



Determinación de técnicas de aprendizaje supervisado para la optimización de
procesos industriales mediante algoritmos de clasificación

Modalidad: Proyecto de Investigación

Alex Giancarlo Figueroa Torres
1095917077

Nicole Stephannie García Ardila
1005150667

UNIDADES TECNOLÓGICAS DE SANTANDER
Facultad de Ciencias Naturales e Ingenierías
Ingeniería Industrial
Bucaramanga (14 de septiembre de 2024)



Determinación de técnicas de aprendizaje supervisado para la optimización de procesos industriales mediante algoritmos de clasificación

Proyecto de Investigación

Alex Giancarlo Figueroa Torres
1095917077

Nicole Stephannie García Ardila
1005150667

Trabajo de Grado para optar al título de
Ingeniero Industrial

DIRECTOR

Victor Alfonso Sanabria Ruiz

Grupo de Investigación de Soluciones Operativas, Logísticas y Desarrollo
Organizacional – SOLYDO

UNIDADES TECNOLÓGICAS DE SANTANDER
Facultad de Ciencias Naturales e Ingenierías
Ingeniería Industrial
Bucaramanga (14 de septiembre de 2024)

F-DC-125

INFORME FINAL DE TRABAJO DE GRADO EN MODALIDAD DE
PROYECTO DE INVESTIGACIÓN, DESARROLLO TECNOLÓGICO,
MONOGRAFÍA, EMPRENDIMIENTO Y SEMINARIO

VERSIÓN: 2.0

Nota de Aceptación

Firma del Evaluador

Firma del Director

DEDICATORIA

Quiero dedicar este proyecto a mi madre, Gloria María Torres, y a mis hermanas, quienes han sido el soporte esencial en mi vida. Agradezco profundamente su constante esfuerzo y apoyo, los cuales han sido cruciales para mi progreso. Asimismo, deseo dedicar este trabajo a mi esposa, Lizeth Gómez, quien siempre ha estado a mi lado, acompañándome en este importante logro.

Alex Figueroa

Este proyecto está dedicado a mis padres, a mis abuelas y a mi hermano. Ellos son la principal razón detrás de mi deseo de crecer como profesional y siempre han estado a mi lado, brindándome apoyo y ánimo constante para alcanzar este objetivo

Nicole García

AGRADECIMIENTOS

Primero, quiero manifestar mi profunda gratitud a Dios por darme la oportunidad de finalizar mis estudios en esta prestigiosa universidad. Asimismo, estoy muy agradecido con todos los profesores de mi carrera, quienes me ofrecieron su sabiduría y apoyo, contribuyendo significativamente a mi desarrollo como ingeniero.

Alex Figueroa

Mi más profundo agradecimiento va dirigido a los profesores que dedicaron su tiempo y esfuerzo para transmitir sus conocimientos, permitiéndome concluir mi carrera en ingeniería. Igualmente, estoy agradecido a Dios por su apoyo y por hacer posible el cumplimiento de esta importante meta y la finalización de mis estudios.

Nicole García

TABLA DE CONTENIDO

RESUMEN EJECUTIVO.....	10
INTRODUCCIÓN.....	12
1. DESCRIPCIÓN DEL TRABAJO DE INVESTIGACIÓN	14
1.1. PLANTEAMIENTO DEL PROBLEMA	14
1.2. JUSTIFICACIÓN.....	15
1.3. OBJETIVOS	16
1.3.1. OBJETIVO GENERAL	16
1.3.2. OBJETIVOS ESPECÍFICOS.....	16
1.4. ESTADO DEL ARTE.....	17
2. MARCO REFERENCIAL	21
2.1. MARCO TEÓRICO	21
2.1.1. APRENDIZAJE SUPERVISADO.....	21
2.1.2. TÉCNICAS DE APRENDIZAJE SUPERVISADO	23
2.1.3. ÁREAS DE APLICACIÓN DEL APRENDIZAJE SUPERVISADO EN LA INDUSTRIA	24
2.1.4. MODELOS PREDICTIVO	26
2.2. MARCO CONCEPTUAL	26
2.3. MARCO LEGAL.....	29
3. DISEÑO DE LA INVESTIGACIÓN.....	30
3.1. TIPO DE INVESTIGACIÓN	30
3.1.1. ENFOQUE	30
3.1.2. MÉTODO.....	31
3.2. FASES DEL TRABAJO DE GRADO	32
4. DESARROLLO DEL TRABAJO DE GRADO	33
4.1. REVISIÓN DOCUMENTAL DE APLICACIONES DE APRENDIZAJE SUPERVISADO EN LA INGENIERIA INDUSTRIAL	33
4.2. COMPARAR LOS DIFERENTES ALGORITMOS DE APRENDIZAJE SUPERVISADO PARA MODELADO DE PROCESOS INDUSTRIALES	34
4.3. INTERPRETACIÓN DE LOS MODELOS INDUSTRIALES DE APRENDIZAJE SUPERVISADO.....	36

F-DC-125

INFORME FINAL DE TRABAJO DE GRADO EN MODALIDAD DE
PROYECTO DE INVESTIGACIÓN, DESARROLLO TECNOLÓGICO,
MONOGRAFÍA, EMPRENDIMIENTO Y SEMINARIO

VERSIÓN: 2.0

4.3.1.	SELECCIÓN DE SOFTWARE DE REGRESIÓN LINEAL	36
4.3.2.	MODELO DE KNN	37
5.	<u>RESULTADOS</u>	<u>39</u>
5.1.	APLICACIONES DE APRENDIZAJE SUPERVISADO EN LA INDUSTRIA	39
5.1.1.	COMPENDIO INVESTIGACIONES RELACIONADAS CON EL APRENDIZAJE SUPERVISADO EN LA INDUSTRIA	39
5.1.2.	ANÁLISIS DE TRABAJOS IDENTIFICADOS	47
5.2.	COMPARACIÓN DE LOS DIFERENTES ALGORITMOS DE APRENDIZAJE SUPERVISADO.....	56
5.2.1.	IDENTIFICACIÓN DE ALGORITMOS DE APRENDIZAJE SUPERVISADO EN LA INDUSTRIA	56
5.2.2.	CARACTERÍSTICAS DE LOS MODELOS DE APRENDIZAJE SUPERVISADO	59
5.2.3.	VENTAJAS Y DESVENTAJAS DE LOS ALGORITMOS DE APRENDIZAJE SUPERVISADO...	60
5.2.4.	CRITERIOS Y EVALUACIÓN DE IDONEIDAD DE ALGORITMOS DE CLASIFICACIÓN DE APRENDIZAJE SUPERVISADO	61
5.3.	INTERPRETACIÓN DE MODELOS PARA PROCESOS INDUSTRIALES DE APRENDIZAJE SUPERVISADO	66
5.3.1.	SOFTWARE RESTUDIO	66
5.3.2.	CARACTERÍSTICAS DE RESTUDIO.....	66
5.3.3.	FUNCIONALIDAD	66
5.3.4.	VENTAJAS.....	67
5.3.5.	REGRESIÓN LINEAL	67
5.3.6.	MODELO KNN	78
6.	<u>CONCLUSIONES</u>	<u>88</u>
7.	<u>RECOMENDACIONES.....</u>	<u>90</u>
8.	<u>REFERENCIAS BIBLIOGRÁFICAS</u>	<u>92</u>

LISTA DE FIGURAS

Figura 1 Tipos de Modelo Aprendizaje Supervisado	22
Figura 2. Fases de la investigación.....	32
Figura 3 Proceso de Revisión Literaria de Aprendizaje Profundo	33
Figura 4 Proceso de Comparación y Evaluación de los Algoritmos de Aprendizaje Supervisado en la Industria.....	34
Figura 5 Selección de Software de Regresión Lineal.....	36
Figura 6 Implementación de Modelo KNN.....	37
Figura 7. Gráfica de Dispersión de los Datos.....	68
Figura 8 Algoritmo de Modelo 1 – Demanda -Tiempo de Producción	68
Figura 9 Demanda-Tiempo de Producción.....	69
Figura 10 Algoritmo de Modelo 2 – Niveles de Inventario – Tiempo de Producción	71
Figura 11 Comportamiento de Niveles_Inventario_tiempo de Producción	72
Figura 12 Algoritmo de Demanda – Niveles de Inventario	74
Figura 13 Comportamiento Demanda Niveles de Inventario	75
Figura 14 Datos Generales.....	78
Figura 15 Datos de Prueba.....	79
Figura 16 Datos de Entrenamiento	80
Figura 17 Datos de Entrenamiento y Prueba	80
Figura 18 Matriz de Predicción	81
Figura 19 Algoritmo de Predicción	81
Figura 20 Matriz de Confusión	82
Figura 21 Verdaderos Positivos	83
Figura 22 Clasificación de FP= Falsos Positivos.....	83
Figura 23 Resultado	84
Figura 24 Datos Reales	84
Figura 25 Nuevos Datos	85
Figura 26 Accuracy.....	85
Figura 27 AUC.....	86
Figura 28 Comparativo	86

LISTA DE TABLAS

Tabla 1 Ventajas y Desventajas de Modelo Supervisado.....	23
Tabla 2 Áreas de Aplicación del Aprendizaje Supervisado	24
Tabla 3. Marco Legal Aplicable Proyecto	29
Tabla 4 Compendio de base Documental	39
Tabla 5. Análisis de los trabajos Identificados.....	47
Tabla 6. Técnicas y Resultados por Autor.....	53
Tabla 7. Principales Algoritmos de Aprendizaje Supervisado Implementados en la Industria	57
Tabla 8. Características de Algoritmos de Aprendizaje Supervisado	59
Tabla 9. Ventajas y Desventajas de cada Algoritmo	60
Tabla 10. Comparación de Algoritmos	64
Tabla 11. Interpretación Cualitativa de cada Criterio.....	64
Tabla 12 Resultados Obtenidos del Algoritmo del Modelo	70
Tabla 13 Resultados de Algoritmo Niveles_Inventario_tiempo de Producción	72
Tabla 14 Resultado Obtenido del Modelo	75
Tabla 15 Comparativo de Modelos Empleados.....	87

RESUMEN EJECUTIVO

Los procesos industriales han tenido un avance importante desde la revolución hasta llegar a la industria 4.0, el uso de sistemas que se basan en análisis de datos e Inteligencia Artificial, se han convertido en una herramienta importante para los procesos industriales. Este proyecto de investigación se enfocó en determinar las estrategias de aprendizaje supervisado que permiten optimizar los procesos relacionados con la industria mediante algoritmos de clasificación.

En la primera etapa se realizó una recisión bibliográfica de las técnicas, metodologías y resultados que se han obtenido con la aplicación de estrategias de aprendizaje supervisado. En la segunda parte, se realizó una comparación de los modelos que son idóneos para procesos de la industrial. Para la última parte de la investigación se llevó a cabo una interpretación de los modelos identificados para lograr predecir las variables clave en procesos industriales como los tiempos de producción, niveles de inventario y demanda de productos que implementan técnicas de entrenamiento y validación mediante algoritmos de clasificación.

Los resultados de la investigación mostraron que El aprendizaje supervisado, con algoritmos como regresión lineal, máquinas de soporte vectorial, árboles de decisión, Naive Bayes y redes neuronales, es clave en diversos sectores industriales. SVM destaca por su alta precisión y manejo de grandes datos, mientras que árboles de decisión y algoritmos como Random Forest y XGBoost son efectivos en tareas específicas. Aunque modelos como las redes neuronales y GBM/XGBoost ofrecen alta precisión y robustez, su entrenamiento es más largo y menos interpretativo. En cambio, la regresión lineal y KNN proporcionan interpretabilidad y

tiempos rápidos, aunque con menor precisión. La regresión lineal se presenta como un modelo robusto para predecir variables continuas en procesos industriales, como demanda y niveles de inventario, superando a KNN en estas tareas.

PALABRAS CLAVE. Algoritmos de predicción y clasificación, Aprendizaje supervisado en cadena de producción, Deep learning, Machine learning.

INTRODUCCIÓN

En la era de la Industria 4.0 y la aproximación hacia la Industrial 5.0, la mejora y optimización de procesos en la industria, se ha convertido en una prioridad global para mejorar la eficiencia, reducir costos y aumentar la competitividad. Una de las herramientas más prometedoras en este ámbito es el aprendizaje supervisado, una técnica de inteligencia artificial que utiliza algoritmos de clasificación para analizar grandes volúmenes de datos y tomar decisiones en los procesos de producción (Bécue, Praça, & Gama, 2021). A nivel internacional, la implementación de estas técnicas ha demostrado ser altamente efectiva, al ofrecer beneficios significativos como la mejora de la calidad del producto, la reducción de tiempos de inactividad, y la minimización de errores humanos (Zeki Murat, Nuhu, & Qasim, 2020).

Para Ramirez *et al.* (2022) los algoritmos de clasificación permiten a las empresas identificar pautas y tendencias en los datos de producción, lo que facilita la detección temprana de problemas y la implementación de soluciones proactivas. En los métodos concernientes a la industria manufacturera, estos algoritmos pueden predecir fallos en maquinaria antes de que ocurran, lo que permite un mantenimiento predictivo y evita costosas interrupciones. En el sector logístico, pueden optimizar las rutas de distribución, lo que reduce costos y mejorando el servicio al cliente. La precisión y la capacidad de adaptación de estos algoritmos los convierten en una herramienta esencial para la modernización y mejora continua de los procesos industriales (Shafiq, Thakre, & Kumar, 2023).

En Colombia, la adopción de metodologías relacionadas con los procesos de aprendizaje supervisado para la optimización de procesos industriales está en una etapa de crecimiento, lo que da muestra de un importante potencial de la

implementación. La industria colombiana, al igual que sus homólogos internacionales, se enfrenta a desafíos como la necesidad de aumentar la productividad y reducir costos operativos (Huertas Mora, 2020). La implementación de algoritmos de clasificación puede ofrecer soluciones a estos retos, al proporcionar una ventaja competitiva en el mercado global. Además, la capacidad de estos algoritmos para adaptarse a diferentes contextos y sectores hace que su aplicación sea versátil y relevante para diversas industrias dentro del país, desde la manufactura hasta la agroindustria (Vesga Acevedo, 2020).

Al tener en cuenta lo expuesto anteriormente, este trabajo de grado tuvo como finalidad determinar las técnicas de aprendizaje supervisado para la optimización de procesos industriales a partir de algoritmos de clasificación. Como parte del diseño metodológico de desarrollo se llevó a cabo la revisión documental de las aplicaciones de estas técnicas a mediante la literatura disponible de documentos, artículos y publicaciones que presentan casos nacionales e internacionales relevantes en la industria. De igual forma, se realizó una comparación de los diferentes algoritmos y una interpretación de los modelos empleados para procesos de inventario, bajo técnicas de entrenamiento y validación de los algoritmos identificados.

1. DESCRIPCIÓN DEL TRABAJO DE INVESTIGACIÓN

1.1. PLANTEAMIENTO DEL PROBLEMA

En el campo de la ingeniería industrial, la optimización de procesos es esencial para garantizar la eficiencia, la calidad y la rentabilidad de las operaciones industriales (Rodrigues *et al.*, 2022). Sin embargo, a pesar de los avances en tecnología y análisis de datos, las Unidades Tecnológicas de Santander no cuenta con investigaciones específicas sobre las técnicas de aprendizaje supervisado y su aplicación en este contexto ha generado una brecha significativa en el aprovechamiento de herramientas que podrían impulsar mejoras sustanciales en los procesos industriales.

Actualmente se desconocen las técnicas y ventajas de la implementación de sistemas con técnicas de aprendizaje supervisado y la creciente tendencia de su uso en el área industrial (Bertolini *et al.*, 2021), esto al dejar de lado este enfoque, al centrarse en optimización de la producción, sin tener en cuenta el potencial de la implementación de estos algoritmos para mejorar los procesos de la cadena de manera específica. Además, el desarrollo de algoritmos de clasificación supervisada requiere habilidades específicas, lo que hace que no se tome en cuenta el potencial de desarrollo de estas técnicas por parte de los grupos de investigación, lo que representa una desventaja competitiva (Fahle *et al.*, 2020).

Al partir lo expuesto previamente se formula la pregunta de investigación:

¿Qué técnicas se han desarrollado bajo estrategias de aprendizaje supervisado para mejorar los procesos de producción en eficiencia, calidad y

rentabilidad en la industria?

1.2. JUSTIFICACIÓN

Al tener en cuenta que la industria tiene una función crucial en la economía y el bienestar social al crear empleos, fomentar avances tecnológicos y proveer productos y servicios fundamentales para la sociedad (Kang et al., 2020). Para investigadores como Zhao et al. (2024), incrementar la productividad y eficiencia en los procesos de fabricación produce un efecto notable en la competitividad de las compañías, lo cual se traduce en un empleo estable y la mejorar de la calidad de vida de los trabajadores. Al utilizar métodos de aprendizaje supervisado en la industria para predecir variables importantes, como la demanda de productos o los tiempos de producción, se mejora el rendimiento, disminuye gastos y eleva la calidad del producto.

Para Jove et al. (2022) la implementación de técnicas de aprendizaje supervisado en la industria trae un impacto significativo en la rentabilidad y la competitividad de las empresas. Al predecir variables clave, como la demanda de productos o los niveles de inventario, las empresas pueden optimizar sus operaciones, reducir los costos de producción, minimizar el desperdicio de recursos y mejorar la eficiencia en general. Esto genera como resultado, mayores márgenes de beneficio, una mayor capacidad para competir en el mercado y una mejor posición para satisfacer las necesidades y expectativas de los clientes (Figueroa et al., 2023).

Para las UTS, el desarrollo de este trabajo de grado se presenta como una oportunidad para la explorar y el desarrollar nuevas metodologías y herramientas. Al investigar y proponer técnicas innovadoras para predecir variables clave en los procesos industriales, se mostrarán las nuevas formas de mejorar la eficiencia y la

productividad. Además, el estudio y la validación de algoritmos de clasificación específicos para la industria generarán nuevos conocimientos y mejores prácticas que puedan ser aplicados en diferentes sectores y contextos industriales. Esto puede fomentar la colaboración entre la industria y la academia, al impulsar la innovación y el avance tecnológico en el campo de la producción industrial.

1.3. OBJETIVOS

1.3.1. OBJETIVO GENERAL

Determinar las técnicas de aprendizaje supervisado para la optimización de procesos industriales, con el fin de mejorar la eficiencia, la calidad y la rentabilidad de las operaciones industriales por medio de algoritmos de clasificación supervisada.

1.3.2. OBJETIVOS ESPECÍFICOS

Realizar una revisión de literatura sobre las aplicaciones del aprendizaje supervisado en ingeniería industrial, identificando casos de estudio, metodologías y resultados relevantes destacando las técnicas de aprendizaje supervisado.

Comparar diferentes algoritmos de aprendizaje supervisado, con el fin de evaluar su idoneidad para el modelado de los procesos industriales

Interpretar modelos de aprendizaje supervisado para predecir variables clave en los procesos industriales, como la demanda de productos, los tiempos de producción o los niveles de inventario, utilizando técnicas de entrenamiento y validación adecuadas por medio de algoritmos de clasificación.

1.4. ESTADO DEL ARTE

Es esencial supervisar el estado del equipo industrial para evitar fallos y optimizar el tiempo de funcionamiento. En el trabajo desarrollado por Rodrigues et al. (2022) se utilizaron técnicas de aprendizaje no supervisado y semi-supervisado para crear modelos predictivos del estado de una máquina industrial. Se buscó identificar cuándo el activo operaba dentro de sus parámetros nominales o estaba fuera de esa zona fue el objetivo principal, lo que indicaba un riesgo de fallo o de operación subóptima. Se demostró que es posible clasificar el estado de la máquina utilizando redes neuronales artificiales. Tres estados, seleccionados mediante el Método del Codo, abarcaban casi toda la variabilidad de los datos estudiados según revelaron el análisis de agrupamiento K- medias y el PCA. Debido a la importancia de la calidad de los lubricantes en el rendimiento y estado de las máquinas, se creó un algoritmo basado en Redes Neuronales para clasificar los lubricantes. Este alcanzó una precisión del 98% comparado con las clasificaciones hechas por expertos humanos. Se encontró una limitación en el estudio, debido a que las clasificaciones existentes solamente trataban los fallos actuales, ya sea a corto o mediano plazo. Con el fin de superar esta limitación, se sugirió llevar a cabo una clasificación a largo plazo en futuras investigaciones.

Actualmente, la inteligencia artificial se aplica con éxito en diversas áreas del sector agrícola, destacando especialmente en la agricultura de precisión. En la investigación de Coppiano y Herrera (2022), se abordan aspectos como la detección de enfermedades en plantas y la predicción de épocas óptimas de siembra. Para alcanzar estos objetivos, es esencial contar con datos precisos y métodos eficaces para evaluar el estado del suelo, considerando factores como las condiciones ambientales, la composición del suelo y su nivel de acidez. Dada la abundancia de cultivos, especialmente de frutas y verduras, resulta crucial intensificar la producción

basándose en las características del suelo destinado a la siembra. El objetivo de este proyecto es predecir la siembra de cultivos en fincas agroecológicas utilizando técnicas de machine learning, en particular máquinas de vectores de soporte (SVM). A través de una API, se procesarán datos ambientales para determinar si el suelo es apto para un cultivo específico. Esta funcionalidad será parte de una aplicación web con modelos de inteligencia artificial basados en SVM, que brindará a los agricultores una herramienta tecnológica para optimizar sus siembras y mejorar su producción futura.

El aprendizaje profundo ha revolucionado la investigación en machine learning en diversos campos, pero la falta de datos de entrenamiento etiquetados sigue siendo un desafío crucial para su rendimiento y generalización. Para superar esta limitación, los investigadores han desarrollado métodos innovadores, destacando el aprendizaje por transferencia y el autoaprendizaje supervisado. La transferencia de aprendizaje aprovecha conocimiento previo de un conjunto de datos extenso y lo aplica a tareas específicas con datos etiquetados limitados, mejorando el rendimiento y la generalización. Por otro lado, el autoaprendizaje supervisado se centra en tareas pretexto que no necesitan anotaciones manuales, permitiendo a los modelos adquirir representaciones valiosas de grandes cantidades de datos sin etiquetar. Aunque estas técnicas han encontrado aplicaciones en campos como el procesamiento de imágenes médicas y el procesamiento del lenguaje natural, también presentan limitaciones, como el desajuste de dominio en el caso del aprendizaje por transferencia. Este artículo desarrollado por Zhao et al. (2024) revisó las aplicaciones recientes de estas técnicas, evaluando sus ventajas y desafíos, y ofrece orientación sobre la mejor técnica a elegir para aplicaciones específicas de aprendizaje profundo, abordando así el desafío de la escasez de datos.

Machine Learning (ML, por sus siglas en inglés) es una subcategoría de la inteligencia artificial que se enfoca en algoritmos capaces de aprender autónomamente a partir de datos de entrada. Durante los últimos diez años, el ML ha avanzado significativamente, como se evidencia en los algoritmos de Aprendizaje Profundo (DL) utilizados en vehículos autónomos y juegos electrónicos de estrategia. Los investigadores cada vez consideran más la aplicación del ML en el campo industrial, con muchos trabajos destacando el ML como un habilitador clave para la evolución de los sistemas de fabricación hacia el nivel de la Industria 4.0. Aunque esto sea así, las aplicaciones industriales continúan teniendo limitaciones y se encuentran restringidas a un reducido número de empresas internacionales. El propósito de esta investigación llevada a cabo por Bertolini et al. (2021) fue ofrecer claridad sobre las reales capacidades y potenciales limitaciones de los algoritmos de ML en la gestión operativa. Se ofrece una revisión completa y organizada para que los profesionales en el campo la comprendan fácilmente. Para esto, los documentos de 2000 en adelante se clasifican según el algoritmo y área de aplicación correspondiente, además se lleva a cabo un análisis de palabras clave para identificar los temas más importantes. El número de publicaciones ha aumentado constantemente, con un notable interés en las técnicas de aprendizaje no supervisado y, sobre todo, en el aprendizaje profundo durante los últimos cinco años. Se han identificado temas emergentes como la planificación y el control de la producción, y el análisis de defectos, además sugiere que el ML se convertirá en algo fundamental en muchos campos de la gestión operativa.

El uso de la Inteligencia Artificial (IA) ha experimentado un fuerte impulso con la cuarta revolución industrial, siendo el Machine Learning (ML) crucial en abordar los desafíos propios de la fabricación, como mejorar y optimizar la calidad de los procesos. En la investigación de Figueroa et al. (2023) se examinó el estado actual del ML y las principales técnicas empleadas, se destaca la importancia y utilidad de

esta herramienta clave en este contexto. En la ingeniería industrial se aplicaron algoritmos de ML en 54 casos que se analizaron. Según los estudios, la utilización de estas técnicas genera resultados destacados en cuanto a precisión y fiabilidad, sobre todo en aspectos críticos como el diagnóstico de fallos, la predicción de defectos y la optimización de sistemas. La eficacia de las técnicas de ML para abordar desafíos industriales se destaca en estos hallazgos, subrayando su relevancia y potencial en la industria contemporánea.

2. MARCO REFERENCIAL

En esta sección se presenta una contextualización teórica y conceptual acerca del aprendizaje supervisado y los algoritmos de clasificación aplicados a la industria, de modos que sea más sencillo entender las temáticas de desarrollo del proyecto de grado.

2.1. MARCO TEÓRICO

Al tener en cuenta que la inteligencia artificial (IA) ha emergido como una fuerza transformadora en diversas industrias, ofreciendo soluciones avanzadas que optimizan procesos, mejoran la eficiencia y reducen costos. Las técnicas de IA, como Deep Learning, los sistemas expertos y Machine Learning,, están revolucionando la manera en que las empresas manejan la producción, la gestión de inventarios y la previsión de la demanda de productos. En el ámbito industrial, la IA se aplica para automatizar y perfeccionar tareas críticas, desde la planificación de la producción hasta la gestión de cadenas de suministro complejas. A continuación, se presentan los principales marcos referenciales para el trabajo de grado.

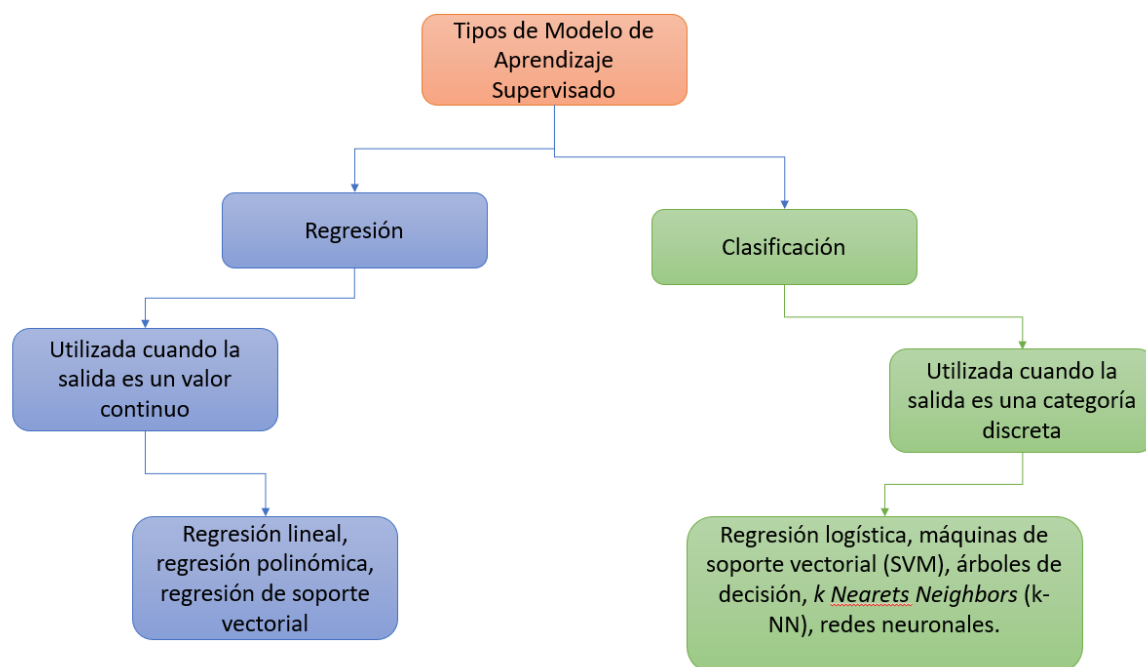
2.1.1. *Aprendizaje Supervisado*

El aprendizaje supervisado es una técnica crucial y versátil en el campo de la AI, poderosa para resolver problemas en diversas disciplinas y aplicaciones. Su habilidad para aprender de ejemplos etiquetados lo hace especialmente efectivo en tareas que cuentan con datos históricos detallados. Una técnica ampliamente utilizada en el campo del aprendizaje automático es el aprendizaje supervisado, siendo una de las más fundamentales (Saravanan & Sujatha, 2018). El uso de este enfoque se fundamenta en un conjunto de datos etiquetados, donde cada ejemplo

del conjunto tiene una entrada y su respectiva salida. Entrenar un modelo en aprendizaje supervisado tiene como objetivo la capacidad de predecir correctamente nuevas salidas, extrapolando desde los datos de entrenamiento (Nasteski, 2020).

Al tomar como base el trabajo de Huertas (2020) existen dos modelos principales de aprendizaje supervisado, las cuales se presentan en la Figura 1.

Figura 1 Tipos de Modelo Aprendizaje Supervisado



Fuente: Autoría Propia con base en Huertas (2020)

Por otra parte, desde la investigación desarrollada por Shafiq et al. (2023) se logró identificar las ventajas y desventajas con las que cuentan los sistemas a los que se les aplica un modelo de aprendizaje supervisado, de modo que a partir de la Tabla 1 se describen estas características de forma resumida.

Tabla 1 Ventajas y Desventajas de Modelo Supervisado

Ventajas y Desventajas del Modelo Supervisado	
Ventajas	Desventajas
Precisión: Si se tienen suficientes datos de entrenamiento representativos, es posible lograr niveles altos de precisión.	Requiere Datos Etiquetados: Etiquetar datos puede ser costoso y laborioso, lo que representa una limitación en la necesidad de contar con datos etiquetados.
Interpretabilidad: Modelos como los árboles de decisión son fáciles de interpretar y comprender.	Sobreajuste: Si el modelo es muy complejo, podría ajustar demasiado los datos de entrenamiento y tener dificultades para generalizar con nuevos datos.
Aplicabilidad: Puede ser utilizado en una gran variedad de problemas prácticos en distintos campos.	Escalabilidad: Algunos algoritmos pueden resultar costosos computacionalmente y no escalan bien con grandes cantidades de datos.

Fuente: Autoría Propia con base en Shafiq *et al.* (2023)

2.1.2. Técnicas de aprendizaje supervisado

Regresión Lineal: Es un método de aprendizaje supervisado que se emplea para modelar la relación entre una variable dependiente continua y una o más variables independientes. Esta técnica, por su simplicidad y fácil interpretación, es una de las más básicas y comunes en el análisis de datos y el aprendizaje automático.

El modelo cuenta con una serie de conceptos básico para su desarrollo, La fórmula matemática para la regresión lineal simple, que involucra una sola variable independiente, es la que se muestra en la ecuación 1. El objetivo del modelo de regresión lineal es encontrar los valores óptimos de los coeficientes (β) que minimicen la suma de los errores cuadráticos (diferencias entre los valores predichos y los valores reales). Esta minimización se logra típicamente utilizando el método de mínimos cuadrados ordinarios (OLS).

$$Y = \beta_0 + \beta_1 X + \varepsilon \quad (1)$$

Variable Dependiente (Y): Representa a la variable que se busca explicar o predecir.

Variables Independientes (X): Hacen referencia a las variables que se emplean para predecir el valor de Y

2.1.3. Áreas de Aplicación del Aprendizaje Supervisado en la Industria

Debido a su simplicidad y capacidad para modelar relaciones entre variables, la regresión lineal se usa ampliamente en varias disciplinas. A continuación, se describen algunas de las principales áreas de aplicación de la regresión lineal en la Tabla 2.

Tabla 2 Áreas de Aplicación del Aprendizaje Supervisado

Áreas de Aplicación del Aprendizaje Supervisado		
Área	Área específica	Descripción
Economía y Finanzas	Predicción de Precios de Activos	Empleada para predecir los precios de acciones, bonos, bienes raíces y otros activos financieros.
	Análisis de Riesgo	Impacto de los factores de riesgo en las decisiones de inversión se evalúa.
	Modelado de Ingresos y Ventas	Predecir ingresos y ventas utilizando variables económicas como el gasto del consumidor, las tasas de interés y el PIB.
Marketing	Análisis de Eficacia Publicitaria	Análisis del efecto de diversas estrategias publicitarias en las ventas o reconocimiento de marca.
	Segmentación de Mercado	Segmentación de mercado utilizando características demográficas y de comportamiento para su identificación.
	Optimización del Precio	Cálculo del precio ideal de productos para maximizar los ingresos.
Salud y Medicina	Predicción de Resultados de Tratamientos	La efectividad de varios tratamientos médicos se evalúa teniendo en cuenta las características del paciente.
	Análisis Epidemiológico	Investigación sobre la conexión entre los factores de riesgo y la aparición de enfermedades.
	Investigación Clínica	Entender la relación entre variables clínicas y resultados de salud a través del modelado de datos clínicos
	Modelado de Sistemas Físicos	Predicción del comportamiento de sistemas físicos y mecánicos mediante la estimación de parámetros.

Ciencias Físicas e Ingeniería	Control de Calidad	Identificar los factores que afectan la calidad del producto a través del análisis de datos de producción.
	Pronóstico de Mantenimiento	Basándose en datos históricos y variables de operación, se pueden predecir fallas en equipos y maquinaria.
Ciencias Sociales	Investigación Demográfica	Investigación de la asociación entre variables demográficas (tales como edad, educación e ingresos) y fenómenos sociales.
	Análisis Político	Análisis de cómo diferentes factores impactan en los resultados electorales.
	Psicología	Relaciones modeladas entre variables psicológicas y comportamientos.
Ambiente y energía.	Modelado Climático:	Investigación de la correlación entre variables climáticas (tales como temperatura y precipitación) y fenómenos del medio ambiente.
	Gestión de Recursos Energéticos	Predicción y optimización del consumo de energía mediante la distribución eficiente de recursos energéticos.
	Evaluación de Impacto Ambiental	Estudio de cómo las actividades humanas afectan al medio ambiente.
Deportes	Análisis de Rendimiento	Análisis de los factores que influyen en el desempeño de deportistas y conjuntos deportivos.
	Predicción de Resultados Deportivos	Creación de predicciones de partidos utilizando estadísticas de desempeño.
	Optimización de Estrategias	Examinar las tácticas y estrategias para aumentar el desempeño en competiciones.
Educación	Evaluación del Rendimiento Académico	Investigación de la correlación entre factores como técnicas pedagógicas, materiales didácticos y logro académico de los alumnos.
	Predicción de Matrícula	Predecir la tasa de matrícula y abandono escolar utilizando variables socioeconómicas y académicas.
	Investigación Educativa	Examinar los datos educativos con el fin de comprender y mejorar los procesos de enseñanza y aprendizaje.
Transporte y Logística	Optimización de Rutas	Predicción y optimización de rutas de transporte para calcular tiempos de viaje.
	Análisis de Tráfico:	Investigamos la conexión entre factores como el tráfico y los tiempos de viaje.
	Gestión de Inventarios	Optimización de niveles de inventario y modelado de la demanda de productos.
Agricultura	Predicción de Rendimiento de Cultivos	Estructuración de la relación entre elementos del clima, técnicas agrícolas y producción de cultivos.
	Gestión de Recursos Agrícolas	Mejora en el uso de recursos como agua y fertilizantes.
	Estudio de la Salud del Suelo	Análisis de los factores que impactan la salud del suelo y la productividad agrícola.

Nota. En tabla se presentan las principales áreas de aplicación en las que se han desarrollado implementaciones del modelo de aprendizaje supervisado. Elaboración propia con base en Figueroa et al. (2023) y Saravanan & Sujatha (2018)

La regresión lineal es una herramienta versátil que encuentra aplicaciones en una amplia variedad de campos. Su capacidad para modelar relaciones entre variables y hacer predicciones precisas la convierte en una técnica invaluable en la toma de decisiones y la optimización de procesos en múltiples disciplinas.

2.1.4. Modelos Predictivo

Los algoritmos diseñados para aprender patrones a partir de datos históricos etiquetados con el fin de hacer predicciones o inferencias sobre nuevos datos son llamados modelos predictivos en aprendizaje supervisado. Estos modelos son esenciales para una amplia gama de usos en inteligencia artificial y aprendizaje automático. Su habilidad para aprender de datos históricos y aplicar ese conocimiento a nuevas situaciones los hace extremadamente valiosos en una variedad amplia de campos, desde la salud y las finanzas hasta la ingeniería y las ciencias sociales. A medida que más datos están disponibles, la importancia y el uso de estos modelos siguen creciendo. Adicionalmente, en la

2.2. MARCO CONCEPTUAL

Gestión de Inventarios

Supervisar y controlar las cantidades de bienes y materiales en los almacenes es lo que se conoce como gestión de inventarios para una empresa. Garantizar la disponibilidad de productos para satisfacer la demanda del cliente y minimizar los costos asociados con el mantenimiento de inventarios es su principal objetivo (Vesga, 2020).

Predicción de Demanda

Uso de modelos de regresión para anticipar la demanda futura de productos y ajustar los niveles de inventario eficazmente (Huertas, 2020).

Optimización de Stock

Este concepto hace referencia a la posibilidad de determinar niveles óptimos de stock clasificando productos según su rotación y valor (Figueroa *et al.*, 2023).

Tiempos de Producción

El lapso de producción en la industria hace referencia al tiempo requerido para finalizar un proceso de fabricación desde su comienzo hasta su conclusión. Constituyen todas las fases, desde la preparación de materiales y la fabricación de componentes hasta el ensamblaje y el acabado del producto final. Es crucial optimizar estos tiempos para mejorar la eficiencia, disminuir gastos y cumplir con la demanda del cliente en tiempo (Adrián *et al.*, 2019).

Estimación de Tiempos

Para Quiroz *et al.* (2022) la estimación de tiempos es el uso de modelos de regresión para predecir los tiempos de producción utilizando datos históricos.

Optimización de Procesos

Identificar los denominados cuellos de botella y áreas de mejora en la línea de producción mediante el uso de técnicas de clasificación.

Cadena de Suministros

El sistema integrado que abarca todas las etapas desde la adquisición de materias primas hasta la entrega del producto final al consumidor es la cadena de suministros. Engloba la coordinación de proveedores, producción, almacenaje, distribución y logística para optimizar la eficiencia y disminuir costos (Sablón *et al.*, 2021).

Rutas de Distribución

Optimización de rutas de distribución y reducción de costos logísticos a través de modelos de clasificación y regresión (Martínez Soto, 2011).

Gestión de Proveedores

Mejorar la gestión de la cadena de suministro a través de clasificar proveedores basándose en su rendimiento y fiabilidad (De Castillo, y otros, 2015).

Predicción de Demanda de Productos

Para (Andrade, Del Río, & Alvear, 2019) el proceso de predicción de la demanda hace parte de estimar la cantidad de productos que los consumidores comprarán en el futuro es lo que implica predecir la demanda de productos. Usa datos del pasado, tendencias de mercado y métodos analíticos para prever las necesidades del cliente, asistiendo a empresas en la planificación de la producción, administración de inventarios y reducción de desperdicios.

2.3. MARCO LEGAL

Tabla 3. Marco Legal Aplicable Proyecto

Marco Legal Aplicable Proyecto	
Normativa	Descripción
Ley de Protección de Datos Personales (Ley 1581 de 2012)	Establecer disposiciones generales para la protección de datos personales en Colombia.
Reglamento General de Protección de Datos (Decreto 1377 de 2013)	Reglamentar parcialmente la Ley 1581 de 2012, detallando aspectos operativos y procedimentales para la protección de datos personales.
Ley de Transparencia y del Derecho de Acceso a la Información Pública Nacional (Ley 1712 de 2014)	Regular el acceso a la información pública y asegurar la transparencia en el tratamiento de datos por entidades públicas y privadas que realizan funciones públicas.

3. DISEÑO DE LA INVESTIGACIÓN

En esta sección se describe el diseño de investigación aplicado a el trabajo de grado al tener en cuenta que el proyecto estuvo enfocado en una investigación de tipo documental para determinar las técnicas de aprendizaje supervisado que se aplican para la optimización de procesos de la industria basados en algoritmos de clasificación supervisada.

3.1. Tipo de Investigación

Al tomar como base a Hernández & Mendoza (2018) el tipo de investigación que aplica a esta investigación es la investigación descriptiva, esto al tener en cuenta que para el autor este tipo de investigación se caracteriza por buscar especificar propiedades, particularidades y características de personas, grupos, comunidades, procesos, objetos o cualquier otro fenómeno que se someta a un análisis. Este tipo de investigación no solo se enfoca en describir variables, sino que también pretende establecer relaciones entre ellas, aunque no llegue al punto de probar hipótesis causales.

3.1.1. Enfoque

Para el enfoque del proyecto la investigación descriptiva permitió recopilar y organizar de manera sistemática la información existente sobre técnicas de aprendizaje supervisado aplicadas a la optimización de procesos industriales. Esto incluye la identificación y catalogación de algoritmos de clasificación, estudios de casos, metodologías, herramientas y marcos teóricos relevantes. Esta fase es crucial para entender el panorama actual y el estado del arte en este campo.

Una de las principales contribuciones de la investigación descriptiva fue que bajo la revisión documental permitió proporcionar descripciones detalladas de las

diversas técnicas y algoritmos de aprendizaje supervisado. Estas descripciones buscaron incluir cómo funcionan los algoritmos, sus aplicaciones específicas en la industria, ventajas, limitaciones y condiciones ideales para su implementación. Una de las principales contribuciones de la investigación descriptiva fue proporcionar descripciones detalladas de las diversas técnicas y algoritmos de aprendizaje supervisado. Estas descripciones pueden incluir cómo funcionan los algoritmos, sus aplicaciones específicas en la industria, ventajas, limitaciones y condiciones ideales para su implementación. Este nivel de detalle es esencial para que los investigadores y profesionales comprendan cómo seleccionar y aplicar estas técnicas en sus propios contextos industriales.

Según Hernández (2006) el enfoque cualitativo es una metodología de investigación que se centra en comprender y describir la realidad desde la perspectiva de los sujetos involucrados, explorando sus experiencias, percepciones, y significados. Este enfoque es holístico, flexible y se adapta a las particularidades del contexto en estudio.

El enfoque cualitativo en esta investigación documental permitió una comprensión más profunda y matizada de cómo las técnicas de aprendizaje supervisado pueden optimizar los procesos industriales. Al centrarse en las experiencias y percepciones de los involucrados, el contexto específico de las aplicaciones y las interrelaciones entre los factores técnicos y humanos, este enfoque permite diseñar estrategias de implementación más efectivas y adaptadas a las realidades del sector industrial.

3.1.2. Método

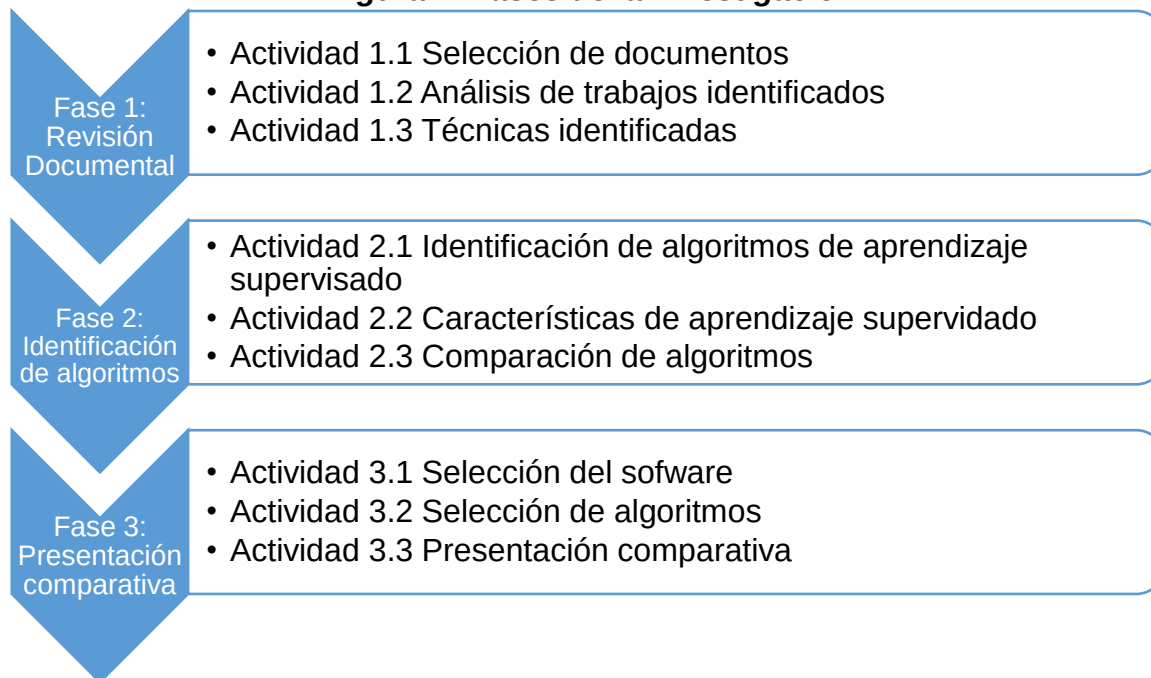
Al tener en cuenta que para Hernández & Torres (2020) el método deductivo es un enfoque de razonamiento lógico que parte de principios generales o premisas

para llegar a conclusiones específicas. Este método se basa en la lógica y la estructura de argumentos en los que, si las premisas son verdaderas, las conclusiones necesariamente también lo serán.

El aporte de este método a la investigación permitió establecer una base teórica sólida y lógica. A partir de principios generales sobre el aprendizaje supervisado y su impacto en la optimización de procesos, se derivan conclusiones específicas y aplicables a casos concretos en la industria, garantizando así que las soluciones propuestas sean coherentes y fundamentadas en conocimientos e investigaciones realizadas en forma previa.

3.2. Fases del trabajo de grado

Figura 2. Fases de la investigación

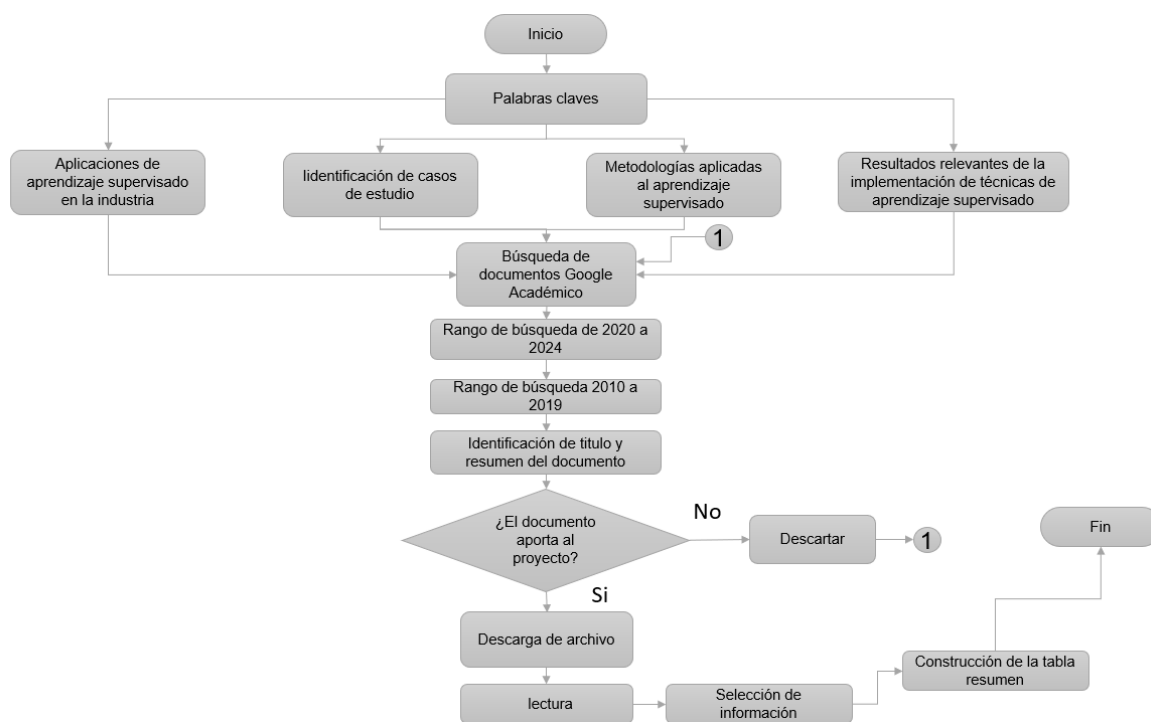


Fuente: Autor

4. DESARROLLO DEL TRABAJO DE GRADO

4.1. REVISIÓN DOCUMENTAL DE APLICACIONES DE APRENDIZAJE SUPERVISADO EN LA INGENIERIA INDUSTRIAL

Figura 3 Proceso de Revisión Literaria de Aprendizaje Profundo

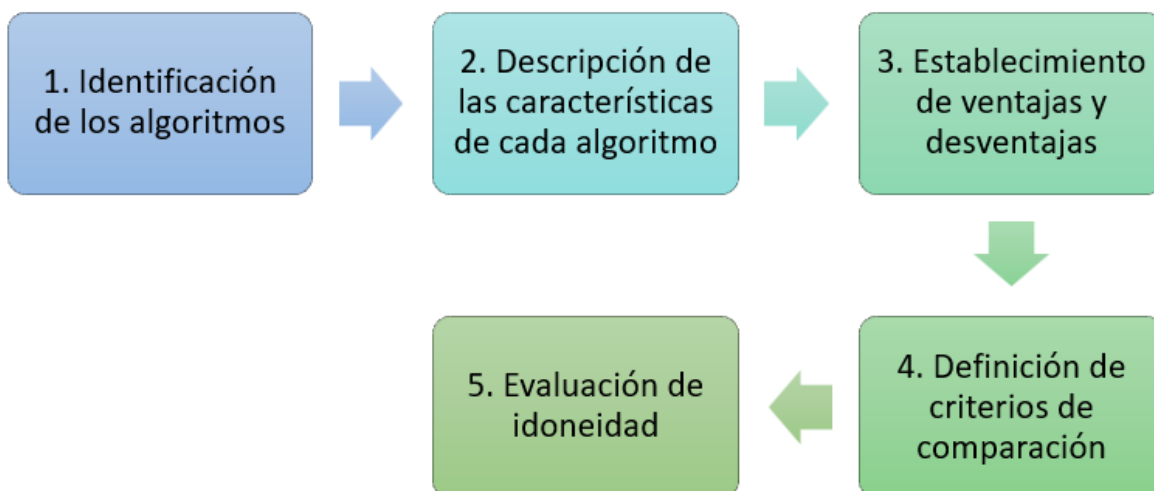


Fuente: Autores

4.2. COMPARAR LOS DIFERENTES ALGORITMOS DE APRENDIZAJE SUPERVISADO PARA MODELADO DE PROCESOS INDUSTRIALES

Para realizar la comparación de los diferentes algoritmos que se emplean en la industria, se tomó como base la revisión documental, de modo que se lograra identificarlos como algoritmos de clasificación y que fueran aplicados a procesos industriales. De esta forma, en el diagrama de la Figura 4 se presentan los pasos empleados para desarrollar la comparativa de evaluación de idoneidad.

Figura 4 Proceso de Comparación y Evaluación de los Algoritmos de Aprendizaje Supervisado en la Industria



Fuente: Autores

Identificación de Algoritmos: Como primer paso se desarrolló una identificación y selección de los algoritmos que se han desarrollado para la industria, esto a partir de las investigaciones identificadas.

Descripción de las características de cada algoritmo: A partir de una revisión bibliográfica presentar de forma cualitativa de cada algoritmo.

Establecimiento de Ventajas y Desventajas: Una vez se describieron los algoritmos, se procedió a realizar una presentación de las desventajas y ventajas de cada algoritmo, de modo que se establecieron perfiles para posterior comparación.

Identificación de aplicaciones en la industria: Con el objetivo de establecer de qué forma los algoritmos se han implementado se presentaron aplicaciones en la industria que se han realizado.

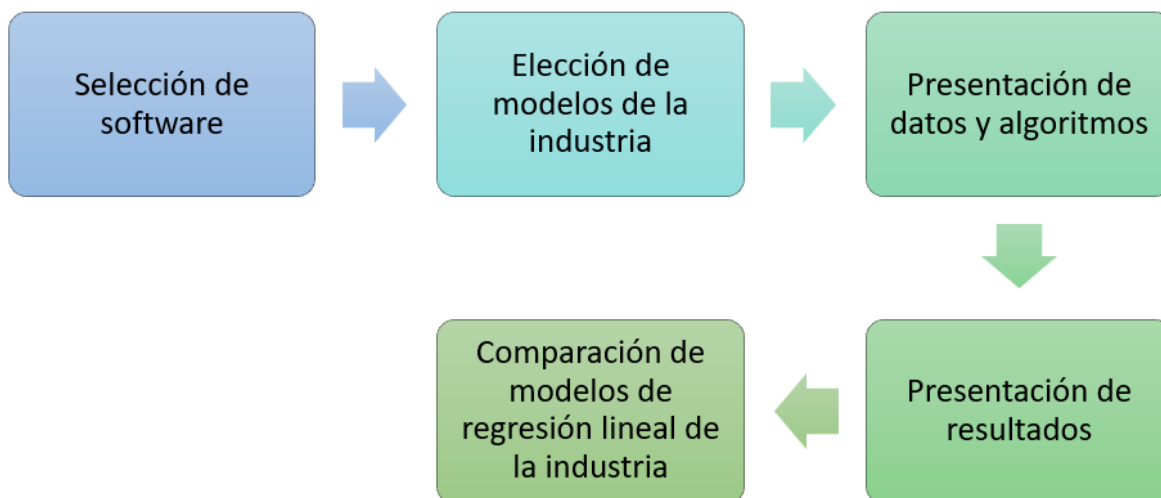
Definición de criterios de comparación: Se presentaron criterios de comparación y una valoración de cada criterio, para lograr una comparación y evaluación de cada algoritmo.

Evaluación de Idoneidad: Con la parametrización de criterios y definición de los valores de evaluación se realizó el proceso de selección de idoneidad del algoritmo.

4.3. INTERPRETACIÓN DE LOS MODELOS INDUSTRIALES DE APRENDIZAJE SUPERVISADO

4.3.1. Selección de Software de Regresión Lineal

Figura 5 Selección de Software de Regresión Lineal



Fuente: Autores

Como se muestra en la Figura 5 como parte de los pasos desarrollados, se tuvo que realizar la selección del software apropiado para trabajar los modelos de regresión lineal.

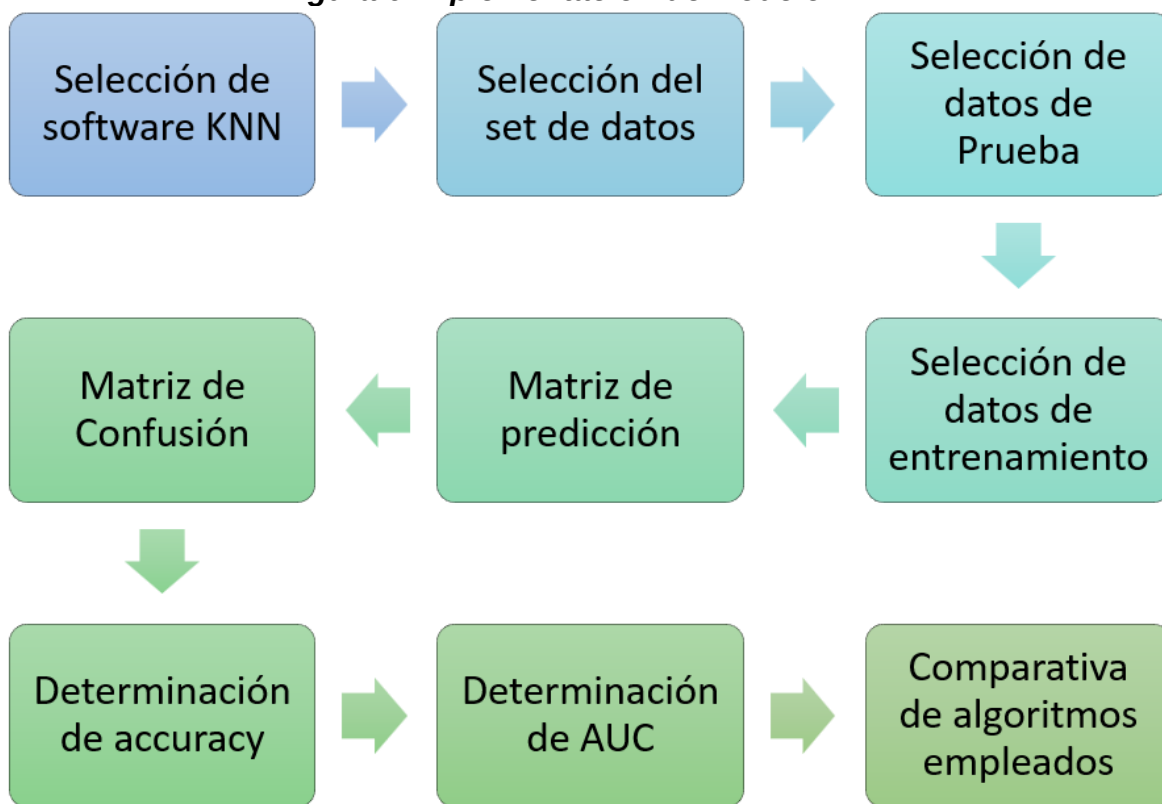
Una vez se seleccionaron los modelos para las etapas de la industria se presentación de los diferentes modelos empleados con sus respectivos datos y algoritmos.

Una vez presentados los algoritmos, se procedió a realizar una presentación de los resultados obtenidos

Por último, se realizó una comparación de los resultados obtenidos para cada algoritmo de regresión lineal empleado.

4.3.2. Modelo de KNN

Figura 6 Implementación de Modelo KNN



Fuente: Autores

Para la segunda parte del objetivo se tomo el modelo KNN y se elaboró una selección de un set de datos generales, el cual se dividió en datos de entrenamiento (70%) y datos de prueba (30%). Como los resultados obtenidos de la prueba

realizada se desarrolló una matriz de predicción y con este resultado se elaboró una matriz de confusión, esto con el fin de determinar el nivel de precisión (accuracy) y la región AUC del modelo KNN.

Como parte final, se realizó una comparativa de los resultados obtenidos con el algoritmo de Regresión Lineal y el algoritmo KNN.

5. RESULTADOS

En esta sección se detallan específicamente cuales fueron los resultados obtenidos en la ejecución del trabajo de grado, documentando cada una de las pruebas y tabulando o graficando si es necesario, los datos obtenidos como respuesta a la solución de la problemática.

5.1. APLICACIONES DE APRENDIZAJE SUPERVISADO EN LA INDUSTRIA

5.1.1. Compendio investigaciones relacionadas con el Aprendizaje Supervisado en la Industria

En la Tabla 4 se presenta una base documenta que presenta estudios, trabajos, artículos y publicaciones relacionada al aprendizaje supervisado al destacar su resumen y los resultados presentado de cada investigación.

Tabla 4 Compendio de base Documental

Compendio de Documentos Relacionados al Aprendizaje Supervisado				
	Título	Resumen	Resultados	Autor
1	An overview of the supervised machine learning methods	En la última década, numerosos métodos de aprendizaje supervisado se han introducido en el campo del aprendizaje automático, convirtiéndose en un área de intensa investigación. Estas técnicas se aplican en el procesamiento y análisis de diversos datos, destacando su capacidad para utilizar datos de entrenamiento anotados con etiquetas de clase en la clasificación. Este documento resume los aspectos fundamentales de varios métodos supervisados y tiene	Estrategias Machine Learning empleadas en aprendizaje supervisado: Árbol de decisión, regresión lineal, Naive Bayes, regresión logística	Nasteski (2018)

		como objetivo principal presentar una visión general del aprendizaje automático y sus técnicas.		
2	Continuous quality control evaluation during manufacturing using supervised learning algorithm for Industry 4.0	Las industrias inteligentes utilizan aprendizaje automático y big data para gestionar la cadena de suministro y aumentar la productividad, enfrentando desafíos en el control de calidad. Estas industrias se basan en el aprendizaje supervisado para mejorar la inspección y el control de parámetros productivos. Usan funciones de kernel y SVM, combinando la función de kernel RBF con redes neuronales para la clasificación y gestión de calidad. Los clasificadores SVM, como C-SVM y V-SVM, se entrenan para manejar múltiples consecuencias en la fabricación. Se evalúan diferentes funciones de kernel y coeficientes gamma, demostrando la superioridad del clasificador SVM en tiempo computacional, medida F1, ROC y precisión. El modelo propuesto busca monitorear la fabricación y controlar fallos.	Este artículo investiga el uso del aprendizaje supervisado para el control de calidad en industrias manufactureras inteligentes, comparando la máquina de soporte vectorial (SVM) con las redes neuronales RBF (RBF NNS) para determinar cuál es más efectiva en identificar y controlar problemas de calidad. Se analizan diversas funciones de kernel para elegir la más eficiente. El principal objetivo es mejorar el clasificador SVM reduciendo errores generalizados y de entrenamiento, cruciales para la toma de decisiones. Los resultados muestran que el modelo RBF es menos preciso que el SVM con C-SVM o V-SVM, los cuales son efectivos en identificar y diagnosticar fallos. Las ventajas del SVM, como mayor velocidad de entrenamiento y capacidad para resolver problemas grandes, lo hacen preferible sobre las redes neuronales RBF. Así, la metodología SVM resulta eficaz para el control de calidad en la Industria 4.0.	Shafiq et al (2023)
3	Risk prediction in life insurance industry using supervised learning algorithms	La evaluación de riesgos es crucial en los seguros de vida para clasificar solicitantes y fijar precios de pólizas. Con el aumento de datos y avances en análisis, el proceso de suscripción puede automatizarse para agilizar la tramitación. Esta investigación mejora la evaluación de riesgos en aseguradoras mediante análisis predictivo, utilizando un conjunto de datos real con más de cien atributos anonimizados. Se realizó una reducción de dimensionalidad con técnicas de Selección de Características Basada en Correlación (CFS) y	Los resultados mostraron que REPTree tuvo el mejor rendimiento con el menor error absoluto medio (MAE) de 1.5285 y el menor error cuadrático medio (RMSE) de 2.027 con CFS, mientras que la Regresión Lineal Múltiple fue la mejor en PCA con MAE de 1.6396 y RMSE de 2.0659.	Boodhun & Jayabalan (2018)

		Análisis de Componentes Principales (PCA). Se aplicaron algoritmos de aprendizaje automático, como Regresión Lineal Múltiple, Red Neuronal Artificial, REPTree y Random Tree, para predecir el nivel de riesgo		
4	Semi-supervised learning for industrial fault detection and diagnosis: A systemic review	La automatización de FDD es crucial en diversas industrias hoy en día, especialmente con enfoques basados en datos y Aprendizaje Automático (ML). Los algoritmos de ML se dividen en métodos supervisados y no supervisados según si las instancias están etiquetadas. Recientemente ha surgido el Aprendizaje Semi-Supervisado (SSL), que combina instancias etiquetadas con otras no etiquetadas para mejorar la precisión en entornos industriales con datos limitados. SSL se ha probado como una solución prometedora para problemas de FDD, aunque no ha habido revisiones sistemáticas hasta este estudio. El artículo organiza la literatura sobre SSL para FDD usando la taxonomía de van Engelen & Hoos, identificando los algoritmos SSL más y menos utilizados, y considerándolos en diversas tareas de detección de fallas y estructuras de datos comunes. También se proponen mejores prácticas para la implementación de SSL en condiciones industriales reales, con el fin de evitar fallos comunes.	Durante la primera década de uso de (Aprendizaje Semi-Supervisado) SSL para (Detección y Diagnóstico de Fallas) FDD, los primeros intentos de aplicar nuevas técnicas de Aprendizaje Automático mostraron algunas inconsistencias. Las más comunes incluyen: considerar técnicas y estrategias de ML que no son SSL (por ejemplo, mezclar métodos supervisados y no supervisados y llamarlos semi-supervisados), utilizar un porcentaje demasiado alto de instancias etiquetadas, lo que dificulta justificar el uso de cualquier método SSL; por ejemplo, en el 25% de las observaciones no estaban etiquetadas, y solo el 10% no lo estaban, mientras que en otras propuestas menos del 50% de las observaciones no estaban etiquetadas, y la presentación de métodos SSL nuevos aunque algo vagos, lo que redujo la posibilidad de repetibilidad experimental.	Ramírez <i>et al.</i> (2023)
	Supervised Machine Learning Algorithms: Classification and Comparison	Supervised Machine Learning (SML) es la búsqueda de algoritmos que razonan a partir de instancias proporcionadas externamente para producir hipótesis generales, las cuales hacen predicciones sobre futuras instancias. La clasificación supervisada es una de las tareas más frecuentemente realizadas por los sistemas inteligentes. Este artículo describe diversas técnicas de clasificación de Aprendizaje Automático	Los resultados muestran que SVM fue el algoritmo con mayor precisión y exactitud. Los algoritmos de clasificación Naïve Bayes y Bosque Aleatorio fueron encontrados como los siguientes más precisos después de SVM, respectivamente. La investigación muestra que el tiempo necesario para construir un modelo y la precisión (exactitud)	Osisanwo <i>et al.</i> (2017)

5		Supervisado (ML), compara varios algoritmos de aprendizaje supervisado y determina el algoritmo de clasificación más eficiente según el conjunto de datos, el número de instancias y variables (características). Se consideraron siete algoritmos diferentes de aprendizaje automático: Naïve Bayes (NB), Tabla de Decisión, Redes Neuronales, Bosque Aleatorio (RF), Máquina de Vectores de Soporte (SVM), (Perceptrón), JRip y Árbol de Decisión (J48) utilizando la herramienta de aprendizaje automático (WEKA). Para implementar los algoritmos, se utilizó un set de datos de Diabetes, que consta de 786 instancias con ocho atributos como variables independientes y uno como variable dependiente para el análisis.	son un factor importante, mientras que la estadística kappa y el Error Absoluto Medio (MAE) son otro factor crucial. Por lo tanto, los algoritmos de ML requieren precisión, exactitud y error mínimo para tener un aprendizaje automático predictivo supervisado.	
6	A review of supervised machine learning algorithms	El aprendizaje automático supervisado implica construir algoritmos que generan patrones e hipótesis utilizando instancias externas para predecir futuras instancias. Los algoritmos de clasificación supervisada buscan categorizar datos basándose en información previa, una tarea común en ciencia de datos.	Se han propuesto diversas técnicas exitosas como las basadas en reglas, lógica, instancias y estocásticas para resolver estos problemas. Este artículo discute la eficacia en términos de precisión, velocidad de aprendizaje, complejidad y riesgo de sobreajuste. Su objetivo principal es ofrecer una comparación general con los algoritmos más avanzados de aprendizaje automático.	Singh et al. (2016)

7	Machine learning and data mining in manufacturing	Este artículo ofrece una revisión exhaustiva sobre cómo aplicar técnicas de aprendizaje automático (ML) para implementar acciones inteligentes en los procesos de manufactura, destacando su potencial beneficioso para abordar los desafíos en la industria. Además, identifica preguntas de investigación significativas aún no abordadas en la literatura reciente. El estudio organiza estudios previos y avances recientes en manufactura en cuatro áreas principales: programación, monitoreo, calidad y falla. Se discuten soluciones existentes en manufactura, incluyendo tareas como clustering, clasificación y regresión, y algoritmos como redes neuronales. Se presentan estadísticas actuales desde diversas perspectivas y se explican las ventajas del uso de técnicas de aprendizaje automático en manufactura, así como estrategias para superar desafíos específicos. Finalmente, se proponen posibles direcciones futuras para la investigación en este campo. Este artículo analizar los aspectos actuales de aplicaciones de aprendizaje automático (ML) y minería de datos (DM) desarrolladas en la industria manufacturera. Se identificaron aplicaciones actuales junto con técnicas adecuadas para realizar las tareas requeridas.	Las aplicaciones de aprendizaje automático probablemente seguirán creciendo a un ritmo más alto, particularmente en la manufactura, debido a que el poder de computación aumenta día a día y la cantidad de datos disponibles es mucho mayor que hace unos años. Las técnicas y tecnologías de big data pueden procesar datos de alta dimensionalidad. Especialmente en lo que respecta a la creciente disponibilidad de datos de manufactura, es muy probable que estas se vuelvan aún más importantes en el futuro. Se presentan brevemente diferentes tipos de aprendizaje automático, como supervisado (clasificación y regresión), no supervisado (clustering, ARM, SPM, detección de anomalías), aprendizaje conjunto y aprendizaje profundo, con sus aplicaciones en el sector manufacturero.	Dogan & Birant (2021)
8	A review of machine learning for the optimization of production processes	La digitalización en la industria manufacturera ha generado un gran interés en integrar métodos de aprendizaje automático y optimización para mejorar los procesos de producción y ahorrar recursos. Este estudio revisa la literatura de 2008 a 2018 sobre estos métodos aplicados a la calidad del producto y la mejora de procesos.	La revisión revela poca correlación entre los datos utilizados, la cantidad de datos, los algoritmos de aprendizaje automático, los optimizadores y los problemas de producción. El artículo discute estas correlaciones, el progreso reciente y los problemas no resueltos en el área.	Weichert <i>et al.</i> (2019)
	Hydrogen production via	Se investigó la producción de hidrógeno limpio a través de la gasificación de biomasa utilizando algoritmos de	El hidrógeno alcanzó una concentración máxima del 35% en volumen, con un valor	Ozbas et al (2019)

9	biomass gasification, and modeling by supervised machine learning algorithms	aprendizaje automático supervisado. Los experimentos se realizaron en un gasificador con un separador ciclónico, aplicando oxígeno puro y aire seco a diferentes caudales para producir gas de síntesis. Se emplearon varios modelos de regresión (lineal, KNN, SVMR y árboles de decisión) para predecir la concentración de H ₂ basándose en parámetros como tiempo, temperatura, CO, CO ₂ , CH ₄ , O ₂ y valor calorífico.	calorífico de aproximadamente 3400 kcal/m ³ . Las tasas de flujo óptimas fueron 0.05 L/min para aire seco y 0.015 L/min para oxígeno puro. Los modelos mostraron una alta precisión con valores de r ² de 0.99, indicando una excelente correlación entre los resultados reales y los predichos.	
10	An integrated decision analytic framework of machine learning with multi-criteria decision making for multi-attribute inventory classification	El estudio desarrolla una metodología híbrida que integra algoritmos de aprendizaje automático con técnicas de toma de decisiones multicriterio (MCDM) para realizar un análisis efectivo de inventario multiatributo. Se utilizan análisis ABC con tres métodos MCDM (peso simple aditivo, proceso jerárquico analítico y VIKOR) para clasificar los artículos de inventario. Luego, se aplican algoritmos como Bayes ingenuo, redes bayesianas, redes neuronales artificiales (ANN) y máquinas de vectores de soporte (SVM) para predecir las clases de los artículos. Un estudio de caso en una empresa automotriz mostró que SVM obtuvo la mejor precisión de clasificación.	Los resultados demostraron que las redes bayesianas, SVM y ANN manejan exitosamente los problemas de datos desequilibrados asociados con la distribución de Pareto, validando la aplicabilidad de los algoritmos de aprendizaje automático en la clasificación de inventario.	Kartal et al. 2016
11	Predicting Stock Exchange using Supervised Learning Algorithms	El estudio del precio del mercado de valores destaca en informática, economía, finanzas y administración. La predicción de acciones mediante Aprendizaje Automático busca determinar el valor futuro del capital en transacciones financieras. El artículo describe la adopción creciente de técnicas de inteligencia artificial para esta tarea. Se enfoca en identificar el mejor modelo de predicción, evaluando SVM, Random Forest, KNN y Naïve Bayes Bernoulli.	Los resultados muestran que SVM es superior para grandes conjuntos de datos, mientras que Random Forest y Naïve Bayes son mejores para conjuntos pequeños. Esta capacidad de predicción promete ser un recurso valioso para las instituciones financieras, ofreciendo soluciones efectivas para los inversores.	Jyothirmayee et al. (2021)
	Aportaciones para el diagnóstico y	El proyecto de tesis doctoral se enfoca en aplicar técnicas de Inteligencia Artificial para resolver problemas de diagnóstico y pronóstico en sectores	La medición de la basicidad del aceite (TBN) en laboratorio es costosa en tiempo y recursos, por lo que se requiere un modelo	Ferreiro (2012)

12	pronóstico en problemas industriales mediante técnicas de clasificación supervisada	industriales como la aeronáutica y la química. Los problemas específicos abordados incluyen la detección y pronóstico del desgaste de frenos de aviones, la detección de rebabas durante el taladrado en las alas de aviones, y la estimación de la basicidad del aceite (TBN) como indicador de degradación. En el sector aeronáutico, se está adoptando un enfoque de mantenimiento basado en la predicción, integrando diagnósticos y pronósticos para evaluar y prever la degradación de componentes críticos como los frenos.	que estime este indicador para determinar la degradación del aceite y decidir si es necesario cambiarlo. Para resolver estos problemas, se emplean técnicas de Clasificación Supervisada, evaluando diversos algoritmos como Redes Bayesianas, árboles de clasificación, reglas de clasificación y algoritmos basados en la distancia. En algunos casos, estos algoritmos se combinan con técnicas como algoritmos evolutivos o análisis estadísticos para mejorar la precisión de los modelos.	
13	Técnicas de clasificación supervisadas para la detección de anomalías en el control de procesos industriales	En la actualidad, la detección de anomalías en procesos industriales es fundamental para optimizarlos y aumentar la eficiencia del proceso productivo, lo que resulta en beneficios significativos para las empresas. Por ello, este artículo implementa cinco técnicas de clasificación supervisada para la detección de anomalías en sistemas industriales. Estas técnicas se entrenaron y validaron utilizando un conjunto de datos que incluía datos etiquetados de funcionamiento normal y anómalo de una planta de control de nivel de líquido en un depósito. Finalmente, se analizaron y compararon los resultados obtenidos para identificar el modelo que ofrece el mejor rendimiento.	Los métodos de árbol de decisión y bosques aleatorios demostraron ser efectivos en la detección de anomalías en la planta de control de nivel, alcanzando valores de AUC superiores al 97% con las mejores configuraciones de hiperparámetros. El método de vecinos más cercanos también obtuvo buenos resultados, con un máximo de AUC del 87.7%. En contraste, las máquinas de vectores de soporte y el discriminante lineal de Fisher no fueron tan efectivos, ya que no superaron el 80% de AUC en ninguna configuración.	Michelena (2012)
14	Clasificación jerárquica de texto con machine learning en la industria petrolera	La clasificación jerárquica de texto es una tarea de aprendizaje automático que permite categorizar grandes volúmenes de documentos mediante una estructura en forma de árbol. Los algoritmos de clasificación de texto en ciencia de datos facilitan la construcción de modelos de aprendizaje automático que predicen la categoría de un documento dentro de una jerarquía de clases predefinidas. Las bibliotecas gratuitas y de código abierto en ciencia de datos	la clasificación jerárquica de texto mediante machine learning sienta las bases para abordar problemas de clasificación documental en diversos sectores, como la industria, instituciones educativas o entidades gubernamentales. Además, el acceso a herramientas gratuitas de ciencia de datos en Python simplifica y alienta el desarrollo de proyectos tecnológicos e	Saldaña (2023)

		disponibles actualmente fomentan el desarrollo de estos proyectos. Este artículo destaca la utilidad de estas herramientas mediante la creación de un modelo de aprendizaje automático supervisado con Python.	investigaciones que requieran la aplicación de la inteligencia artificial	
15	A Machine Learning-Based Comparative Approach to Predict the Crop Yield Using Supervised Learning With Regression Models	La agricultura es crucial para la existencia humana y el crecimiento nacional, proporcionando alimentos, materias primas y empleos. En India, es la principal fuente de ingresos para la mayoría y ha evolucionado con tecnología moderna. Integrar tecnología en la agricultura es vital. El aprendizaje automático, que analiza grandes volúmenes de datos, ayuda a identificar oportunidades y riesgos. Este proyecto de investigación desarrolla un modelo de aprendizaje automático para predecir la producción agrícola, utilizando seis modelos de regresión. El Random Forest Regressor destacó, con un error absoluto medio (MAE) de 468.16 y una puntuación de validación cruzada de 0.6087, superando a los demás modelos.	Esta investigación se centró en la adaptación de técnicas de Machine Learning (ML) para estimar el rendimiento de los cultivos, utilizando seis modelos de regresión supervisada: Regresión Lineal, Árbol de Decisión, Gradient Boosting, Random Forest, XGBoost y Voting. Los modelos de Random Forest y XGBoost fueron los más precisos, aunque XGBoost presentó sobreajuste. Tras ajustar los parámetros, Random Forest superó a XGBoost, logrando un MAE de 468.16, un MSE de 825.29, un R2 de 0.7952 y una validación cruzada de 0.6087. Además, se destacó el uso de técnicas de Inteligencia Artificial (IA), Deep Learning (DL) y Visión por Computadora (CV) para analizar imágenes del campo y de los cultivos, permitiendo identificar enfermedades y maleza, y separarlos de los cultivos sanos lo antes posible.	Panigrahi et al 2023
16	Application of artificial intelligence and machine learning for BIM: review	El control de calidad es fundamental en los flujos de trabajo del Modelado de Información de Construcción (BIM). A lo largo del ciclo de vida del edificio, es vital seguir sus indicadores. El BIM, que maneja grandes volúmenes de datos, requiere herramientas avanzadas para su análisis, desde procesamiento de imágenes hasta big data. La inteligencia artificial (IA) y el aprendizaje automático (ML) han demostrado ser eficaces en la automatización de procesos y extracción	La revisión destaca el objetivo del BIM y la integración de la IA en él, fundamental en proyectos de construcción y paisajismo. La IA en BIM y los Gemelos Digitales mejora la industria. Los sistemas de "resolución de problemas" y "aprendizaje" aseguran alta precisión en la detección de patrones y predicciones con grandes volúmenes de datos. En informática, la IA estudia "agentes	Bassir et al (2023)

		de datos útiles en diversas industrias. Además de seguir indicadores, la IA y el ML pueden predecir cuándo y dónde realizar mantenimiento y control de calidad. Este artículo revisa las aplicaciones de IA y ML en BIM y discute sugerencias y desafíos adicionales, proporcionando una comprensión amplia de su impacto en la construcción y estudios futuros.	inteligentes" que perciben y comprenden su entorno, maximizando sus probabilidades de éxito. Con el avance tecnológico, la implementación efectiva de IA se ha vuelto más accesible, pero persisten retos éticos y de conocimiento en AECO. La investigación en IA, BIM y Gemelos Digitales promete mejorar las tecnologías de construcción y paisajismo actuales.
--	--	--	--

Nota. La tabla presenta un compendio de investigaciones de aprendizaje supervisado aplicado a la industria

5.1.2. Análisis de Trabajos Identificados

Al tener en cuenta la información seleccionada mediante la revisión documental, a continuación, se presenta un análisis de los trabajos en investigaciones tomadas como base. Cabe resaltar que se incluyen las investigaciones del estado del arte, de modo que se buscó destacar las técnicas empleadas, el área de la industria en el que se aplicó y las metodología u objetivo del estudio.

Tabla 5. Análisis de los trabajos Identificados

Análisis de los trabajos Identificados de Aprendizaje Supervisado en la Industria					
Nº	Título	Técnicas	Área	Metodología / Resumen	Autor
1	An overview of the supervised machine learning methods	Árbol de decisión, regresión lineal, Naive Bayes, regresión logística	Procesamiento y análisis de datos	Resumen general de métodos supervisados, describiendo fundamentos y aplicaciones	Nasteski (2018)

2	Continuous quality control evaluation during manufacturing using supervised learning algorithm for Industry 4.0	SVM, redes neuronales RBF	Control de calidad en manufactura inteligente	Comparación de SVM y RBF en control de calidad, evaluando funciones de kernel y coeficientes gamma.	Shafiq et al. (2023)
3	Risk prediction in life insurance industry using supervised learning algorithms	Regresión Lineal Múltiple, Red Neuronal Artificial, REPTree, Random Tree	Evaluación de riesgos en seguros de vida	Reducción de dimensionalidad con CFS y PCA, aplicación de varios algoritmos para predecir riesgos.	Boodhun & Jayabalan (2018)
4	Semi-supervised learning for industrial fault detection and diagnosis: A systemic review	Algoritmos semi-supervisados	Detección y Diagnóstico de Fallas (FDD)	Revisión sistemática de la literatura sobre SSL para FDD.	Ramírez et al. (2023)
5	Supervised Machine Learning Algorithms: Classification and Comparison	Tabla de Decisión, Bosque Aleatorio, Naïve Bayes, SVM, Redes Neuronales, JRip, J48	Clasificación de datos	Comparación de varios algoritmos usando un conjunto de datos de Diabetes.	Osisanwo et al. (2017)
6	A review of supervised machine learning algorithms	Algoritmos basados en reglas, lógica, instancias y estocásticas	Clasificación de datos	Revisión general de algoritmos supervisados evaluando precisión, velocidad y complejidad.	Singh et al. (2016)
7	Machine learning and data mining in manufacturing	Clustering, clasificación, regresión, SVM, redes neuronales	Procesos de manufactura	Revisión de técnicas y algoritmos en manufactura.	Dogan & Birant (2021)
8	A review of machine learning for the optimization of production processes	Diversos algoritmos de aprendizaje automático y optimización	Optimización de procesos de producción	Revisión de literatura sobre métodos de aprendizaje automático aplicados a la calidad del producto y mejora de procesos.	Weichert et al. (2019)
9	Hydrogen production via biomass gasification, and modeling by supervised machine learning algorithms	Regresión lineal, KNN, SVMR, árboles de decisión	Producción de hidrógeno limpio	Modelado de la concentración de H2 utilizando varios modelos de regresión basados en parámetros experimentales.	Ozbas et al. (2019)

10	An integrated decision analytic framework of machine learning with multi-criteria decision making for multi-attribute inventory classification	Naïve Bayes, redes bayesianas, ANN, SVM	Clasificación de inventario	Integración de algoritmos de aprendizaje automático con técnicas MCDM para la clasificación de inventario en una empresa automotriz.	Kartal et al. (2016)
11	Predicting Stock Exchange using Supervised Learning Algorithms	SVM, Random Forest, KNN, Naïve Bayes Bernoulli	Predicción del mercado de valores	Evaluación de varios algoritmos de predicción para determinar el modelo más efectivo en la predicción de precios de acciones.	Jyothirmayee et al. (2021)
12	Aportaciones para el diagnóstico y pronóstico en problemas industriales mediante técnicas de clasificación supervisada	Redes Bayesianas, árboles de clasificación, reglas de clasificación, algoritmos basados en la distancia	Diagnóstico y pronóstico en sectores industriales	Aplicación de técnicas de inteligencia artificial para detectar y pronosticar problemas en aeronáutica y química.	Ferreiro (2012)
13	Técnicas de clasificación supervisadas para la detección de anomalías en el control de procesos industriales	Árbol de decisión, bosques aleatorios, vecinos más cercanos, SVM, discriminante lineal de Fisher	Detección de anomalías en procesos industriales	Entrenamiento y validación de técnicas supervisadas para identificar el mejor modelo en la detección de anomalías.	Michelena (2012)
14	Modelo de planeación de inventarios para E-commerce, utilizando herramientas de inteligencia artificial para hacer pronósticos de demanda y clasificación de inventarios	Regresión Lineal Múltiple, Bosques Aleatorios, SVM, K-means, Clustering Jerárquico, DBSCAN, Modelo de Mezcla Gaussiana	Planeación de inventarios en E-commerce	Comparación de métodos de pronóstico y clasificación de inventarios, destacando el uso de inteligencia artificial sobre técnicas tradicionales.	Vesga (2021)
15	Algoritmo conceptual para la clasificación supervisada en el reconocimiento lógico combinatorio de patrones	Algoritmo ACC basado en agrupamiento conceptual	Toma de decisiones en diversas áreas del conocimiento	Desarrollo y validación experimental del algoritmo ACC, comparado con algoritmos clásicos de RLCP.	Figueroa (2017)

16	Algoritmos de aprendizaje supervisado utilizando datos de monitoreo de condiciones: Un estudio para el pronóstico de fallas en máquinas	Redes neuronales profundas LSTM, Gradient Boosting, XGBoost	Detección de fallas en activos físicos	Implementación de algoritmos supervisados en clasificación y regresión, ajustando arquitecturas de redes neuronales y comparando modelos.	Huertas (2020)
17	Artificial intelligence applications in supply chain: A descriptive bibliometric analysis and future research directions	Aprendizaje supervisado, no supervisado y por refuerzo	Cadenas de suministro	Revisión sistemática de la literatura y análisis bibliométrico de 136 artículos sobre aplicaciones de IA en la cadena de suministro.	Richi et al. (2021)
18	Clasificación jerárquica de texto con machine learning en la industria petrolera	Clasificación jerárquica de texto	Industria petrolera	Desarrollo de un modelo de clasificación jerárquica de texto utilizando Python y bibliotecas de código abierto.	Saldaña (2023)
19	A Machine Learning-Based Comparative Approach to Predict the Crop Yield Using Supervised Learning With Regression Models	Regresión Lineal, Árbol de Decisión, Gradient Boosting, Random Forest, XGBoost, Voting	Predicción de rendimiento agrícola	Análisis comparativo de modelos de regresión supervisada para predecir el rendimiento de cultivos.	Panigrahi et al. (2023)
20	Estudio y análisis de algoritmos de clasificación supervisada para la identificación de enfermedades en plantas de papa	Algoritmos supervisados para el análisis de imágenes	Agricultura (enfermedades en plantas de papa)	: Implementación de algoritmos de aprendizaje supervisado para identificar enfermedades en plantas mediante imágenes.	Luna (2022)
21	Automatic Risk Assessment for an Industrial Asset Using Unsupervised and Supervised Learning	Aprendizaje supervisado y no supervisado, redes neuronales artificiales, análisis de agrupamiento K-medias, PCA.	Supervisión y mantenimiento predictivo en activos industriales	Evaluación de diferentes técnicas de aprendizaje para la detección de fallas y la evaluación de riesgos en la gestión de activos.	Rodrigues et al. (2022)
22	A Survey on Supervised Learning Techniques for the Classification of Emotions in Text	Clasificación supervisada	Análisis de emociones en texto	Revisión exhaustiva de técnicas de aprendizaje supervisado para la clasificación de emociones en texto.	Ribeiro et al. (2018)

23	Machine Learning for Financial Risk Management: A Survey	Diversos algoritmos supervisados	Gestión de riesgos financieros	Revisión de aplicaciones de aprendizaje supervisado en la gestión de riesgos financieros, con énfasis en la evaluación de modelos.	Petropoulos et al. (2021)
24	Machine Learning Approaches in Predictive Maintenance of Aircraft Systems	Regresión lineal, Random Forest, SVM, KNN	Uso de infraestructura de carga de vehículos eléctricos	Aplicación y comparación de varios algoritmos supervisados para predecir el uso de infraestructura de carga de vehículos eléctricos.	Dou et al. (2021)
25	Applying Supervised Machine Learning Algorithms for Predicting Electric Vehicle Charging Infrastructure Usage	Regresión lineal, Random Forest, SVM, KNN	Regresión lineal, Random Forest, SVM, KNN	Aplicación y comparación de varios algoritmos supervisados para predecir el uso de infraestructura de carga de vehículos eléctricos.	Johnson et al. (2021)

Fuente: Autores

Al tener en cuenta cada uno de los trabajos identificados con relación a aprendizaje supervisado que se muestra en la Tabla 5, a continuación, se presenta un análisis de los resultados identificados:

Los resultados mostraron que las técnicas Populares son SVM, Random Forest, y Redes Neuronales son las técnicas más frecuentemente utilizadas en diversas aplicaciones industriales.

Las áreas de aplicación técnicas de aprendizaje supervisado abarcan trabajos en una amplia gama de áreas industriales, que incluyen manufactura, evaluación de riesgos, predicción de fallas, gestión de inventarios, y predicción del mercado de valores.

Como parte de las metodologías implementadas, la mayoría de los estudios emplean enfoques comparativos para evaluar la efectividad de diferentes algoritmos de aprendizaje supervisado en sus respectivas áreas de aplicación.

Al hablar de la innovación y desarrollo, se observa un creciente interés en integrar técnicas de aprendizaje supervisado con otros enfoques, como el aprendizaje no supervisado y la toma de decisiones multicriterio, para mejorar la precisión y eficiencia de los modelos.

Tabla 6. Técnicas y Resultados por Autor

Técnicas y Resultados por Autor			
Nº	Técnica	Resultado	Autor
1	Árbol de decisión, regresión lineal, Naive Bayes, regresión logística	Estas técnicas son fundamentales en el aprendizaje supervisado y se utilizan ampliamente para diversas tareas de clasificación y regresión.	Nasteski (2018)
2	SVM (máquina de soporte vectorial), RBF NNs (Redes Neuronales RBF)	VM es más eficaz y precisa que las RBF NNs para el control de calidad en industrias manufactureras. SVM ofrece mayor velocidad de entrenamiento y capacidad para resolver problemas a gran escala.	Shafiq et al. (2023)
3	REPTree, Regresión Lineal Múltiple.	REPTree tuvo mejor rendimiento con CFS, mientras que la Regresión Lineal Múltiple fue más efectiva con PCA.	Boodhun & Jayabalan (2018)
4	Aprendizaje Semi-Supervisado (SSL)	Los primeros intentos de SSL mostraron inconsistencias, como la mezcla incorrecta de métodos supervisados y no supervisados y el uso de un alto porcentaje de instancias etiquetadas.	Ramírez et al. (2023)
5	SVM, Naïve Bayes, Bosque Aleatorio.	SVM tuvo la mayor precisión y exactitud, seguido por Naïve Bayes y Bosque Aleatorio. La precisión y el error mínimo son cruciales para los algoritmos de Machine Learning.	Osisanwo et al. (2017)
6	Técnicas basadas en reglas, lógica, instancias y estocásticas	Se discuten la precisión, velocidad de aprendizaje, complejidad y riesgo de sobreajuste de estas técnicas	Singh et al. (2016)
7	Aprendizaje supervisado (clasificación y regresión), no supervisado (clustering, ARM, SPM, detección de anomalías).	El aprendizaje automático en manufactura es cada vez más relevante gracias a la creciente disponibilidad de datos y poder de computación.	Dogan & Birant (2021)
8	Diversos algoritmos de aprendizaje automático.	Se discuten las correlaciones entre datos, algoritmos y problemas de producción, así como problemas no resueltos en la investigación.	Weichert et al. (2019):
9	Modelos de alta precisión con r^2 de 0.99.	Alta correlación entre resultados reales y predichos en la medición de concentración de hidrógeno.	Ozbas et al. (2019)

10	Redes bayesianas, SVM, ANN.	Estos algoritmos manejan con éxito problemas de datos desequilibrados asociados con la distribución de Pareto.	Kartal et al. (2016)
11	SVM, Random Forest, Naïve Bayes.	SVM es superior para grandes conjuntos de datos, mientras que Random Forest y Naïve Bayes son mejores para conjuntos pequeños.	Jyothirmayee et al. (2021)
12	Redes Bayesianas, árboles de clasificación, reglas de clasificación, algoritmos basados en la distancia.	Evaluación de diversos algoritmos para medir la basicidad del aceite (TBN) en laboratorios.	Ferreiro (2012)
13	Árbol de decisión, Bosques aleatorios, vecinos más cercanos, SVM, discriminante lineal de Fisher.	Los métodos de árbol de decisión y bosques aleatorios fueron los más efectivos en la detección de anomalías.	Michelena (2012)
14	Algoritmo de clasificación supervisada ACC.	El algoritmo ACC demostró eficacia en la validación experimental y comparación con otros algoritmos tradicionales.	Figueroa (2017)
15	Redes neuronales profundas (LSTM), Gradient Boosting, XGBoost.	LSTM y algoritmos de conjunto demostraron rendimiento excelente en problemas secuenciales.	Huertas (2020)
16	Clasificación jerárquica de texto.	Herramientas gratuitas de ciencia de datos en Python simplifican la aplicación de IA en diversos sectores.	Saldaña (2023)
17	Regresión Lineal, Árbol de Decisión, Gradient Boosting, Random Forest, XGBoost, Voting.	Random Forest y XGBoost fueron los más precisos en la estimación del rendimiento de cultivos.	Panigrahi et al. (2023)
18	IA en BIM y Gemelos Digitales	IA mejora la industria de construcción y paisajismo, aunque persisten retos éticos y de conocimiento.	Bassir et al. (2023)
19	Aprendizaje supervisado y no supervisado	Clasificación del estado de máquinas industriales utilizando redes neuronales artificiales.	Rodrigues et al. (2022)
20	SVM.	IA en agricultura de precisión, especialmente para la detección de enfermedades y predicción de siembra.	Coppiano & Herrera (2022)
21	Aprendizaje por transferencia, autoaprendizaje supervisado	Estas técnicas abordan el desafío de la falta de datos de entrenamiento etiquetados.	Zhao et al. (2024)
22	Aprendizaje Profundo, aprendizaje no supervisado.	ML es crucial en la evolución de sistemas de manufactura hacia la Industria 4.0.	Bertolini et al. (2021)
23	Diversos algoritmos de ML.	Los algoritmos de ML son efectivos en el diagnóstico de fallos, predicción de defectos y optimización de sistemas en la industria.	Figueroa et al. (2023)

Fuente: Autores

Al tener en cuenta la Tabla 6, las estrategias de aprendizaje supervisado, como SVM, árboles de decisión, regresión lineal, Naive Bayes, y redes neuronales,

son ampliamente utilizadas y efectivas en diversos campos, incluyendo la manufactura, agricultura, detección de fallos y predicción de rendimientos. SVM destaca por su precisión y capacidad para manejar grandes conjuntos de datos, mientras que los árboles de decisión y los algoritmos de conjunto como Random Forest y XGBoost son valorados por su rendimiento en tareas específicas. La integración de IA y técnicas de ML continúa creciendo, ofreciendo mejoras significativas en múltiples industrias.

5.2. COMPARACIÓN DE LOS DIFERENTES ALGORITMOS DE APRENDIZAJE SUPERVISADO

Al tomar como base los trabajos previamente identificados a continuación se buscó realizar una comparación entre los diferentes algoritmos de aprendizaje supervisado, con el fin de evaluar su idoneidad para el modelado de los procesos industriales.

En términos generales en el trabajo de Nasteski (2018) y Singh et al. (2016) se presentó una visión general de los métodos supervisados, describiendo sus fundamentos y aplicaciones. Otros estudios como Shafiq et al. (2023) y Boodhun & Jayabalan (2018) aplican y comparan técnicas específicas en contextos prácticos, como el control de calidad y la evaluación de riesgos.

Al tomar en cuenta a Shafiq et al. (2023) y Osisanwo et al. (2017) ofrecieron comparaciones detalladas y cuantitativas de algoritmos específicos, evaluando su desempeño en tareas prácticas. Por otra parte, las investigaciones de Ramírez et al. (2023) y Richi et al. (2021) realizaron revisiones sistemáticas de la literatura, al proporcionar una evaluación cualitativa de los algoritmos y su aplicación en diversas.

5.2.1. Identificación de Algoritmos de Aprendizaje Supervisado en la Industria

Para evaluar la idoneidad de diferentes algoritmos de aprendizaje supervisado en el modelado de procesos industriales, se deben considerar varios factores, como la precisión, la eficiencia, la capacidad de manejo de datos, la interpretabilidad y la robustez del modelo. A continuación, se presenta una comparación de algunos algoritmos comunes de aprendizaje supervisado,

destacando sus características, ventajas y desventajas en el contexto de procesos industriales, a partir de la información desarrollada por Andrade *et al.*, 2019)

La Tabla 7 se presentan cada uno de los modelos de aprendizaje supervisado que han sido implementados en la industria.

Tabla 7. Principales Algoritmos de Aprendizaje Supervisado Implementados en la Industria

Nº	Algoritmo	Descripción	Autor
1	Regresión Lineal	Es una metodología de análisis predictivo que busca definir la relación entre una variable dependiente, que toma valores continuos, y una o más variables independientes. Esto se logra trazando una línea recta que mejor se ajuste a los datos disponibles. El criterio utilizado para determinar el mejor ajuste es la minimización de la suma de los cuadrados de las diferencias entre los valores observados en el conjunto de datos y los valores que el modelo predice. Aplicaciones: Predicción de demanda de productos. Estimación de tiempos de producción	(Nasteski, An overview of the supervised machine learning, 2017)
2	Árboles de Decisión	Es un algoritmo de aprendizaje automático basado en el método de conjunto, que construye múltiples árboles de decisión de manera aleatoria para mejorar la precisión y robustez de las predicciones. En este enfoque, cada árbol realiza una predicción individual y la clase final es determinada por un proceso de votación, donde la clase que reciba más votos es seleccionada como la predicción final del modelo. Aplicaciones: Clasificación de inventarios. Detección de anomalías en procesos de producción.	(Osisanwo, Akinsola, Awodale, O, & O., 2017)
3	Bosques Aleatorios (Random Forest)	Un bosque aleatorio es un algoritmo de aprendizaje conjunto que construye múltiples árboles de decisión de manera aleatoria y los combina para producir una predicción más precisa y robusta. Cada árbol en el bosque vota por una clase y la clase con más votos es la predicción del modelo.	(Islam & Hassanzadeh, 2020)

4	Máquinas de Soporte Vectorial (SVM)	Las máquinas de soporte vectorial (SVM) son algoritmos de aprendizaje supervisado que se utilizan para clasificación y regresión. Funcionan encontrando el hiperplano que mejor separa las diferentes clases en el espacio de características, maximizando el margen entre las clases. Aplicaciones: Clasificación de niveles de inventario. Predicción de demanda.	(Figueroa Donayre, Guillén Guevara, & Flores, 2023)
5	Redes Neuronales Artificiales (ANN)	Las redes neuronales artificiales (ANN) son modelos de aprendizaje inspirados en la estructura y el funcionamiento del cerebro humano. Están formadas por capas de nodos (neuronas) interconectados, donde cada conexión tiene un peso ajustable. Las ANN son capaces de aprender y modelar relaciones complejas y no lineales a partir de los datos. Aplicaciones: Predicción de demanda de productos Optimización de tiempos de producción.	(Kang, Catal, & Tekinerdogan, 2020)
6	Gradient Boosting Machines (GBM) y XGBoost	Gradient Boosting Machines (GBM) y su implementación optimizada XGBoost son métodos de aprendizaje conjunto que crean un modelo predictivo a partir de un conjunto de modelos base débiles, normalmente árboles de decisión. Cada modelo nuevo se ajusta para corregir los errores residuales de los modelos anteriores, mejorando así la precisión general. Aplicaciones: Predicción de demanda. Clasificación de inventarios.	(Islam & Hassanzadeh, 2020)
7	K-Nearest Neighbors (KNN)	El algoritmo K-Nearest Neighbors (KNN) es un método de clasificación y regresión basado en la similitud de las muestras. Para clasificar una nueva muestra, el algoritmo busca las K muestras de entrenamiento más cercanas en el espacio de características y asigna la clase más común (o promedio para regresión) de estos vecinos cercanos a la nueva muestra. Aplicaciones: Clasificación de inventarios. Predicción de demanda.	(Cebekhulu, Onumanyi, & Sherrin, 2022)

Fuente: Autores

Como se puede determinar, las investigaciones realizadas presentan de forma concreta cada uno de los métodos que identifican a cada algoritmo para aplicaciones en la industria. La elección e implementación de cada algoritmo dependerá de la finalidad y características del problema a resolver. En la Tabla 8 se presentan los algoritmos seleccionados de aprendizaje supervisado.

5.2.2. Características de los modelos de Aprendizaje Supervisado

Tabla 8. Características de Algoritmos de Aprendizaje Supervisado

Características de Algoritmos de Aprendizaje Supervisado			
Nº	Algoritmo	Características	Autor
1	Regresión Lineal y Regresión Lineal Múltiple	Modelos simples que asumen una relación lineal entre las variables independientes y la dependiente. Fácil de interpretar y explicar.	(van Engelen & Hoos, 2020)
2	Árboles de Decisión	Árboles de decisión dividen los datos en subconjuntos homogéneos basados en características.	(Jijo & Abdulazeez, 2021)
3	Bosques Aleatorios (Random Forest)	combina múltiples árboles para mejorar la precisión y reducir el sobreajuste.	(Kiangala & Wang, 2021)
4	Máquinas de Soporte Vectorial (SVM)	Encuentra el hiperplano óptimo que separa las clases en el espacio de características. Puede utilizar diferentes funciones de kernel para manejar relaciones no lineales.	(Bertolini, Mezzogori, & Neroni, 2021)
5	Redes Neuronales Artificiales (ANN)	Modelos complejos inspirados en el cerebro humano, capaces de aprender patrones complejos. Capaces de manejar datos no lineales y de alta dimensionalidad.	(Pavan & Akshit, 2021)
6	Gradient Boosting Machines (GBM) y XGBoost	Construyen árboles de decisión secuencialmente, corrigiendo los errores del árbol anterior. Muy precisos y eficientes en muchos tipos de datos.	(Praveen, Prateek, Pradyumna, & Pragathi, 2020)
7		Algoritmo basado en la distancia que clasifica una muestra en función de las etiquetas de sus vecinos más cercanos.	(Raizada & Saini, 2021)

	K-Nearest Neighbors (KNN)	Simple y fácil de entender.	
--	------------------------------	-----------------------------	--

5.2.3. Ventajas y desventajas de los Algoritmos de Aprendizaje Supervisado

Tabla 9. Ventajas y Desventajas de cada Algoritmo

	Algoritmo	Ventajas	Desventajas	Autor
1	Regresión Lineal	Rápido de entrenar y evaluar. Buena interpretabilidad.	Limitado a relaciones lineales; puede no capturar la complejidad de los datos industriales. Sensible a outliers.	(Mahesh, 2018)
2	Árboles de Decisión y Bosques Aleatorios (Random Forest)	Buena capacidad para capturar relaciones no lineales. Manejo de datos categóricos y numéricos. Robustez frente a outliers.	Menos interpretables que modelos lineales. Requieren más tiempo de entrenamiento.	(Jijo & Abdulazeez, 2021)
3	Máquinas de Soporte Vectorial (SVM)	Alta precisión en problemas de clasificación. Eficiente en espacios de alta dimensionalidad.	Requiere tuning de hiperparámetros. Menos interpretables.	(Cervantes, García Lamont, Mazahua, Lisbeth, & Lopez, 2020)
4	Redes Neuronales Artificiales (ANN)	Alta precisión para problemas complejos. Flexible y capaz de manejar grandes volúmenes de datos.	Largo tiempo de entrenamiento. Difícil de interpretar (caja negra).	(Bertolini, Mezzogori, & Neroni, 2021)
5	Gradient Boosting Machines (GBM) y XGBoost	Alta precisión y eficiencia. Capacidad para manejar datos no lineales.	Requiere tuning de hiperparámetros. Mayor tiempo de entrenamiento en comparación con árboles individuales	(Aziz, Patah Akhir, Aziz, Jaafar, & Hasan, 2020)
6	K-Nearest Neighbors (KNN)	No requiere suposiciones sobre la distribución de los datos. Fácil de implementar.	Ineficiente en grandes volúmenes de datos. Sensible a la elección del número de vecinos (K) y la escala de las características.	(Bansal, Goyal, & Choudhary, 2022)

Fuente: Autores

5.2.4. Criterios y Evaluación de Idoneidad de Algoritmos de Clasificación de Aprendizaje supervisado

Para comparar los algoritmos de aprendizaje supervisado en la industria, se utilizaron varios parámetros o criterios clave. Estos criterios fueron seleccionados al tener en cuenta investigaciones como la de Ensafi & Amin (2022), Feizabadi (2022) y Ahmed & Chakraborty (2024), los cuales tuvieron en cuenta características que son fundamentales para evaluar la idoneidad y efectividad de los algoritmos en aplicaciones industriales.

5.2.4.1 Descripción de Criterios

Precisión

La precisión mide qué tan bien un modelo predice resultados correctos. Es crucial en aplicaciones industriales donde decisiones basadas en predicciones incorrectas pueden tener costos significativos.

Interpretabilidad

La interpretabilidad se refiere a la facilidad con la que los humanos pueden entender cómo un modelo hace sus predicciones. En entornos industriales, es importante poder explicar las decisiones del modelo a los interesados y comprender las variables que influyen en el resultado.

Complejidad del Modelo

Este criterio evalúa la cantidad de recursos computacionales necesarios para entrenar y ejecutar el modelo. En industrias con grandes volúmenes de datos, los modelos menos complejos pueden ser preferibles por su eficiencia y velocidad.

Robustez

La robustez se refiere a la capacidad del modelo para manejar datos ruidosos y outliers. En aplicaciones industriales, los datos suelen ser imperfectos, por lo que los modelos robustos son esenciales para obtener predicciones fiables.

Capacidad de Generalización

La capacidad de generalización indica cómo se desempeña el modelo en datos no vistos durante el entrenamiento. Es crucial para asegurar que el modelo no esté sobre ajustado y pueda aplicarse efectivamente en diferentes situaciones industriales.

Flexibilidad

La flexibilidad se refiere a la capacidad del algoritmo para manejar distintos tipos de datos y problemas. En la industria, los datos pueden variar ampliamente, por lo que los modelos versátiles son valiosos.

Escalabilidad

La escalabilidad evalúa cómo se comporta el modelo al aumentar la cantidad de datos. Es importante en entornos industriales donde los volúmenes de datos pueden ser muy grandes y seguir creciendo.

Tiempo de Entrenamiento y Predicción

Este criterio considera el tiempo necesario para entrenar el modelo y hacer predicciones. En aplicaciones industriales donde las decisiones deben tomarse

rápidamente, los modelos con tiempos de entrenamiento y predicción más cortos son preferidos.

5.2.4.2 Justificación de los Criterios

Precisión y Capacidad de Generalización: Cruciales para asegurar que el modelo haga predicciones precisas y fiables en datos no vistos.

Interpretabilidad: Importante para la aceptación y confianza de los interesados en la industria, así como para el cumplimiento de regulaciones.

Complejidad del Modelo y Tiempo de Entrenamiento y Predicción: Afectan la viabilidad práctica del uso del modelo en términos de recursos y tiempo.

Robustez: Garantiza que el modelo pueda manejar datos reales que suelen tener ruido e imperfecciones.

Flexibilidad y Escalabilidad: Aseguran que el modelo pueda adaptarse a diferentes tipos de problemas y volúmenes de datos, manteniendo su efectividad a medida que crece la escala de datos.

Para lograr evaluar de forma comparativa cada algoritmo, se emplearon 3 valores para cada criterio, entre los valores de bajo, medio y alto. Al basar la comparación en estos criterios, se obtiene una evaluación integral de la idoneidad de cada algoritmo para su uso en procesos industriales, permitiendo una mejor selección del modelo adecuado para tareas específicas como la predicción de demanda, tiempos de producción y niveles de inventario. En la Tabla 10 se presenta la evaluación desarrollada y de igual forma, una descripción e interpretación.

Tabla 10. Comparación de Algoritmos

Algoritmo	Precisión	Interpretabilidad	Tiempo de Entrenamiento	Robustez	Manejo de No Linealidad
Regresión Lineal	Media	Alta	Bajo	Baja	No
Árboles de Decisión	Media	Media	Medio	Alta	Sí
Random Forest	Alta	Media	Alto	Alta	Sí
SVM	Alta	Baja	Alto	Alta	Sí
Redes Neuronales	Muy Alta	Baja	Muy Alto	Alta	Sí
GBM / XGBoost	Muy Alta	Media	Alto	Alta	Sí
KNN	Media	Alta	Bajo	Baja	Sí

Fuente: Autores

Para evaluar y comparar los algoritmos de aprendizaje supervisado en la industria, se utiliza una escala cualitativa de "Alta", "Media" y "Baja" para cada criterio. A continuación, en la Tabla 11, se describe cómo se interpreta esta escala para cada uno de los criterios seleccionados:

Tabla 11. Interpretación Cualitativa de cada Criterio

Criterio	Alta	Media	Baja
Precisión	Produce predicciones muy exactas con bajo error.	Produce predicciones razonablemente precisas con errores moderados.	Produce predicciones inexactas con errores relativamente altos.
Interpretabilidad	Modelo fácil de entender y explicar.	Modelo con cierta interpretabilidad, requiere esfuerzo adicional.	Modelo difícil de interpretar y explicar.
Complejidad del Modelo	Requiere muchos recursos computacionales y tiempo.	Complejidad moderada con requerimientos razonables de recursos.	Modelo simple, requiere pocos recursos y es rápido de entrenar.
Robustez	Maneja bien datos ruidosos y outliers.	Moderadamente robusto, puede verse afectado por ruido y outliers.	Sensible a datos ruidosos y outliers, reduce precisión y efectividad.

Capacidad de Generalización	Generaliza bien a nuevos datos, evitando el sobreajuste.	Capacidad de generalización razonable, con posible sobreajuste.	Tiende a sobreajustarse, no generaliza bien a nuevos datos.
Flexibilidad	Muy flexible, maneja una amplia variedad de datos y problemas.	Moderadamente flexible, adecuado para ciertos tipos de datos.	Rígido, solo funciona bien para tipos específicos de datos.
Escalabilidad	Escala bien con el aumento del volumen de datos, manteniendo rendimiento.	Escala razonablemente, rendimiento puede degradarse con más datos.	No escala bien, muestra degradación significativa en rendimiento.
Tiempo de Entrenamiento y Predicción	Requiere mucho tiempo para entrenar y predecir.	Tiempo de entrenamiento y predicción razonable.	Rápido de entrenar y predecir.

Fuente: Autores

Esta comparación muestra que no existe un único algoritmo que sea el mejor para todas las aplicaciones. La elección del algoritmo adecuado depende de las necesidades específicas del problema industrial, como la necesidad de precisión, interpretabilidad, y recursos computacionales disponibles. Por ejemplo, para predicciones muy precisas en problemas complejos, los modelos avanzados como las Redes Neuronales y los modelos de Boosting son preferibles. Para aplicaciones que requieren interpretabilidad y rapidez, los modelos más simples como la regresión lineal y los árboles de decisión son más adecuados.

5.3. INTERPRETACIÓN DE MODELOS PARA PROCESOS INDUSTRIALES DE APRENDIZAJE SUPERVISADO

5.3.1. SOFTWARE RESTUDIO

En un entorno en constante evolución, el proceso de evaluación y mejora continua se ha convertido en una necesidad estratégica para individuos y organizaciones por igual. El restudio, un enfoque integral para la revisión y actualización de conocimientos y habilidades, emerge como una herramienta esencial en este contexto. Este estudio se centra en desglosar el concepto de restudio, explorando sus características clave, funcionalidades y las ventajas que ofrece.

5.3.2. Características de Restudio

El restudio se distingue por su enfoque sistemático y adaptativo, diseñado para abordar las áreas de conocimiento que requieren reforzamiento. Entre sus características más destacadas se encuentran la personalización del contenido revisado, la integración de nuevas metodologías educativas y la capacidad para identificar y llenar lagunas en el aprendizaje previo. Además, el restudio suele incorporar mecanismos de retroalimentación continua para asegurar que los objetivos de aprendizaje se cumplan de manera efectiva.

5.3.3. Funcionalidad

La funcionalidad del restudio radica en su habilidad para ajustar el proceso de aprendizaje a las necesidades específicas del usuario. Esto incluye la revisión de materiales previamente aprendidos, la incorporación de nuevos conocimientos y la aplicación práctica a través de ejercicios y estudios de caso. Herramientas digitales y plataformas educativas modernas facilitan esta funcionalidad al ofrecer

recursos interactivos y técnicas de evaluación que permiten un seguimiento detallado del progreso.

5.3.4. Ventajas

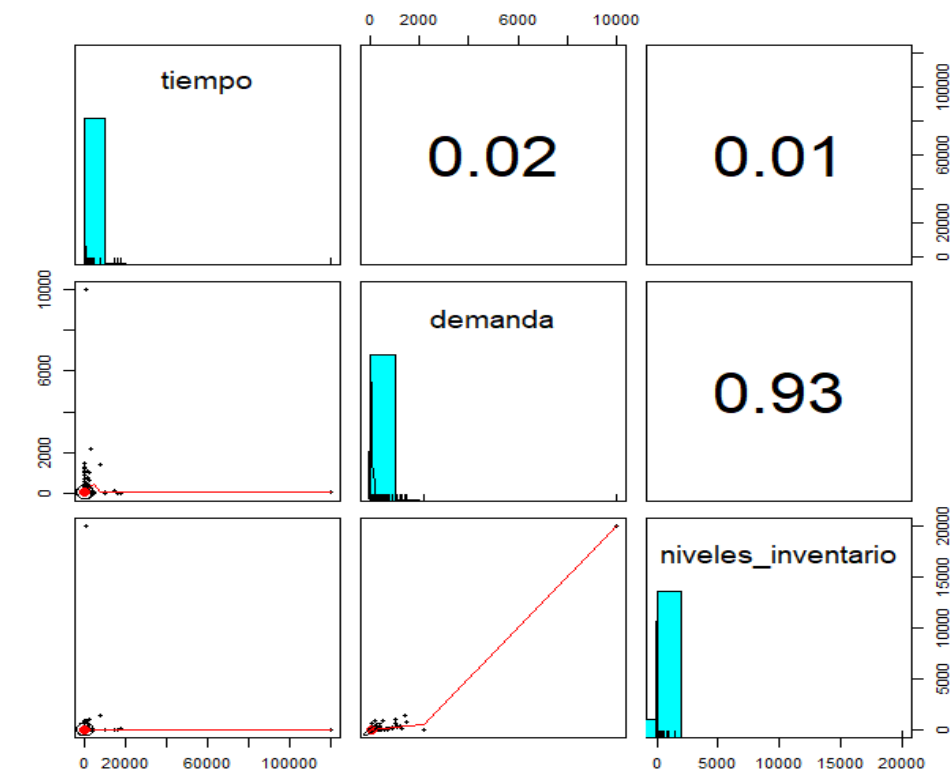
El restudio ofrece múltiples ventajas que contribuyen a un aprendizaje más eficaz y sostenido. Entre las más significativas se encuentran la mejora continua en la comprensión de conceptos clave, la capacidad para adaptarse a cambios en el campo de estudio y el refuerzo de habilidades prácticas. Además, el restudio promueve la retención a largo plazo del conocimiento y fomenta la confianza en la aplicación de nuevas habilidades. Este enfoque también permite a los profesionales mantenerse actualizados en sus respectivas áreas, lo que puede ser un factor determinante en el éxito y la relevancia en un mercado laboral competitivo.

Se tomaron dos algoritmos de aprendizaje supervisado, acompañados de un conjunto de datos sintéticos los cuales fueron sacados de una plataforma de inteligencia artificial (mostly) para ser ingresados en Restudio, plataforma digital de programación al cual nos ayudaría analizar los datos, integrar valores, visualizar y manipular los datos.

5.3.5. Regresión Lineal

Se empezó con la visualización una relación entre las variables del conjunto de datos. se creó una matriz de dispersión, la cual lleva histogramas y la correlación de las variables, todo esto para completar un panel de gráficos que se muestra en la Figura 7.

Figura 7. Gráfica de Dispersión de los Datos



Fuente: Autores

Dentro del modelo se tomaron tres modelos de nuestro conjunto de datos para así, realizar el estudio y saber cuál es la opción más óptima para el modelo.

Datos y resultados

Modelo 1= DEMANDA-TIEMPO DE PRODUCCIÓN

Figura 8 Algoritmo de Modelo 1 – Demanda -Tiempo de Producción

```
library(tidyverse)
library(boot)
library(car)
library(QuantPsyc)
library(ggplot2)
library(psych)
library(latticeExtra)

proyecto1 <- read_excel("proyecto1.xlsx")
view(proyecto1)

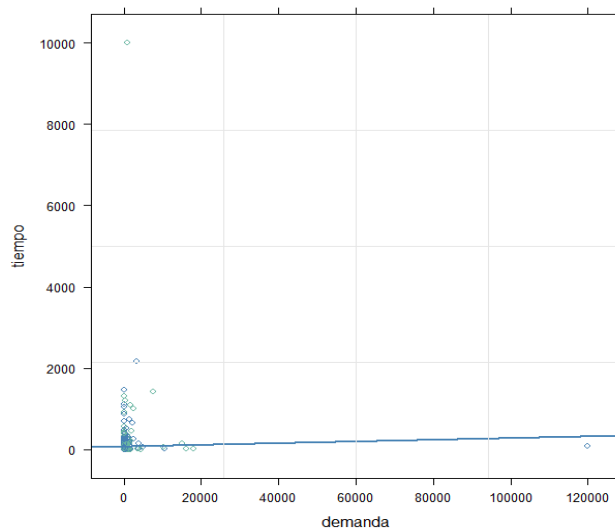
##regression
pairs.panels(proyecto1)
reg <- lm(proyecto1$demanda ~ proyecto1$tiempo)
summary(reg)

xyplot(proyecto1$demanda ~ proyecto1$tiempo, col=c("steelblue", "#69b6a2"),
       lwd=2, strip = TRUE, xlab = "demanda",
       ylab = "tiempo", panel = function(x,y,...)
       {panel.xyplot(x,y,...)
        panel.lmline(x,y,...)
        panel.grid()})

#correlaciond e pearson
sqrt(0.2786)
```

Fuente: Autores

Figura 9 Demanda-Tiempo de Producción



Fuente: Autores

Resultados obtenidos

Tabla 12 Resultados Obtenidos del Algoritmo del Modelo

Descripción	Valor
Coeficientes	
- Intercepto	348.97 (p < 0.01)
- Pendiente de Demanda	0.2722 (p = 0.4462)
Estadísticas de Ajuste	
- R-cuadrado	0.0005816
- R-cuadrado ajustado	-0.0004198
- Error estándar residual	3948
F-estadístico	0.5808 (p = 0.4462)

Fuente: Autores

Intercepto: El valor de 348.97 indica el valor estimado de la variable dependiente cuando la demanda es 0. El p-valor es muy pequeño ($p < 0.01$), lo que indica que el intercepto es estadísticamente significativo.

Pendiente de Demanda: El coeficiente de 0.2722 indica la relación entre la demanda y la variable dependiente. Sin embargo, su p-valor (0.4462) indica que no es estadísticamente significativo, lo que sugiere que la demanda no tiene un impacto significativo en el modelo.

R-cuadrado: Un valor extremadamente bajo (0.0005816) sugiere que el modelo explica prácticamente nada de la variabilidad en la variable dependiente.

R-cuadrado ajustado: Un valor negativo (-0.0004198) implica que el modelo no solo no explica la variabilidad, sino que es peor que un modelo que solo usa la media de los datos.

F-estadístico: Un valor de 0.5808 con un p-valor de 0.4462 también refuerza la falta de significancia global del modelo.

Modelo 2= NIVELES DE INVENTARIO_ TIEMPO DE PRODUCCIÓN

Figura 10 Algoritmo de Modelo 2 – Niveles de Inventario – Tiempo de Producción

```
library(tidyverse)
library(boot)
library(car)
library(QuantPsyc)
library(ggplot2)
library(psych)
library(latticeExtra)

proyecto1 <- read_excel("proyecto1.xlsx")
view(proyecto1)

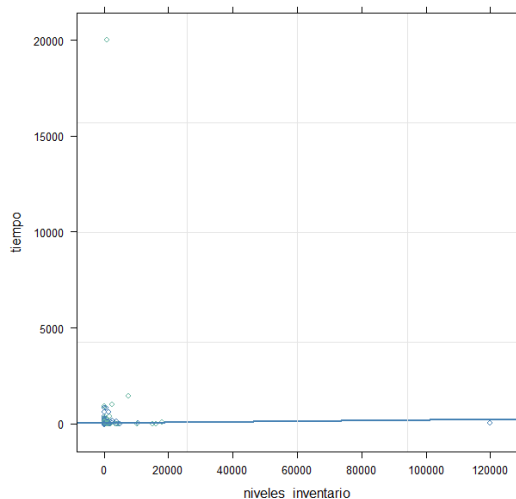
##regression
pairs.panels(proyecto1)
reg <- lm(proyecto1$niveles_inventario ~ proyecto1$tiempo)
summary(reg)

xyplot(proyecto1$niveles_inventario ~ proyecto1$tiempo, col=c("steelblue", "#69b6a2",
  lwd=2, strip = TRUE, xlab = "niveles_inventario",
  ylab = "tiempo", panel = function(x,y,...)
  {panel.xyplot(x,y,...)
  panel.lmline(x,y,...)
  panel.grid()})

#correlacion de pearson
sart(0.2786)
```

Fuente: Autores

Figura 11 Comportamiento de Niveles_Inventario_tiempo de Producción



Fuente: Autores

Resultados obtenidos

Tabla 13 Resultados de Algoritmo Niveles_Inventario_tiempo de Producción

Descripción	Valor
Coeficientes	
- Intercepto	368.01 (p < 0.01)
- Pendiente de Niveles de Inventario	0.05897 (p = 0.7635)
Estadísticas de Ajuste	
- R-cuadrado	9.072e-05
- R-cuadrado ajustado	-0.0009112
- Error estándar residual	3949
F-estadístico	0.09055 (p = 0.7635)

Fuente: Autores

Intercepto: El valor del intercepto es 368.01, lo que significa que cuando los niveles de inventario son cero, el valor de la variable dependiente (el resultado que se está modelando) será aproximadamente 368.01. Dado que el p-valor es menor a 0.01, este coeficiente es estadísticamente significativo, lo que sugiere que el intercepto tiene un impacto importante en el modelo.

Pendiente de Niveles de Inventario: El coeficiente es 0.05897, lo que indica que, por cada unidad adicional en los niveles de inventario, la variable dependiente aumentará en 0.05897 unidades. Sin embargo, el p-valor es 0.7635, lo que sugiere que este coeficiente no es estadísticamente significativo. En otras palabras, los niveles de inventario no parecen tener un impacto relevante en el resultado modelado.

R-cuadrado: Un valor de $9.072 \times 10^{-59.072} \times 10^{-5}$ (o 0.00009072) indica que el modelo explica menos del 0.01% de la variabilidad en la variable dependiente. Esto sugiere que el modelo no está capturando la relación entre las variables de manera efectiva.

R-cuadrado ajustado: El valor negativo (-0.0009112) indica que, cuando se ajusta por el número de variables en el modelo, el ajuste es incluso peor que si solo se utilizara la media para predecir el resultado. Esto refuerza la idea de que el modelo no es adecuado para predecir la variable dependiente.

Error estándar residual: El valor de 3949 refleja la variabilidad de los residuos o la diferencia entre los valores observados y los predichos por el modelo. Un error estándar residual tan alto indica que las predicciones del modelo se desvían considerablemente de los valores reales.

F-estadístico: Un F-estadístico de 0.09055 y un p-valor de 0.7635 indican que el modelo completo no es estadísticamente significativo. Esto implica que el modelo no tiene poder explicativo para la variable dependiente.

Modelo 3= DEMANDA- NIVELES DE INVENTARIO

Figura 12 Algoritmo de Demanda – Niveles de Inventario

```
library(tidyverse)
library(boot)
library(car)
library(quantpsyc)
library(ggplot2)
library(psych)
library(latticeExtra)

proyecto1 <- read_excel("proyecto1.xlsx")
view(proyecto1)

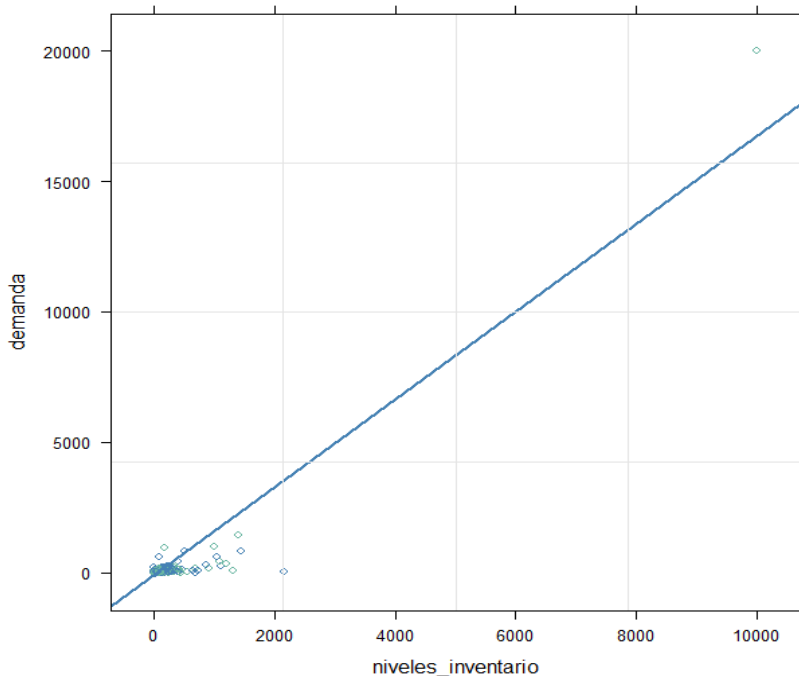
##regression
pairs.panels(proyecto1)
reg <- lm(proyecto1$niveles_inventario ~ proyecto1$demanda)
summary(reg)

xyplot(proyecto1$niveles_inventario ~ proyecto1$demanda, col=c("steelblue", "#69b6a2"),
        lwd=2, strip = TRUE, xlab = "niveles_inventario",
        ylab = "demanda", panel = function(x,y,...)
        {panel.xyplot(x,y,...)
        panel.lmline(x,y,...)
        panel.grid()})

#correlacion de pearson
sqrt(0.2786)
```

Fuente: Autores

Figura 13 Comportamiento Demanda Niveles de Inventario



Fuente: Autores

Resultados obtenidos

Tabla 14 Resultado Obtenido del Modelo

Descripción	Valor
Coefficientes	
- Intercepto	56.55 (p < 0.001)
- Pendiente de Niveles de Inventario	0.50767 (p < 0.001)
Estadísticas de Ajuste	
- R-cuadrado	0.8566
- R-cuadrado ajustado	0.8564
- Error estándar residual	132.5
F-estadístico	5960 (p < 2.2e-16)

Fuente: Autores

Intercepto: El valor de 56.55 indica que cuando los niveles de inventario son cero, el valor estimado de la variable dependiente es 56.55. Dado que el p-valor es extremadamente bajo ($p < 0.001$), este coeficiente es estadísticamente significativo, lo que sugiere que el intercepto tiene un impacto importante en el modelo.

Pendiente de Niveles de Inventario: El coeficiente de 0.50767 indica que, por cada unidad adicional en los niveles de inventario, la variable dependiente aumenta en 0.50767 unidades. El p-valor es muy bajo ($p < 0.001$), lo que indica que este coeficiente es altamente significativo y que los niveles de inventario tienen un impacto positivo y significativo en la variable dependiente.

R-cuadrado: El valor de 0.8566 indica que el modelo explica el 85.66% de la variabilidad en la variable dependiente, lo que sugiere un ajuste muy bueno del modelo.

R-cuadrado ajustado: El valor de 0.8564, muy cercano al R-cuadrado, indica que el ajuste del modelo sigue siendo muy alto incluso cuando se ajusta por el número de variables incluidas.

Error estándar residual: El valor de 132.5 indica que la variabilidad de los residuos (diferencias entre los valores observados y los predichos) es relativamente baja, lo que sugiere que las predicciones del modelo son razonablemente precisas.

F-estadístico: El valor extremadamente alto del F-estadístico (5960) con un p-valor extremadamente bajo ($p < 2.2e-16$) indica que el modelo es globalmente significativo, lo que refuerza la idea de que hay una relación estadísticamente significativa entre los niveles de inventario y la variable dependiente.

Comparación de Modelos Regresión Lineal

Utilizando los resultados de los 3 modelos pasamos a realizar una comparación entre ellos para mirar cual será el mejor predictor.

1. Coeficientes

- **Modelo 1:** El coeficiente para demanda no es significativo ($p = 0.4462$), lo que indica que demanda no es un buen predictor del tiempo.
- **Modelo 1:** El coeficiente para niveles_inventario tampoco es significativo ($p = 0.7635$) cuando se usa para predecir tiempo, lo que sugiere que niveles_inventario no tiene un impacto significativo sobre el tiempo en este contexto.
- **Modelo 3:** Ambos coeficientes (intercepto y pendiente de niveles_inventario) son altamente significativos ($p < 0.001$), indicando una fuerte relación entre niveles_inventario y demanda.

2. RECOMENDACIONES:

- **Modelo 1:** El modelo no explica bien la variabilidad en tiempo (R-cuadrado ~ 0.0006), lo que sugiere que la relación entre tiempo y demanda es muy débil.
- **Modelo 2:** Similar al Modelo 1, el ajuste del modelo es muy bajo (R-cuadrado ~ 0.00009), indicando que niveles_inventario no explica bien la variabilidad en tiempo.
- **Modelo 3:** El modelo muestra un ajuste excelente con un alto R-cuadrado de 0.8566, lo que sugiere que niveles_inventario explica una gran parte de la variabilidad en demanda.

3. ERROR ESTANDAR:

- **Modelo 1 y Modelo 1:** Ambos modelos tienen errores estándar residuales altos (3948 y 3949, respectivamente), indicando una alta variabilidad no explicada en tiempo.

- **Modelo 3:** El error estándar residual es significativamente menor (132.5), lo que refleja una mejor capacidad del modelo para predecir demanda.

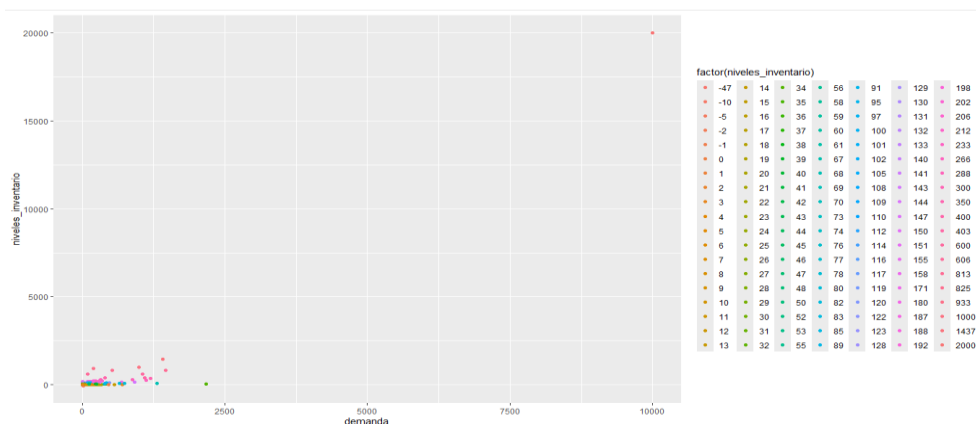
5.3.6. Modelo KNN

El algoritmo de k vecinos más cercanos, también conocido como KNN o k-NN, es un método no paramétrico de clasificación supervisada que se basa en la proximidad para realizar clasificaciones o predicciones sobre la agrupación de un punto de datos individual

En este estudio, se implementó un algoritmo de K-Nearest Neighbors (KNN) con el propósito de evaluar la capacidad predictiva del modelo en relación con los niveles de inventario y la demanda. El objetivo central es predecir los niveles de inventario empleando un conjunto de datos sintéticos, lo que permitirá medir la efectividad del algoritmo en un entorno controlado.

Base de Datos General (Data Set)

Figura 14 Datos Generales

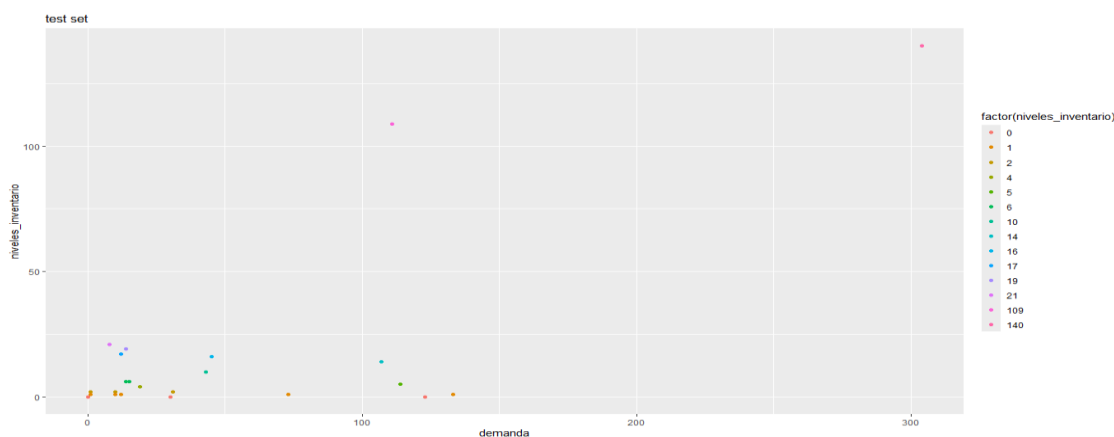


Fuente: Autores

Para ejecutar el algoritmo KNN, primero se dividieron los datos en un conjunto de entrenamiento (70%) y un conjunto de prueba (30%). Posteriormente, se genera una gráfica de dispersión que permite visualizar las variables de demanda y niveles de inventario, como se muestra en la Figura 14.

Datos de Prueba

Figura 15 Datos de Prueba



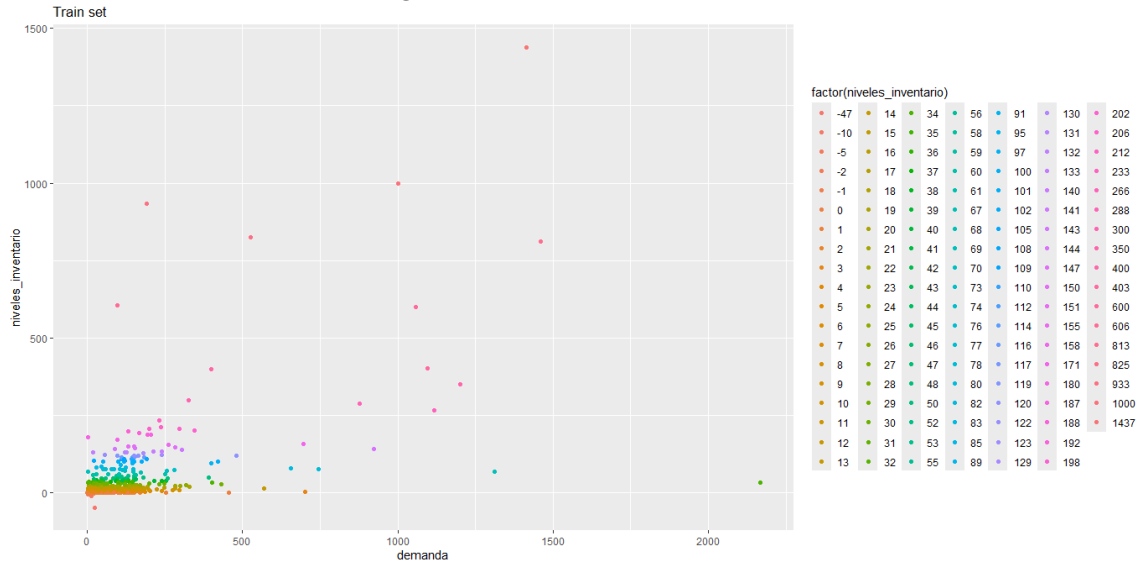
Fuente: Autores

La Figura 15 muestran los datos de prueba, "test set", el cual visualiza el conjunto de datos de prueba (30%). En este gráfico, los puntos estarán coloreados según los niveles de inventario. Este análisis nos permitirá observar cómo se distribuyen en función de la demanda y, además, identificar posibles patrones o agrupaciones relevantes.

Datos de Entrenamiento

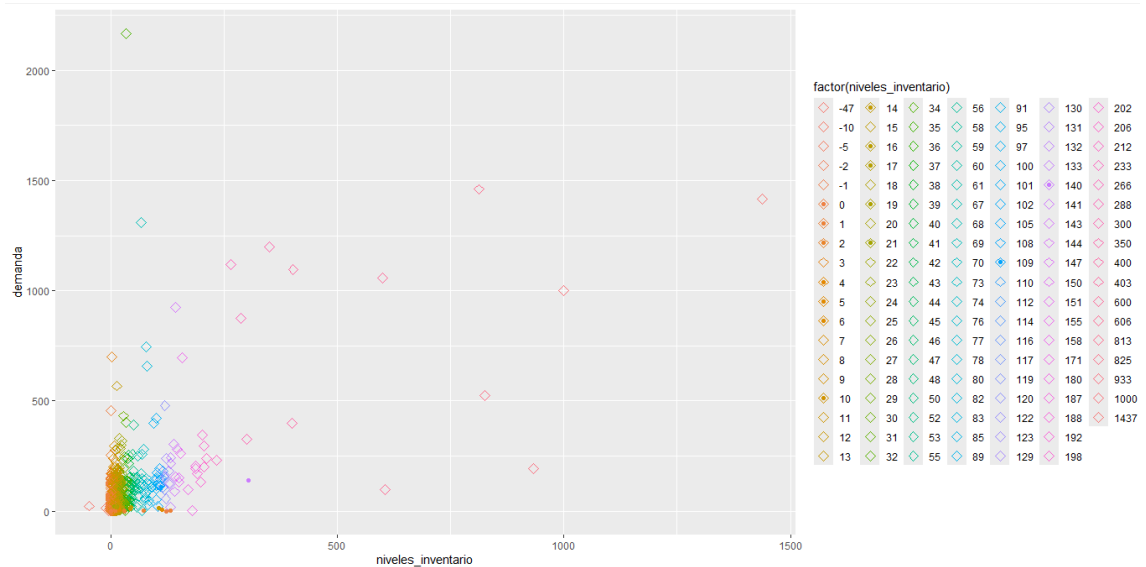
El paso siguiente fue determinar los datos de entrenamiento, denominados "train set", en la el conjunto de datos de entrenamiento (70%). En la Figura 16 se puede observar la distribución de los datos en función de las variables demanda y niveles de inventario.

Figura 16 Datos de Entrenamiento



Fuente: Autores

Figura 17 Datos de Entrenamiento y Prueba



Fuente: Autores

En la Figura 17 se observa el resultado de los datos de dispersión, en el que se superponen los datos de entrenamiento, representados con rombos rellenos (◆), y los datos de prueba,

presentados con círculos por defecto (◊). Esto permitirá comparar visualmente las distribuciones y analizar las relaciones entre las variables en ambos conjuntos de datos.

Matriz de Predicción

Figura 18 Matriz de Predicción

```
333      45      1      2
[ reached 'max' / getOption("max.print") -- omitted 1773 rows ]
> table(resul1,d_train$niveles_inventario)

resul1 -47 -10  -5  -2  -1   0   1   2   3   4   5   6   7   8   9  10  11  12  13  14  15  16  17  18
-47    0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
-10    0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
-5     1   0   3   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
-2     0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
-1     0   0   0   0   7   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
0      0   0   1   0   1   1 200   8   2   0   0   1   0   0   0   1   0   0   0   0   1   0   0
1      0   0   0   0   0   0 23 306  13   4   1   4   0   0   0   0   0   2   0   0   0   0   0

resul1  19  20  21  22  23  24  25  26  27  28  29  30  31  32  34  35  36  37  38  39  40  41  42  43
-47    0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
-10    0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
-5     0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
-2     0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
-1     0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
0      0   0   0   0   0   0   0   0   0   0   0   0   0   0   2   0   0   0   0   0   0   0
1      0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0

resul1  44  45  46  47  48  50  52  53  55  56  58  59  60  61  67  68  69  70  73  74  76  77  78  80
-47    0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
-10    0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
-5     0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
-2     0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
-1     0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
0      0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
1      0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0

resul1  82  83  85  89  91  95  97 100 101 102 105 108 109 110 112 114 116 117 119 120 122 123 129 130
-47    0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
-10    0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
-5     0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
-2     0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
-1     0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
```

Fuente: Autores

En la Figura 18 se presenta el resultado de la matriz de confusión, la cual proporciona el número de veces que el modelo predijo cada clase en comparación con la clase verdadera. Esto permitió evaluar la precisión y el desempeño del modelo.

Figura 19 Algoritmo de Predicción

```
> ##PRECISION##
> sum(resul1==d_train$niveles_inventario)/nrow(d_train)*100
[1] 81.67142
```

Fuente: Autores

A continuación, se procederá a calcular la precisión del modelo para evaluar el desempeño de los clasificadores. Para ello, se compararán las predicciones del modelo con las etiquetas verdaderas, obteniendo un vector de valores TRUE o FALSE que indica si la predicción es correcta o incorrecta, respectivamente. Además, se sumarán los valores TRUE, que representan las predicciones correctas, y se calculará la proporción de predicciones correctas en relación con el número total de observaciones.

El resultado obtenido es 81.67142, lo cual indica que el modelo KNN tiene una precisión del 81.67%. Esto sugiere que las predicciones realizadas por el modelo coinciden con las etiquetas reales del conjunto de datos de entrenamiento. Una precisión alta indica que el modelo clasifica correctamente las observaciones.

Matriz de Confusión

Figura 20 Matriz de Confusión

		Actual Values	
		Negative (n)	Positive (s)
Predicted Values	Negative (n)	TN	FN
	Positive (s)	FP	TP

Fuente: Autores

En el algoritmo se va a clasificar TP= verdaderos positivos

F-DC-125

INFORME FINAL DE TRABAJO DE GRADO EN MODALIDAD DE
PROYECTO DE INVESTIGACIÓN, DESARROLLO TECNOLÓGICO,
MONOGRAFÍA, EMPRENDIMIENTO Y SEMINARIO

VERSIÓN: 2.0

Figura 21 Verdaderos Positivos

```
> ##VERDADEROS POSITIVOS##
> tp
[1] 1 18 26 31 32 43 44 45 47 49 55 60 68 78 81 85 90 92 94 99
[21] 101 110 115 121 128 132 137 145 148 149 154 161 165 181 184 185 192 200 221 229
[41] 237 242 243 248 253 259 264 272 275 288 290 308 322 332 333 339 354 355 371 383
[61] 384 385 393 402 406 408 409 420 424 428 436 437 459 465 469 477 481 485 486 488
[81] 490 493 499 502 510 511 513 539 554 556 564 572 574 577 583 593 594 607 609 614
[101] 615 634 639 640 644 664 671 685 687 689 703 717 719 724 727 728 744 745 746 747
[121] 749 754 757 759 776 777 780 795 799 801 816 817 822 829 835 843 846 848 858 871
[141] 874 875 885 917 925 932 936 943 948 953 955 959 962 976 979 982 993 1000 1001 1003
[161] 1024 1028 1032 1033 1043 1046 1047 1053 1062 1075 1076 1083 1092 1096 1099 1100 1129 1130 1138 1153
[181] 1166 1171 1174 1179 1180 1183 1191 1192 1200 1216 1232 1242 1253 1256 1264 1271 1273 1276 1284 1288
[201] 1293 1294 1302 1308 1312 1337 1342 1345 1358 1363 1367 1382 1388 1389 1400 1405 1422 1424 1434 1435
[221] 1450 1451 1452 1453 1455 1465 1467 1475 1486 1487 1489 1492 1501 1504 1507 1509 1512 1524 1529 1534
[241] 1541 1543 1559 1561 1567 1592 1593 1613 1628 1632 1638 1647 1652 1659 1668 1674 1678 1700 1703 1704
[261] 1714 1721 1722 1725 1739 1751 1758 1769 1775 1777 1783 1794 1795 1796 1797 1808 1812 1826 1834 1844
[281] 1845 1852 1866 1872 1880 1884 1888 1889 1891 1893 1898 1902 1903 1914 1915 1932 1945 1957 1975 1976
[301] 1978 1983 1984 1994 2001 2005 2008 2013 2038 2042 2043 2048 2060 2064 2067 2074 2076 2085 2091 2102
```

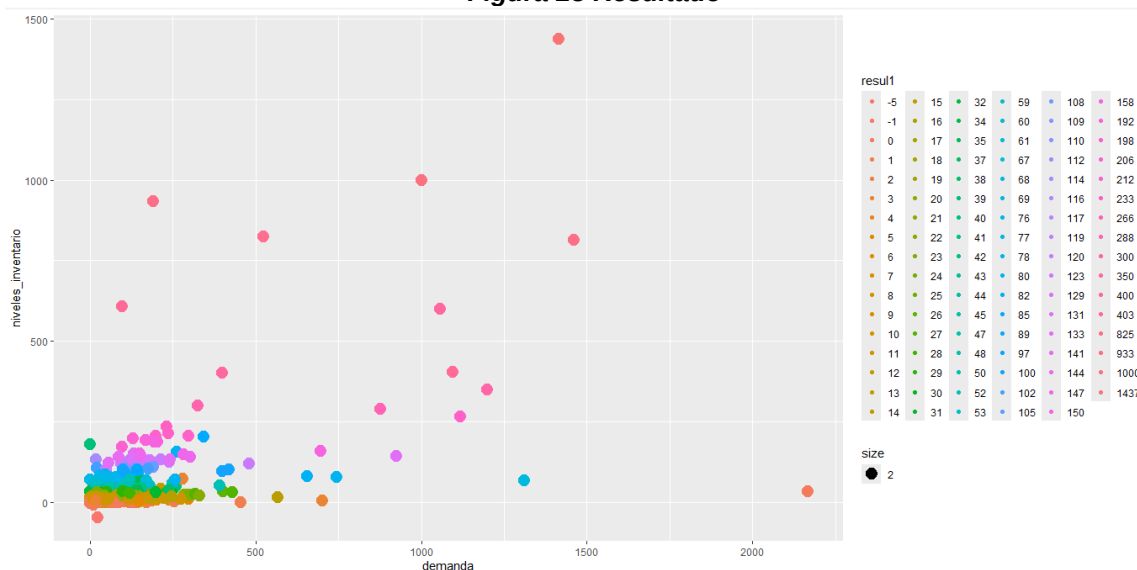
Fuente: Autores

Figura 22 Clasificación de FP= Falsos Positivos

```
> ##FALSOS POSITIVOS##
> fp
[1] 65 92 97 131 165 177 269 279 286 326 409 431 451 465 518 532 653 667 688 737
[21] 753 897 908 930 944 996 1005 1021 1033 1041 1072 1113 1128 1170 1227 1272 1351 1390 1474 1482
[41] 1488 1496 1542 1588 1591 1611 1767
```

Fuente: Autores

Figura 23 Resultado

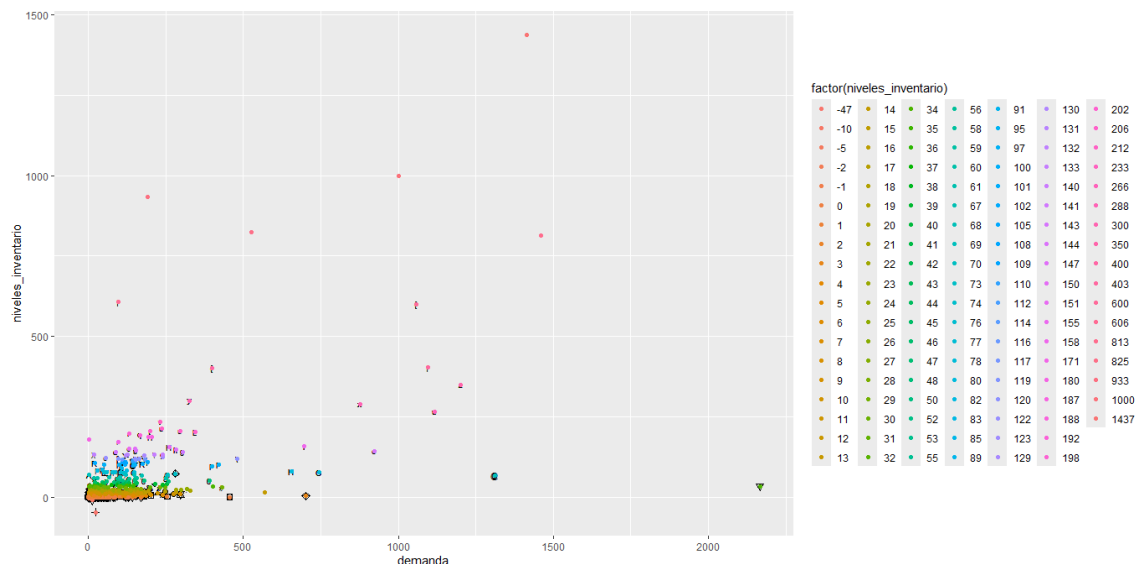


Fuente: Autores

Como resultado, se presentará una gráfica que muestra cómo se distribuyen las predicciones del modelo en función de las características de los datos (demanda y niveles de inventario).

A continuación, en la Figura 24 se muestra el resultado de datos reales, esto se logró mediante la función KNN, se especificó el conjunto de entrenamiento, el conjunto de prueba, las etiquetas y el número de vecinos (k). Con estos parámetros, se obtuvieron las predicciones para los datos de prueba, basadas en las características del conjunto de entrenamiento.

Figura 24 Datos Reales



Fuente: Autores

Figura 25 Nuevos Datos

```
> ##CLASIFICACION DE NUEVOS DATOS##
> y_pred=knn(d_train[,-2],d_test[,-2],cl=d_train$niveles_inventario,k=3)
> y_pred
[1] 10 0 5 2 21 0 119 1 1 2 6 2 5 16 1 4 1 1 19 0 15 14 144 1
131 Levels: -47 -10 -5 -2 -1 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 ... 1437
> |
```

Fuente: Autores

A continuación, mediante la función KNN, se especificará el conjunto de entrenamiento, el conjunto de prueba, las etiquetas y el número de vecinos (k). Con estos parámetros, se obtuvieron las predicciones para los datos de prueba, basadas en las características del conjunto de entrenamiento.

Figura 26 Accuracy

```
> ##VALORES PREDICHOS Y REALES##
> Accuracy(y_pred,y_true=d_test$niveles_inventario)
[1] 0.7083333
```

Fuente: Autores

En la Figura 26 para evaluar el rendimiento del modelo KNN, se van a calcular dos métricas, la precisión (Accuracy) y el área bajo la curva (AUC). La precisión del modelo es del 70.83%, lo que indica que el modelo clasificó correctamente aproximadamente el 70.83% de las observaciones en el conjunto de prueba. Esta métrica representa la proporción de predicciones correctas entre todas las predicciones realizadas.

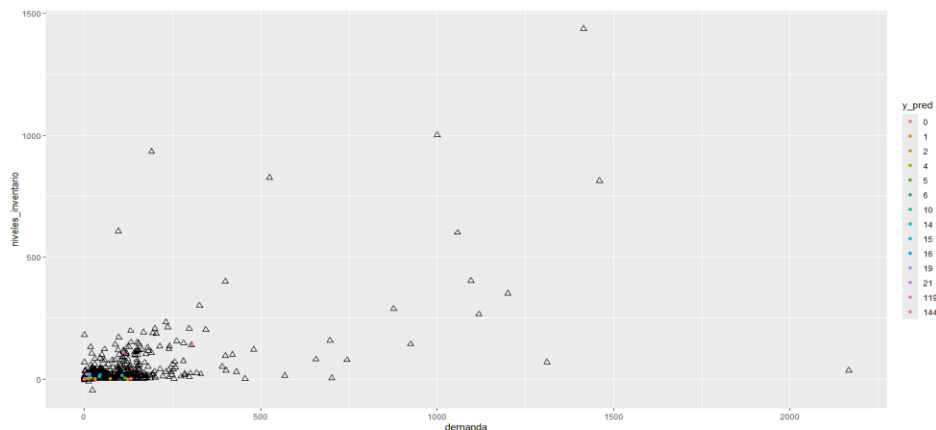
Figura 27 AUC

```
> AUC(y_pred =y_pred,y_true = d_test$niveles_inventario)
[1] 0.8666667
```

Fuente: Autores

Como se muestra en la Figura 27 el AUC (Área Bajo la Curva) es 0.87, lo que sugiere que el modelo tiene una buena capacidad para discriminar entre las clases. Este valor indica un rendimiento sólido, con una alta capacidad para distinguir entre las clases positivas y negativas.

Figura 28 Comparativo



Fuente: Autores

En la Figura 28 permite comparar visualmente los datos reales del conjunto de entrenamiento con las predicciones del modelo KNN en el conjunto de prueba. Los puntos del conjunto de entrenamiento estarán representados en azul, mientras que las predicciones del modelo en el conjunto de prueba se mostrarán en colores aleatorios, facilitando la comparación de cómo se alinean las predicciones con los datos reales.

El modelo KNN ha mostrado una precisión del 70.83%, lo cual refleja la proporción de predicciones correctas respecto al total de observaciones. Además, el AUC (Área Bajo la Curva) de 0.87 indica que el modelo tiene una excelente capacidad de discriminación, con una alta precisión en la diferenciación entre clases. Estos resultados conjuntamente ofrecen una visión clara de la efectividad del modelo en la clasificación de datos.

Tabla 15 Comparativo de Modelos Empleados

REGRESION LINEAL	KNN
Proporciona valores continuos, predice volúmenes exactos y realiza predicciones precisas	Modelo de clasificación, óptimo para predecir categorías o clases. Además, para clasificar niveles de inventario en categorías, no proporciona valores continuos.

Fuente: Autores

6. CONCLUSIONES

Al tener en cuenta los tipos de modelos de aprendizaje se pudo establecer que las estrategias de aprendizaje supervisado, como SVM, árboles de decisión, regresión lineal, Naive Bayes y redes neuronales, son herramientas clave y efectivas en diferentes sectores industriales. Entre ellas, SVM se destaca por su alta precisión y manejo de grandes volúmenes de datos, mientras que los árboles de decisión y algoritmos como Random Forest y XGBoost sobresalen en tareas específicas. Además, la creciente integración de inteligencia artificial y técnicas de aprendizaje automático sigue impulsando mejoras significativas en diversas industrias, lo que subraya su importancia para la innovación y optimización de procesos.

Dentro de la investigación se pudo establecer que existen varios algoritmos en términos de precisión, interpretabilidad, tiempo de entrenamiento, robustez y manejo de no linealidad. Los algoritmos más precisos, como Redes Neuronales y GBM/XGBoost, tienen una mayor robustez y pueden manejar no linealidad, pero a costa de un tiempo de entrenamiento alto y menor interpretabilidad. En contraste, algoritmos como la Regresión Lineal y KNN ofrecen una mayor interpretabilidad y tiempos de entrenamiento más rápidos, pero son menos robustos y no tan precisos. Árboles de Decisión y Random Forest logran un balance medio, combinando buena robustez y manejo de no linealidad con tiempos de entrenamiento moderados y precisión aceptable.

Al tener en cuenta en el modelo de Regresión Lineal, se pudo establecer que ninguno de los modelos que predicen tiempo (Modelos 1 y 2) muestra una buena capacidad predictiva o una relación significativa con sus variables predictoras. Por

otra parte, el Modelo 3, que predice demanda a partir de niveles_inventario, es claramente el modelo más robusto y efectivo. La alta significancia de los coeficientes y el alto R-cuadrado indican que niveles_inventario es un predictor fuerte de demanda, y el modelo tiene un buen ajuste.

Para predecir variables clave en los procesos industriales, como la demanda de productos, los tiempos de producción o los niveles de inventario, el modelo de regresión lineal es el más adecuado, ya que está diseñado para predecir valores continuos, como el volumen de inventario. Este modelo proporciona predicciones precisas y significativas, mostrando una fuerte relación entre las variables predictivas y la variable objetivo. Aunque el KNN es útil para clasificación, no es el mejor para predecir valores continuos. En este caso, la regresión lineal ofrece un enfoque más robusto y efectivo para cumplir con el objetivo de predecir variables clave en los procesos industriales.

7. RECOMENDACIONES

Para maximizar el impacto de las estrategias de aprendizaje supervisado en sectores industriales, se recomienda seleccionar el modelo más adecuado según el tipo de tarea y la naturaleza de los datos. Por ejemplo, SVM debe emplearse cuando se requiera alta precisión y se trabajen con grandes volúmenes de datos, mientras que árboles de decisión, Random Forest, y XGBoost son más adecuados para tareas específicas que requieren un balance entre precisión, velocidad y robustez. Además, es importante seguir explorando la integración de inteligencia artificial para aprovechar las mejoras en la optimización de procesos y fomentar la innovación continua en la industria.

Al seleccionar un algoritmo para su aplicación en proyectos, es fundamental equilibrar precisión, interpretabilidad y tiempo de entrenamiento según las necesidades específicas. Si la prioridad es la precisión y robustez en escenarios con datos no lineales, se deben preferir modelos como Redes Neuronales o GBM/XGBoost, aunque impliquen un mayor tiempo de entrenamiento y menor interpretabilidad. En casos donde la interpretabilidad y tiempos de respuesta rápidos sean más importantes, como en sistemas de toma de decisiones rápidas, algoritmos como Regresión Lineal o KNN serán más adecuados. Finalmente, para lograr un equilibrio entre precisión, robustez y tiempo de entrenamiento, Árboles de Decisión o Random Forest son opciones recomendables, especialmente en tareas donde el manejo de no linealidad es necesario, pero se requieren tiempos de cómputo moderados.

Para predecir variables clave en los procesos industriales, como la demanda de productos, los tiempos de producción o los niveles de inventario, el modelo de

regresión lineal es el más adecuado, ya que está diseñado para predecir valores continuos, como el volumen de inventario. Este modelo proporciona predicciones precisas y significativas, mostrando una fuerte relación entre las variables predictivas y la variable objetivo. Aunque el KNN es útil para clasificación, no es el mejor para predecir valores continuos. En este caso, la regresión lineal ofrece un enfoque más robusto y efectivo para cumplir con el objetivo de predecir variables clave en los procesos industriales.

8. REFERENCIAS BIBLIOGRÁFICAS

- Adrián, A., Del Río, C., & Alvear, D. (2019). Estudio de Tiempos y Movimientos para Incrementar la Eficiencia en una Empresa de Producción de Calzado. *Revista Información Tecnológica*, 30(3), 1-16. Obtenido de https://www.scielo.cl/scielo.php?pid=S0718-07642019000300083&script=sci_arttext
- Ahmed, S., & Chakraborty, R. K. (2024). Chapter eleven - Predictive big data analytics for supply chain demand forecasting. *Computational Intelligence Techniques for Sustainable Supply Chain Management*, 301-330. Obtenido de <https://www.sciencedirect.com/science/article/abs/pii/B978044318464200011X>
- Andrade, A. M., Del Río, C. A., & Alvear, D. L. (2019). Estudio de Tiempos y Movimientos para Incrementar la Eficiencia en una Empresa de Producción de Calzado. *Revista Información Tecnológica*, 30(3), 83-94. Obtenido de <https://www.scielo.cl/pdf/infotec/v30n3/0718-0764-infotec-30-03-00083.pdf>
- Aziz, N., Patah Akhir, E. A., Aziz, I. A., Jaafar, J., & Hasan, M. H. (2020). A Study on Gradient Boosting Algorithms for Development of AI Monitoring and Prediction Systems. *2020 International Conference on Computational Intelligence (ICCI)*, 1-12. Obtenido de <https://ieeexplore.ieee.org/abstract/document/9247843>
- Bansal, M., Goyal, A., & Choudhary, A. (2022). A comparative analysis of K-Nearest Neighbor, Genetic, Support Vector Machine, Decision Tree, and Long Short Term Memory algorithms in machine learning. *Decision Analytics Journal*, 3(1), 1-12. Obtenido de <https://www.sciencedirect.com/science/article/pii/S2772662222000261>
- Bécue, A., Praça, I., & Gama, o. (2021). Artificial intelligence, cyber-threats and Industry 4.0: challenges and opportunities. *Artificial Intelligence Review*, 54, 3849–3886. Obtenido de <https://link.springer.com/article/10.1007/s10462-020-09942-2>
- Bertolini, M., Mezzogori, D., & Neroni, M. Z. (2021). Machine Learning for industrial applications: A comprehensive literature review. *Expert Systems with Applications*, 175, 1-12. Obtenido de <https://www.sciencedirect.com/science/article/abs/pii/S095741742100261X>
- Cebekhulu, E., Onumanyi, A. J., & Sherrin, J. I. (2022). Performance Analysis of Machine Learning Algorithms for Energy Demand–Supply Prediction in Smart Grids. *Sustainability*, 14(5), 1-10. Obtenido de <https://www.mdpi.com/2071-1050/14/5/2546>
- Cervantes, J., Garcia Lamont, F., Mazahua, R., Lisbeth, & Lopez, A. (2020). A comprehensive survey on support vector machine classification: Applications, challenges and trends. *Neurocomputing*, 408, 189-215. Obtenido de <https://www.sciencedirect.com/science/article/abs/pii/S0925231220307153>
- Coppiano Marín, A. D., & Herrera Vargas, C. J. (2022). Desarrollo de aplicativo web basado en máquinas de vectores de soporte(svm) de aprendizaje supervisado para la predicción en la recomendación de cultivos mediante datos ambientales para fincas

- agroecológicas del cantón La Maná, provincia de Cotopaxi. *Repositorio Institucional Universidad Técnica de Cotopaxi*, 1-86. Obtenido de <https://repositorio.utc.edu.ec/handle/27000/8457>
- De Castillo, P., Vivaldini, M., Pires, S., Rita, A., Argoud, A. R., & Carmargo Junior, J. B. (2015). La gestión de proveedores de segundo nivel en la cadena de suministros. Un estudio en la industria automotriz en Brasil. *Invenio: Revista de investigación académica*, 34, 109-115. Obtenido de <https://dialnet.unirioja.es/servlet/articulo?codigo=5132263>
- Ensafi, Y., & Amin, S. H. (International Journal of Information Management Data Insights). Time-series forecasting of seasonal items sales using machine learning – A comparative analysis. *International Journal of Information Management Data Insights*, 2(1), 1-16. Obtenido de <https://www.sciencedirect.com/science/article/pii/S2667096822000027>
- Fahle, S., Prinz, C., & Kuhlentötter, B. (2020). Systematic review on machine learning (ML) methods for manufacturing processes – Identifying artificial intelligence (AI) methods for field application. *Procedia CIRP*, 93, 413-418. Obtenido de <https://www.sciencedirect.com/science/article/pii/S2212827120307435>
- Feizabadi, J. (2022). Machine learning demand forecasting and supply chain performance. *International Journal of Logistics Research and Applications*, 1-15. Obtenido de <https://www.tandfonline.com/doi/full/10.1080/13675567.2020.1803246>
- Figueroa Donayre, E. M., Guillén Guevara, M. L., & Flores, R. H. (2023). Machine Learning en la Industria 4.0: Análisis de su relevancia y aplicaciones. *Revista Científica de Ingeniería Nawparisun*, 4(5), 1-13. Obtenido de <https://unaj.edu.pe/revista/index.php/vpin/article/view/272>
- Hernández Sampieri, R. (2006). Tipo de Investigación. *Capítulo III - Marco Metodológico poerw*, 1-10. Obtenido de <http://virtual.urbe.edu/tesispub/0094597/cap03.pdf>
- Hernández Sampieri, R., & Torres Mendoza, P. C. (2020). Metodología de la Investigación: Rutas cuantitativas, cualitativas y mixta. *Mc Graw Hill*, 1-753. Obtenido de https://d1wqtxts1xzle7.cloudfront.net/64591365/Metodolog%C3%ADa_de_la_investigaci%C3%B3n._Rutas_cuantitativa_cualitativa_y_mixta-libre.pdf?1601784484=&response-content-disposition=inline%3B+filename%3DMETODOLOGIA_DE_LA_INVESTIGACION_LAS_RUTA.pdf&Expires=
- Hernández Sampieri, R., Fernández Collado, C., & Baptista Lucio, M. d. (2016). Metodología de la Investigación 6ta Edición. *Editorial McGraw Hill*, 1-634. Obtenido de <https://drive.google.com/file/d/0B7fKI4RAT39QeHNzTGh0N19SME0/view?resourcekey=0-Tg3V3qROROH0Aw4maw5dDQ>
- Hernández, S. R., & Mendoza Torres, C. P. (2018). Metodología de la Investigación: Las rutas cuantitativa, cualitativas y Mixta. *Editorial Mc Graw Hill*, 1-753. Obtenido de https://d1wqtxts1xzle7.cloudfront.net/64591365/Metodolog%C3%ADa_de_la_investigaci%C3%B3n._Rutas_cuantitativa_cualitativa_y_mixta-libre.pdf?1601784484=&response-content-disposition=inline%3B+filename%3DMETODOLOGIA_DE_LA_INVESTIGACION_LAS_RUTA.pdf&Expires=

- Huertas Mora, A. (2020). Algoritmos de aprendizaje supervisado utilizando datos de monitoreo de condiciones: Un estudio para el pronóstico de fallas en máquinas. *Repositorio Universidad Santo Tomas*, 1-77. Obtenido de <https://repository.usta.edu.co/bitstream/handle/11634/29886/2020alexanderhuertas.pdf?sequence=1&isAllowed=y>
- Islam, S., & Hassanzadeh, A. (2020). Prediction of probable backorder scenarios in the supply chain using Distributed Random Forest and Gradient Boosting Machine learning techniques. *Journal of Big Data*, 7, 1-22. Obtenido de <https://link.springer.com/article/10.1186/s40537-020-00345-2>
- Jijo, B. T., & Abdulazeez, A. M. (2021). Classification Based on Decision Tree Algorithm for Machine Learning. *Journal of Applied Science and Technology Trends*, 2(1), 20-28. Obtenido de <https://www.jastt.org/index.php/jasttpath/article/view/65>
- Jove, E., Urda, D., & Herrero, Á. (2022). Avances recientes en la aplicación de la ciencia de datos a la ciberseguridad industrial. *Tecnología de las telecomunicaciones*, 1-2. Obtenido de https://gicap.ubu.es/publications/2021/PDF/2021_c01_Recent%20advances%20in%20the%20application%20of%20data%20science%20to%20industrial%20cybersecurity.pdf
- Kang, Z., Catal, C., & Tekinerdogan, B. (2020). Machine learning applications in production lines: A systematic literature review. *Computers & Industrial Engineering*, 1-15. Obtenido de <https://www.sciencedirect.com/science/article/pii/S036083522030485X>
- Kiangala, S. K., & Wang, Z. (2021). An effective adaptive customization framework for small manufacturing plants using extreme gradient boosting-XGBoost and random forest ensemble learning algorithms in an Industry 4.0 environment. *Machine Learning with Applications*, 4, 1-15. Obtenido de <https://pdf.sciencedirectassets.com/777839/1-s2.0-S2666827021X00023/1-s2.0-S2666827021000050/main.pdf?X-Amz-Security-Token=IQoJb3JpZ2luX2VjEFlaCXVzLWVhc3QtMSJHMEUCIG9fYWZKbeXMNI%2FeV1KpROpEc4bpGe5RFkKuWtn5pmg%2FAiEAxkO0iBL%2BJtcP1ozHLbrNsNcPjE0GN0Ug0TkBcA>
- Mahesh, B. (2018). Machine Learning Algorithms - A Review. *International Journal of Science and Research (IJSR)*, 1-6. Obtenido de https://www.researchgate.net/profile/Batta-Mahesh/publication/344717762_Machine_Learning_Algorithms_-_A_Review/links/5f8b2365299bf1b53e2d243a/Machine-Learning-Algorithms-A-Review.pdf?eid=5082902844932096t
- Martínez Soto, M. (2011). Desarrollo de un modelo de gestión del conocimiento en la cadena de suministro de la industria agroalimentaria. *Repositorio Universidad Politécnica de Madrid*, 1-12. Obtenido de <https://oa.upm.es/6199/>
- Nasteski, V. (2017). An overview of the supervised machine learning. *Horizont*, 1-11. Obtenido de https://www.researchgate.net/profile/Vladimir-Nasteski/publication/328146111_An_overview_of_the_supervised_machine_learning_methods/links/5c1025194585157ac1bba147/An-overview-of-the-supervised-machine-learning-methods.pdf
- Nasteski, V. (2020). An overview of the supervised machine learning. *Repositorio Universidad de Partizanska - Macedonia*, 4(5), 1-11. Obtenido de <https://www.researchgate.net/profile/Vladimir>

- Nasteski/publication/328146111_An_overview_of_the_supervised_machine_learning_methods/links/5c1025194585157ac1bba147/An-overview-of-the-supervised-machine-learning-methods.pdf
- Osisanwo, F., Akinsola, J., Awodale, O., O, H. J., & O., O. (2017). Supervised Machine Learning Algorithms: Classification and Comparison. *International Journal of Computer Trends and Technology (IJCTT)*, 48(3), 128-138. Obtenido de <https://www.researchgate.net/publication/318338750>
- Pavan, D., & Akshit, K. (2021). A Comprehensive Comparative Study of Artificial Neural Network (ANN) and Support Vector Machines (SVM) on Stock Forecasting. *Annals of Data Science*, 1-15.
- Praveen, K. B., Prateek, B., Pradyumna, K., & Pragathi, B. (2020). Inventory Management using Machine Learning. *International Journal of Engineering Research & Technology (IJERT)*, 9(6), 1-4. Obtenido de https://www.researchgate.net/profile/Madhuri-Javaregowda/publication/342529859_Inventory_management_using_Machine_Learning/links/5f00bc48299bf18816008a64/Inventory-management-using-Machine-Learning.pdf
- Quiroz Flores, J. C., Rios del Castillo, P., & Guia Espinoza, R. (2022). Modelo de producción de la industria peruana. *Revista Venezolana de Gerencia: RVG*, 590-611. Obtenido de <https://dialnet.unirioja.es/servlet/articulo?codigo=8890698>
- Raizada, S., & Saini, J. R. (2021). Comparative Analysis of Supervised Machine. *(IJACSA) International Journal of Advanced Computer Science and Applications*, 12(11), 1-9. Obtenido de https://d1wqtxts1xzle7.cloudfront.net/86566701/Paper_12-Comparative_Analysis_of_Supervised_Machine_Learning_Techniques-libre.pdf?1653672024=&response-content-disposition=inline%3B+filename%3DComparative_Analysis_of_Supervised_Machi.pdf&Expires=1719577778&
- Ramirez Asis, E., Vilchez Carcamo, J., & Thakar, C. M. (2022). A review on role of artificial intelligence in food processing and manufacturing industry. *Materials today Proceedings*, 2462-2465.
- Rodrigues, J. A., Martins, A., Mendes, M., Torres Farinha, J., Mateus, R. J., & Cardoso, A. J. (2022). Automatic Risk Assessment for an Industrial Asset Using Unsupervised and Supervised Learning. *Journal Energies*, 15(24), 1-15. Obtenido de <https://www.mdpi.com/1996-1073/15/24/9387>
- Sablón-Cossío, N., Orozco Crespo, E., Pulido Rojano, A., Acevedo Urquiaga, A. J., & Ruiz Cedeño, S. d. (2021). Análisis de integración de la cadena de suministros en la industria textil en Ecuador. Un caso de estudio. *Ingeniare. Revista chilena de ingeniería*, 29(1), 1-15. Obtenido de https://www.scielo.cl/scielo.php?pid=S0718-33052021000100094&script=sci_arttext&lng=pt
- Saravanan, R., & Sujatha, P. (2018). A State of Art Techniques on Machine Learning Algorithms: A Perspective of Supervised Learning Approaches in Data Classification. *Second International Conference on Intelligent Computing and Control Systems (ICICCS)*, 1-12. Obtenido de <https://ieeexplore.ieee.org/abstract/document/8663155>
- Shafiq, M., Thakre, K., & Kumar, D. (2023). Continuous quality control evaluation during manufacturing using supervised learning algorithm for Industry 4.0. *The International*

F-DC-125

**INFORME FINAL DE TRABAJO DE GRADO EN MODALIDAD DE
PROYECTO DE INVESTIGACIÓN, DESARROLLO TECNOLÓGICO,
MONOGRAFÍA, EMPRENDIMIENTO Y SEMINARIO**

VERSIÓN: 2.0

- Journal of Advanced Manufacturing Technology*, 1-12. Obtenido de <https://link.springer.com/article/10.1007/s00170-023-10847-x>
- van Engelen, J. E., & Hoos, H. H. (2020). A survey on semi-supervised learning. *Machine Learning*, 109, 373-440. Obtenido de <https://link.springer.com/article/10.1007/s10994-019-05855-6>
- Vesga Acevedo, D. N. (2020). Modelo de planeación de inventarios para E-commerce, utilizando herramientas de inteligencia artificial para hacer pronósticos de demanda y clasificación de inventarios. *Repositorio Universidad de los Andes*, 1-26. Obtenido de <https://repositorio.uniandes.edu.co/server/api/core/bitstreams/001dfff1-a84c-4cf3-9bd2-dd4d42063ec2/content>
- Zeki Murat, Ç., Nuhu, A., & Qasim, Z. (2020). Machine Learning in Predictive Maintenance towards Sustainable Smart Manufacturing in Industry 4.0. *Sustainability*, 12(19), 1-12. Obtenido de <https://www.mdpi.com/2071-1050/12/19/8211>
- Zhao, Z., Alzubaidi, L., & Zhang, J. (2024). A comparison review of transfer learning and self-supervised learning: Definitions, applications, advantages and limitations. *Expert Systems with Applications*, 242, 1-13. Obtenido de <https://www.sciencedirect.com/science/article/pii/S0957417423033092>