleado en segundos	0.09
Clasificaciones correctas	312
Clasificaciones Incorrectas	0
% Clasificaciones correctas	100%
% Clasificaciones Incorrectas	0%
Total de instancias	312

Matriz de confusión			
Clasificación de Riesgo	a	b	c
a = Medio	102	0	0
b = Bajo	0	196	0
c = Alto	0	0	14

## RandomForest

Es un método de aprendizaje automático basado en ensambles. Construye múltiples árboles de decisión independientes y los combina para realizar predicciones. Cada árbol se construye utilizando una muestra aleatoria de las instancias de entrenamiento y un subconjunto aleatorio de características. Durante la clasificación, cada árbol vota por la etiqueta de clase y la clase más frecuente se selecciona como predicción final. RandomForest es capaz de manejar datos complejos y ruidosos, y puede capturar relaciones no lineales entre las características.

Se obtuvieron los siguientes resultados:

Clasificador	RandomForest
Tiempo empleado en segundos	0.13
Clasificaciones correctas	308
Clasificaciones Incorrectas	4
% Clasificaciones correctas	98.72%
% Clasificaciones Incorrectas	1.28%
Total de instancias	312

Matriz de confusión			
Clasificación de Riesgo	a	b	c
a = Medio	102	0	0
b = Bajo	0	196	0
c = Alto	4	0	10

## J48

El algoritmo J48, también conocido como C4.5, en el software Weka es un método de aprendizaje automático basado en árboles de decisión. Construye un árbol de decisión utilizando

una estrategia de particionamiento recursivo. En cada paso, selecciona la característica que mejor divide el conjunto de datos según ciertos criterios de ganancia de información, como la entropía o el índice de Gini. Luego, el árbol se expande creando ramas para cada valor posible de la característica seleccionada. Durante la clasificación, las instancias descienden por el árbol siguiendo las ramas correspondientes a sus características, y se asigna la etiqueta de clase final del nodo hoja alcanzado.

Se obtuvieron los siguientes resultados:

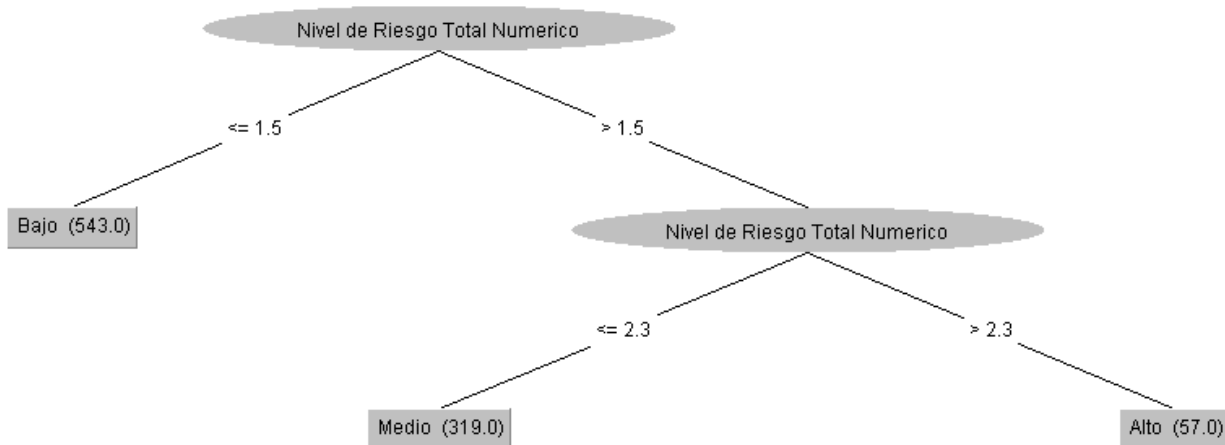
Clasificador	J48
Tiempo empleado en segundos	0.01
Clasificaciones correctas	312
Clasificaciones Incorrectas	0
% Clasificaciones correctas	100%
% Clasificaciones Incorrectas	0%
Total de instancias	312

Matriz de confusión			
Clasificación de Riesgo	a	b	c
a = Medio	102	0	0
b = Bajo	0	196	0
c = Alto	0	0	14

Al comparar los resultados podemos ver que el clasificador J48 (en weka de software libre) es el adecuado para la clasificación de riesgo de rotación en los datos suministrados por el Sondeo Nacional 2023 y cumpliendo las reglas aplicadas al nivel de riesgo por cada instrumento interpretado. Tanto por su precisión como por su velocidad al generar el modelo.

Clasificador	ZeroR	DecisionStump	DecisionTable	RandomForest	J48
Tiempo empleado en segundos	0	0	0.09	0.13	0.01
Clasificaciones correctas	196	298	312	308	312
Clasificaciones Incorrectas	116	14	0	4	0
% Clasificaciones correctas	62.82%	95.51%	100%	98.72%	100%
% Clasificaciones Incorrectas	37.18%	4.49%	0%	1.28%	0%
Total de instancias	312	312	312	312	312

El algoritmo J48 arrojó el árbol de decisión que generó para la clasificación del riesgo de rotación en el 34% de la data suministrada.



### 10. Análisis de costos

La estimación de costos del proyecto se determina en pesos colombianos. a continuación en las siguientes tablas se describe los costos asociados a la operación por tarea de OSEA.

TAREA	DESARROLLO OSEA				
Recurso	Nombre	Cant	Duración (sem)	Costo (sem)	Total costos
Lider de proyecto	Ingeniero lider desarrollo software	1	8	\$1,125,000	\$9,000,000
Personal operativo	Ingeniero de sistema/desarrollador	3	8	\$950,000	\$22,800,000
Bienes de capital (equipos)	Computador portatil	4	8	\$45,000	\$1,440,000
Base de datos	Amazon RDS	20 GB	8	\$0	\$0
Machine Learning	WEKA	1	8	\$0	\$0
	<b>DURACIÓN TAREA</b>		<b>8</b>	<b>COSTO TAREA</b>	<b>\$33,240,000</b>

Entregables	Especificación de requisitos
	Módulo de funcionamiento
	Módulo de implementación
	Diseño detallado del software
	Adquisición servicio en la nube
	Adquisición machine learning
	Código fuente, prototipo
	Software del aplicativo
Manual técnico y de usuario	

<b>COSTO TOTAL DE TAREAS</b>	<b>\$33,240,000</b>
------------------------------	---------------------

COSTO INVERSIÓN		
Concepto	Item	Costo
Software	Machine Learning WEKA	\$0
Almacenamiento Cloud 2 meses	20 GB	\$0
<b>TOTAL COSTO INVERSIÓN</b>		<b>\$0</b>

COSTO PROYECTO	
Costo inversión	\$0
Costo operación	\$33,240,000
IVA (19%)	\$6,315,600
<b>TOTAL</b>	<b>\$39,555,600</b>

Como desarrolladores del proyecto y después de determinar los costos asociados al desarrollo y la operación del software de predicción OSEA, estimamos una utilidad del 20% y un 10% por concepto de imprevistos. A continuación se relaciona el valor total del software.

VALOR SOFTWARE OSEA	
Costo proyecto	\$39,555,600
Utilidad 20%	\$7,911,120
<b>SUBTOTAL</b>	<b>\$47,466,720</b>
Imprevistos (10%)	\$4,746,672
<b>TOTAL</b>	<b>\$52,213,392</b>

## 11. Conclusiones

Para proponer, diseñar e implementar un algoritmo predictivo para disminuir y evitar la rotación y fuga de talentos de las empresas que contratan los servicios de GSH, a partir del análisis de los datos otorgados por los empleados mediante la encuesta Sondeo Nacional 2023 fue necesario investigar sobre las variables causantes de las renunciaciones o búsqueda de otros empleos por parte de los colaboradores en general, para contrastar con la información capturada por las encuestas y encontrar instrumentos de clasificación que nos ayudaran a agrupar las partes del cuestionario en diferentes variables de riesgo de rotación involuntario para realizar una predicción del fenómeno e identificar cuáles de estos aspectos están relacionados con el fenómeno de rotación.

Estos cuestionarios que se emplean suelen ser extensos y la recopilación de la información puede llegar a ser extremadamente difícil si no se realiza mediante un método que facilite su recopilación y organización, por ejemplo una entrevista en sitio o el llenado del cuestionario a mano para luego ser digitalizado impondría horas y espacio para errores por lo cual el método de encuesta en línea puede ser la mejor manera de obtener la información ya que con el anonimato los encuestados tienden a ser más honestos en sus respuestas.

La mayoría de herramientas en línea para conducir tus propias encuestas como Microsoft forms, Google forms o Survey Monkey, tienen la ventaja de almacenar los resultados, mostrar algunas gráficas y análisis simples así como la posibilidad de descargar la información en hojas de cálculo para realizar las operaciones que se consideren necesarias, incluso se pueden crear reglas que alimenten las tablas de una base de datos para manejar históricos en caso de que las preguntas de la encuesta no cambien.

Al relacionar las variables obtenidas por el Sondeo Nacional 2023 de los colaboradores colombianos, en instrumentos de valoración de riesgo de rotación laboral pudimos aplicar algoritmos de clasificación que al ser entrenados pueden predecir en este caso hasta con un 100% de certeza el nivel de riesgo de rotación en los colaboradores consultados por el Sondeo Nacional 2023.

Al probar los algoritmos ZeroR, DecisionStump, DescisionTable, RandomForest y J48, pudimos identificar que para este caso en particular el J48 es el mas eficaz, debido a su rapidez y nivel de certeza, no obstante otros algoritmos de clasificación y predicción pueden mostrar mejores resultados con otra data, otra forma de probar algoritmos eficaces para la tarea sería tener valores históricos y contrastar el nivel de riesgo con los porcentajes de rotación, además de realizar pruebas del modelo con otra data para revisar el nivel de certeza con esta nueva información.

## 12. Referencias

- A. (2021, 15 junio). *Fuga y retención del talento*. APERHU. <https://aperhu.pe/fuga-y-retencion-del-talento/>
- A. (2022, 17 noviembre). *Estrategias de retención de talento humano*. Accom. <https://www.wecom.com/tendencias/estrategias-de-retencion-de-talento-humano/>
- Bidanda, Bopaya M., ed. 2023. “INTRODUCTION TO INDUSTRY 4.0 AND INDUSTRIAL INTERNET OF THINGS (IIoT).” Chap. 27.1 in *Maynard's Industrial and Systems Engineering Handbook*. 6th ed. New York: McGraw.
- Bifet, A., Holmes, G., Kirkby, R., & Pfahringer, B. (2010). *MOA: Massive Online Analysis*. *Journal of Machine Learning Research*, 11(52), 1601–1604. <http://jmlr.org/papers/v11/bifet10a.html>
- Calders, T., & Custers, B. (2013). *What Is Data Mining and How Does It Work?* (Vol. 3). [https://link.springer.com/chapter/10.1007/978-3-642-30487-3\\_2](https://link.springer.com/chapter/10.1007/978-3-642-30487-3_2)
- Casas Anguita, J., Repullo Labrador, J. R., & Donado Campos, J. (2003). *La encuesta como técnica de investigación. Elaboración de cuestionarios y tratamiento estadístico de los datos (I)*. *Atencion primaria*, 31(8), 527–538. [https://doi.org/10.1016/s0212-6567\(03\)70728-8](https://doi.org/10.1016/s0212-6567(03)70728-8)
- Durán, C. B. (2022, 13 julio). *¿Por qué a las empresas se les hace tan difícil retener el talento?* Forbes Colombia. <https://forbes.co/2022/07/13/capital-humano/por-que-a-las-empresas-se-les-hace-tan-dificil-retener-el-talento>

El Nuevo Siglo. (s. f.). *La fuga de talento de las empresas, una amenaza para el desarrollo.*

<https://www.elnuevosiglo.com.co/articulos/09-09-2021-la-fuga-de-talento-de-las-empresas-una-amenaza-para-el-desarrollo>

F. Martínez-Plumed et al., "CRISP-DM Twenty Years Later: From Data Mining Processes to Data Science Trajectories," in *IEEE Transactions on Knowledge and Data Engineering*, vol. 33, no. 8, pp. 3048-3061, 1 Aug. 2021, doi: 10.1109/TKDE.2019.2962680.

<https://ieeexplore.ieee.org/abstract/document/8943998/citations#citations>

IMTERIN GROUP.(2020, noviembre 30) *Fuga de Talento: causas, consecuencias y soluciones.*

<https://interimgrouphr.com/blog/fuga-talento-causas-soluciones/#:~:text=%C2%BFQu%C3%A9%20es%20la%20Fuga%20de,vida%20laboral%20en%20otra%20empresa.>

*La fuga de talento supone a las empresas una pérdida de hasta un 20% del salario anual del trabajador.*

(2020, 31 enero). Equipos y talento. <https://www.equiposytalento.com/noticias/2020/01/31/la-fuga-de-talento-supone-a-las-empresas-una-perdida-de-hasta-un-20-del-salario-anual-del-trabajador>

Nasr, M. M., Shaaban, E., Samir, A., & Nasr, M. (2019). *A proposed Model for Predicting Employees' Performance Using Data Mining Techniques: Egyptian Case Study.* In *Article in International Journal of Computer Science and Information Security.* <https://www.researchgate.net/publication/331165269>

Redactora, J. (2022, abril 28). ▷ *¿Qué son los Bancos de Información y cómo funcionan?* 2023.

MiDinero.info. <https://midinero.info/que-son-los-bancos-de-informacion-y-como-funcionan/>

S. (2023, 28 enero). *No todo es dinero. Estas son las claves para evitar la fuga de talentos de una empresa.* Semana.com últimas Noticias de Colombia y el Mundo.

<https://www.semana.com/hablan-las-marcas/articulo/no-todo-es-dinero-estas-son-las-claves-para-evitar-la-fuga-de-talentos-de-una-empresa/202300/>

Web, R. (2022, 14 marzo). *¿Cómo y por qué debes evitar la fuga de talento?* Recursos Humanos, el Portal del Capital Humano | Guía Empresas Recursos Humanos - Infocapitalhumano.pe es un portal especializado en recursos humanos en Perú, que brinda la más completa información del y para el sector de recursos humanos. <https://infocapitalhumano.pe/recursos-humanos/noticias-y-movidas/como-y-por-que-debes-evitar-la-fuga-de-talento/>

8 apps for field data collection (teamscope, ODK, etc). (2023, abril 2). [https://www.teamscopeapp.com.translate.google/mobile-data-collection-guide/7-mobile-data-collection-apps-for-field-research? x\\_tr\\_sl=en& x\\_tr\\_tl=es& x\\_tr\\_hl=es& x\\_tr\\_pto=rq](https://www.teamscopeapp.com.translate.google/mobile-data-collection-guide/7-mobile-data-collection-apps-for-field-research? x_tr_sl=en& x_tr_tl=es& x_tr_hl=es& x_tr_pto=rq)

11 Most Popular Machine Learning Software Tools in 2023. (2023.). Retrieved April 1, 2023, from <https://www.softwaretestinghelp.com/machine-learning-tools/>

Elpais (2022, 12 noviembre) *¿Renunciaría a su trabajo? La reveladora respuesta de los colombianos a esta pregunta.* <https://www.elpais.com.co/economia/renunciaria-a-su-trabajo-la-reveladora-respuesta-de-los-colombianos-a-esta-pregunta.html>

Portafolio (2016, 02 mayo) ¿Qué es lo que aburre a los empleados y los lleva a renunciar?

<https://www.portafolio.co/economia/empleo/renuncia-gente-495127>

García, M. (2023, 18 mayo) Fuga de talento, síndrome de burnout, ¿Quién se encarga de esto?

<https://www.linkedin.com/pulse/fuga-de-talento-s%C3%A9ndrome-burnout-qui%C3%A9n-se-encarga-esto-garcia/>

Miravalles, J. Cuestionario de Maslach Burnout Inventory.

<http://www.javiermiravalles.es/sindrome%20burnout/Cuestionario%20de%20Maslach%20Burnout%20Inventory.pdf>

Los 4 factores que fomentan la fuga de talento, 2022, [https://www.linkedin.com/news/story/los-4-](https://www.linkedin.com/news/story/los-4-factores-que-fomentan-la-fuga-de-talento-6026154/)

[factores-que-fomentan-la-fuga-de-talento-6026154/](https://www.linkedin.com/news/story/los-4-factores-que-fomentan-la-fuga-de-talento-6026154/)