



Escuela de Posgrados

CONSTRUCCIÓN DE UN MODELO PREDICTIVO PARA LA ASIGNACIÓN DE TAREAS A EQUIPOS TRABAJO DEL SECTOR TECNOLÓGICO

Cindy Vanessa Alvarez Arbelaez
Laura Isabel Quintero Riaño
Maicol Steven Parra Molina

Trabajo de Grado presentado como requisito para optar al título de:

Especialista en Big Data e Inteligencia de Negocios

Asesor: Ingrid Durley Torres Pardo
Docente

Universidad Católica Luis
Amigó Facultad de Ingenierías
y Arquitectura
Especialización en Big Data e Inteligencia de
Negocios Medellín, Colombia

2024

Dedicatoria

A nuestros padres, que con su ejemplo de superación, amor, paciencia y apoyo nos han guiado durante nuestra vida para cumplir nuestros objetivos educativos y profesionales. Nos han inspirado a perseguir nuestros sueños sin olvidarnos del presente y de disfrutar nuestra familia. A nuestras parejas que nos apoyaron y nos motivaron con la mejor actitud para culminar esta meta. Con todo nuestro cariño y reconocimiento dedicamos este trabajo a ustedes.

Agradecimientos

Agradecemos al personal docente de la Especialización Big data e Inteligencia de negocios 2023-2024 de la Universidad Católica Luis Amigo por transmitirnos el mejor conocimiento para culminar este posgrado y aplicarlo en nuestras labores profesionales, contribuyendo a nuestro crecimiento tanto personal como profesional. Agradecemos a nuestros compañeros por hacer este camino más divertido, por todas las lecciones aprendidas, horas compartidas y enseñanzas.

Resumen

La asignación de tareas a los equipos de trabajo del sector tecnológico es un factor crucial para optimizar la productividad y el desempeño organizacional, los modelos predictivos basados en técnicas de aprendizaje supervisado ofrecen una solución prometedora para abordar esta actividad, generando eficiencia, entregas oportunas y un equilibrio de equipo. En el caso de estudio presentado se construyó un modelo predictivo para la asignación de tareas a los equipos del área de tecnología, aplicando machine learning (ML) y técnicas de clasificación, aprovechando datos históricos de la compañía en cuanto al personal con las variables relevantes. Para el desarrollo del trabajo se aplicó la metodología CRISP-DM para recopilar, preparar y modelar los datos históricos y relevantes de los equipos de trabajo. Se utilizaron algoritmos de clasificación supervisados (Árboles de Decisión, K-Vecinos Cercanos (KNN), Máquinas de Vectores de Soporte (SVM), y Regresión Logística) para crear los modelos predictivos. El modelo SVM obtuvo las métricas de precisión más altas (53%), demostrando su capacidad para predecir la asignación de tareas con mayor precisión que los otros modelos. Además, los modelos lograron identificar variables que influyen en la capacidad ocupacional de los equipos, como la experiencia y el tipo de tarea. La evaluación del modelo SVM reveló que la mayoría de las predicciones fueron correctas, pero aún había margen de mejora. Los resultados obtenidos indican viabilidad del modelo, pero invitan a la compañía a recopilar más datos y considerar factores adicionales como económicos, emocionales entre otros, para mejorar el rendimiento del modelo.

Palabras clave:

Asignación de tareas

Machine learning (ML)

Modelos de clasificación

SVM

Equipos de trabajo

Tabla de Contenido

1. Introducción	9
2. Planteamiento del Problema	11
3. Justificación	13
4. Marco de Referencias	15
4.1 Marco Teórico	15
4.2 Marco Conceptual	23
4.3 Marco Normativo	26
5. Antecedentes	27
6. Objetivos	29
a. Objetivo General	29
b. Objetivos Específicos	29
7. Viabilidad	30
8. Metodología	31
9. Resultados	34
10. Conclusiones	51
11. Recomendaciones	53
12. Referencias	54

Lista de Figuras

Figura 1 Planteamiento del problema con Ishikawa.....	12
Figura 2 Descripción de las variables disponibles.....	36
Figura 3 Aplicación de funciones Python para análisis de datos.	37
Figura 4 Variables finales para trabajar los modelos.....	38
Figura 5 Representación de variable objetivo y transformación a variable numérica.....	38
Figura 6 Aplicación del Split para los modelos.....	39
Figura 7 Instanciado de modelos de clasificación.....	39
Figura 8 Entrenamiento de modelos sin hiper parámetros.....	40
Figura 9 Predicción de modelos sin hiper parámetros.....	40
Figura 10 GridSearch modelo Árbol de decisión.....	41
Figura 11 GridSearch modelo KNN.....	41
Figura 12 GridSearch modelo SVM.....	42
Figura 13 GridSearch modelo Regresión Logística.....	42
Figura 14 Entrenamiento y predicción mejores parámetros Árbol de decisión.....	43
Figura 15 Entrenamiento y predicción mejores parámetros KNN.....	43
Figura 16 Entrenamiento y predicción mejores parámetros SVM.....	44
Figura 17 Entrenamiento y predicción mejores parámetros Regresión logística.....	44
Figura 18 Resultados Accuracy de los modelos sin hiper parámetros.....	45
Figura 19 Resultados Recall de los modelos sin hiper parámetros.....	45
Figura 20 Resultados Precisión de los modelos sin hiper parámetros.....	46
Figura 21 Resultados evaluación final modelo Árbol de decisión.....	46
Figura 22 Resultados evaluación final modelo KNN.....	47
Figura 23 Resultados evaluación final modelo SVM.....	47

Figura 24 Resultados evaluación final modelo Regresión logística.48
Figura 25 Matriz de confusión SMV49

Lista de Tablas

Tabla 1 Fases CRISP-DM aplicadas en el objetivo 1.	31
Tabla 2 Fases CRISP-DM aplicadas en el objetivo 2.	32
Tabla 3 Fases CRISP-DM aplicadas en el objetivo 3.	32
Tabla 4 Entendimiento del negocio.	34
Tabla 5 Resultado métricas de los modelos de clasificación.	48

1. Introducción

El área de tecnología se ha convertido en una de las áreas más predominantes y significativas del entorno empresarial por la evolución constante del mercado que ahora se torna más digital y dinámico. Esta área, se centra en crear, vigilar y controlar soluciones tecnológicas para la mejora de productos y servicios, teniendo en cuenta la variable de tiempo para ser cada vez más competitivos y eficientes en el mercado con entregas oportunas a los clientes internos y externos.

El presente trabajo describe el proceso de construcción de un modelo predictivo para la asignación de tareas a equipos de trabajo del área de tecnología de una compañía nacional financiera, aplicando Machine Learning, modelos de clasificación y basándose en la metodología CRIPS-MD, con el fin de mejorar la productividad de los equipos de trabajo, asignando balanceadamente las tareas al personal idóneo, obteniendo como resultado mejora en el bienestar del empleado, eficiencia en el proceso y contribuyendo a la productividad total de la compañía, ya que actualmente se evidencian retrasos en las entregas de las soluciones tecnológicas a los clientes internos y externos, lo que está impactando el crecimiento del negocio, status nacional y valor, afectando de esta manera los recursos y presupuestos asignados.

Con la creación del modelo supervisado (de clasificación), se asignan mejor las tareas a los equipos de trabajo teniendo en cuenta datos históricos y comportamientos. Este enfoque permitirá a la empresa anticiparse y asignar recursos de manera más eficiente, optimizando los tiempos de entrega y reduciendo retrasos.

La cantidad de datos puede ser limitada pero igualmente se buscará aprovechar al máximo los datos disponibles y aplicar técnicas avanzadas de modelado para desarrollar una solución efectiva que contribuya positivamente a la compañía. En el desarrollo de este

trabajo se encuentran las aplicaciones de los modelos de ML para la asignación de tareas a los equipos de trabajo del área de tecnología con la metodología, resultados y comentarios respectivos.

2. Planteamiento del Problema

Actualmente la empresa para la cual se está realizando esta intervención desarrolla soluciones financieras digitales y el equipo a cargo de los desarrollos es el área de tecnología, esta área recibe las necesidades del negocio de acuerdo a una planeación trimestral, la cual se divide en entregas quincenales por sprint; cada equipo de tecnología tiene asignadas unas tareas para la ejecución y cierre dentro de cada sprint.

Los equipos de trabajo tienen una planeación definida e inician con la ejecución de tareas de acuerdo a dicha planeación, sin embargo los equipos de trabajo se ven afectados por el ingreso de nuevas tareas que son requerimientos no planeados y que no se tiene dimensionados dentro del alcance inicial de los proyectos, como por ejemplo los requerimiento legales por parte de los entes regulatorios, corrección de vulnerabilidades identificadas por ciberseguridad, incidentes por fallas en los servicios que ya se encuentran en producción y que están afectando a los clientes.

Todos estos requerimientos deben ser atendidos con una prioridad inmediata impactando la planeación inicialmente pactada con el negocio. En esta empresa no se aplican modelos predictivos para la asignación efectiva de las tareas a los equipos, los cuales ayudarían a tener en cuenta variables predeterminadas y/o analizar de forma proactiva como se encuentra la capacidad de los equipos, como se han atendido estos requerimientos anteriormente, cual es el equipo que ya ha atendido este tipo de tareas de acuerdo a su conocimiento y experiencia.

La ausencia de modelos o metodologías de trabajo impacta directamente la capacidad ocupacional de las personas que hacen parte de los equipos de tecnología generando una sobrecarga laboral cuando no se asignan de forma efectiva y con un análisis previo estos nuevos requerimientos, también generan retrasos en los proyectos y aumento de los tiempos de entrega de las necesidades planteadas inicialmente con el negocio y los clientes.

La asignación ineficiente de tareas está impactando las entregas oportunas de valor del negocio y su crecimiento, así mismo afecta la administración de los recursos y presupuesto generando sobrecostos para la organización debido al incumplimiento de los proyectos. A continuación, se ilustra el diagrama de Ishikawa que sirve para identificar el problema que motivo para el trabajo de grado y la relación entre los factores y causas para el mismo.

Fuente: Elaboración propia

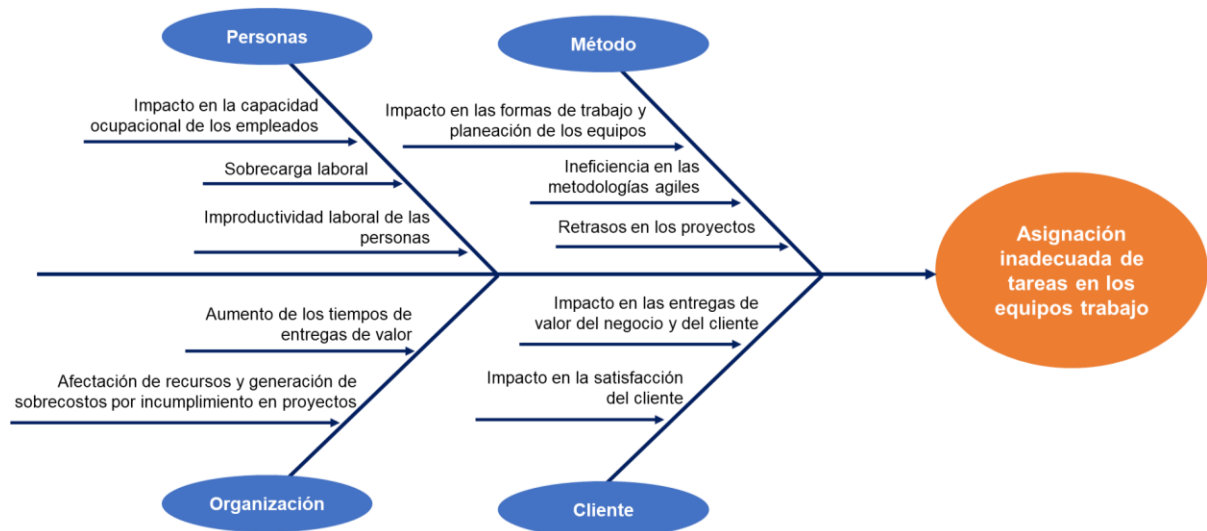


Figura 1 Planteamiento del problema con Ishikawa.

3. Justificación

La gestión eficiente de los equipos de trabajo en las áreas de tecnología es necesaria para mantener la competitividad y la eficacia operativa, es por esto que dicha asignación de tareas a los equipos de trabajo debe realizarse en una forma adecuada y efectiva; lo anterior, constituye uno de los retos más comunes que enfrentan actualmente las organizaciones que trabajan bajo metodologías ágiles y dividen su trabajo por líneas de negocio buscando aumentar la productividad y calidad de los productos y servicios para los clientes.

Una adecuada asignación de tareas en los equipos de trabajo asegura que las organizaciones puedan satisfacer la demanda de sus clientes tanto internos como externos de manera oportuna, cumpliendo con los objetivos en los plazos establecidos y contribuyendo al éxito general de las organizaciones de bienes y servicios. Para realizar una asignación de tareas de forma óptima y una buena gestión de equipos en las organizaciones es necesario analizar variables que impacte dicha productividad, datos históricos, tiempos de ejecución de tareas y experiencia de las personas que las atienden.

Actualmente, se observa que muchas actividades o tareas que ejecutan los empleados en las organizaciones tienen alto impacto en el buen desarrollo o prestación de un bien o servicio, por lo tanto, una buena gestión en estas tareas en los equipos se vuelve crucial al momento de definir tiempos de entrega, personal necesario, costos y demás variables para desarrollar un proyecto en una empresa de bienes o servicios.

Así mismo se evidencia que las organizaciones en la actualidad están aplicando el machine learning para mejorar la eficiencia operativa y el comportamiento de los empleados de acuerdo a la asignación de tareas lo que influye directamente en la productividad empresarial. Utilizar estas tecnologías permite la toma de decisiones más oportuna y la prestación de servicios a tiempo a los clientes, lo que conduce a una producción más rápida y efectiva de productos y servicios sin importar el tipo de empresa (Ramachandran et al., 2022).

La aplicabilidad del machine learning ayuda a las empresas a priorizar estrategias y decisiones relacionadas con la asignación de tareas comprendiendo mejor qué aspectos de la asignación de tareas son cruciales para mejorar el rendimiento de los empleados en las empresas y cómo pueden asignar las tareas de una manera más precisa, aumentando así la productividad y el bienestar de los empleados (Pap et al., 2022).

Es por esto que la falta de un método sistemático y predictivo para dicha asignación de tareas puede desencadenar retrasos en la entrega de proyectos, improductividad laboral, sobrecarga en los empleados e insatisfacción del cliente.

Esta intervención empresarial, tiene como objetivo analizar esta problemática y aplicar diferentes técnicas de aprendizaje supervisado con los algoritmos de clasificación que nos permita construir un modelo predictivo para la asignación óptima de tareas a los equipos de trabajo de tecnología de acuerdo a las variables entregadas con el negocio.

Esta intervención permitirá a la empresa tener disponible una propuesta para asignar las tareas de una forma más precisa, optimizando la carga de trabajo de los equipos, disminuyendo los cuellos de botella y usando los recursos disponibles efectivamente.

4. Marco de Referencias

4.1 Marco Teórico

En diferentes industrias el área de tecnología se ha convertido en una de las áreas más significativas e importantes ya que las empresas se están adaptando a la constante evolución de soluciones tecnológicas para mejorar los productos, servicios y procesos. De acuerdo a esto y a la inmediatez que demanda el mercado surge la necesidad de implementar y aplicar metodologías ágiles de trabajo donde se han convertido en pilares fundamentales en los equipos de trabajo del área de tecnología por su flexibilidad, entrega iterativa, trabajo en equipo y mejora continua. Por definición, las metodologías ágiles son aquellas que permiten adaptar la forma de trabajo a las condiciones del proyecto, consiguiendo flexibilidad e inmediatez en la respuesta para amoldar el proyecto y su desarrollo a las circunstancias específicas del entorno. (IBS, 2024)

Las metodologías ágiles son bien adoptadas por las *pequeñas* y medianas empresas (PYME) debido a que les permiten tener procesos organizados, repetibles y mejorables sin una alta inversión de presupuesto y de tiempo en su implementación. Las metodologías más utilizadas en la industria actualmente es SCRUM (Deemer et al., 2010). Las metodologías ágiles ayudan a que los equipos de trabajo del área de tecnología se puedan ajustar a cambios en los requerimientos de los proyectos, realizando entregas iterativas, cumpliendo en la brevedad posible el tiempo pactado y a su vez permite controlar adecuadamente presupuestos y/o costos designados a las diferentes soluciones y/o proyectos.

En la actualidad, aún se usan métodos tradicionales para la asignación de tareas a los equipos de trabajo, basados en decisiones administrativas o experiencia de los líderes con el personal para distribuir las tareas entre los miembros del equipo, lo que puede ocasionar insatisfacción

laboral de algunos empleados, pues en muchos casos puede notarse un desbalance en la carga de tareas y esto puede llevar a ineficiencia operativa. Con el conocimiento del machine learning ahora las empresas pueden aprovechar las ventajas de este para analizar datos históricos y tomar decisiones más acertadas para asignar tareas a los equipos de trabajo, manteniendo y creando un excelente ambiente laboral y por lo tanto alta productividad.

La satisfacción laboral también es un factor importante para predecir el rendimiento laboral, compromiso laboral y el compromiso organizacional, en otras palabras, cuando los empleados están satisfechos es más probable que los empleados rindan y se comprometan con sus organizaciones. ¿Por qué son importantes los empleados satisfechos? En teoría, se supone que los empleados satisfechos se sienten obligados para perseguir y alcanzar metas desafiantes y continuar con la organización. La satisfacción en lugar de trabajo empuja a los empleados a realizar un esfuerzo adicional, como sugiere la norma de reciprocidad. La satisfacción promueve que los empleados se comprometan con la organización porque podrían "crear sus propios recursos". Así mismo la calidad de vida laboral es una situación favorable que mejora la satisfacción de los empleados al cumplir con recompensas adecuadas, seguridad laboral y oportunidades de crecimiento. La calidad laboral designa la comprensión más amplia que tiene un individuo de su trabajo, implica sentimientos positivos de los empleados, posibilidades de desarrollo, uso de las propias habilidades, futuro laboral significativo y un equilibrio entre el trabajo y la familia (Rimi et al., 2023).

Con el machine learning y su aplicación de modelos las empresas obtienen ventajas de precisión y eficiencia en la asignación de tareas, generando mejores tiempos de entrega, respuestas oportunas a los clientes internos y externos, bienestar laboral y mejora en los recursos presupuestales.

El machine learning además de mejorar la eficiencia operativa y el comportamiento de los

empleados de acuerdo a la asignación de tareas, influye directamente en la productividad empresarial. Esta tecnología permite la toma de decisiones más oportuna y la prestación de servicios a tiempo a los clientes, lo que conduce a una producción más rápida y efectiva de productos y servicios sin importar el tipo de empresa (Ramachandran et al., 2022).

Además, la aplicabilidad del machine learning en las empresas permite identificar factores críticos que influyen en la capacidad ocupacional de los empleados, factores como la organización del trabajo y la innovación son especialmente relevantes, y el machine learning ayuda a las empresas a priorizar estrategias y decisiones relacionadas con la asignación de tareas. Al comprender mejor qué aspectos de la asignación de tareas son cruciales para el rendimiento de los empleados, las empresas pueden asignar tareas de manera más precisa, aumentando así la productividad y el bienestar de los empleados (Papet al., 2022).

En la actualidad además de la asignación de tareas, el machine learning se aplica y/o utiliza para evaluar criterios salariales en los empleados. Analiza datos como formación, experiencia, rendimiento y habilidades para determinar el valor salarial y la clasificación de los empleados en la organización. Esta evaluación precisa mediante un modelo permite asignar tareas de manera equilibrada y de acuerdo a las fortalezas de cada empleado, maximizando el rendimiento de estos contribuyendo así al éxito del proyecto y a la eficiencia de la organización. Adicional se destaca la importancia de la equidad y la transparencia en la gestión del talento humano (Loyarte-López & García-Olaizola, 2022).

Para contemplar equidad y transparencia en los modelos de machine learning es importante la medición y resultados de los indicadores de rendimiento estratégicos. Estos indicadores permiten identificar las fortalezas y debilidades de los empleados a través del tiempo, lo que facilita la asignación de tareas que se ajusten a sus habilidades y experiencia. Esto puede mejorar la productividad y la eficiencia de los empleados, y contribuir al logro de los objetivos

de la organización. Por lo cual la recopilación de datos de los empleados en cuanto a indicadores de rendimiento son un insumo principal al momento de aplicar machine learning (Al Ramahi et al., 2023).

Es fundamental tener en cuenta la eliminación de sesgos al utilizar datos históricos para asignar tareas a los empleados de acuerdo al análisis de la capacidad en cualquier empresa. La preferencia por la comodidad en lugar de la competencia puede influir en las asignaciones, lo que afecta el rendimiento. Por lo tanto, la limpieza de datos históricos y el uso de medidas de éxito ayudan a identificar y eliminar asignaciones fallidas retrospectivamente, mejorando así el rendimiento de los modelos de machine learning y garantizando una asignación justa y precisa de tareas (Tecimer et al., 2022).

En las empresas el machine learning es una herramienta valiosa debido a su capacidad para abordar problemas complejos, manejar datos de alta dimensión y adaptarse a entornos cambiantes. Sin embargo, enfatiza la importancia de elegir el algoritmo adecuado y realizar pruebas en un entorno realista para determinar su idoneidad para un centro de trabajo específico, puesto que una empresa de manufactura generalmente está constituida por varios procesos y áreas de fabricación. La selección del algoritmo adecuado depende de diversos factores y variables, incluyendo los datos disponibles y la configuración de parámetros (Wuest et al., 2016).

Por ejemplo, en plataformas de microtareas en línea, el Machine learning se ha aplicado para asignar tareas de manera efectiva a los trabajadores, mediante un enfoque innovador llamado "MCMC-TA" (Markov Chain Monte Carlo Task Assignment) que utiliza técnicas de modelado estadístico y simulación para asignar tareas de manera efectiva a los trabajadores en estas plataformas y que cada empleado pueda consultar su asignamiento actual o estatus. El algoritmo se basa en la estimación de la calidad de los trabajadores y la identificación de los

candidatos adecuados para tareas específicas. Esta técnica mejoró la asignación de tareas mediante la aplicación de algoritmos como Gaussian Mixture Models (GMM) y Markov Chain Monte Carlo (MCMC) (Moayedikia et al., 2020).

Finalmente, el machine learning ha transformado la gestión de recursos humanos, ha permitido al personal de recursos humanos de todo tipo de empresas poder analizar grandes conjuntos de datos para identificar patrones y tendencias en el desempeño laboral. Esto facilita una asignación de tareas más equitativa y eficiente en los equipos de trabajo. Sin embargo, es esencial utilizar el ML de manera transparente y evitar resultados sesgados para asegurar una toma de decisiones justa en el ámbito laboral. En el futuro, se espera que las tecnologías continúen evolucionando y tengan un impacto aún más profundo en las funciones de recursos humanos (Indarapu et al., 2023).

El aprendizaje automático tiene algunos modelos, entre ellos se encuentra el aprendizaje supervisado de clasificación o también conocido como métodos de clasificación, el cual se basa en tener un conjunto de datos de entrenamiento etiquetados los cuales contienen muestras normales como anómalas para construir un modelo predictivo (Omar et al, 2013). También (Cunningham et al., 2008), en su libro Técnicas de aprendizaje automático para multimedia indican que el aprendizaje supervisado implica aprender un mapeo entre conjuntos de variables de entrada y de salida, donde las variables de salida son las etiquetas para predecir los datos deseados. Además, en los modelos de clasificación las variables dependientes (salida) corresponden a atributos que indican a qué clase en particular pertenece una muestra (González et al., 2015). Los algoritmos supervisados de clasificación más comunes son redes neuronales (NN), máquinas de vectores de soporte (SVM), k-vecinos más cercanos (k-NN), redes bayesianas (BN), bosques aleatorios (RF) y árbol de decisión (DT).

De acuerdo con (Deng, 2020), los árboles de decisión (DT) son modelos predictivos que utilizan una estructura de árbol para dividir el espacio de características en regiones más simples y representar las relaciones no lineales entre las variables de entrada y la variable objetivo. Estos modelos son ampliamente utilizados en aprendizaje automático debido a su capacidad para manejar datos complejos y su interpretabilidad. Esta técnica permite la construcción de modelos predictivos precisos y fácilmente interpretables, lo que los convierte en una herramienta valiosa en diversos campos de aplicación. “Una ventaja de los árboles de decisión es que pueden sintetizar una gran cantidad de datos en una estructura de datos pequeña y compacta que contiene la información esencial que necesitamos para la clasificación” (Salzberg, 1998).

Las máquinas de soporte vectorial (SVM, por sus siglas en inglés) son un conjunto de algoritmos de aprendizaje supervisado desarrollados por Vapnik y Cortés en el año 1995 y su equipo AT&T, que han surgido como métodos relacionados con problemas de clasificación y regresión. Su buen desempeño ha llevado a su uso en una gran variedad de problemas. Algunos investigadores han utilizado SVM para solucionar problemas de clasificación y regresión relacionados a la predicción de series de tiempo, mostrando ser mejores en comparación a otras metodologías tradicionales como modelos econométricos y en algunos casos a modelos de aprendizaje de máquina. La idea detrás de las SVM es que, a partir de unos inputs de entrada al modelo, se etiquetan las clases y se entrena una SVM construyendo un modelo que sea capaz de predecir la clase de los nuevos datos que se introduzcan al modelo. La SVM representa en un eje de coordenadas los vectores de entrenamiento, separando las clases por un espacio lo más grande posible. Cuando nuevos datos son introducidos al modelo, estos se colocan sobre el mismo eje y en función de la cercanía de los grupos antes separados, los cuáles serán clasificados en una u otra clase.

El modelo de regresión K-Vecinos más cercanos (KNN) es un método fácil de entender y

utiliza los K puntos de datos más cercanos (más similares en características de entrada) en el conjunto de datos de entrenamiento para estimar el valor de una nueva observación. La regresión KNN se utiliza comúnmente en una variedad de aplicaciones predictivas (Naghibi y Dashtpajardi, 2017). El modelo de regresión KNN opera determinando la distancia entre una nueva observación y todas las observaciones presentes en los datos de entrenamiento.

Una de las principales ventajas de la regresión KNN es su simplicidad e interpretabilidad. Sin embargo, seleccionar un valor adecuado para K es crucial, ya que elegir valores demasiado pequeños o grandes puede resultar en un ajuste excesivo o insuficiente, respectivamente. Además, el rendimiento de la regresión KNN puede verse afectado en datos de alta dimensión o cuando los datos tienen una estructura compleja (Bishop y Nasrabadi, 2006).

La regresión logística (RL) es uno de los instrumentos estadísticos más expresivos y versátiles de que se dispone para el análisis de datos. Su origen se remonta a la década de los sesenta (Confield, Gordon y Smith 1961); su uso se universaliza y expande desde principios de los ochenta debido, especialmente, a las facilidades informáticas con que se cuenta desde entonces. En los últimos años se ha verificado una presencia muy marcada de esta técnica por su gran utilidad. Los métodos de regresión de variable dependiente cualitativa abarcan diferentes modelos que tratan de explicar y predecir una característica cualitativa a partir de los datos de otras variables conocidas, bien cuantitativas o cualitativas que actúan como variables explicativas. La RL consiste en obtener una función logística de las variables independientes que permita clasificar a los individuos en una de las dos subpoblaciones o grupos establecidos por los dos valores de la variable dependiente. (Revista nefrología, 2023)

La regresión logística es un instrumento estadístico de análisis bivariado o multivariado, de uso tanto explicativo como predictivo. Resulta útil su empleo cuando se tiene una variable dependiente dicotómica (un atributo cuya ausencia o presencia se ha puntuado con los valores

cero y uno, respectivamente) y un conjunto de n variables predictoras o independientes, que pueden ser cuantitativas (que se denominan covariables o covariadas) o categóricas. En este último caso, se requiere que sean transformadas en variables ficticias o simuladas (“dummy”), (Bellmanciencias, 2020)

Uno de los rasgos característicos esenciales de la regresión logística que la diferencia de la regresión convencional es la variable dependiente categórica. Las diferencias entre los análisis significan tanto la evaluación de parámetros como el resumen de los supuestos a la vez. El principio de funcionamiento del modelo de regresión logística se basa en la estimación de valores, al igual que la regresión lineal. La diferencia significativa entre la regresión logística y la regresión lineal es el tipo de conjunto de datos. En la regresión logística, las variables están en forma categórica. Los valores de las variables dependientes se obtienen utilizando la posible ocurrencia en regresión logística. (Ahmadini et al., 2022).

El conocimiento de estas técnicas permitiría al lector la comprensión de los modelos aplicados, poder aplicar sus evaluaciones y entender los resultados.

Para aplicar correctamente el machine learning, en este caso, se debe tener en cuenta la recopilación de datos relevantes sobre los empleados y las tareas asignadas. Esto puede incluir información sobre experiencia, habilidades, conocimientos, nivel de formación, historial laboral, complejidad de tareas que se le asignan, entre otras. Es esencial contar con un conjunto de datos completo y de calidad para parametrizar el modelo de machine learning. Posteriormente, se debe continuar con la preparación de los datos donde se incluye la eliminación de valores atípicos y coordinación de variables. Luego corresponde seleccionar el algoritmo de machine learning adecuado, entrenar el modelo para que aprenda patrones y relaciones de los datos para crear las predicciones. Y finalmente se puede pasar a la validación del modelo evaluando el rendimiento en un conjunto de datos o realizando pruebas,

para finalmente de acuerdo a resultados ajustar lo necesario y proceder con la implementación en tiempo real.

El machine learning está demostrando ser una herramienta fuerte en la gestión de equipos, especialmente en lo que respecta a la asignación de tareas. Empleando algoritmos de aprendizaje supervisado se pueden obtener resultados a futuro, identificando patrones de desempeño, distribución de cargas de trabajo, capacidad ocupacional, entre otros.

4.2 Marco Conceptual

Para una mejor comprensión del desarrollo de los objetivos del trabajo de grado se relacionan algunos conceptos:

Aprendizaje supervisado: Estos algoritmos cuentan con un aprendizaje previo basado en un sistema de etiquetas asociadas a unos datos que les permiten tomar decisiones o hacer predicciones como la que se busca en este trabajo. (Iberdrola,2024)

Scrum: Se define a scrum como una colección de procesos para la gestión de proyectos, que permite centrarse en la entrega de valor para el cliente y la potenciación del equipo para lograr su máxima eficiencia, dentro de un esquema de mejora continua. Scrum es un marco de trabajo iterativo e incremental para el desarrollo de proyectos y se estructura en ciclos de trabajo llamados Sprints. Esos son iteraciones de 1 a 4 semanas y se suceden una detrás de la otra, (Alfronzo, 2014).

Machine Learning: Es una disciplina del campo de la Inteligencia Artificial utilizada en este trabajo para identificar a través de algoritmos, patrones en datos masivos y elaborar las predicciones propuestas para la problemática planteada. Este aprendizaje permite a los computadores realizar tareas específicas de forma autónoma, es decir, sin necesidad de ser

programados, (Iberdrola, 2024).

Metodología CRISP: Esta metodología integra todas las tareas necesarias para el desarrollo de este estudio, desde la fase inicial de comprensión del problema hasta el despliegue del modelo predictivo propuesto como solución a través del uso de técnicas de Machine Learning, incluye descripciones de las fases normales de un proyecto, las tareas necesarias en cada fase y una explicación de las relaciones entre las tareas, (IBM, 2021).

Google colab: Es un documento ejecutable que permitirá escribir, ejecutar y compartir código arbitrario de Python en el navegador, para realizar las tareas de aprendizaje automático y análisis de los datos disponibles para este trabajo, (Agrega, 2024).

Python: Lenguaje de programación utilizado para el desarrollo de este trabajo, ampliamente utilizado en la ciencia de datos y Machine Learning (ML). Es eficiente y fácil de aprender, además se puede ejecutar en muchas plataformas diferentes, (AWS, 2024).

GridSearch: Técnica utilizada para encontrar los mejores hiperparámetros para un modelo de aprendizaje automático. Se realiza una búsqueda exhaustiva a través de una cuadrícula de combinaciones predefinidas de hiperparámetros y se evalúa el rendimiento del modelo utilizando validación cruzada para determinar cuáles son los mejores valores de hiperparámetros, (AWS, 2024).

Recall: Indica el porcentaje de muestras clasificadas correctamente sobre el número de muestras que pertenecen a esta clase. Un bajo recall indica que no se están detectando muestras de esa clase, (thedataschools, 2024).

Accuracy: Indica el porcentaje de aciertos del modelo, es decir, la cantidad de muestras clasificadas correctamente sobre el total de muestras, (thedataschools, 2024).

Precision: Indica el porcentaje de muestras clasificadas correctamente sobre el número total de muestras clasificadas en esa clase. Una baja precisión indica que hay un alto número de falsos positivos, (thedataschools, 2024).

F1 – Score: Media armónica entre la precisión y el recall, de forma que indica un desempeño general del modelo para esa clase, (Uniandes, 2024).

Variable dicotómica: es una variable cualitativa o categórica que toma únicamente dos valores. En estadística, las variables dicotómicas se codifican habitualmente para la realización de cálculos con los valores 0 (no) y 1 (sí), (ikusmira, 2024).

Variables dummy: es aquella que toma el valor 1 o 0 para indicar la presencia o ausencia de una cierta característica o condición. Se utiliza comúnmente en análisis estadísticos y modelos de regresión para incluir información categórica en un modelo matemático. (msanchez, 2024)

4.3 Marco Normativo

Para efectos de este trabajo de grado se debe tener en cuenta la Ley de Protección de Datos Personales o Ley 1581 de 2012 debido a que la compañía otorgó permisos sobre la base de datos, la cual almacena información de sus empleados y proyectos de innovación. En conjunto con la empresa se firmó un acuerdo de confidencialidad, en el cual se prohíbe la publicación de este trabajo de grado.

5. Antecedentes

Actualmente se encuentran varias investigaciones relacionadas con la aplicabilidad de modelos predictivos enfocados a la asignación de tareas a los equipos de trabajo. Según el artículo “Machine learning and role of artificial intelligence in optimizing work performance and employee behavior” la aplicación del machine learning para creación de modelos predictivos mejora los resultados laborales y el comportamiento de los empleados, estos aumentan la eficiencia humana, reducen el trabajo repetitivo y mejoran la eficiencia laboral (Ramachandran et al., 2022). El artículo informa que las empresas tienen un mejor desempeño cuando los algoritmos pueden aprovechar eficazmente grandes volúmenes de datos. Además, las soluciones basadas en modelos prometen mejorar la productividad y ayudar a los profesionales de recursos humanos a seleccionar el personal más capacitado.

En otra investigación llamada “Generating balanced workload allocations in hospitals” se aborda la importancia de asignar de manera equitativa la carga de trabajo entre diferentes grupos de empleados para crear condiciones laborales sostenibles. El estudio propone equilibrar la carga de trabajo en diferentes áreas del hospital optimizando el problema de programación de admisión de pacientes. Se introduce un nuevo objetivo de equidad que considera tanto el equilibrio espacial (entre áreas del hospital) como el temporal (entre días en el período de planificación). El problema resultante, con dos objetivos, se resuelve mediante un algoritmo de búsqueda en el espacio de criterios exactos. Utilizando datos del mundo real, los resultados demuestran cómo es posible generar asignaciones de carga de trabajo equitativas espacial y temporalmente al minimizar el nuevo objetivo de equidad propuesto (Smet, 2023).

En la investigación “On the task assignment with group fairness for spatial crowdsourcing”

el problema principal es la asignación de tareas, donde este se vuelve crucial para el funcionamiento efectivo del sistema. Esta investigación afirma que el equilibrio laboral es un aspecto crítico en la asignación de tareas, ya que influye en la participación y satisfacción de los trabajadores y el éxito del proyecto. Por lo tanto, esta investigación tuvo como objetivo maximizar la calidad de la asignación de tareas a los empleados mientras cumple el equilibrio grupal y otras restricciones. Para esto desarrolla un algoritmo novedoso para resolver este problema de manera eficiente. Al crearlo, realizan experimentos exhaustivos utilizando conjuntos de datos ficticios y reales, donde los resultados muestran que la propuesta arrojada mejora significativamente el equilibrio laboral, logrando el objetivo propuesto de asignar las tareas de manera adecuada teniendo en cuenta criterios de tareas, ubicaciones, funciones del empleado, utilidad de la tarea, descripción del trabajador, entre otros, alcanzando la productividad deseada en la compañía, (Han, 2023). También de acuerdo a la investigación “A dynamic task assignment model for aviation emergency rescue based on multi-agent reinforcement learning” se encuentra la importancia de asignar mediante un modelo las tareas a los rescatistas de emergencia área al momento de atender desastres naturales en China, donde las pérdidas humanas y materiales generalmente son significativas. Esta investigación propone un modelo de asignación de tareas dinámico basado en aprendizaje para asignar lógicamente las tareas de rescate de emergencia, donde modela cada lugar de rescate como un agente y el equipo de rescate como los recursos del agente, parametrizándole el entorno del problema, en este caso los fenómenos históricos registrados. Según el escenario adoptado los resultados muestran que el modelo de asignación de tareas optimiza eficazmente tanto el tiempo de respuesta de la tarea como la tasa de finalización de la misma, obteniendo una estrategia optimizada y prometedor para estas situaciones, (Shen, 2023).

6. Objetivos

a. **Objetivo General**

Construir un modelo predictivo que permita la asignación de las tareas a los equipos trabajo del sector de la tecnología de acuerdo a su capacidad y conocimientos

b. **Objetivos Específicos**

- Recolectar los datos disponibles de los equipos teniendo en cuenta las variables identificadas en la RSL de trabajo con los que cuenta la empresa.
- Realizar la preparación de los datos obtenidos aplicando normalización y limpieza de datos.
- Construir modelos predictivos evaluando su desempeño para elegir el modelo más óptimo.

7. Viabilidad

Los recursos necesarios para llevar a cabo esta intervención son un equipo de trabajo con conocimientos en aplicación de modelo de machine learning, un computador con una memoria RAM mínimo de 16 GB, acceso a herramientas de software libres para analítica de datos, acceso a las bases de datos de asignación de tareas Azure Devops y la información de las personas asignadas a los equipos de trabajo, herramientas ofimáticas Microsoft y conexión a internet mínimo de 100 megas.

El alcance de esta intervención es proponer y entregar un modelo para la empresa que permita asignar de forma óptima y eficiente las tareas dentro de los equipos del área de tecnología por medio de la identificación de variables que incidan en la distribución de dichas tareas de acuerdo a criterios establecidos como la capacidad, conocimientos y/o productividad de los equipos.

Las implicaciones que se generan en el desarrollo de esta intervención es una solicitud por parte del área jurídica sobre no revelar el nombre de la empresa para la cual entregaremos la propuesta, también hacer un buen uso de las bases de datos proporcionadas por la empresa teniendo en cuenta buenas prácticas de la seguridad de la información.

La propuesta que se entregará a la empresa tiene consecuencias positivas dado que una de las ventajas de los modelos predictivos es que busca aprovechar los datos históricos y variables relevantes para toma decisiones en cuanto a la asignación de tareas.

Con esta intervención la empresa tendrá disponible un método más eficiente para la asignación de las tareas lo que le permitirá tener una mayor productividad, reducción de costos y mejorar los tiempos de entrega de los proyectos comprometidos con el negocio.

Esta intervención permitirá a la empresa tener disponible una propuesta para asignar las tareas de una forma más precisa, optimizando la carga de trabajo de los equipos, disminuyendo los cuellos de botella y usando de forma más óptima los recursos disponibles.

8. Metodología

En este apartado se describen las fases de la Metodología CRISP-DM aplicadas en cada uno de los objetivos específicos para el desarrollo de este trabajo de grado, dando contexto sobre las actividades realizadas en cada una de las fases y como impactaron los objetivos específicos.

Metodología CRISP-DM:

- Fase I entendimiento del negocio.
- Fase II identificación de los datos.
- Fase III preparación de los datos.
- Fase IV modelado.
- Fase V evaluación.
- Fase VI despliegue.

Tabla 1 Fases CRISP-DM aplicadas en el objetivo 1.

Fase I entendimiento del negocio	En esta fase nos enfocamos en comprender el negocio, cuál es la misión de la organización, cuáles son los procesos de negocio impactados, entender cómo trabaja el equipo de tecnología y cuáles son los datos disponibles para abordar el desarrollo del proyecto.
---	---

Tabla 2 Fases CRISP-DM aplicadas en el objetivo 2.

<p>Fase II identificación de los datos</p>	<p>En esta fase se extraen los datos con los que se van a trabajar y se realiza un análisis exhaustivo sobre la calidad y la relevancia de los datos para responder al problema del negocio.</p>
<p>Fase III preparación de los datos</p>	<p>En esta fase se realiza la limpieza y transformación de los datos para obtener una información consistente, estandarizada y útil para construir los modelos.</p>

Tabla 3 Fases CRISP-DM aplicadas en el objetivo 3.

<p>Fase IV modelado</p>	<p>En esta fase se aplica el modelo de aprendizaje automático y se exploran diversas técnicas de aprendizaje supervisado de clasificación con el fin de encontrar las mejores métricas de acuerdo a las necesidades del negocio.</p>
<p>Fase V evaluación</p>	<p>En esta fase se realiza la evaluación de los modelos, comparando los resultados respecto a las métricas predeterminadas, buscando seleccionar el mejor modelo, en este punto se pueden tomar decisiones de regresar a etapas anteriores y verificar si se deben realizar ajustes en el modelado. La</p>

	<p>evaluación de los modelos se realiza en dos líneas, la primera con los hiper parámetros por defecto y luego con los mejores hiper parámetros encontrados a través del método Grid search.</p> <p>Las métricas utilizadas para la evaluación con el Recall, Accuracy y Precisión.</p>
<p>Fase VI despliegue</p>	<p>Esta fase nos enfocamos en el análisis de los resultados de los modelos de aprendizaje supervisado por clasificación que se implementaron, con el fin de entender los resultados de las métricas y seleccionar el modelo que más aporte valor de acuerdo al problema. Este proyecto no tiene como alcance la implementación, dado que se considera necesaria analizar con mayor profundidad los datos disponibles y recolectar la información de nuevas variables para que los resultados del modelo puedan satisfacer las necesidades del negocio.</p>

9. Resultados

En este apartado se describe de forma clara y concisa del paso a paso que se abordó para desarrollar y cumplir con el objetivo general y cada uno de los objetivos específicos.

A continuación, se detalla el desarrollo de cada uno de los objetivos específicos y las actividades realizadas en las fases de la metodología CRISP-DM.

Objetivo específico 1: Recolectar los datos disponibles de los equipos teniendo en cuenta las variables identificadas en la RSL de trabajo con los que cuenta la empresa.

En este objetivo se aplicó la Fase I Entendimiento del negocio.

Descripción de las actividades realizadas en la Fase I Entendimiento del negocio:

- **Entendimiento del negocio:**

Tabla 4 Entendimiento del negocio.

Descripción de la organización.	La organización para la cual estamos desarrollando este trabajo de grado es una empresa que desarrolla soluciones financieras digitales con una amplia oferta de servicios financieros y no financieros, que les permiten a los usuarios acceder a diferentes servicios de forma digital y ágil.
Misión de la organización.	Acompañar a los usuarios y clientes en la cotidianidad para que puedan lograr sus metas financieras y tener un aliado digital disponible para sus necesidades.
Procesos de negocio impactados.	Gestión de la tecnología y formas de trabajo

	de los equipos.
--	-----------------

- **Formas de trabajo del equipo de tecnología:**

El área de tecnología para la cual se está realizando esta intervención tiene a cargo los desarrollos de las soluciones financieras para los clientes y se divide en varios equipos de trabajo que reciben las necesidades del negocio de acuerdo a una planeación trimestral, la cual se divide en entregas quincenales por sprint para entregar valor de forma iterativa. El equipo a cargo de los desarrollos se divide en diferentes grupos de trabajo que tienen asignadas unas tareas para la ejecución y cierre dentro de cada sprint.

Los equipos de trabajo tienen una planeación definida e inician con la ejecución de tareas de acuerdo a dicha planeación, sin embargo, se ven afectados por el ingreso de nuevas tareas que son requerimientos no planeados y la atención de esta demanda es la que se busca mejorar con esta intervención.

Para la dinámica de trabajo en el área de tecnología se toma como referencia los marcos de trabajo ágiles como Scrum y Kanban con el fin de dividir el trabajo en diferentes entregables para entregas de valor continuas, actualmente los equipos trabajan en modalidad híbrida y se tiene una cultura de trabajo fresca, cómoda, aunque algo acelerada para dar cumplimiento a las necesidades del cliente y usuarios.

- **Revisión y análisis de los datos disponibles para el desarrollo de los modelos:**

Los datos disponibles para el desarrollo de este proyecto se presentan en la siguiente imagen, cada una de las variables contiene su explicación.

Columna	Descripción	Ejemplo
ID	Numero de identificación	1791206
Work Item Type	Nombre de la iteración	User Story
Tipo de Tarea	Nombre del grupo de tarea a ejecutar	Backend
Assigned To	Nombre de la persona asignada a la tarea	Edison Andres Guirales Arredondo
Señority	Nivel de cargo de la persona para la tarea	Junior
Modalidad	Modalidad o lugar donde trabaja	Hibrido
Genero	Tipo de Genero	M
State	Estado en el que se encuentra la tarea	New
Tags	Etiquetas que asignan a la tarea	Acompañamiento
Equipo_Trabajo	Nombre del equipo que atiende la tarea	Datos
OriginalEstimate	Horas que tarda la tarea	2
CompletedWork	Horas para completar la tarea	2
Nivel_Esfuerzo	Nivel de complejidad de la tarea	2
CreatedDate	Fecha de creación de la tarea	8/09/2020 12:05:06 p. m.
ClosedDate	Fecha de cierre de la tarea	29/09/2020 9:08:01 a. m.
Categoría HA	Nombre de categoría que se asigna a la tarea	Priority

Figura 2 Descripción de las variables disponibles.

Objetivo específico 2: Realizar la preparación de los datos obtenidos aplicando normalización y limpieza de datos.

En este objetivo se aplicó la Fase II identificación de los datos y la Fase III preparación de los datos.

Descripción de las actividades realizadas en la Fase II identificación de los datos:

- **Extracción y análisis de los datos por medio de Python:**

Se realiza la extracción de las bases de datos, desde la herramienta de Azure Devops donde se encuentra el listado de tareas que han atendido todos los equipos de tecnología y la segunda base de datos se tenía en Excel donde se encuentra el nivel de cargo, experiencia y modalidad de trabajo de las personas que ejecutaron las tareas. Ambas bases de datos se unificaron por medio de Excel y se importaron en formato de CSV para el cargue en Googlecolab y así poder iniciar con el análisis de los datos.

Para el análisis de los datos que tenemos cargados en CSV se utilizan estas funciones de

Python `d.info()` y `d.head()` que nos ayudan a conocer un poco mejor la información o los datos disponibles con los que vamos a trabajar en el proyecto.

```
d.info()
d.head(3)
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8617 entries, 0 to 8616
Data columns (total 17 columns):
#   Column                Non-Null Count  Dtype
---  -
0   ID                     8610 non-null   float64
1   Work Item Type        8610 non-null   object
2   Tipo_Tarea            8610 non-null   object
3   AssignedTo            8583 non-null   object
4   Anos_Experiencia     8610 non-null   object
5   Seniority             8610 non-null   object
6   Modalidad            8610 non-null   object
7   Genero                8610 non-null   object
8   State                 8610 non-null   object
9   Tags                  3125 non-null   object
10  Equipo_Trabajo       8610 non-null   object
11  OriginalEstimate     0 non-null     float64
12  CompletedWork        0 non-null     float64
13  Nivel_Esfuerzo      8610 non-null   object
14  CreatedDate          8610 non-null   object
15  ClosedDate           8610 non-null   object
16  Categoria_HA         28 non-null     object
dtypes: float64(3), object(14)
memory usage: 1.1+ MB
```

	ID	Work Item Type	Tipo_Tarea	AssignedTo	Anos_Experiencia	Seniority	Modalidad	Genero	State	Tags	Equipo_Trabajo
0	1825154.0	User Story	Front	Esteban Sierra Munera	1_a_2_anos	Semiseniior	Remoto	M	Closed	NaN	Desarrollo_Interno
1	1825170.0	User Story	Infraestructura	Esteban Sierra Munera	1_a_2_anos	Semiseniior	Remoto	M	Closed	NaN	Desarrollo_Interno
2	1828107.0	User Story	Analitica	Milena Correa Rivillas	1_a_2_anos	Semiseniior	Presencial	F	Closed	NaN	Desarrollo_Interno

Figura 3 Aplicación de funciones Python para análisis de datos.

Descripción de las actividades realizadas en la Fase III preparación de los datos:

- **Preparación de los datos usando Python:**

En esta actividad se realiza la preparación de datos para obtener una data más apropiada para construir los modelos, identificando cuales son las variables que nos agregan valor, verificando si existen datos nulos, si hay datos que deben ser imputados, entre otras transformaciones necesarias para dejar la data con calidad.

Se realiza la eliminación de variables que no aportan en el modelo, estas variables son retiradas del dataset dado que la información que contienen no agrega valor para el negocio ni para el problema estudiado en esta intervención.

Se realiza eliminación de datos nulos y valores de N/A que no aportan valor al modelo para obtener las variables finales para trabajar.

```

#Limpieza de datos
d.drop(['ID', 'Tags', 'State', 'work Item Type', 'originalEstimate', 'completedWork', 'CreatedDate', 'ClosedDate', 'Categoria_HA', 'AssignedTo'], axis='columns',
[16] dfinal.info()

<class 'pandas.core.frame.DataFrame'>
Index: 8583 entries, 0 to 8609
Data columns (total 7 columns):
#   Column                Non-Null Count  Dtype
---  ---
0   Tipo_Tarea            8583 non-null   object
1   Anos_Experiencia      8583 non-null   object
2   Seniority              8583 non-null   object
3   Modalidad              8583 non-null   object
4   Genero                 8583 non-null   object
5   Equipo_Trabajo        8583 non-null   object
6   Nivel_Esfuerzo        8583 non-null   object
dtypes: object(7)
memory usage: 536.4+ KB

```

Figura 4 Variables finales para trabajar los modelos.

Se aplica la función de Label_encoder para transformar los nombres de la variable objetivo a tipo numérico y proceder con la construcción de los modelos.

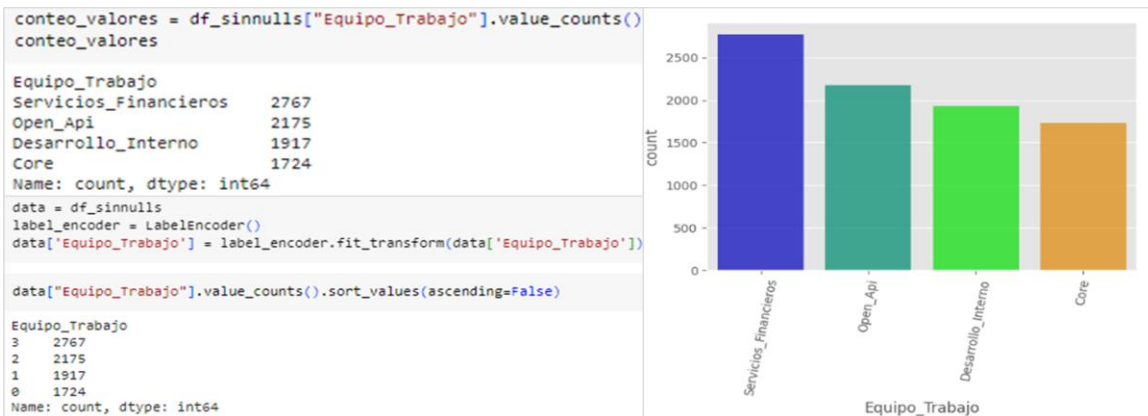


Figura 5 Representación de variable objetivo y transformación a variable numérica.

Objetivo específico 3: Construir modelos predictivos evaluando su desempeño para elegir el modelo óptimo.

En este objetivo se aplicó la Fase IV modelado, Fase V evaluación y Fase VI despliegue.

Descripción de las actividades realizadas en la fase IV modelado:

- **Realizar el Split para dividir los datos:**

En esta actividad se dividen los datos en un conjunto de entregamiento y evaluación para el modelo.

```
[30] #Split
      X_train, X_test, Y_train, Y_test = train_test_split(X, Y, test_size=0.3, random_state=77)
```

Figura 6 Aplicación del Split para los modelos.

- **Instanciar los modelos de clasificación que se utilizarán:**

En esta actividad se procede a instanciar los modelos de clasificación a construir Árbol de decisión, K-Vecinos Más Cercanos (KNN), Máquinas de Vectores de Soporte (SVM) y Regresión Logística.

```
Model1Tree = DecisionTreeClassifier()
```

```
Model2KNN = KNeighborsClassifier()
```

```
Model3SMV = SVC()
```

```
Model4RegLog = LogisticRegression()
```

Figura 7 Instanciado de modelos de clasificación.

- **Realizar entrenamiento de modelos de clasificación sin asignación de hiper parámetros:**

Usando la función .fit se realiza el entrenamiento de los modelos, en este ejercicio no se asignan hiper parámetros.

```
Model1Tree.fit(X_train, Y_train)
```

```
DecisionTreeClassifier  
DecisionTreeClassifier()
```

```
Model2KNN.fit(X_train, Y_train)
```

```
KNeighborsClassifier  
KNeighborsClassifier()
```

```
Model3SMV.fit(X_train, Y_train)
```

```
SVC  
SVC()
```

```
Model4RegLog.fit(X_train, Y_train)
```

```
LogisticRegression  
LogisticRegression()
```

Figura 8 Entrenamiento de modelos sin hiper parámetros.

- **Realizar predicción de los modelos sin hiper parámetros:**

Se aplica la función .Predict obteniendo los resultados de la predicción para cada uno de los modelos de clasificación.

```
Y_pred1 = Model1Tree.predict(X_test)  
Y_pred1
```

```
array([3, 2, 3, ..., 2, 1, 3])
```

```
Y_pred2 = Model2KNN.predict(X_test)  
Y_pred2
```

```
array([2, 2, 1, ..., 2, 3, 3])
```

```
Y_pred3 = Model3SMV.predict(X_test)  
Y_pred3
```

```
array([2, 2, 3, ..., 2, 1, 3])
```

```
Y_pred4 = Model4RegLog.predict(X_test)  
Y_pred4
```

```
array([0, 1, 2, ..., 2, 2, 2])
```

Figura 9 Predicción de modelos sin hiper parámetros.

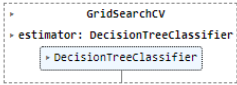
- **Aplicar GridSearch para encontrar los mejores hiper parámetros para cada uno de los modelos:**

Aplicación de GridSearch para modelo de Árbol de decisión encontrando los mejores hiper parámetros.

```

Model1tree_hiper = {'criterion':['gini','entropy','log_loss'],'min_samples_split':[2,5,10],'min_samples_leaf':[1, 2, 4], 'max_depth':[2,3,4,5,6,7,8,9,10,11,12,13,14],
Model1Tree_1 = DecisionTreeClassifier()
Model1Tree_1gvs = GridSearchCV(Model1Tree_1,Model1tree_hiper, cv=7)
Model1Tree_1gvs.fit(X_train, Y_train)

```



```

Model1Tree_1gvs.best_params_
{'criterion': 'entropy',
 'max_depth': 8,
 'min_samples_leaf': 4,
 'min_samples_split': 2}

```

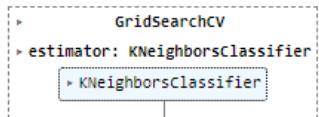
Figura 10 GridSearch modelo Árbol de decisión.

Aplicación de GridSearch para modelo de KNN encontrando los mejores hiper parámetros.

```

Model2KNN_hiper = {'n_neighbors':[3, 5, 7, 9, 10], 'metric':['euclidean', 'manhattan', 'chebyshev']}
Model2KNN_1 = KNeighborsClassifier()
Model2KNN_1gvs = GridSearchCV(Model2KNN_1,Model2KNN_hiper, cv=5)
Model2KNN_1gvs.fit(X_train, Y_train)

```



```

Model2KNN_1gvs.best_params_
{'metric': 'euclidean', 'n_neighbors': 9}

```

Figura 11 GridSearch modelo KNN.

Aplicación de GridSearch para modelo de SMV encontrando los mejores hiper parámetros.

```
Model3SMV_hiper = {'C': [0.1, 0.5, 1.0], 'kernel': ['linear', 'rbf'], 'gamma': ['scale', 'auto']}  
Model3SMV_1 = SVC()  
Model3SMV_1gvs = GridSearchCV(Model3SMV_1, Model3SMV_hiper, cv=5)  
Model3SMV_1gvs.fit(X_train, Y_train)  
  
└─ GridSearchCV  
  └─ estimator: SVC  
    └─ SVC  
  
Model3SMV_1gvs.best_params_  
  
{'C': 1.0, 'gamma': 'scale', 'kernel': 'rbf'}
```

Figura 12 GridSearch modelo SVM.

Aplicación de GridSearch para modelo de Regresión logística encontrando los mejores hiper parámetros.

```
Model3LR_hiper = {'C': [0.1, 0.5, 1.0], 'penalty': ['l1', 'l2'], 'solver': ['liblinear', 'saga']}  
Model3LR_1 = LogisticRegression()  
Model3LR_1gvs = GridSearchCV(Model3LR_1, Model3LR_hiper, cv=5)  
Model3LR_1gvs.fit(X_train, Y_train)  
  
└─ GridSearchCV  
  └─ estimator: LogisticRegression  
    └─ LogisticRegression  
  
Model3LR_1gvs.best_params_  
  
{'C': 0.5, 'penalty': 'l1', 'solver': 'saga'}
```

Figura 13 GridSearch modelo Regresión Logística.

- **Realizar entrenamiento y predicción para cada modelo con los mejores hiper parámetros:**

Luego de aplicar el Gridsearch para cada uno de los modelos de clasificación se realiza nuevamente el entrenamiento y la predicción de los modelos utilizando los mejores hiper parámetros.

Entrenamiento y predicción para Árbol de decisión.

```
best_params_Tree = {'criterion': 'entropy',  
                    'max_depth': 8,  
                    'min_samples_leaf': 4,  
                    'min_samples_split': 2,}
```

```
BestModeloTree = DecisionTreeClassifier(**best_params_Tree)  
BestModeloTree.fit(X_train, Y_train)
```

```
DecisionTreeClassifier  
DecisionTreeClassifier(max_depth=8, min_samples_leaf=4)
```

```
Y_pred_BestModelTree = BestModeloTree.predict(X_test)  
Y_pred_BestModelTree
```

```
array([2, 2, 3, ..., 2, 1, 3])
```

Figura 14 Entrenamiento y predicción mejores parámetros Árbol de decisión.

Entrenamiento y predicción para KNN.

```
best_params_KNN = {'metric': 'euclidean', 'n_neighbors': 9}
```

```
BestModeloKNN = KNeighborsClassifier(**best_params_KNN)  
BestModeloKNN.fit(X_train, Y_train)
```

```
KNeighborsClassifier  
KNeighborsClassifier(metric='euclidean', n_neighbors=9)
```

```
Y_pred_BestModelKNN = BestModeloKNN.predict(X_test)  
Y_pred_BestModelKNN
```

```
array([2, 2, 3, ..., 2, 1, 3])
```

Figura 15 Entrenamiento y predicción mejores parámetros KNN.

Entrenamiento y predicción para SMV.

```
best_params_SMV = {'C': 1.0, 'gamma': 'scale', 'kernel': 'rbf'}
```

```
BestModeloSMV = SVC(**best_params_SMV)  
BestModeloSMV.fit(X_train, Y_train)
```

```
▾ SVC  
SVC()
```

```
Y_pred_BestModelSMV = BestModeloSMV.predict(X_test)  
Y_pred_BestModelSMV
```

```
array([2, 2, 3, ..., 2, 1, 3])
```

Figura 16 Entrenamiento y predicción mejores parámetros SVM.

Entrenamiento y predicción para Regresión logística.

```
best_params_RL = {'C': 0.5, 'penalty': 'l1', 'solver': 'saga'}
```

```
BestModeloRL = LogisticRegression(**best_params_RL)  
BestModeloRL.fit(X_train, Y_train)
```

```
▾ LogisticRegression  
LogisticRegression(C=0.5, penalty='l1', solver='saga')
```

```
Y_pred_BestModelRL = BestModeloRL.predict(X_test)  
Y_pred_BestModelRL
```

```
array([0, 1, 3, ..., 2, 2, 2])
```

Figura 17 Entrenamiento y predicción mejores parámetros Regresión logística.

Descripción de las actividades realizadas en la fase V evaluación:

- **Realizar evaluación de los modelos sin hiper parámetros:**

Para la evaluación de los modelos se utilizan las métricas Accuracy, Recall y Precisión obteniendo estos resultados:

Accuracy: esta métrica mide la proporción de predicciones correctas respecto al total de

predicciones realizadas.

Recall (Sensibilidad): es la capacidad de un modelo de clasificación para identificar correctamente todos los casos positivos.

Precisión: esta métrica mide la capacidad de un modelo para evitar clasificar incorrectamente instancias negativas como positivas.

Evaluación con la métrica Accuracy.

```
accuracy1 = accuracy_score(Y_test, Y_pred1)
accuracy1
0.5277669902912622
```

```
accuracy2 = accuracy_score(Y_test, Y_pred2)
accuracy2
0.4679611650485437
```

```
accuracy3 = accuracy_score(Y_test, Y_pred3)
accuracy3
0.5386407766990291
```

```
accuracy4 = accuracy_score(Y_test, Y_pred4)
accuracy4
0.3533980582524272
```

Figura 18 Resultados Accuracy de los modelos sin hiper parámetros.

Evaluación con la métrica Recall.

```
Recall1 = recall_score(Y_test, Y_pred1, average='macro')
Recall1
0.504711676286684
```

```
Recall2 = recall_score(Y_test, Y_pred2, average='macro')
Recall2
0.46269602999386117
```

```
Recall3 = recall_score(Y_test, Y_pred3, average='macro')
Recall3
0.510385737017834
```

```
Recall4 = recall_score(Y_test, Y_pred4, average='macro')
Recall4
0.31784007360882927
```

Figura 19 Resultados Recall de los modelos sin hiper parámetros.

Evaluación con la métrica Precisión.

```
Precision1 = precision_score(Y_test, Y_pred1, average='macro')  
Precision1
```

```
0.531986495769593
```

```
Precision2= precision_score(Y_test, Y_pred2, average='macro')  
Precision2
```

```
0.46212197383265646
```

```
Precision3 = precision_score(Y_test, Y_pred3, average='macro')  
Precision3
```

```
0.5624489496200621
```

```
Precision4 = precision_score(Y_test, Y_pred4, average='macro')  
Precision4
```

```
0.3386101959974642
```

Figura 20 Resultados Precisión de los modelos sin hiper parámetros.

- **Realizar evaluación de los modelos con los mejores hiper parámetros:**

En esta actividad se realiza nuevamente la evaluación de los modelos utilizando los mejores hiper parámetros y las métricas Accuracy, Recall y Precisión obteniendo estos resultados:

Resultado de evaluación para modelo Árbol de decisión.

```
accuracyarbol = accuracy_score(Y_test, Y_pred_BestModelTree)  
accuracyarbol
```

```
0.5370873786407767
```

```
recallarbol = recall_score(Y_test, Y_pred_BestModelTree, average='macro')  
recallarbol
```

```
0.5087569356347893
```

```
precisionarbol = precision_score(Y_test, Y_pred_BestModelTree, average='macro')  
precisionarbol
```

```
0.5562118223583548
```

Figura 21 Resultados evaluación final modelo Árbol de decisión.

Resultado de evaluación para modelo KNN.

```
accuracyKNN = accuracy_score(Y_test, Y_pred_BestModelKNN)  
accuracyKNN
```

```
0.48854368932038833
```

```
recallKNN = recall_score(Y_test, Y_pred_BestModelKNN, average='macro')  
recallKNN
```

```
0.4710983547572617
```

```
precisionKNN = precision_score(Y_test, Y_pred_BestModelKNN, average='macro')  
precisionKNN
```

```
0.4811624928606878
```

Figura 22 Resultados evaluación final modelo KNN.

Resultado de evaluación para modelo SVM.

```
accuracySVM = accuracy_score(Y_test, Y_pred_BestModelSMV)  
accuracySVM
```

```
0.5386407766990291
```

```
recallSVM = recall_score(Y_test, Y_pred_BestModelSMV, average='macro')  
recallSVM
```

```
0.510385737017834
```

```
precisionSMV = precision_score(Y_test, Y_pred_BestModelSMV, average='macro')  
precisionSMV
```

```
0.5624489496200621
```

Figura 23 Resultados evaluación final modelo SVM.

Resultado de evaluación para modelo Regresión logística.

```
accuracyRL = accuracy_score(Y_test, Y_pred_BestModelRL)
accuracyRL
```

0.3561165048543689

```
recallRL = recall_score(Y_test, Y_pred_BestModelRL, average='macro')
recallRL
```

0.31870515589655823

```
precisionRL = precision_score(Y_test, Y_pred_BestModelRL, average='macro')
precisionRL
```

0.33727627866508664

Figura 24 Resultados evaluación final modelo Regresión logística.

Descripción de las actividades realizadas en la fase VI despliegue:

- **Análisis de los resultados luego de la evaluación de los modelos**

En esta actividad analizamos los resultados de evaluación de los modelos de aprendizaje supervisado construidos, donde se logra evidenciar en la siguiente tabla los valores de las métricas obtenidas sin y con hiper parámetros.

Tabla 5 Resultado métricas de los modelos de clasificación.

Modelo	Métricas sin hiper parámetros			Métricas con hiper parámetros		
	Accuracy	Recall	Precisión	Accuracy	Recall	Precisión
Árbol de decisión	0,52	0,50	0,53	0,53	0,50	0,55
KNN	0,46	0,46	0,46	0,48	0,47	0,48
SMV	0,53	0,51	0,56	0,53	0,51	0,56
Regresión logística	0,35	0,31	0,33	0,35	0,31	0,31

En esta actividad se descartan los resultados del modelo de Regresión logística dado que los resultados son menores frente a los demás modelos.

Finalmente, analizando los resultados de las métricas se evidencia que el mejor modelo es el de Máquina de vector de soporte y con este procedimos a construir la matriz de confusión.

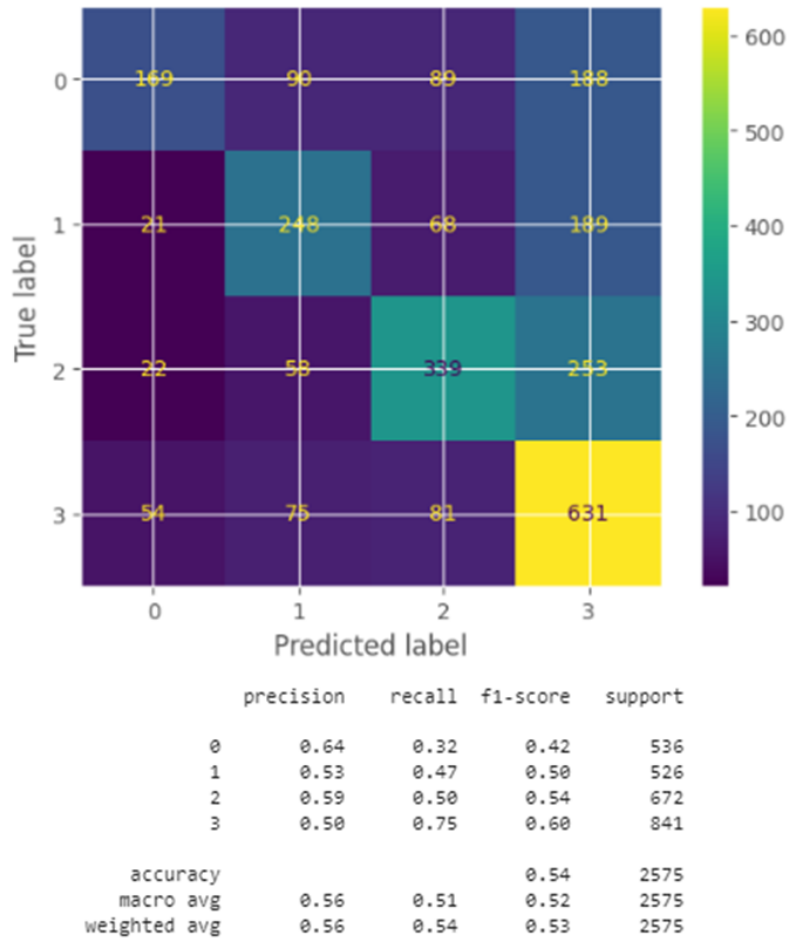


Figura 25 Matriz de confusión SMV

A partir de la matriz de confusión podemos analizar los siguientes puntos como resultado del modelo de SMV.

1. Observamos que el modelo SVM tuvo un buen rendimiento en la clasificación de las clases 0, 1, 2 y 3, ya que la mayoría de los casos fueron correctamente clasificados como se evidencia en la diagonal principal de la matriz de confusión.
2. En cuanto a los errores clasificación también observamos en la matriz fuera de la diagonal principal, por ejemplo, en la intersección de la fila 0 y la columna 1 el modelo clasificó incorrectamente 90 instancias que realmente pertenecían a la clase 0 poniéndola como clase 1 y lo mismo ocurre en otras intersecciones, lo que indica que

el modelo no es perfecto y comete errores en la predicción.

3. Para analizar más a detalle el desempeño del modelo se debe tomar cada clase por separado dado que esto nos permitirá identificar qué clases son más difíciles de predecir y pueden requerir un análisis y ajuste adicionales. Por ejemplo, la clase 2 parece ser la más difícil de predecir, ya que tiene el mayor número de falsos positivos y falsos negativos en comparación con las demás.

Luego de la implementación de todos los modelos mencionados, se logra evidenciar a través de la métrica Accuracy del modelo SMV que el 53% de las predicciones realizadas por este modelo son correctas y es un resultado que nos invita a continuar en un refinamiento de los datos utilizados para este proyecto, analizar a detalle las necesidades del negocio y potenciar los insumos para lograr unos mejores resultados.

10. Conclusiones

Durante el desarrollo de este trabajo de grado, se cumplieron los tres objetivos establecidos, se logró recolectar los datos de los equipos de trabajo para realizar la intervención y preparar los datos aplicando técnicas de normalización y limpieza de datos, incluyendo la identificación y corrección de valores atípicos, la gestión de datos faltantes y la normalización de variables para garantizar la consistencia y calidad del conjunto de datos, logrando finalmente construir modelos predictivos utilizando técnicas de aprendizaje supervisado por clasificación, evaluando su desempeño a través de las métricas como la precisión, el recall y la exactitud, lo que nos permitió comparar y analizar diferentes enfoques de modelado y seleccionar aquel que demostró el mejor rendimiento en la predicción de la asignación de tareas a los equipos de trabajo de tecnología.

Con los resultados obtenidos de los modelos predictivos podemos concluir que es necesario robustecer los datos usados en esta intervención con el fin de mejorar el rendimiento de dichos modelos, la calidad y cantidad de datos son fundamentales por lo cual es necesario recopilar más datos relevantes y detallados sobre los miembros del equipo, como por ejemplo las habilidades, experiencias pasadas, la capacidad de los miembros del equipo para aprender nuevas tecnologías, factores emocionales como el estado de ánimo de los miembros del equipo, las preferencias individuales de los miembros del equipo en términos de los tipos de proyectos o tecnologías, así mismo nuevas variables sobre las tareas y equipos de trabajo para mejorar la capacidad del modelo para hacer predicciones precisas por ejemplo la disponibilidad de los recursos y de cada miembro del equipo, la carga de trabajo actual de los equipos, las necesidades específicas del proyecto o tareas como el tiempo de entrega, el alcance del trabajo y las tecnologías requeridas.

Finalmente podemos concluir que el desarrollo de modelos de machine learning es un proceso iterativo y de mejora continua. A medida que recopilas más datos y obtienes retroalimentación

del negocio y del rendimiento de los modelos se puede continuar iterando y mejorando para que sea más precisa y efectiva la asignación de tareas a los equipos de trabajo.

11. Recomendaciones

En las investigaciones actuales sobre asignación de tareas a los equipos de trabajo se evidencia que están enfocadas en distribuir las tareas a los equipos de trabajo sin tener en cuenta factores como costos del proyecto. Por lo cual se sugiere para estudios futuros considerar factores económicos en cuanto a costo de los proyectos o actividades a realizar para tener unos modelos predictivos con alcance óptimo de recursos financieros.

Además, consideramos relevante la inclusión de un factor de compatibilidad en nuestro modelo. Este factor se refiere a la capacidad del equipo para trabajar de manera eficaz con el cliente interno. Al tomar en cuenta la compatibilidad, el modelo se entrena de manera más precisa, lo que resulta en una asignación de tareas más efectiva.

Asimismo, para futuros estudios, es apropiado considerar factores emocionales, incluir el estado de ánimo de los empleados al desarrollar modelos predictivos no solo es un imperativo ético, sino también una estrategia inteligente. Las empresas que incorporan este factor tienen la oportunidad de construir equipos más comprometidos, productivos y resilientes, lo que contribuye al éxito a largo plazo de la organización. Al integrar el componente humano en los modelos predictivos, las empresas están mejor preparadas para identificar factores emocionales que puedan influir en el ambiente laboral y en la productividad en general. De esta manera, pueden tomar medidas proactivas para mejorar el bienestar laboral y optimizar aún más la asignación de tareas a los equipos de trabajo

12. Referencias

Aguilar-Rodríguez, I. E., Bernal-Torres, C. A., Artieda-Cajilema, C. H., & Tapia-Andino, G. F. (2023). Smart working and base technologies in corporate performance: New directions in emerging firms. *Asia Pacific Management Review*, 28(3), 358–369. <https://doi.org/10.1016/j.apmr.2022.12.008>

Al Ramahi, A. M., Davies, A., & Al Kaabi, F. (2023). Measuring the impact of intellectual capital performance management indicators through the lens of organizational capital: A Police Agency case study. *Policing: A Journal of Policy and Practice*, 17. <https://doi.org/10.1093/police/paad050>

Atikur Rahaman, Md., Dilip Taru, R., Gupta, A., Prajapat, V., & Abdul Latif Mahmud, Md. (2023).

Factors influencing employee performance and their impact on productivity: A study of commercial banks in Bangladesh. *Banks and Bank Systems*, 18(2), 127–136. [https://doi.org/10.21511/bbs.18\(2\).2023.11](https://doi.org/10.21511/bbs.18(2).2023.11)

Beltran Beatriz, (s.f.). La disciplina denominada Minería de Datos. <http://bbeltran.cs.buap.mx/NotasMD.pdf>

Baigi, K., & Stewart, W. F. (2015). Headache and migraine. a leading cause of absenteeism. *Handbook of Clinical Neurology*, 131, 447–463. <https://doi.org/10.1016/B978-0-444-62627-1.00025-1>

Becker, J., Knackstedt, R., & Pöppelbuß, J. (2009). Developing Maturity Models for IT Management. *Business & Information Systems Engineering*, 1(3), 213–222. <https://doi.org/10.1007/s12599-009-0044-5>

Chang, V., Mou, Y., Xu, Q. A., & Xu, Y. (2022). Job satisfaction and turnover decision of employees in the Internet sector in the US. *Enterprise Information Systems*. <https://doi.org/10.1080/17517575.2022.2130013>

Cozzani, V., Antonioni, G., Khakzad, N., Khan, F., Taveau, J., & Reniers, G. (2013). Quantitative Assessment of Risk Caused by Domino Accidents. *Domino Effects in the Process Industries: Modelling, Prevention and Managing*, 208–228. <https://doi.org/10.1016/B978-0-444-54323-3.00010-5>

Datascientest, (2023). ¿Qué es un algoritmo?. <https://datascientest.com/es/que-es-un-algoritmo> Fields, D., & Bocarnea, M. C. (2004).

Organizational Behavior. *Encyclopedia of Social Measurement*, 951–956. <https://doi.org/10.1016/B0-12-369398-5/00528-4>

Gurubhagavatula, I., Barger, L. K., Barnes, C. M., Basner, M., Boivin, D. B., Dawson, D., Drake, C. L., Flynn-Evans, E. E., Mysliwiec, V., Daniel Patterson, P., Reid, K. J., Samuels, C., Shattuck, N. L., Kazmi, U., Carandang, G., Heald, J. L., & Van Dongen, H. P. A. (2021).

Guiding principles for determining work shift duration and addressing the effects of work shift duration on performance, safety, and health: Guidance from the American Academy of Sleep Medicine and the Sleep Research Society. *Sleep*, 44(11). <https://doi.org/10.1093/sleep/zsab161>

Hwang, Y., Al-Arabi, M., Rouibah, K., & Chung, J. Y. (2016). Toward an integrative view for the leader-member exchange of system implementation. *International Journal of Information Management*, 36(6), 976–986. <https://doi.org/10.1016/j.ijinfomgt.2016.06.008>

Iberdrola, (s.f.). ¿Qué es el 'machine learning' y para qué sirve?. [https://www.iberdrola.com/innovacion/machine-learning-aprendizaje-automatico#:~:text=El%20Machine%20Learning%20es%20una,elaborar%20predicciones%20\(an%C3%A1lisis%20predictivo\)](https://www.iberdrola.com/innovacion/machine-learning-aprendizaje-automatico#:~:text=El%20Machine%20Learning%20es%20una,elaborar%20predicciones%20(an%C3%A1lisis%20predictivo))

IBM, (s.f.). ¿Qué es Business Intelligence?. <https://www.ibm.com/mx-es/topics/business-intelligence>

IBM, (s.f.). ETL (Extract, Transform, Load). <https://www.ibm.com/topics/etl>

IBM, (s.f.). What are virtual machines (VMs)?. <https://www.ibm.com/topics/virtual-Machines>

IBM, (2023). Direct Access Storage Devices (DASDs). <https://www.ibm.com/docs/en/aix/7.2?topic=subsystem-direct-access-storage-devices-dasds>

Indarapu, S. R. K., Vodithala, S., Kumar, N., Kiran, S., Reddy, S. N., & Dorthi, K. (2023). Exploring human resource management intelligence practices using machine learning models. *Journal of High Technology Management Research*, 34(2). <https://doi.org/10.1016/j.hitech.2023.100466>

Khan, S. M., & Abbas, J. (2022). Mindfulness and happiness and their impact on employee creative performance: Mediating role of creative process engagement. *Thinking Skills and Creativity*, 44. <https://doi.org/10.1016/j.tsc.2022.101027>

Kisielnicki, J., & Misiak, A. M. (2017). EFFECTIVENESS of AGILE COMPARED to WATERFALL IMPLEMENTATION METHODS in IT PROJECTS: ANALYSIS BASED on BUSINESS INTELLIGENCE PROJECTS. *Foundations of Management*, 9(1), 273–286. <https://doi.org/10.1515/fman-2017-0021>

Limpo, L., & Junaidi, J. (2022). Influence of empowering and ethical leadership on employees' job satisfaction, performance, and organization commitment. *Humanities and Social Sciences Letters*, 11(1), 22–36. <https://doi.org/10.18488/73.v11i1.3241>

Loyarte-López, E., & García-Olaizola, I. (2022). Machine Learning Based Method for Deciding Internal Value of Talent. *Applied Artificial Intelligence*, 36(1). <https://doi.org/10.1080/08839514.2022.2151160>

Machado, G. V., Cunha, Í., Pereira, A. C. M., & Oliveira, L. B. (2019a). DOD-ETL: distributed on-demand ETL for near real-time business intelligence. *Journal of Internet Services and Applications*, 10(1), 1–15. <https://doi.org/10.1186/S13174-019-0121-Z/FIGURES/8>

Machado, G. V., Cunha, Í., Pereira, A. C. M., & Oliveira, L. B. (2019b). DOD-ETL: distributed on-demand ETL for near real-time business intelligence. *Journal of Internet Services and Applications*, 10(1). <https://doi.org/10.1186/s13174-019-0121-z>

Maghsoudi, M., & Nezafati, N. (2023). Navigating the acceptance of implementing business intelligence in organizations: A system dynamics approach. *Telematics and Informatics Reports*, 11. <https://doi.org/10.1016/j.teler.2023.100070>

Moayedikia, A., Ghaderi, H., & Yeoh, W. (2020). Optimizing microtask assignment on crowdsourcing platforms using Markov chain Monte Carlo. *Decision Support Systems*, 139. <https://doi.org/10.1016/j.dss.2020.113404>

Nvidia, (s.f.). What Is Stream Processing?. <https://www.nvidia.com/en-us/glossary/data-science/stream-processing/>

Nundlall, C., & Nagowah, S. D. (2022). Task allocation and coordination process in distributed agile software development: an ontology based approach. *Information*

Technology and Management, 23(3), 167–192. <https://doi.org/10.1007/s10799-022-00365-9>

Nyathi, M., & Kekwaletswe, R. (2023). Realizing employee and organizational performance gains through electronic human resource management use in developing countries. *African Journal of Economic and Management Studies*, 14(1), 121–134. <https://doi.org/10.1108/AJEMS-11-2021-0489>

Olier David. (2023, August). *Smart Working: What is it?*
<https://pandorafms.com/blog/smart-working/>

Ontotext, (s.f.). What are Ontologies?.
<https://www.ontotext.com/knowledgehub/fundamentals/what-are-ontologies/#:~:text=An%20ontology%20is%20a%20formal,better%20sense%20of%20their>

Pap, J., Mako, C., Illessy, M., Kis, N., & Mosavi, A. (2022b). Modeling Organizational Performance with Machine Learning. *Journal of Open Innovation: Technology, Market, and Complexity*, 8(4). <https://doi.org/10.3390/joitmc8040177>

Parra, X., Tort-Martorell, X., Alvarez-Gomez, F., & Ruiz-Viñals, C. (2022). Chronological Evolution of the Information-Driven Decision-Making Process (1950–2020). *Journal of the Knowledge Economy*. <https://doi.org/10.1007/s13132-022-00917-y>

Peoplehum, (s.f.). What are Employee Incentives?.
<https://www.peoplehum.com/glossary/incentives>

Ragazou, K., Passas, I., Garefalakis, A., & Zopounidis, C. (2023). Business intelligence model empowering SMEs to make better decisions and enhance their competitive advantage. *Discover Analytics*, 1(1). <https://doi.org/10.1007/s44257-022-00002-3>

Ramachandran, K. K., Apsara Saleth Mary, A., Hawladar, S., Asokk, D., Bhaskar, B., & Pitroda, J. R. (2022). Machine learning and role of artificial intelligence in optimizing work performance and employee behavior. *Materials Today: Proceedings*, 51, 2327–2331. <https://doi.org/10.1016/j.matpr.2021.11.544>

Rao Pabolu, V. K., Shrivastava, D., & Kulkarni, M. S. (2022). A Dynamic System to Predict an Assembly Line Worker's Comfortable Work-Duration Time by Using the Machine Learning Technique. *Procedia CIRP*, 106, 270–275. <https://doi.org/10.1016/j.procir.2022.02.190>

Rasul, I., Rogger, D., & Williams, M. J. (2021). Management, Organizational Performance, and Task Clarity: Evidence from Ghana's Civil Service. *Journal of Public Administration Research and Theory*, 31(2), 259–277. <https://doi.org/10.1093/jopart/muaa034>

Rimi, N. N., Rubel, M. R. B., & Kee, D. M. H. (2023). QUALITY OF WORK LIFE AND EMPLOYEE WORK OUTCOMES: A HIERARCHICAL MODEL WITH MEDIATION ANALYSIS. *International Journal of Business and Society*, 24(1), 421–439. <https://doi.org/10.33736/ijbs.5625.2023>

Sadatnya, A., Sadeghi, N., Sabzekar, S., Khanjani, M., Tak, A. N., & Taghaddos, H. (2023). Machine learning for construction crew productivity prediction using daily work reports. *Automation in Construction*, 152. <https://doi.org/10.1016/j.autcon.2023.104891>

Santhosh, R., & Mohanapriya, M. (2021). Generalized fuzzy logic based performance prediction in data mining. *Materials Today: Proceedings*, 45, 1770–1774. <https://doi.org/10.1016/j.matpr.2020.08.626>

Sharma, V., & Bala, M. (2020). An Improved Task Allocation Strategy in Cloud using

Modified K-means Clustering Technique. *Egyptian Informatics Journal*, 21(4), 201–208.

<https://doi.org/10.1016/j.eij.2020.02.001>

Tan, Y. (2014). Risk management and performance in the Chinese banking sector.

Performance, Risk and Competition in the Chinese Banking Industry, 65–139.

<https://doi.org/10.1533/9781780634463.65>

Tarigan, J., Cahya, J., Valentine, A., Hatane, S., & Jie, F. (2022). Total reward system, job

satisfaction and employee productivity on company financial performance: evidence from

Indonesian Generation Z workers. *Journal of Asia Business Studies*, 16(6), 1041–1065.

<https://doi.org/10.1108/JABS-04-2021-0154>

Tecimer, K. A., Tüzün, E., Moran, C., & Erdogmus, H. (2022). Cleaning ground truth data in

software task assignment. *Information and Software Technology*, 149.

<https://doi.org/10.1016/j.infsof.2022.106956>

Wuest, T., Weimer, D., Irgens, C., & Thoben, K. D. (2016). Machine learning in

manufacturing: Advantages, challenges, and applications. *Production and Manufacturing*

Research, 4(1), 23–45. <https://doi.org/10.1080/21693277.2016.1192517>

Deng, H., Runger, G., & Tuv, E. (2020). A Comprehensive Review of Decision Tree Learning.

IEEE Transactions on Knowledge and Data Engineering, 33(2), 356-372.