TESIS
UNIVERSIDAD CIENTÍFICA DEL SUR

UNIVERSIDAD CIENTÍFICA DEL SUR

FACULTAD DE INGENIERÍA

CARRERA PROFESIONAL DE INGENIERÍA DE SISTEMAS EMPRESARIALES

"IMPLEMENTACIÓN DE MACHINE LEARNING PARA LA PREDICCIÓN DE DEMANDA EN LA FERRETERÍA VARGAS"

Tesis para optar el título profesional de:

INGENIERO DE SISTEMAS EMPRESARIALES

Presentado por:

Nayeli Jimena Huacho Taipe (0000-0002-1517-8827)

D'alessandro Armando Roque Barreto (0000-0003-1901-7251)

Asesor:

Wilver Auccahuasi Aiquipa (0000-0001-8820-4013)

Lima – Perú

2026

Tesis publicada con autorización del autor
No olvide citar esta tesis

ACTA DE SUSTENTACIÓN DE TESIS

Lima, 30 de julio del 2026

Los integrantes del Jurado de tesis:

Presidente	**Manuel Antonio Pereyra Acosta**
Miembro 1	**Martin Isidro Velasquez Medina**
Miembro 2	**José Luis Espinoza Melgarejo**

Se reúnen para evaluar la tesis titulada:
"Implementación de machine learning para la predicción de demanda en la ferretería Vargas"
con código de registro **PRE-12-2025-00797.**

Presentado por los bachilleres:
Nayeli Jimena Huacho Taipe
D'alessandro Armando Roque Barreto

Para optar al título profesional de
Ingeniero de Sistemas Empresariales

Asesorado por:
Wilver Auccahuasi Aiquipa

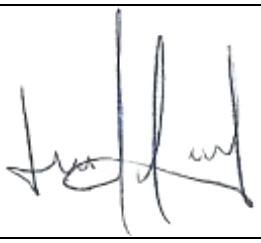
Luego de haber evaluado el informe final de tesis y evaluado el desempeño del estudiante de la Carrera de **Ingeniería de Sistemas Empresariales** en la sustentación, concluyen de manera unánime (X) por mayoría simple ( ) calificar a:

Tesista:	**Nayeli Jimena Huacho Taipe**		
Nota (en letras):	**Dieciséis**		
Aprobado ( )	Aprobado - Muy buena (X)	Aprobado - Sobresaliente ( )	Desaprobado ( )

Tesista:	**D'alessandro Armando Roque Barreto**		
Nota (en letras):	**Dieciséis**		
Aprobado ( )	Aprobado - Muy buena ( )	Aprobado - Sobresaliente ( )	Desaprobado ( )

Los miembros del jurado firman en señal de conformidad.

	
Manuel Antonio Pereyra Acosta	Wilver Auccahuasi Aiquipa
Presidente del Jurado	*Asesor*

	
Martin Isidro Velasquez Medina	José Luis Espinoza Melgarejo
Miembro 1	*Miembro 2*

ANEXO: FORMATO DE DECLARACIÓN DE ORIGINALIDAD DE CONTENIDO DE INFORME FINAL DE TESIS O TRABAJO DE INVESTIGACIÓN

UNIVERSIDAD CIENTÍFICA DEL SUR

DECLARACIÓN DE ORIGINALIDAD DE CONTENIDO DEL INFORME FINAL DE TESIS / TRABAJO DE INVESTIGACIÓN [1]

Lima, 31 de Julio del 2026

Señor,
Dr. Percy Mayta Tristán

Vicerrector de Investigación
Universidad Científica del Sur
Presente. –

De nuestra consideración,

Los abajo firmantes – Nayeli Jimena Huacho Taipe y D'alessandro Armando Roque Barreto, estudiantes/egresados del Programa académico de Ingeniería de Sistemas Empresariales de la Universidad Científica del Sur, y el asesor de tesis Dr. Wilver Auccahuasi Aiquipa – declaramos que este informe final de tesis/trabajo de investigación titulado "IMPLEMENTACIÓN DE MACHINE LEARNING PARA LA PREDICCIÓN DE DEMANDA EN LA FERRETERÍA VARGAS", que presentamos para optar el título profesional/grado académico **de Ingeniero de Sistemas Empresariales**, es original.

En consecuencia, declaramos que el informe no contiene plagio ni total ni parcial; que, cuando se utilizó información de fuentes externas, se reconoció su autoría mediante la citación correspondiente; y que los resultados obtenidos son producto íntegro de nuestra investigación y no han sido falseados ni fabricados. Todo ello en cumplimiento del Código de ética en la Investigación, Reglamento de Disciplina de Pregrado y Posgrado, Reglamento de Propiedad Intelectual, Normas y procedimientos de los trabajos de investigación para la obtención de títulos profesionales y grados académicos [2], que afirmamos conocer en su totalidad.

Declaramos, además, que no se han utilizado herramientas de inteligencia artificial (IA) generativa, de forma parcial ni total, en ninguna fase del desarrollo de la investigación ni en la redacción del informe final de tesis/trabajo de investigación. Sí declaro haber utilizado herramientas de IA no generativa para corrección ortográfica y de estilo, conforme a la normativa de la Universidad [3]. Todo el material ha sido revisado por el/los autor(es), quien(es) asume(n) la responsabilidad de su contenido.

Por ello, en caso de identificarse alguna situación de plagio, falsificación o fabricación de resultados, uso indebido de IA generativa o uso no declarado de IA, nos sometemos al proceso de investigación que establezca la Universidad y aceptamos las sanciones que pudieran corresponder. También declaramos conocer las normativas institucionales que regulan las sanciones aplicables en estos casos, incluido el de la compra y venta de tesis.

Firmamos en conformidad con lo declarado,

Firma tesista
DNI: 72198403

Firma tesista
DNI: 72683683

Firma del asesor de tesis
DNI:43375865

[1] En conformidad con la Resolución del Consejo Directivo N° 084-2022-SUNEDU/CD
[2] Los documentos están en: https://repositorio.cientifica.edu.pe/handle/20.500.12805/3652
[3] GEA-GIE-POL-05 Política del uso de inteligencia artificial

INFORME DE REVISIÓN DE ORIGINALIDAD

Título del documento evaluado. Implementación de machine learning para la predicción de demanda en la ferretería Vargas

Autores. Nayeli Jimena Huacho Taipe; D'alessandro Armando Roque Barreto

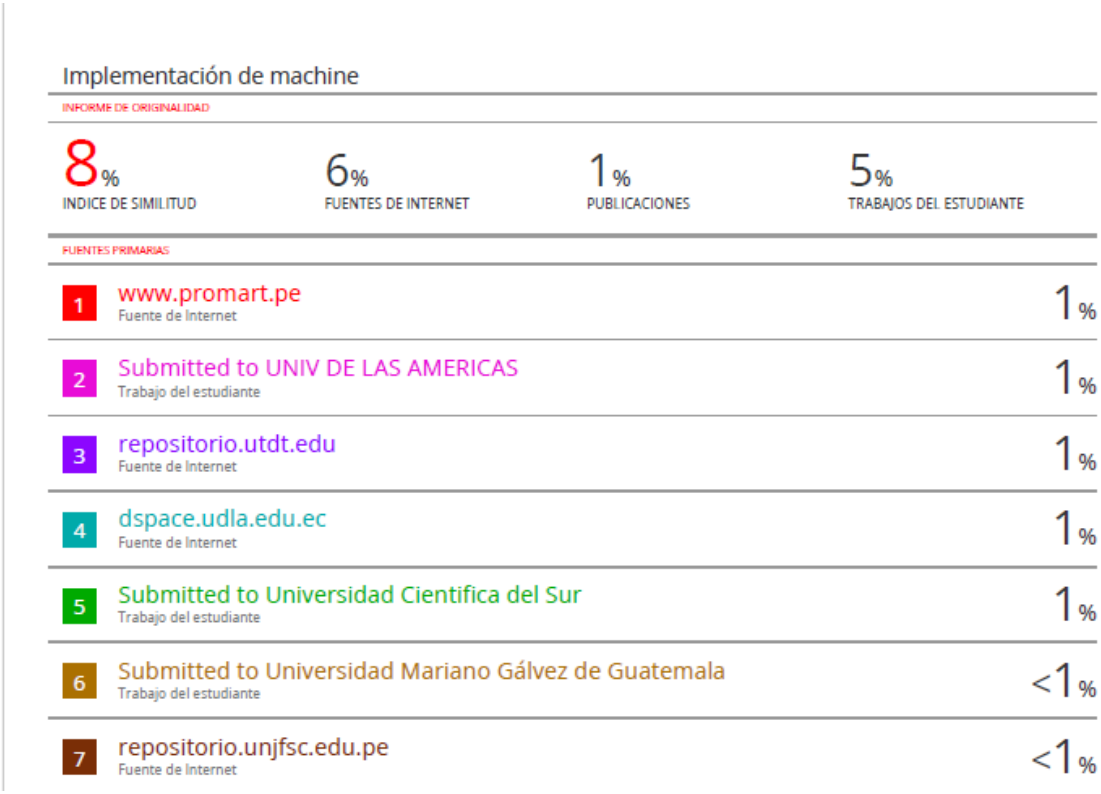
Mecanismo de revisión de originalidad. Evaluación con Turnitin (ID 3014299786).

Código de registro. PRE-12-2025-00797

Resultado de la evaluación. 8%

Revisado por. Magaly Kelly Guerrero Huaracallo.

Comentarios sobre la revisión. Filtros usados: excluir fuentes de menos de 12 palabras.



Implementación de machine

INFORME DE ORIGINALIDAD

8%	6%	1%	5%
INDICE DE SIMILITUD	FUENTES DE INTERNET	PUBLICACIONES	TRABAJOS DEL ESTUDIANTE

FUENTES PRIMARIAS

1	www.promart.pe Fuente de Internet	1%
2	Submitted to UNIV DE LAS AMERICAS Trabajo del estudiante	1%
3	repositorio.utdt.edu Fuente de Internet	1%
4	dspace.udla.edu.ec Fuente de Internet	1%
5	Submitted to Universidad Cientifica del Sur Trabajo del estudiante	1%
6	Submitted to Universidad Mariano Gálvez de Guatemala Trabajo del estudiante	<1%
7	repositorio.unjfsc.edu.pe Fuente de Internet	<1%

TESIS
UNIVERSIDAD CIENTÍFICA DEL SUR

Dedicatoria

A Dios, por guiarme en cada paso de este camino, acompañarme en los momentos más difíciles y llenarme de su fortaleza cuando más lo necesité.

A mi mamá y a mi abuela Juliana, por su amor incondicional y por cada sacrificio que hicieron por mí. Este logro es también suyo. A mi hermana, por su apoyo constante e incondicional, por estar siempre presente y creer en mí incluso cuando yo dudaba.

A cada uno de ustedes, les dedico este logro con todo el cariño y la gratitud que llevo en el corazón. Gracias por ser mi fortaleza, mi motivación y mi luz.

Nayeli Jimena Huacho Taipe

A Dios por iluminar mi avance en esta etapa académica, a mi padre Mauro por apoyarme en todo momento, por su presencia y darme su respaldo absoluto en toda mi estancia universitaria, esto jamás lo hubiera conseguido sin su ayuda. A mi madre por su amor incondicional y apoyo emocional en toda esta travesía, a mi hermano Jesús por apoyarme a su manera y ayudarme a tomar decisiones, y reflexionar acerca de mi carrera y mi futuro.

En última instancia a mi compañera con la que comparto esta investigación, un trayecto compartido muy gratificante que dará frutos.

Como todo lo bueno en la vida, este resultado nace del esfuerzo conjunto, cada uno en sus aportaciones en pequeña o en gran medida, es autor de mi éxito, ojalá me sigan acompañando. Por ello, les dedico este logro con toda mi gratitud.

D'alessandro Armando Roque Barreto

Tesis publicada con autorización del autor
No olvide citar esta tesis

TESIS
UNIVERSIDAD CIENTÍFICA DEL SUR

Agradecimientos

A Dios, por ser nuestra guía y sostén espiritual a lo largo de todo este proceso.

A los docentes de nuestra universidad, por compartir con generosidad su conocimiento y experiencia, contribuyendo de manera significativa a nuestra formación profesional.

A nuestro asesor de tesis, el Dr. Auccahuasi, por acompañarnos y orientarnos durante cada etapa de esta investigación.

A la ferretería Vargas y a su propietaria, por abrirnos sus puertas y permitirnos llevar a cabo nuestra investigación llevado a un escenario real.

Finalmente, a la Universidad Científica del Sur, por ser una institución comprometida con sus valores y con la excelencia académica.

Tesis publicada con autorización del autor
No olvide citar esta tesis

TESIS
UNIVERSIDAD CIENTÍFICA DEL SUR

Índice general

Dedicatoria ... ii
Agradecimientos ... iii
Índice general ... 1
Índice de figuras ... 3
Índice de tablas ... 4
Índice de anexos ... 5
Resumen ... 6
Abstract ... 7
Introducción ... 8
CAPÍTULO I: PLANTEAMIENTO DEL PROBLEMA ... 10
1.1. Descripción de la realidad problemática ... 10
1.2. Formulación del problema ... 13
1.2.1. Problema general ... 13
1.2.2. Problemas específicos ... 13
1.3. Justificación de la investigación ... 14
1.3.1. Justificación Metodológica ... 14
1.3.2. Justificación Teórica ... 14
1.3.3. Justificación Práctica ... 14
1.3.4. Justificación Económica ... 15
1.4. Limitaciones de la investigación ... 15
1.5. Viabilidad de la investigación ... 16
CAPÍTULO II: MARCO TEÓRICO ... 18
2.1. Antecedentes de la investigación ... 18
2.1.1. Internacional ... 18
2.1.2. Nacional ... 28
2.2. Bases teóricas ... 33
2.2.1. Primeras variables: Machine Learning ... 33
2.2.2. Segunda variable: Predicción de demanda ... 38
2.3. Objetivos de la investigación ... 42
2.3.1. Objetivo general ... 42
2.3.2. Objetivos específicos ... 42
2.4. Formulación de hipótesis ... 42
2.4.1. Hipótesis general ... 42
2.4.2. Hipótesis específicas ... 42

Tesis publicada con autorización del autor
No olvide citar esta tesis

TESIS
UNIVERSIDAD CIENTÍFICA DEL SUR

CAPÍTULO III: DISEÑO METODOLÓGICO 43
3.1. Diseño de la investigación 43
3.2. Tipo de Investigación 43
3.3. Nivel de Investigación 44
3.4. Enfoque 44
3.5. Población 45
3.6. Muestra 45
3.7. Muestreo 46
3.8. Operacionalización de variables 46
3.9. Técnicas para la recolección de datos 47
3.9.1. Técnica 47
3.9.2. Instrumentos (Validez y Confiabilidad) 47
3.10. Técnica para el procesamiento y análisis de datos 49
3.11. Aspectos éticos 50
CAPÍTULO IV: RESULTADOS 52
4.1. Ajuste: Parámetros e Hiperparámetros 53
4.2. Indicadores de precisión 56
4.3. Análisis Multicriterio SCORING 57
4.4. Resultados del Sistema 59
4.5. Análisis de Viabilidad Económica 66
4.5.1. Supuestos económicos 66
4.5.2. Flujo de caja económico proyectado 68
4.5.3. Costo de Oportunidad del capital (COK) 69
4.5.1. Evaluación final 71
CAPÍTULO V: DISCUSIÓN, CONCLUSIONES Y RECOMENDACIONES 74
5.1. Discusión 74
5.2. Conclusiones 77
5.3. Recomendaciones 80
Referencias bibliográficas 82
Abreviaturas 89
Anexos 93

Tesis publicada con autorización del autor
No olvide citar esta tesis

TESIS
UNIVERSIDAD CIENTÍFICA DEL SUR

Índice de figuras

Figura 1: Comparación de Modelos Primera Quincena Agosto 60
Figura 2: Comparación de Modelos Segunda Quincena Agosto 62
Figura 3: Comparación de Modelos Primera Quincena Septiembre 63
Figura 4: Comparación de Modelos Segunda Quincena Septiembre 64
Figura 5: Evolución de Incidencias de Quiebres de stock 65
Figura 6: Evolución de Inventario Obsoleto 66
Figura 7: Ficha para observar la eficiencia de abastecimiento 95
Figura 8: Ficha para observar la reducción de excedentes 95
Figura 9: Método SCORING 1 96
Figura 10: Método SCORING 2 97
Figura 11: Método SCORING 3 98
Figura 12: Fotografía de la propuesta de colaboración 99
Figura 13: Fotografía de la obtención del registro de ventas 100
Figura 14: Fotografía del código para definir atributos 101
Figura 15: Fotografía del código para división de datos train y test 102
Figura 16: Fotografía del código para modelo XGBoost 103
Figura 17: *Fotografía del código para modelo Redes Neuronales* 104
Figura 18: Fotografía del código para modelo Random Forest 105
Figura 19: Código para el módulo controlador predictor 106
Figura 20: Código para el módulo controlador loader 107
Figura 21: Formulario para el ingreso de parámetros 108
Figura 22: Salida de resultados en formato tabla 109
Figura 23: Salida de resultados en formato gráfico de barras 109
Figura 24: Flujo de proceso antes del sistema 111
Figura 25: Flujo de proceso después del sistema 111

Tesis publicada con autorización del autor
No olvide citar esta tesis

Índice de tablas

Tabla 1: Resumen de operacionalización de Variables 46
Tabla 2: Selección de Hiperparámetros para el modelo XGBoost 54
Tabla 3: Selección de Hiperparámetros para el modelo RandomForest 55
Tabla 4: Selección de Hiperparámetros para el modelo MLP Redes Neuronales 56
Tabla 5: Evaluación de Precisión de Modelos 57
Tabla 6: Criterios Ponderados 58
Tabla 7: Puntos de satisfacción de los modelos 58
Tabla 8: Resultados del Método SCORING 59
Tabla 9: Predicción de la Primera Quincena de Agosto 60
Tabla 10: Predicción de la Segunda Quincena de Agosto 61
Tabla 11: Predicción de la Primera Quincena de Septiembre 62
Tabla 12: Predicción de la Segunda Quincena de Septiembre 63
Tabla 13: Evolución de Incidencias de Quiebres de stock 65
Tabla 14: Evolución de Inventario Obsoleto 66
Tabla 15: Costos Operativos 67
Tabla 16: Proyección de Costos y Beneficios 68
Tabla 17: Variabilidad del Costo de Oportunidad 69
Tabla 18: Flujo de Caja descantado (COK Variable) 70
Tabla 19: Indicadores Financieros 71
Tabla 20: Comparación TIR vs COK por año 72
Tabla 21: Análisis por escenarios 72
Tabla 22: Matriz de Consistencia 93
Tabla 23: Operacionalización de Variables 94

Tesis publicada con autorización del autor
No olvide citar esta tesis

Índice de anexos

Anexo 1: Matriz de consistencia 93
Anexo 2: Matriz de operacionalización 94
Anexo 3: Fichas de Observación para los indicadores 95
Anexo 4: Análisis Multicriterio SCORING 96
Anexo 5: Evidencias de ejecución de la Investigación 99
Anexo 6: Evidencias de la construcción del programa 101
Anexo 7: Evidencias del funcionamiento del programa 108
Anexo 8: Constancia emitida por la institución donde se realizó la investigación 110
Anexo 9: Flujo de procesos 111
Anexo 10: Informe de aprobación de Tesis del Asesor 112
Anexo 11: Constancia de Consentimiento Informado 113

Tesis publicada con autorización del autor
No olvide citar esta tesis

TESIS
UNIVERSIDAD CIENTÍFICA DEL SUR

Resumen

En el proyecto tomamos como base a la ferretería Vargas, en donde se presentaban deficiencias críticas con respecto a la gestión de la logística al momento de abastecer la tienda. Por ello se puso como objetivo el "Implementar un sistema de Machine Learning con funciones de predicción de demanda, del año 2025". La metodología empleada fue del tipo cuantitativa con diseño experimental y con un enfoque aplicativo dada la naturaleza de la investigación. Los resultados lograron demostrar las validaciones puntuales de las hipótesis, el cómo el algoritmo XGBoost mejoró significativamente la eficiencia del abastecimiento, se redujo el quiebre de stock, por ejemplo, en la segunda quincena de septiembre de 2025, el modelo registró apenas un 0.76% de quiebre frente al 38% del método tradicional. Se concluyó que el uso de algoritmos funciona mejor en campos complejos que alteran la normalidad de las ventas, como lo sucedido en la pandemia, el buscar métodos es crucial para adaptarse y asegurar la continuidad operativa del negocio.

Palabras Clave: Machine Learning, predicción, demanda, gestión de inventarios, quiebre de stock, abastecimiento, variabilidad.

Tesis publicada con autorización del autor
No olvide citar esta tesis

TESIS
UNIVERSIDAD CIENTÍFICA DEL SUR

Abstract

In this project, we used Vargas Hardware Store as the case study, where critical deficiencies were identified in logistics management during the store's replenishment process. Therefore, the objective was to implement a Machine Learning system with demand forecasting capabilities for the year 2025. The methodology adopted was quantitative, with an experimental design and an applied research approach, given the nature of the study. The results provided clear validation of the research hypotheses, demonstrating how the XGBoost algorithm significantly improved replenishment efficiency and reduced stockouts. For example, during the second half of September 2025, the model recorded a stockout rate of only 0.76%, compared to 38% under the traditional method. It was concluded that the use of algorithms performs better in complex environments that disrupt normal sales patterns, such as those experienced during the pandemic. Identifying and implementing advanced methods is crucial for adapting to changing conditions and ensuring business continuity.

Keywords: Machine Learning, forecasting, demand, inventory management, stockouts, replenishment, variability.

Tesis publicada con autorización del autor
No olvide citar esta tesis

Introducción

En el Perú, las ferreterías son muy comunes debido a su función esencial de proveer materiales y herramientas al mercado peruano. Tiene una demanda constante y esto ayuda que sea flexible a lo cambiante del mercado, esa es la razón por la que muchas mypes prefieren este negocio para iniciar sus operaciones. Esto da cierta relevancia a nuestro tema, pues tenemos un negocio presente el cual buscamos mejorar, teniendo por una parte al negocio y en la otra la tecnología que se usará, siendo esta el machine learning, el cual permite automatizar algunos procesos, predecir algoritmos y poder optimizar el uso de recursos. Varias Mypes no se pueden dar el mismo lujo que una empresa grande de permitirse gastos que pueden desbalancear sus ganancias.

El valor de este proyecto radica en buscar mejorar la eficiencia de uno de los procesos más importantes de este negocio, el poder predecir la demanda y el control de inventario, de esta forma disminuir errores y disminuir los gastos al tratar de resolver estos, así como asegurarse la disponibilidad de varias de las mercaderías que se vende.

Se ubicó una ferretería llamada “Vargas” y se especificó que casi no se cuenta con algún sistema que ayude el negocio en sus actividades, esto significa que no muchos empresarios de ferreterías toman como prioridad el cubrir varios de sus procesos con innovadoras herramientas tecnológicas, cosa que se puede notar por algunos errores del inventario, el cual no se nota cuando hay un material que falta abastecer o ver patrones de compra para determinar qué productos son los más demandados.

El machine learning puede ser una herramienta muy poderosa para atender estas dificultades, pues su algoritmo puede permitir el analizar datos de venta para sí

Tesis publicada con autorización del autor
No olvide citar esta tesis

determinar la demanda de futuros productos, así como el correcto abastecimiento y poder reducir el inventario obsoleto dentro del negocio, y con esta información se pueden llegar a mejores decisiones y mejorar las ganancias.

Este proyecto se dirige a las personas que se desempeñan en la ferretería, y están dispuestas a brindarnos la información necesaria para llevar a cabo el proyecto y trabajar de manera colaborativa, además de permitirnos implementar la solución en un entorno favorable para obtener resultados.

Finalmente, según las particularidades de este informe, el capítulo I expone la problemática del estudio, junto con sus preguntas, justificaciones, limitaciones y viabilidad. En el capítulo II, se presentan los antecedentes referenciales, los fundamentos teóricos y las hipótesis planteadas. En el capítulo III, se encuentra la metodología aplicada en la investigación, incluyendo los instrumentos utilizados. En el capítulo IV, se presentan los resultados obtenidos y en el capítulo V, la discusión, las conclusiones y las recomendaciones. Asimismo, se incorporan las referencias bibliográficas y los anexos que respaldan el desarrollo del presente estudio.

Tesis publicada con autorización del autor
No olvide citar esta tesis

CAPÍTULO I: PLANTEAMIENTO DEL PROBLEMA

1.1. Descripción de la realidad problemática

A nivel mundial, particularmente en países de bajos y medianos ingresos, existen industrias o negocios que dentro de todo concepto se consideran tradicionales, negocios que tienen mucha resistencia al cambio o innovaciones. Según Vargas (2022), en una nota técnica de Agronegocios toma como base el sector agroalimentario mexicano, ya que según estudios anteriores a este se demostró que el sector agroalimentario no le da su debida importancia a la adopción de nuevas tecnologías. Sacando de esta manera resultados basado en las unidades de producción con uso de tecnologías informáticas en las tareas agropecuarias; saca como resultados que el uso de computadora para la agricultura digital tiene un promedio nacional del 5.57%, son solo 4 estados de México quienes toman una importante consideración de estas, haciendo que el promedio sea bajo.

De igual forma Vargas no solo hace hincapié que este problema es a causa de una subestimación de los recursos tecnológicos, sino que habla de cómo esa subestimación es muchas veces causada por una ignorancia sobre el tema. En República Dominicana, Vargas & Abreu (2023) mencionan de manera directa de cómo el no estimar de manera correcta los inventarios afecta la rentabilidad en empresas que se dedican directamente al comercio, como ejemplo muestran el caso de la Empresa “Repuestos Alba SRL”, y de que lo indispensables es tener una buena administración de inventarios o un equipo competente en el área de logística. Para dar énfasis a este punto, la investigación muestra una de las utilidades percibidas más desastrosas en un cierre del periodo 2020, a causa de la mercancía obsoleta que generó un gasto de RD$ 310,498.84, una variación del 26%, una cuarta parte de la utilidad neta se quedó como gasto que ya no se puede

Tesis publicada con autorización del autor
No olvide citar esta tesis

recuperar. En la revista China Journal of Digital Economy, Liu et al. (2025), habla de cómo después de la pandemia del COVID-19 muchos negocios enfrentaron dificultades para continuar en operaciones, su rentabilidad está en riesgo. Muchos negocios optaron por tener quiebres de stock, lo que significa no gastar recursos en sobreestimar inventario, prefirieron subestimar. Pero lejos de arreglar el problema de los gastos, lo que ocasionó es que los clientes se fueran con la competencia de esas tiendas, porque seguidamente no encontraban lo que querían. Al no atender la demanda de manera correcta no solo se puede generar pérdidas de gasto, sino el arriesgar la fidelidad del cliente.

A nivel de América Latina, Inostroza (2025) traslada esa problemática a un contexto de salud pública en Chile, si bien no va con nuestro trabajo, lo importante aquí es como el Hospital del Salvador enfrenta desafíos para garantizar el abastecimiento de los insumos clínicos. Inostroza hecha completamente a lo rudimentario y a la burocracia, menciona una desconexión tecnológica, esto limita la planificación para un volumen crítico de 1957 insumos. Lo que demuestra que un mal abastecimiento puede afectar más que solo el sector del comercio ya que aquí afecta directamente a la salud pública. Si vemos lo que señala en su estudio Vesga (2020), resaltó el impacto económico de esto, en donde se le dio un valor cercano al 30 % de los activos de una empresa. Esto porque se tienen variables relacionadas con los pronósticos de lo que entra y sale con mayor frecuencia, así al momento de comprar las mercaderías, saber cuáles productos llenar más y cuáles disminuir o retirar su stock.

A nivel nacional, el Perú está presenciando una transformación notable en el escenario empresarial. Según la tesis de Aguirre et al. (2021), se discute el impacto significativo en la integración de la Inteligencia Artificial como táctica en el

Tesis publicada con autorización del autor
No olvide citar esta tesis

ámbito de los negocios. Esta adopción no solo está redefiniendo las operaciones comerciales, sino que también está configurando el futuro de las empresas peruanas. El 23% de los CEO peruanos ya están implementando cambios en el propósito de sus organizaciones, mientras que otro 35% está considerando hacerlo. Este cambio es un indicativo positivo para las organizaciones peruanas, ya que sugiere un movimiento hacia la mejora de sus servicios y finalmente, el contentamiento de sus consumidores

Durante la pandemia, la inteligencia artificial ha sido la tecnología más utilizada por las organizaciones, según un artículo publicado en el diario El Peruano (El Peruano, 2021). Los profesionales ven ventajas en la implementación de la IA en sus negocios o emprendimientos, ya que facilita la automatización en sus procesos de ventas. La inteligencia artificial se puede implementar en diversas fases del control de inventario al igual que en su ciclo de facturación. Además, la reducción de costos es un beneficio adicional, especialmente para las empresas que buscan formas de reducir gastos. Según Aguirre et al. (2021), El 21% de los profesionales capacitados en usar las tecnologías de información han abierto el camino para el correcto uso de las inteligencias artificiales. Asimismo, el 43% de estos especialistas señala que debido a la pandemia del COVID-19 ellos como empresarios han acelerado la implementación de inteligencia artificial. Este dato evidencia el impacto significativo y la importancia de la inteligencia artificial en el entorno empresarial actual, que cada vez es más creciente y globalizado.

Desde una perspectiva local, la ferretería Vargas se enfrenta a una realidad desafiante debido a la falta de tener un control de inventario y la falta de predecir la demanda. Esta situación ha provocado la pérdida de productos, desorden en el almacén, al no poder saber la cantidad que se va a vender o el stock que queda en

Tesis publicada con autorización del autor
No olvide citar esta tesis

almacén, lo que ha resultado en pérdidas financieras significativas y un descenso en el contentamiento de los clientes. La ausencia de un sistema dispuesto a un seguimiento y control de inventario ha dificultado que la empresa mantenga un registro preciso de sus productos, lo que a menudo resulta en la sobreestimación o subestimación de su inventario.

1.2. Formulación del problema

1.2.1. Problema general

¿De qué manera la implementación de machine learning optimiza los métodos de predicción de la demanda de productos, en la ferretería "Vargas", año 2025?

1.2.2. Problemas específicos

Problema Específico 1:

¿En qué medida la predicción de demanda mediante algoritmos machine learning optimiza el abastecimiento de productos de la ferretería Vargas, en 2025?

Problema Específico 2:

¿En qué medida la predicción de demanda mediante algoritmos machine learning influye en la reducción de inventario excedente en la ferretería Vargas, en 2025?

Tesis publicada con autorización del autor
No olvide citar esta tesis

1.3. Justificación de la investigación

1.3.1. Justificación Metodológica

Este trabajo se justifica por medio de un enfoque cuantitativo, debido al trabajar con las predicciones del machine learning, se obtendrán resultados numéricos y probabilísticos, al evaluar las propiedades de un conjunto de datos. El modelo hipotético-deductivo facilita el poder formular y probar una hipótesis, al tener un análisis sistemático de los datos, dando más validez a los resultados que se obtendrán.

1.3.2. Justificación Teórica

Este estudio se justifica ya que busca cubrir ciertos vacíos existentes. En el control de inventarios en el aspecto que hay pocos trabajos que exploran el contexto de las ferreterías. Según Silva (2021, el machine learning, se basa en el uso de algoritmos para procesar grandes cantidades de información y realizar predicciones o sugerencias precisas. Al aprovechar la inteligencia artificial (IA) para optimizar procesos, pronosticar la demanda y reducir errores, esta tecnología revolucionará la gestión logística y aumentará la eficiencia operativa y la satisfacción del cliente en el competitivo mercado.

1.3.3. Justificación Práctica

Este trabajo se justifica porque los resultados tendrán aplicaciones prácticas que brindarán una mejora en el ámbito de gestión de logística para la ferretería "Vargas". Al poder desarrollar un modelo de ML el negocio podrá predecir las demanda o agotamientos de insumos, permitiéndole ser más efectivo al momento de comprar las mercaderías, así como también dar un mejor control al abastecimiento de productos que se pueden estar agotando.

Tesis publicada con autorización del autor
No olvide citar esta tesis

1.3.4. Justificación Económica

Esta herramienta de machine learning facilita la optimización de la previsión de la demanda y gestión de compras. Ayuda a reducir los costes operativos, las pérdidas por sobreproducción o desabastecimiento y mejorando la rotación de productos. Esto mejora la eficiencia del negocio, mayor satisfacción del cliente y da cabida a la rentabilidad.

1.4. Limitaciones de la investigación

Las limitaciones que se consideraron indispensables para la presentación del trabajo son las siguientes:

- **Limitación en el acceso a la muestra**

Por la naturaleza de la investigación y la disponibilidad de datos, la muestra está dedicada únicamente a los productos de la categoría de construcción en los últimos 5 años (2020-2024), para poder realizar la observación para la aplicación de machine learning.

Si bien la muestra se limita por fechas predilectas, esto no compromete la utilidad futura del sistema, el software ha sido armado bajo un enfoque escalable. La ferretería va a seguir generando nuevos registros históricos de venta, un modelo de Machine Learning requiere lo que es reentrenar continuamente, garantizando la vigencia y adaptación continua en la precisión de las precisiones a lo largo del tiempo.

Tesis publicada con autorización del autor
No olvide citar esta tesis

- **Limitaciones temporales**

Las limitaciones temporales están dedicadas a poder evaluar las fichas de observación en dos tiempos, antes de la implementación del Machine Learning y después de la implementación de Machine Learning.

Para mitigar esta restricción se decidieron aplicar las fichas de observación y las de salida se aplicarán en períodos de alta transaccionalidad, para garantizar el flujo de datos denso. De esta manera nuestro trabajo será representativo para evidenciar la optimización y qué diferencias existen entre un método y los algoritmos.

1.5. Viabilidad de la investigación

La viabilidad del proyecto se considera alta, ya que Vargas Hardware ha expresado interés en mejorar la gestión de inventario y ventas mediante el uso de técnicas de aprendizaje automático espacial. Esta iniciativa busca optimizar la eficiencia operativa y fortalecer el posicionamiento de la empresa en el mercado local. El proyecto se basa en estrategias basadas en datos históricos reales. Se utilizó Python como herramienta de desarrollo principal en el entorno de Visual Studio Code.

- **Viabilidad técnica, económica y operativa:** La ferretería cuenta con el equipo básico necesario para implementar el modelo, incluyendo una computadora compatible con Python y Visual Studio Code. También se utilizaron bibliotecas de código abierto como Pandas, Scikit-Learn y Matplotlib, lo que permitió desarrollar el sistema sin costosas licencias. Esto hace que la implementación sea sostenible y viable, incluso para pequeñas empresas.

Tesis publicada con autorización del autor
No olvide citar esta tesis

- **Oportunidades y retorno de la inversión:** Después de la implementación de machine learning mejoró significativamente la planificación de compras, reducir las pérdidas por artículos no vendidos y aumentar la disponibilidad de productos de alta demanda. Esto facilitó una mejor toma de decisiones estratégicas y permite al equipo centrarse en actividades de mayor valor, lo que repercute positivamente en las ventas y la eficiencia del negocio.
- **Escalabilidad y mantenimiento:** El sistema desarrollado se adapta al crecimiento de la ferretería y permite la integración de nuevos datos de ventas y productos a medida que las operaciones se expanden. Además, el modelo puede actualizarse periódicamente para mejorar su precisión y posteriormente integrarse con otras herramientas, como hojas de cálculo o sistemas de gestión empresarial, manteniendo así la flexibilidad en la gestión de la información.

Tesis publicada con autorización del autor
No olvide citar esta tesis

CAPÍTULO II: MARCO TEÓRICO

2.1.Antecedentes de la investigación

2.1.1. Internacional

En el trabajo de investigación "Machine Learning para predecir volúmenes operacionales de las líneas de negocio de Cenabast.", Madariaga (2021), quien en su investigación desarrollada en la universidad de chile, tuvo como objetivo el implementar machine learning para predecir los volúmenes operacionales de las líneas de negocio de Cenabast, para este estudio se hizo datos históricos de las compras que fueron realizada entre 2022 a 2021, analizando 441 medicamentos que fueron seleccionados por homogeneidad, para este análisis se aplicó seis modelos de machine learning: regresión lineal, árbol de decisión, bosque aleatorio, Ridge, Lasso y Elastic Net. lo cual mostró como resultado que el modelo de regresión lineal obtuvo un mejor desempeño, con un coeficiente determinado del 72,76%, por otro lado, el árbol de decisión con hiperparámetros ajustados alcanzó un 73,14%. como conclusión este estudio destaca la viabilidad de utilizar técnicas de machine learning para las planificaciones de compras en el sector público, lo cual esto genera ahorros significativos y ayuda a optimizar la gestión de recursos, esta investigación demuestra que la aplicación de un modelo predictivo mejora la toma de decisión, tomando en cuenta los datos históricos y es apta para grandes volúmenes de datos históricos.

En otro trabajo realizado en Colombia, por Manrique y Camargo (2022), en su investigación realizada por la Fundación Universitaria Konrad Lorenz, con el objetivo de predecir la demanda de teléfonos inteligentes introductorios en el mercado colombiano utilizando modelos de aprendizaje automático Este estudio

Tesis publicada con autorización del autor
No olvide citar esta tesis

utilizó datos históricos de ventas de dispositivos móviles vendidos entre enero y junio de 2022, proporcionados por una empresa de telecomunicaciones. La muestra consistió en 419 registros y 12 variables, incluyendo precio, especificaciones del producto y cantidad vendida. También se emplearon modelos supervisados como árboles de decisión, bosques aleatorios, máquinas de vectores de soporte, redes neuronales y XGBoost, que se evaluaron mediante métricas como RMSE, MAE y MAPE. Los resultados mostraron que el modelo XGBoost tuvo el mejor rendimiento, destacando variables como el precio del dispositivo, la marca y el tamaño de la batería como las más influyentes en la predicción. En conclusión, los investigadores destacaron la eficacia del aprendizaje automático para pronosticar la demanda de nuevos productos, superando a los métodos tradicionales, y recomendaron actualizaciones periódicas de los modelos para incorporar los cambios tecnológicos y las preferencias de los consumidores. Este estudio proporciona evidencia relevante para la aplicación de técnicas de aprendizaje automático en la gestión de compras, demostrando su utilidad en la optimización de inventarios y la planificación estratégica. Este enfoque puede adaptarse a otros sectores, como las tiendas de artículos para el hogar, para mejorar la precisión de las previsiones de demanda y la eficiencia de la cadena de suministro.

Según Romano (2025), en su investigación titulada *"Deep Learning para predicción de demanda retail a gran escala: un modelo inteligente por producto para la totalidad de sus tiendas"*, señaló que la predicción de demanda en la industria del retail enfrenta desafíos clave en contextos de alta granularidad, donde miles de combinaciones producto-tienda deben gestionarse de manera simultánea. Indicó que el sistema Smart Replenishment de Scanntech utiliza actualmente un

Tesis publicada con autorización del autor
No olvide citar esta tesis

método basado en el promedio de ventas de los últimos 30 días, el cual ha demostrado ser inexacto, generando problemas de sobrestock y desabastecimiento. Por ello, consideró que las empresas deben adoptar modelos robustos y escalables de deep learning para mejorar la precisión de los pronósticos.

El objetivo de su estudio fue optimizar la predicción de demanda dentro del sistema Smart Replenishment mediante el desarrollo de una estrategia de modelado basada en deep learning, entrenando un único modelo por producto que integrará la información de todas las tiendas asociadas. La variable independiente fue el modelo de deep learning, operacionalizado en cuatro arquitecturas: Simple FeedForward (MLP), DeepAR, Temporal Fusion Transformer (TFT) y WaveNet. La variable dependiente correspondió a la precisión de la predicción de demanda, evaluada a través de tres dimensiones: Error Cuadrático Medio Raíz (RMSE), mejora porcentual respecto al benchmark y ranking medio de desempeño. El enfoque fue cuantitativo, con un diseño no experimental, longitudinal y comparativo.

La población estuvo conformada por datos históricos de ventas de productos de las categorías de refrigerantes y cervezas en supermercados de São Paulo, Brasil, entre diciembre de 2022 y noviembre de 2024. La muestra incluyó 839.178 transacciones, 30 tiendas, 154 productos y 1.319 combinaciones producto-tienda, segmentadas en tres clústeres según nivel de demanda: moderada, alta y media-alta. Para la recolección de datos se utilizó la base histórica de ventas con variables de fecha, código de tienda, código de producto, nombre de producto y cantidad vendida, aplicando análisis mediante Python con bibliotecas GluonTS, PyTorch Lightning y optimización bayesiana con TPE.

Tesis publicada con autorización del autor
No olvide citar esta tesis

Los resultados mostraron que el modelo Simple FeedForward superó sistemáticamente al benchmark de promedio de 30 días en todos los clústeres. En demanda alta ganó en el 88,9% de las combinaciones, con una mejora mediana del 17,3% en RMSE (30.709 frente a 38.957 del benchmark). En demanda media-alta obtuvo victorias en el 85,2% de los casos, con una mejora del 12,7% (15.902 frente a 19.163). En demanda moderada alcanzó un 55,7% de victorias, con una mejora del 6,6% (6.287 frente a 6.627). Además, superó a XGBoost en los tres clústeres, con tiempos de entrenamiento significativamente menores: 3.764 segundos frente a los 59.000 de DeepAR, 51.000 de TFT y 50.000 de WaveNet. La simulación en Smart Replenishment durante 10 días mostró que Simple FeedForward habría evitado la pérdida de 7,1 millones de unidades en ventas, reduciendo la ruptura de stock de 101,5 millones a 94,4 millones de unidades, lo que representó un ahorro económico estimado de 3,7 millones de dólares en comparación con el método actual. Se concluyó que la estrategia de modelado a nivel de producto con arquitectura Simple FeedForward constituye la solución más robusta, precisa, eficiente y escalable para la predicción de demanda a gran escala en retail, superando tanto a métodos tradicionales como a algoritmos de machine learning consolidados como XGBoost.

Esta investigación ofrece una metodología directamente aplicable al sector ferretero, donde se manejan miles de referencias distribuidas en múltiples sucursales. Al entrenar un único modelo por producto, permite capturar patrones generales de manera eficiente y escalable, lo que resulta clave para artículos de alta rotación como cemento, pinturas o electrodos, así como para productos especializados de baja demanda.

Tesis publicada con autorización del autor
No olvide citar esta tesis

En otro trabajo de investigación "Aplicación de Técnicas de Aprendizaje Supervisado para la Predicción de Ventas en Hentel, mediante la Implementación de Modelos de Machine Learning con Python y Pandas", Cazares (2025), tuvo como objetivo principal desarrollar un modelo de predicción de precios para la empresa Hentel, dedicada a la comercialización de productos tecnológicos, mediante la aplicación de técnicas de aprendizaje supervisado.

Para ello, se utilizó un conjunto de datos históricos con más de 6000 registros de ventas comprendidos entre los años 2014 y 2018, que incluía variables como precio del producto (variable objetivo), peso del producto (weight_lbs), tiempo en el mercado (days_on_market), condición (prices_condition), disponibilidad (prices_availability), indicador de oferta (prices_isSale) y marca (brand). Se aplicaron tres modelos de machine learning: regresión lineal, Random Forest y XGBoost, implementados en Python con las librerías Pandas, Scikit-learn y SHAP. Los resultados mostraron que los modelos Random Forest y XGBoost superaron ampliamente a la Regresión Lineal, con coeficientes de determinación (R^2) de 0,733 y 0,741 respectivamente, frente a un R^2 de 0,446 de la regresión lineal. En cuanto al error absoluto medio (MAE), Random Forest obtuvo 17.526,91 (mejor que XGBoost con 18.169,88), mientras que en RMSE, XGBoost fue ligeramente superior (35.266,73 vs 35.761,23 de Random Forest). El análisis de importancia de variables reveló que Random Forest prioriza el peso del producto y el tiempo en mercado, mientras que XGBoost otorgó mayor relevancia a la marca (LG, Razer, Apple, Samsung).

Como conclusión, el estudio destaca que Random Forest es el modelo óptimo para Hentel por su equilibrio entre precisión (R^2=0,733) y facilidad de interpretación, permitiendo optimizar la estrategia de fijación de precios (pricing dinámico),

Tesis publicada con autorización del autor
No olvide citar esta tesis

mejorar la competitividad y reducir errores en la toma de decisiones. Esta investigación demuestra que la implementación de modelos de machine learning supervisado puede transformar significativamente la planificación estratégica en empresas del sector tecnológico, promoviendo una cultura basada en el análisis de datos.

Flores y Guerra (2023), en su investigación titulada *"Análisis comparativo de modelos predictivos de Machine Learning para la proyección de ventas de productos Nestlé en supermercados TIA en Ecuador", como* objetivo principal desarrollar y aplicar modelos de Redes Neuronales Artificiales (RNA) y Random Forest para la proyección de ventas (Sell Out) de productos Nestlé en los supermercados TIA en Ecuador, comparando la eficacia de ambos métodos en la precisión de las predicciones. Para ello, se utilizó una base de datos histórica de ventas diarias comprendida entre enero y julio de 2023, agregada a nivel mensual por local, ciudad y región, proveniente de los sistemas de Nestlé Ecuador (datalake en Azure con procesamiento en Databricks).

Las variables incluidas fueron: año, mes, local (TIA), ciudad, región, subcadena, unidades vendidas, venta neta, totalSo (variable objetivo: venta realizada por Nestlé a TIA), stock y totalStock. Se aplicaron dos modelos: Random Forest (optimizado con OOB error y grid search, con 131 árboles y profundidad sin delimitar) y Redes Neuronales Artificiales (perceptrón multicapa con arquitectura optimizada mediante búsqueda grid, función de activación ReLU, 15 neuronas en primera capa oculta y 10 en la segunda, batch_size de 25 y 25 épocas). Los resultados mostraron que Random Forest superó significativamente a las RNA, con un Error Cuadrático Medio (MSE) de 393,51 frente a 1278,84, y un coeficiente de determinación (R^2) del 95% frente al 84%. Además, Random Forest

Tesis publicada con autorización del autor
No olvide citar esta tesis

fue mucho más eficiente en tiempo de ejecución (1,31 segundos vs 95 segundos de RNA) y en calibrado de hiperparámetros (50 segundos vs 3242 segundos). El análisis de importancia de variables reveló que las características más influyentes fueron: unidades vendidas, stock, totalStock y mes. Como conclusión, los autores destacan que Random Forest es el modelo óptimo para la predicción de ventas en Nestlé Ecuador por su alta precisión (R^2=95%), eficiencia computacional y robustez, recomendando su implementación como herramienta principal para la optimización de inventarios, planificación estacional y mejora de la relación con el cliente clave TIA. Esta investigación demuestra que la adopción de técnicas avanzadas de machine learning permite a las organizaciones anticipar la demanda, reducir ineficiencias en la cadena de suministro y fomentar una cultura de innovación y toma de decisiones basada en datos.

La implementación de machine learning para la predicción de demanda en una ferretería permitirá que el negocios pueda anticipar con mayor precisión las necesidades de inventario de productos como herramientas, materiales de construcción, ferretería en general y artículos de alta rotación, considerando variables estacionales como temporada de lluvias, feriados o campañas escolares, comportamiento histórico de ventas, promociones, días festivos y factores externos como el comportamiento de la competencia o el mercado de la construcción. Al aplicar modelos como Random Forest, XGBoost o Redes Neuronales, la ferretería podría optimizar sus niveles de stock, reducir el sobreabastecimiento o el desabastecimiento, mejorar la planificación de compras y aumentar la rentabilidad, todo ello mediante una toma de decisiones basada en datos y no en la intuición.

Tesis publicada con autorización del autor
No olvide citar esta tesis

Martínez y Alarcón (2025), en su investigación titulada *"Aplicación de modelo predictivo para análisis de ventas aplicado a una cafetería mediante la implementación de técnicas de aprendizaje automático y series temporales"*, señalaron que las cafeterías familiares en Colombia enfrentan una fuerte presión competitiva debido a la llegada de cadenas internacionales con sistemas avanzados de análisis de datos. La ausencia de herramientas objetivas para anticipar la demanda ha llevado a que las decisiones se tomen con base en la intuición, lo que genera sobreproducción, desperdicios y desabastecimiento. Por ello, los autores sostienen que estos pequeños negocios deben incorporar metodologías de ciencia de datos para competir de manera efectiva.

El objetivo del estudio fue desarrollar un modelo predictivo de ventas que integrará técnicas de series temporales, específicamente ARIMA, junto con algoritmos de aprendizaje automático supervisado como Regresión Lineal y Random Forest, siguiendo la metodología SEMMA. La variable dependiente fue el nivel de ventas, entendida como la predicción de cantidad de productos e ingresos. Las variables independientes incluyeron características temporales como día de la semana, mes, trimestre y fin de semana, además de transformaciones cíclicas; características del producto como precio unitario, costo y código; y características del cliente representadas mediante un identificador codificado. El enfoque fue cuantitativo, con un diseño no experimental, transversal y correlacional.

La población correspondió al conjunto de transacciones históricas de una cafetería familiar, con una muestra de 65.193 registros obtenidos durante 111 días entre mayo y agosto de 2025. Para la recolección de datos se utilizó un archivo Excel

Tesis publicada con autorización del autor
No olvide citar esta tesis

con el historial de transacciones, y el análisis se realizó en Python mediante las bibliotecas pandas, scikit-learn y statsmodels.

Los resultados evidenciaron que el modelo Random Forest alcanzó un coeficiente de determinación (R^2) de 0,978 en la predicción de ingresos, lo que indica que explicó el 97,8% de la variabilidad de las ventas. Asimismo, la Regresión Lineal obtuvo un R^2 de 1,000 en la predicción de cantidades. Estos hallazgos permitieron concluir que los algoritmos de aprendizaje automático superan significativamente a los modelos tradicionales de series temporales como ARIMA, ofreciendo ventajas para la optimización de inventarios, la planificación del personal y el diseño de estrategias comerciales basadas en datos.

Kamarthi et al. (2024), en su investigación titulada *"Hierarchical Industrial Demand Forecasting with Temporal and Uncertainty Explanations"*, señalaron que la previsión de demanda industrial jerárquica enfrenta desafíos significativos debido a que los modelos de aprendizaje profundo de última generación funcionan como cajas negras que no ofrecen explicaciones sobre sus predicciones. Esta limitación reduce la confianza y la adopción por parte de los actores involucrados en entornos industriales, como la planificación de producción y la gestión de inventarios. Por ello, los autores sostienen que resulta cada vez más necesario desarrollar métodos de interpretabilidad adaptados a estructuras jerárquicas y salidas probabilísticas.

El objetivo del estudio fue proponer un método novedoso de interpretabilidad para la previsión jerárquica probabilística de series temporales a gran escala, adaptando técnicas genéricas de interpretabilidad a los retos de las estructuras jerárquicas y la incertidumbre. La variable independiente fue el método de interpretabilidad

Tesis publicada con autorización del autor
No olvide citar esta tesis

HIEREINTERPRET, con dos modificaciones principales: la aproximación de subárbol y la alternativa determinista basada en cuantiles. La variable dependiente correspondió a la precisión de la explicación, evaluada mediante dos dimensiones: Importance Accuracy Score (IAS) y External Variable Detection Accuracy (EVDA). El enfoque fue cuantitativo, con un diseño no experimental, transversal y correlacional.

La población estuvo conformada por datos de demanda de una gran empresa química (Dow) junto con otros conjuntos de datos públicos como M5, Tourism-L y Wiki. La muestra incluyó más de 10.000 productos y series temporales distribuidas en hasta 12 niveles jerárquicos, además de 1.000 datasets semisintéticos generados para la evaluación. Para la recolección de datos se utilizaron registros históricos de ventas mensuales entre 2018 y 2023, así como indicadores económicos externos, aplicando técnicas de generación de series sintéticas con anomalías asignadas a distintos niveles jerárquicos.

Los resultados mostraron una mejora promedio del 62% en la detección de dependencias entre variables y del 12,3% en dependencias de la variable objetivo para pronósticos determinísticos, además de mejoras entre el 10% y el 25% en pronósticos probabilísticos. Estos hallazgos evidencian que la aproximación de subárbol y la adaptación por cuantiles permiten explicar con alta precisión modelos jerárquicos de forecasting. Se concluyó que el método HIEREINTERPRET identifica de manera efectiva tendencias históricas relevantes, indicadores económicos clave que influyen en las predicciones y cambios en la varianza e incertidumbre de los pronósticos, lo que facilita una toma de decisiones más informada y estratégica en la planificación de la demanda industrial.

Tesis publicada con autorización del autor
No olvide citar esta tesis

2.1.2. Nacional

En la tesis “Sistema basado en Machine Learning para la mejora de gestión de inventario en A&M Confecciones”, Gonzales y Pérez (2023) realizó un trabajo con el propósito de optimizar el control de inventario en una empresa textil con el objetivo de implementar un sistema en machine learning para incurrir la demanda que puedan llegar a tener sus productos y mejorar la gestión de estos mismos dentro de sus almacenes. Para este propósito, se ha operacionalizado la variable independiente como el sistema basado en aprendizaje automático. La variable dependiente es la gestión de inventarios. Esta investigación tuvo un diseño preexperimental y un enfoque de modo cuantitativo. La muestra estuvo conformada por 30 registros, los métodos utilizados para recopilar los datos incluyeron la observación y el uso de formularios de registro. Producto de esta investigación se revelaron que la adopción de un sistema de machine learning condujo a un incremento del 20% en la exactitud de las proyecciones de demanda, así como a una disminución del 15% en los gastos vinculados al sobrante de inventario. En tal sentido se pudo afirmar que la aplicación de tecnologías basadas en machine learning resulta ser extremadamente ventajosa para optimizar el manejo de inventarios en compañías textiles. En conclusión, se ha evidenciado una optimización notable en el control de inventarios de A&M Confecciones, dicha mejora facilita la solución a dificultades tales como el desequilibrio logístico, la gestión de envío.

En otro trabajo realizado por Bernaola y Varillas (2022), diseñó y evaluó un sistema de predicción utilizando aprendizaje automático para mejorar el control de los inventarios. Para ello ha operacionalizado su variable independiente: que es machine learning. Y la variable dependiente Gestión de Inventario. Esta

Tesis publicada con autorización del autor
No olvide citar esta tesis

investigación tuvo un diseño preexperimental de corte transversal de nivel correlacional de enfoque cuantitativo. La muestra estuvo compuesta por 190 nombres de productos de la ferretería, organizados en 30 divisiones distintas. Se seleccionaron 30 tipos de categorías mediante un proceso de muestreo aleatorio simple para realizar el análisis. Los instrumentos que emplearon para la recolección de los datos fueron aplicados la técnica del fichaje. Producto de esta investigación los resultados indicaron un avance considerable en la precisión de las predicciones de inventario, con un aumento del 25% en la exactitud en comparación con los métodos tradicionales de gestión de inventario. Además, se observó una reducción del 20% en los costos asociados con el exceso de inventario y una disminución del 30% en las pérdidas por falta de stock. En tal sentido se pudo afirmar que estos hallazgos resaltan el potencial del machine learning para optimizar la gestión de inventario y elevar la eficiencia operativa en empresas del sector minorista.

Este estudio se alinea estrechamente ya que demuestra el potencial de las técnicas de machine learning para optimizar la gestión de inventario, lo cual es fundamental para mejorar la eficiencia operativa y reducir costos en empresas similares. Incorporar estos hallazgos en la investigación del proyecto permitirá mejorar prácticas identificadas en la investigación anterior, esto agiliza la creación e implementación del sistema propuesto, al tiempo que aumenta su eficacia en la predicción de demanda y gestión de compras.

Montalvo (2022) analizó el impacto del machine learning en el coeficiente de rotación de inventarios y su relevancia para la gestión logística empresarial. Para ello ha operacionalizado su variable independiente: Machine learning. Y la variable dependiente control de logística. Este estudio se llevó a cabo utilizando

Tesis publicada con autorización del autor
No olvide citar esta tesis

un diseño experimental de tipo transversal, con un enfoque cuantitativo y un nivel correlacional. La muestra estuvo conformada por una población de 1415 documentos que reflejan la rotación de inventario, originados de los pedidos hechos al almacén. Se seleccionó una muestra de 343 productos que presentaron la mayor frecuencia de rotación diaria, organizados en 28 fichas de registro a lo largo de cuatro semanas de siete días. De forma paralela, se evaluó el nivel de cumplimiento en los despachos, partiendo de 1146 productos entregados. De estos, se tomó una muestra de 327 productos despachados, igualmente distribuidos en 28 fichas de registro para el mismo periodo de cuatro semanas. Los instrumentos que emplearon para la recolección de los datos fue la técnica es la de fichaje y las fichas de registro. Producto de esta investigación los datos revelan que la tasa de movimiento de inventario tuvo un incremento significativo aumento, pasando del 26.5% al 53.25%. De manera similar, el nivel de cumplimiento en los despachos mejoró significativamente, elevándose del 56.3% al 80.72%. En tal sentido se pudo afirmar que la adopción de tecnologías de machine learning ha tenido un impacto positivo y considerable en la optimización de la gestión logística de la chicharronería Don Jovi S.A.C., lo que permite afirmar su eficacia en el proceso.

Esta investigación refuerza la idea de que el machine learning es una herramienta valiosa para empresas de diversos sectores que buscan mejorar su eficiencia operativa y reducir costos, ya que al enfocarse no solo en el impacto del machine learning en el coeficiente de predicción de demanda y su gestión de compras, sino que también en su rotación de inventarios, lo cual proporciona una visión más completa del potencial del machine learning para optimizar la gestión logística en las empresas.

Tesis publicada con autorización del autor
No olvide citar esta tesis

Según Flores (2021), en su trabajo realizó un estudio para optimizar la gestión de compras en ferreterías, específicamente en MAYTA Hardware. Se desarrolló un modelo predictivo para determinar qué productos comprar para mantener un inventario suficiente. Este modelo busca optimizar el proceso de compra, reducir costos innecesarios y mejorar la calidad del servicio al cliente para evitar el exceso de inventario y la falta de stock. En este estudio, el modelo predictivo fue la variable independiente, mientras que el nivel óptimo de inventario fue la variable dependiente.

La investigación aplicada y experimental empleó la metodología CRISP-DM, un reconocido método de minería de datos que incluye fases como la comprensión del negocio y de los datos, la preparación, el modelado, la evaluación y la implementación. La muestra incluyó 168 productos, seleccionados de un total de 298 artículos vendidos entre 2020 y 2021. Para obtener esta información, se escanearon documentos de venta físicos en hojas de cálculo de Excel. Posteriormente, se depuraron los datos, se estandarizaron los nombres de los productos, se eliminaron los documentos incompletos y se codificaron las variables.

En la fase de análisis se utilizó la herramienta WEKA, que permitió la aplicación de diversos algoritmos de clasificación, como J48, RandomForest, RandomTree y JRip. Entre los hallazgos clave, el algoritmo J48 demostró ser el más eficaz para predecir qué productos comprar en ferreterías, seguido del modelo RandomTree. Ambos enfoques permiten tomar decisiones informadas basadas en datos históricos, lo que contribuye a una gestión de inventario más eficiente. Gracias a este modelo predictivo basado en minería de datos y aprendizaje automático, MAYTA Hardware logró mejorar significativamente su proceso de compras y, al

Tesis publicada con autorización del autor
No olvide citar esta tesis

considerar la tasa de rotación y las ganancias generadas por cada artículo, identificar con gran precisión qué productos reordenar, cuáles evaluar y cuáles evitar.

Finalmente tenemos un artículo realizado por Andrade et al. (2024), titulado "Técnicas de Machine Learning aplicadas a la interpretación de datos de mercado" que tuvo el objetivo de explorar técnicas de inteligencia artificial en distintos sectores empresariales, para la toma de decisiones estratégicas referentes a la demanda de productos o servicios. El trabajo define sus variables de la siguiente manera, la variable independiente es representada por las variables de regresión, por otro lado, su variable independiente es la eficiencia empresarial, donde sus dimensiones se dividen en sectores importantes del país. El diseño es caso múltiple y de enfoque mixto y de alcance correlacional, para la población y muestra se segmento una cantidad importante de clientes de distintos sectores empresariales. Los resultados fueron que los modelos de regresión lograron reducir hasta un 5% los costos de inventario y la satisfacción del cliente aumentó un 10%, esto al mejorar la predicción de la demanda junto a otras optimizaciones de las operaciones. Finalizando, colocaron como conclusiones que el avance es significativo, y estos modelos de clasificación y predicción ayudan a la segmentación de los clientes, permitiendo conocer los productos y servicios de su interés, y con ello se aumentó su letalidad a la empresa.

Tesis publicada con autorización del autor
No olvide citar esta tesis

2.2. Bases teóricas

2.2.1. Primeras variables: Machine Learning

En 1995 Vladimir Vapnik publicaría su libro más importante, llamado "The Nature of Statistical Learning Theory", considerado la biblia del aprendizaje automático. En este libro incluiría sus teorías desarrolladas, de manera personal o en compañía de otros matemáticos y estadistas, siendo el más importante Alexey Chervonenkis.

Entre sus teorías más importantes están la Teoría del Aprendizaje Estadístico en base a formulas, esta plantea inducir una función predictiva en base a un conjunto de datos observados.

$$(x_1, y_1), \dots, (x_l, y_l) \subset X \times Y$$

En el escenario de predecir ventas o demandas, el sistema busca la aproximación de una función objetivo desconocida basándose en un histórico conjunto de datos. Su núcleo de esta teoría se sostiene sobre el Principio de Minimización del Riesgo Empírico (ERM). Su mismo nombre indica que busca minimizar el riesgo real de las funciones de decisión, siendo su fórmula matemática la siguiente.

$$R(w) = \int L\left(y, f(x, w)\right) dP(x, y)$$

La primera parte conformada por L (…) cuantifica la discrepancia entre el valor real y la predicción estimada. Y el dP(…) representa la función de distribución de probabilidad, la cual en un principio es desconocida.

Otra teoría que destacar es, Teoría Vapnik – Chervonenkis (también llamada Teoría VC por las iniciales). Tanto el estadístico Vladimir Vapnik y el matemático Alexey Chervonenkis la desarrollaron antes de publicar el

Tesis publicada con autorización del autor
No olvide citar esta tesis

libro a partir de los años 60, con el objetivo de saber cómo un algoritmo es capaz de acertar el futuro con datos del pasado.

Esta teoría es el corazón de su libro, y hablan de que para que un modelo de predicción sea efectivo, no solo tiene que memorizar los datos de entrenamiento, sino que también debe tener la capacidad de extrapolar ese conocimiento a nuevos conjuntos de datos. Lo que justifica porque se debe separar lo que son los datos de entrenamiento y los datos de prueba.

Su aporte matemático más importante es establecer con una probabilidad de al menos $1-n$, el error esperado en el futuro (R) este acotado superiormente por el riesgo empírico (R_{emp}), por lo que dependiendo de la complejidad del modelo se recibirá términos de penalización. En la fórmula aparte se tiene la l como tamaño de la muestra y el h como la dimensión VC que representa la complejidad matemática del algoritmo escogido.

$$R(w) \leq R_{emp}(w) + \sqrt{\frac{h\left(\ln\left(\frac{2l}{h}\right)+1\right) - \ln\left(\frac{\eta}{4}\right)}{l}}$$

Lo presente establece los principios matemáticos que permiten a los algoritmos el acomodar y desarrollar patrones a partir de conjuntos de datos empíricos sin la necesidad de una programación explícita.

Tesis publicada con autorización del autor
No olvide citar esta tesis

2.2.1.1. *Dimensión 1: RandomForest Regressor*

El bosque aleatorio se trata de un algoritmo de machine learning que trabaja bajo el modelo de regresión que busca ajustar varios árboles de decisión. Su historia va desde Robert Schapire (1990) quien introdujo formalmente el concepto de aprendizaje en conjunto. Pero quien desarrolló la técnica y el algoritmo de Bosques aleatorios fue Leo Breiman (2001), el postula que la combinación de múltiples árboles de decisión individuales genera un predictor fuerte con una varianza significativamente menor.

2.2.1.1.1. Análisis multicriterio

Keeney y Raiffa(1993) mencionan que para evaluar un problema se debe tener conciencia si los criterios de un problema son independientes o no. El método de análisis multicriterio SCORING no es más que el calcular mediante una suma ponderada de sus calificaciones individuales. Significa que se multiplica la puntuación de cada criterio por el peso de importancia, para que finalmente se sume en una calificación única que nos permite organizar las distintas alternativas.

"If the attributes are additive independent, then the multiattribute value function is given by"

$$v(x) = \sum_{i=1}^{n} w_i \, v_i(x_i)$$

Tesis publicada con autorización del autor
No olvide citar esta tesis

2.2.1.1.2. Búsqueda en cuadricula

Para evaluar la robustez del método de puntuación (SCORING) se utiliza el Análisis del Espacio de Pesos. Según los lineamientos de Erkut y Tarimcilar (1991) en lugar de asumir los pesos como valores estáticos, se llevó a cabo una exploración sistemática en red sobre el espacio de ponderaciones permisibles.

Sí trasladamos este concepto a modelos con regiones de estabilidad (modelos machine learning), de esta manera aseguramos que la alternativa seleccionada como óptima sea vigente ante variaciones controladas.

Para formalizar esta teoría a la optimización de hiperparámetros, se tienen a los autores Hutter et .al (2011) que hablan de la optimización de HPO como un problema de caja negra, principalmente lo definieron como un problema de optimización combinatoria al tener múltiples entradas.

- **Indicador 1: Raíz del error cuadrático medio (RMSE)**

  En una medida que se usa para calcular diferencias entre valores predichos y los valores observados, estos son dados por un estimador.

- **Indicador 2: Error Absoluto Medio (MAE)**

  Toma el valor absoluto de los posibles errores, sin importar el tamaño todos pesan por igual.

Tesis publicada con autorización del autor
No olvide citar esta tesis

- **Indicador 3: Coeficiente de determinación (R^2)**

  Se trata de otra métrica centrada en evaluar los modelos de regresión, especialmente entre las variables dependiente como la demanda independiente como lo serían el precio, las promociones, etc.

2.2.1.2. Dimensión 2: XGBoost Regressor

Se trata de otros algoritmos ML que hacen implementación de la potenciación de gradiente, Jerome Friedman (2001) terminó de desarrollar su Teoría de Gradient Boosting, y a posterior Tianqi Chen y Carlos Guestrin (2016) desarrollarían los algoritmos de impulso de Gradiente extremo (XGBoost), funciona como un sistema escalable de árboles de decisión propulsados por gradiente. No solo utilizan el modelo original de Friedman, sino que existe una optimización formal de segundo orden. La función objetivo del algoritmo, retrata la suma de una función perdida que mide la diferencia entre el valor real y el predicho, y un término de regularización que penaliza la complejidad del modelo.

$$\mathcal{L}^{(t)} = \sum_{i=1}^{n} l\left(y_i, \hat{y}_i^{(t)}\right) + \sum_{k=1}^{t} \Omega\left(f_k\right)$$

Tesis publicada con autorización del autor
No olvide citar esta tesis

2.2.1.3. Dimensión 3: Red Neuronal para regresión

Es un modelo computacional que trata de simular el comportamiento de un cerebro humano mediante nodos llamadas neuronas que desplazan entradas y salidas. Las redes neuronales se sustentan sobre el Teorema de Aproximación Universal desarrollado por Cybenko (1989) y extendido por Hornik (1991). El Teorema habla de que una red neuronal con una cantidad finita de neuronas es capaz de aproximar cualquier función continua $f(x)$ con un grado de precisión arbitrario. La salida de una red neural $F(x)$ se define matemáticamente de la siguiente manera:

$$F(x) = \sum_{i=1}^{N} v_i \, \sigma(w_i \cdot x + b_i)$$

En esta fórmula se observan que x es el vector de la entrada, N es el número total de neuronas, w_i es el vector de pesos, b_i es el sesgo de la i–ésima neurona oculta, v_i es el peso que conecta la i-ésima neurona oculta con la neurona de salida, y finalmente σ representa la función de activación.

2.2.2. Segunda variable: Predicción de demanda

La teoría que fundamenta la predicción de la demanda en esta investigación es la Teoría del Suavizado Exponencial Triple de Holt-Winters, desarrollada por Holt (1957) y Winters (1960), la cual postula que toda serie de demanda puede descomponerse en tres componentes latentes: nivel basal (L_t), tendencia (T_t) y estacionalidad (S_t), los cuales se actualizan recursivamente mediante ponderaciones exponenciales que asignan mayor peso a los datos

Tesis publicada con autorización del autor
No olvide citar esta tesis

recientes, bajo el supuesto de que el futuro próximo es función del nivel y la tendencia actuales ajustados por el factor estacional.

El modelo se expresa formalmente mediante las ecuaciones: $L_t = \alpha \cdot (Y_t/S_{t-m}) + (1-\alpha) \cdot (L_{t-1} + T_{t-1})$ para el nivel; $T_t = \beta \cdot (L_t - L_{t-1}) + (1-\beta) \cdot T_{t-1}$ para la tendencia; $S_t = \gamma \cdot (Y_t/L_t) + (1-\gamma) \cdot S_{t-m}$ para la estacionalidad; y el pronóstico a k periodos se obtiene mediante $F_{t+k} = (L_t + k \cdot T_t) \cdot S_{t+k-m}$, donde Y_t es la demanda observada, m es el número de periodos del ciclo estacional, y α, β, γ son constantes de suavizado entre 0 y 1 optimizadas para minimizar el error cuadrático medio.

Esta teoría es ampliamente utilizada en la literatura académica y en la práctica empresarial por su capacidad para modelar series con tendencia y estacionalidad, reconociendo además que la predicción exacta es imposible, por lo que el pronóstico se interpreta como una aproximación probabilística que reduce la incertidumbre en la planificación de inventarios y la toma de decisiones operativas.

2.2.2.1 Dimensión 1: Eficiencia del Abastecimiento

La teoría del Lote Económico de Compras (Modelo EOQ) Ford Whitman Harris (1913), establece que la eficiencia en el abastecimiento se logra al equilibrar dos costos contrapuestos: el costo de realizar los pedidos y el costo de almacenar el inventario. Si se compra en cantidades muy pequeñas, los costos de transporte y gestión se disparan por la alta frecuencia de pedidos; pero si se compra en exceso, el dinero se estanca y aumentan los costos de bodega. Por lo tanto, la eficiencia perfecta consiste en hallar el

Tesis publicada con autorización del autor
No olvide citar esta tesis

punto exacto de equilibrio que minimice ambos gastos simultáneamente.

Para calcular matemáticamente este abastecimiento óptimo, se utiliza la fórmula $Q^* = \sqrt{\frac{2DS}{H}}$, donde Q^* es la cantidad exacta de unidades que se deben solicitar en cada pedido. En esta ecuación, D representa la demanda anual del producto, S es el costo fijo por procesar cada orden de compra y H es el costo de mantener una sola unidad guardada en el inventario durante un año; permitiendo así predecir el tamaño de lote ideal para no generar desabastecimiento ni sobrecostos.

- **Indicador 1: Porcentaje de quiebres de stock.**

  Este indicador mide la frecuencia con la que un producto se agota. Una previsión de la demanda más precisa debería reducir este porcentaje, ya que permite planificar mejor la reposición.

  Fórmula del cálculo

$$\%Quiebre\ de\ Stock = \frac{Número\ de\ productos\ en\ quiebre}{Total\ de\ productos\ demandados} \times 100$$

2.2.2.2. Dimensión 2: Reducción de excedentes de inventario

La teoría del sistema Just-in-Time (Justo a Tiempo) de Taiichi Ohno (1938), postula que cualquier acumulación o excedente de inventario es una ineficiencia crítica que oculta fallas operativas, inmoviliza capital y genera costos innecesarios de almacenamiento. Bajo este enfoque, las empresas deben transitar

Tesis publicada con autorización del autor
No olvide citar esta tesis

de un sistema basado en producir por si acaso, a uno basado estrictamente en el tirón (*pull*) de la demanda del cliente, donde los materiales se reciben y procesan únicamente en la cantidad exacta y en el momento justo en que se necesitan, eliminando de raíz el desperdicio del sobrestock.

Para gestionar y predecir matemáticamente esta reducción de excedentes, se utiliza la Ley de Little expresada como $WIP = \lambda \times CT$, donde WIP representa el inventario acumulado o estancado en el sistema. En esta ecuación, λ es la tasa de salida o demanda real del cliente y CT es el tiempo de ciclo que toma procesar y entregar el producto; demostrando de manera matemática que la única forma sostenible de disminuir el exceso de inventario (WIP) sin afectar el nivel de servicio es reduciendo drásticamente los tiempos de proceso y entrega (CT).

- **Indicadores: Porcentaje de Inventario Obsoleto**

  Este indicador de Porcentaje de Inventario obsoleto o estancado mide la proporción de mercancía en stock que no ha tenido movimiento (ventas) en un período determinado, reflejando ineficiencias en la gestión de inventario y posibles pérdidas para la ferretería.

  Fórmula del cálculo:

$$\%Inventario\ Obsoleto\ = \frac{Inventario\ no\ rotado\ en\ X\ meses}{Inventario\ total} \times 100$$

Tesis publicada con autorización del autor
No olvide citar esta tesis

2.3. Objetivos de la investigación

2.3.1. Objetivo general

Implementar machine learning para la predicción de demanda de productos, aplicado a la ferretería "Vargas", año 2025

2.3.2. Objetivos específicos

Objetivo Específico 1: Determinar como la implementación de machine learning mejora la capacidad de eficiencia del abastecimiento de productos, aplicado a la ferretería "Vargas", año 2025

Objetivo Específico 2: Determinar como la implementación de machine learning mejorará la reducción de excedentes de inventario de los productos, aplicado a la ferretería "Vargas", año 2025

2.4. Formulación de hipótesis

2.4.1. Hipótesis general

Al implementar machine learning mejorará significativamente la predicción de demanda de productos en la ferretería "Vargas", año 2025

2.4.2. Hipótesis específicas

Hipótesis Específica 1: Al implementar machine learning mejorará el cálculo de la eficiencia del abastecimiento de los productos de la ferretería "Vargas", año 2025

Hipótesis Específica 2: Al implementar machine learning mejorará el cálculo de los excedentes de inventario de los productos de la ferretería "Vargas", año 2025

Tesis publicada con autorización del autor
No olvide citar esta tesis

CAPÍTULO III: DISEÑO METODOLÓGICO

3.1.Diseño de la investigación

Según los autores González y Rivera (2024), en su libro "Investigación cuantitativa. Claves para estudiantes universitarios", escribe que los diseños de investigación se pueden clasificar dependiendo de ciertas características, la que nos compete es según la acción de los investigadores que se dividen en el tener solo que observar o el intervenir al momento de evaluar los resultados. Siguiendo estos pasos podremos decir que nuestra investigación es experimental pues se evaluarán resultados antes y después de la implementación de un modelo de machine learning.

3.2. Tipo de Investigación

Según González y Rivera (2024), la investigación cuantitativa se define por la recopilación y el análisis de datos numéricos para responder a las preguntas de estudio. Bajo esta premisa la presente investigación se clasifica dentro del tipo cuantitativo. Este enfoque permite medir variables y evaluar cambios a través de indicadores estadísticos.

Siguiendo esta misma línea la investigación es de tipo experimental. Se evaluarán los resultados antes y después de la implementación del modelo de machine learning. Este diseño permite comparar el estado inicial de los procesos de la ferretería con el estado posterior a la intervención tecnológica. De este modo se puede determinar si el modelo genera una mejora efectiva en los beneficios operativos o económicos del negocio.

Tesis publicada con autorización del autor
No olvide citar esta tesis

3.3. Nivel de investigación

El nivel de investigación se define como aplicativo. Según Espinoza y Ochoa (2021), el nivel aplicativo tiene como propósito mejorar la sociedad o evaluar la producción de algún producto o proceso. En el presente estudio se ha detectado una oportunidad de mejora dentro de una ferretería. Esta oportunidad puede estar relacionada con la predicción de demanda, la optimización de inventarios o la atención al cliente. A partir de esta detección se implementa un modelo de machine learning como herramienta de solución. Luego mediante una comparación de resultados antes y después de dicha implementación se evalúan los beneficios obtenidos como la reducción de tiempos, el aumento de precisión o el ahorro de costos.

Esta dinámica es precisamente la que caracteriza a una investigación de nivel aplicativo. Se genera conocimiento que se traduce en una mejora tangible dentro de una organización o proceso específico. Por lo tanto, el nivel aplicativo justifica plenamente el enfoque experimental del estudio. La medición pre y post intervención constituye el mecanismo central para validar si el modelo de machine learning produce o no los beneficios esperados en la ferretería.

3.4. Enfoque

Según los autores González y Rivera (2024), en su libro "Investigación cuantitativa. Claves para estudiantes universitarios", define que el enfoque de investigación cuantitativa está basado en el paradigma postpositivista, este enfoque permite estudiar fenómenos que pueden ser observados, medidos y cuantificados. En tal sentido, en este trabajo se adoptó un enfoque cuantitativo para la implementación de machine learning

Tesis publicada con autorización del autor
No olvide citar esta tesis

3.5. Población

Según los autores González y Rivera (2024), en su libro "Investigación cuantitativa. Claves para estudiantes universitarios", define la población como elementos de un conjunto que comparte características similares, En nuestra investigación, participa la población conformada por todos los productos registrados en las ventas de la Ferretería "Vargas" durante los últimos cinco años (2020–2024).

3.6. Muestra

Según los autores González y Rivera (2024), en su libro "Investigación cuantitativa. Claves para estudiantes universitarios" define que la muestra es un subconjunto dentro de la población que tiene un valor más representativo, que es seleccionado para ser estudiado en características más significativas. Por consiguiente, la muestra comprende 109 productos de construcción. Los datos de ventas se recopilaron durante un período de cinco años (2020-2024) y representan aproximadamente 5700 registros de ventas. El criterio de inclusión fue que los productos de construcción debían haber estado registrados en las tiendas “Vargas" durante los años especificados. Se excluyeron los productos vendidos en menos de 30 unidades durante el período de análisis para evitar distorsiones causadas por picos de ventas significativos de ciertos artículos, como los ladrillos, que podrían distorsionar la demanda general.

Tesis publicada con autorización del autor
No olvide citar esta tesis

3.7. Muestreo

Según los autores González y Rivera (2024), en su libro "Investigación cuantitativa. Claves para estudiantes universitarios" define que el muestreo tiene diversas formas de ser estudiado según las características del proyecto, una de estas es donde el investigador por conveniencias selecciona a los participantes según su facilidad de acceso y disponibilidad. Por consiguiente, esta investigación tiene un muestreo por conveniencia, debido a que se selecciona productos de una categoría de construcción en los años (2020–2024), los cuales están fácilmente disponibles en los registros proporcionados por la ferretería "Vargas"

3.8. Operacionalización de variables

Tabla 1: *Resumen de operacionalización de Variables*

VARIABLES	DIMENSIÓN	INDICADOR	FÓRMULAS
Machine Learning	RandomForest Regressor	Raíz del error cuadrático medio (RMSE)	$RMSE = \sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{n}}$
	XGBoost Regressor	Error Absoluto Medio (MAE)	$MAE = \frac{\sum \lvert y_i - \hat{y}_i \rvert}{n}$
	Red Neuronal para regresión Multicapa	Coeficiente de determinación (R^2)	$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \overline{y_i})^2}$
Predicción de demanda	Eficiencia del Abastecimiento	Porcentaje de quiebres de stock	$= \frac{Número\ de\ productos\ en\ quiebre}{Total\ de\ productos\ demandados} \times 100$
	Reducción de excedentes de inventario	Porcentaje de Inventario Obsoleto o Estancado	$= \frac{Inventario\ no\ rotado\ en\ X\ meses}{Inventario\ total} \times 100$

Fuente: *Elaboración Propia*

Tesis publicada con autorización del autor
No olvide citar esta tesis

3.9. Técnicas para la recolección de datos

Según los autores González y Rivera (2024), en su libro "Investigación cuantitativa. Claves para estudiantes universitarios", define que las técnicas e instrumentos permiten reunir una cantidad de datos válidas para el tipo de estudio, pues de esta dependerá las respuestas del proyecto. La técnica responde la manera y la herramienta el formato de los ítems que estén en concordancia con nuestras variables.

3.9.1. Técnica

Según los autores González y Rivera (2024) en su libro "Investigación cuantitativa. Claves para estudiantes universitarios" explica que la técnica de Evaluación o análisis incluye lo que es el procesamiento de información donde se construyen grandes dominios, continuando se les da un cálculo donde se analizan sus medidas y se plantea la búsqueda de patrones de comportamiento de dicha información.

3.9.2. Instrumentos (Validez y Confiabilidad)

Las fichas de registro visibles en el ***Anexo 3***, propone para observar la reducción de excedentes va más allá de ser una simple tabla repetida, ya que constituye una hoja resumen más completa que integra tanto datos cuantitativos como información contextual relevante para el posterior procesamiento estadístico o mediante machine learning. En esta versión mejorada, el investigador no solo registra por cada producto el stock inicial, la demanda real, el excedente calculado y su porcentaje correspondiente, sino que además añade columnas como el stock sugerido por el modelo predictivo y una sección de metadatos donde se consigna si hubo

Tesis publicada con autorización del autor
No olvide citar esta tesis

promociones, eventos inusuales (como feriados o lluvias), días de mayor excedente o la tasa de rotación del producto. De este modo, la ficha de registro deja de ser un mero instrumento de observación puntual y se convierte en una herramienta de acopio de información estructurada que permite, al final del período de estudio, comparar el comportamiento real del inventario con las predicciones del algoritmo, identificar patrones de sobreabastecimiento y evaluar si la implementación de machine learning logra efectivamente reducir los excedentes.

Según los autores González y Rivera (2024) en su libro "Investigación cuantitativa. Claves para estudiantes universitarios" de diversas herramientas para el procesamiento de la información, pero ante todas estas las que más destaca para nuestro proyecto es la construcción de bases de datos, que el libro define como una compilación de datos organizados listos para su análisis.

- Validez: Nuestra investigación se refleja en el uso de un modelo de algoritmo machine learning, por lo que para garantizar el éxito se usará una base de datos como el formato para registrar los datos históricos de ventas provenientes de los sistemas de gestión de la ferretería.
- Confiabilidad: Esta se debe asegurar mediante la replicabilidad del proceso de análisis, debido a que trabajaremos con los modelos de machine learning se puede decir que pueden ser entrenados y evaluados múltiples veces, en las que se darán resultados consistentes en una igualdad de condiciones.

Tesis publicada con autorización del autor
No olvide citar esta tesis

3.10. Técnica para el procesamiento y análisis de datos

Según González y Rivera (2024), el análisis de datos de las fichas de observación permite el registro estructurado de una manera sistemática de los comportamientos, eventos o situaciones corroboradas en un contexto determinado, lo que facilita la obtención de resultados a preguntas de investigaciones cuantitativas. Se está considerando a las siguientes técnicas para el sistema propuesto:

a. **Monitoreo y reentrenamiento continuo:** Propone un enfoque basado en el flujo MLOps (Operaciones de Machine Learning). Se definen indicadores de rendimiento y evalúan que las métricas no caigan por debajo de un umbral aceptable.

   El reentrenamiento será programado, el sistema se entrenará automáticamente cada mes incorporando las nuevas ventas registradas en la base de datos PostgreSQL. Con esto aseguramos que el modelo se adapte a las nuevas tendencias, y se mantendrá un entrenamiento de los últimos 6 años, y de esta manera el sistema no queda obsoleto

b. **Automatización del flujo:** Se tendrá un pipeline de base de datos para la arquitectura del sistema. El usar PostgreSQL nos permitirá el eliminar la manipulación manual para el cambio de valores. Además de asignar un control de acceso de roles para tener funciones restringidas a personal de la ferretería, por decir el personal de venta solo podrá registrar una venta con cada transacción de venta.

c. **Cold Start:** Se utilizo la estrategia de Cold Start para tener una correcta gestión de los productos entrantes. Al registrar un nuevo producto que no

Tesis publicada con autorización del autor
No olvide citar esta tesis

tiene ventas históricas, se clasificara por similitud de ventas y categoría de otros productos, y así no tener un arranque frío.

3.11. Aspectos éticos

Según los autores González y Rivera (2024), los aspectos éticos en las investigaciones cuantitativas son fundamentales para garantizar la integridad de los datos verificables, ya que buscan proteger la dignidad, los derechos, el bienestar de los participantes durante la investigación. Para esta investigación se trabajará exclusivamente con los registros de ventas de la ferretería de los últimos 5 años (2020-2024).

Esta investigación no requiere la aprobación del comité de ética, ya que no implica la participación directa de personas ni la recopilación de datos personales sensibles. El estudio se centra exclusivamente en el análisis de datos históricos de ventas de la ferretería "Vargas", los datos utilizados son de carácter comercial y se tratarán de con los principios de la gestión de la información.

- Consentimiento Informado: La ferretería proporcionará consentimiento de usar su información personal y su registro de ventas entre otros datos los cuales se usarán para esta investigación.
- Confidencialidad y Privacidad: En este estudio, se garantiza la confidencialidad y el uso ético de la información proporcionada por la Ferretería Vargas, incluyendo registros de ventas y otros datos sensibles. Se seguirán estrictamente los principios éticos para proteger la privacidad de la información obtenida, asegurando que sea utilizada exclusivamente con propósitos académicos y de investigación.
- Responsabilidad Social y Ambiental: El estudio se lleva a cabo conforme a los principios de responsabilidad social para promover el uso ético de la

Tesis publicada con autorización del autor
No olvide citar esta tesis

tecnología y evitar impactos negativos en el entorno laboral y el medio ambiente. Los resultados del estudio buscan generar beneficios sostenibles para la industria de la construcción y su comunidad.

Tesis publicada con autorización del autor
No olvide citar esta tesis

CAPÍTULO IV: RESULTADOS

En esta sección se explica el funcionamiento del programa, los procedimientos que le acompañaron y los resultados obtenidos. Para evaluar el desempeño de los modelos, primero se tienen que identificar las variables que serán entrenadas. Esto podemos dividirlo mediante la ingeniería de características. Las columnas de nuestros archivos que entrarán al entrenamiento son las siguientes:

- **Fecha:** Mediante la ingeniería de características, se dividió las columnas el */Día/ /Mes/ /Año/ /Quincena/* y se agregó la variable de /Días festivos/ como variable externa.
- **Identificador de producto:** Tanto el */ID/* como el */Nombre/* están como columnas para señalar directamente a qué producto estamos prediciendo la demanda.
- **Categoría**: La ferretería cuenta con una clasificación de ciertos materiales; no son muchos, pero muestran tendencias en ciertos materiales.
- **Cantidad:** Existe una columna de */Cantidad/* que indica exactamente la demanda de tal producto en un rango de fecha preestablecido; es la variable que los algoritmos buscan predecir.

Aparte de explicar el funcionamiento y la salida del sistema, este capítulo tiene un apartado para explicar los costos y beneficios del proyecto, evaluando la rentabilidad económica considerando los cambios del mercado.

Tesis publicada con autorización del autor
No olvide citar esta tesis

4.1.Ajuste: Parámetros e hiperparámetros

En el campo del desarrollo de algoritmos de Machine Learning se tiende a trabajar con variables ajustables, estas pueden tener características de tener un origen interno o externo. Los parámetros se aprenden y actualizan a partir del entrenamiento de los datos, aparte de que no se puede configurar o tratar de modificar de manera directa. Por otro lado, los hiperparámetros se definen de manera manual y no se obtienen de manera automática, se tiene que introducir por algún conocedor de datos, pueden estar en constante configuración con tal de mejorar el rendimiento del modelo.

Existen varias técnicas de optimización que buscaN la mejor combinación de hiperparámentros. Para calcular los hiperparámetros de nuestros 3 modelos se decidió usar el Grid Search (búsqueda en cuadrícula), es un método que se define por ser exhaustivo al probar distintas combinaciones.

Su proceso va de la siguiente manera, primero se define la cuadrícula conde contiene los nombres de los hiperparámetros y valores a probar dentro de ellos; luego se hace una evaluación de cada intersección y para esto se entrena los modelos configurados; por último, se compara los resultados y se valida a un ganador mediante sus puntuaciones finales.

Se, decidió este método por dos razones muy esenciales en este trabajo de investigación:

- No se deja opciones sin probar, a cada parámetro le da un valor y busca el mejor resultado iterando con los demás resultados.
- En el ecosistema Python se tiene a la biblioteca scikit-learn, donde implementa esta técnica a través de la clase GridSearchCV

Aparte de tener estas ventajas el lenguaje Python cuenta con la capacidad de usar librerías para llamar a indicadores y evaluar resultados del entrenamiento (MAE, RMSE, R^2) .

Tesis publicada con autorización del autor
No olvide citar esta tesis

Para cumplir con la exigencia de detallar las "categorías y valores probados", se realizó una Tabla de Espacio de Búsqueda por cada modelo complejo para dar un orden visual a la investigación.

a) Hiperparámetros del XGBoost

En la *Figura 13* en anexos se puede ver la configuración del modelo XGBoost, y tiene el siguiente desglose de parámetros:

- n_estimators: Muestra el número de árboles totales a construir, secuenciales;
- max_depth: Es la profundidad máxima de los árboles, que tanto puede crecer individualmente, sirve para ajustar la complejidad del modelo y el riesgo de ruido.
- learning_rate: Controla le velocidad con la que el modelo aprende de los errores de los árboles anteriores a este.
- subsample: Sirve para determinar el porcentaje de filas de datos que se seccionan aleatoriamente para entrenar cada árbol.

Tabla 2: *Selección de hiperparámetros para el modelo XGBoost*

Hiperparámetro	Valores Evaluados (Grid)	Valores Óptimos
n_estimators	[100, 200, 300]	200
max_depth	[4, 6, 8]	6
learning_rate	[0.01, 0.05, 0.1]	0.05
subsample	[0.7, 0.8, 0.9]	0.8

Fuente: *Elaboración Propia*

Tesis publicada con autorización del autor
No olvide citar esta tesis

b) Hiperparámetros del RandomForest

En el caso de RandomForest, se trata de un modelo que tiene al Bagging como principal diferenciador. Mantiene muchos atributos similares al anterior, por lo cual se le añadió un agregado:

- min_samples_split: Es el número mínimo de muestras requeridas para abrir un nodo interno.

***Tabla 3**: Selección de hiperparámetros para el modelo RandomForest*

Hiperparámetro	Valores Evaluados (Grid)	Valores Óptimos
n_estimators	[50, 100, 150]	100
max_depth	[4, 6, 8, 10]	8
min_samples_split	[2, 5, 10]	5

***Fuente:** Elaboración Propia*

c) Hiperparámetros de Redes Neuronales Multicapa

Para el modelo de Red neuronal, presente en los anexos como *Figura 14*, y contiene ajustes bastante diferenciables y contienen nuevos tipos de valores.

- hidden_layer_sizes (Estructura de capas ocultas): Define cuántas capas ocultas tiene la red y cuántas neuronas hay en cada una.
- solver (Optimizador): El algoritmo encargado de actualizar los pesos de las neuronas para minimizar el error durante el entrenamiento (sgd para Descenso de Gradiente Estocástico o adam).
- alpha (Penalización de Regularización L_2): Evita el sobreajuste aplicando una penalización a los pesos excesivamente grandes, obligando a la red a mantener un comportamiento más suave y generalizable.

Tesis publicada con autorización del autor
No olvide citar esta tesis

- max_iter (Máximo de iteraciones / Épocas): El límite máximo de ciclos de entrenamiento (épocas) que la red ejecutará para converger antes de detenerse.

Tabla 4: *Selección de hiperparámetros para el modelo MLP Redes Neuronales*

Hiperparámetro	Valores Evaluados (Grid)	Valores Óptimos
hidden_layer_sizes	[(32, 16), (64, 32), (100,)]	(64, 32)
solver	['sgd', 'adam']	'sgd'
alpha	[0.0001, 0.001, 0.01]	0.001
max_iter	[200, 500]	500

Fuente: *Elaboración Propia*

4.2. Indicadores de precisión

Para evaluar la precisión de los distintos modelos se tiene que conceptualizar e indicadr cada uno.

- **El Error Absoluto Medio (MAE):** Te indica que tan alejado esta la predicción del valor real, esto medido por las mismas unidades de las variables. Por decir, el mejor valor es el XGBoost ya que su MAE significa que tal modelo tiende a fallar por 1.25 unidades.
- **La Raiz del Error Cuadrático Medio (RMSE):** Este también mide la magnitud del error, pero eleva las diferencias al cuadrado antes de promediarlas. Tiene que ser cercano al valor del MAE.
- **Coeficiente de Determinación (R2):** Mide la proporción de la variabilidad entre un valor de 0 a 1(0% a 100%). Mientras el valor es más cercano mejor, pues significa que el modelo puede explicar en su mayoría la variabilidad del comportamiento histórico.

Tesis publicada con autorización del autor
No olvide citar esta tesis

Tabla 5: *Evaluación de Precisión de Modelos*

MODELOS	MAE	RMSE	R2
XGBoost	1.25	1.45	0.95
Red Neuronal Multicapa	1.38	1.66	0.63
RandomForest	1.27	1.49	0.90
Regresión Lineal	1.84	1.67	0.73

Fuente: *Elaboración Propia*

Con esta explicación se puede decir que el mejor modelo es el XGBoost, la justificación técnica radica que sus valores son los óptimos. Tiene el menor error promedio a diferencia de los demás, también tiene una mayor consistencia ante errores grandes y posee una excelente capacidad d eajuste y generalización al explicar el 95% de varianza de los datos.

4.3. Análisis multicriterio SCORING

Al sacar los mejores valores para los hiperparámetros de cada modelo, se tiene que evaluar cual de estos modelos es el óptimo para el sistema que deseamos desarrollar. Para esto usamos el Método SCORING, un análisis multicriterio basado en puntajes, su procedimiento tiene 7 etapas puntuales para identificar las distintas alternativas y están visibles en los *Anexos*.

Este apartado será para explicar de manera resumida sus procedimientos más importantes, se inicia definiendo la meta.

"Seleccionar el modelo de Machine Learning para su sistema de predicción de demanda de productos de la ferretería Vargas."

Luego de evaluar los objetivos se definen los criterios y se les coloca un ponderado:

Tesis publicada con autorización del autor
No olvide citar esta tesis

- **Precisión:** El criterio más importante pues busca minimizar el error de predicción. Se usaron indicadores como el MAE/RMSE/R2
- **Eficiencia Técnica:** Un modelo más rápido facilita las pruebas y futuras actualizaciones. Sus indicadores para tomar son el Tiempo/RAM
- **Vía Despliegue:** Tiene que ver con la compatibilidad del algoritmo con la estructura del proyecto, al usar un modelo MVC y api's para Postgres o Google Drive
- **Facilidad de ajuste:** La optimización es importante, pues un modelo en constante mantenimiento o ajuste de los hiperparámetros ocasionan cuellos de botella.

Tabla 6: *Criterios Ponderados*

Criterio	Peso
Precisión	40%
Eficiencia Técnica	20%
Vía de despliegue	30%
Facilidad de Ajuste	10%

Fuente: *Elaboración Propia*

Luego se establece el nivel de satisfacción que alcanzó cada modelo, en su respectivo indicador. En este caso va de 1 a 5. Donde 1 = muy bajo; 5= muy alto.

Tabla 7: *Puntos de satisfacción de los modelos*

Modelo	Precisión	Eficiencia Técnica	Vía de despliegue	Facilidad de Ajuste
Regresión Lineal	2	5	5	5
Random Forest	4	2	4	4
XGBoost	5	4	4	3
MLP	2	3	3	2

Fuente: *Elaboración Propia*

Tesis publicada con autorización del autor
No olvide citar esta tesis

Tabla 8: *Resultados del Método SCORING*

Modelo	Precisión (40)	Eficiencia Técnica (20)	Vía de despliegue (30)	Facilidad de Ajuste (10)	Total
Regresión Lineal	80	100	150	50	380
Random Forest	160	40	120	40	360
XGBoost	200	80	120	30	430
MLP	80	60	90	20	250

Fuente: *Elaboración Propia*

De acuerdo con el método SCORING, XGBoost obtuve el mayor puntaje, y por ende es la alternativa más adecuada para el sistema. Cumpliendo con los principales rasgos de evaluación.

4.4. Resultados del Sistema

Para retratar la funcionalidad del sistema, se trabajaron 2 tipos de predicción, uno usando el modelo XGBoost y el método tradicional que la ferretería anteriormente usaba para realizar sus compras. Se colocaron 10 productos clave para revisar los cambios, estos 10 productos se escogieron no por ser los más caros o ser los más populares comercialmente, se escogieron porque representan las principales familias de productos dentro de los materiales de construcción.

Tesis publicada con autorización del autor
No olvide citar esta tesis

Tabla 9: *Predicción de la Primera Quincena de Agosto*

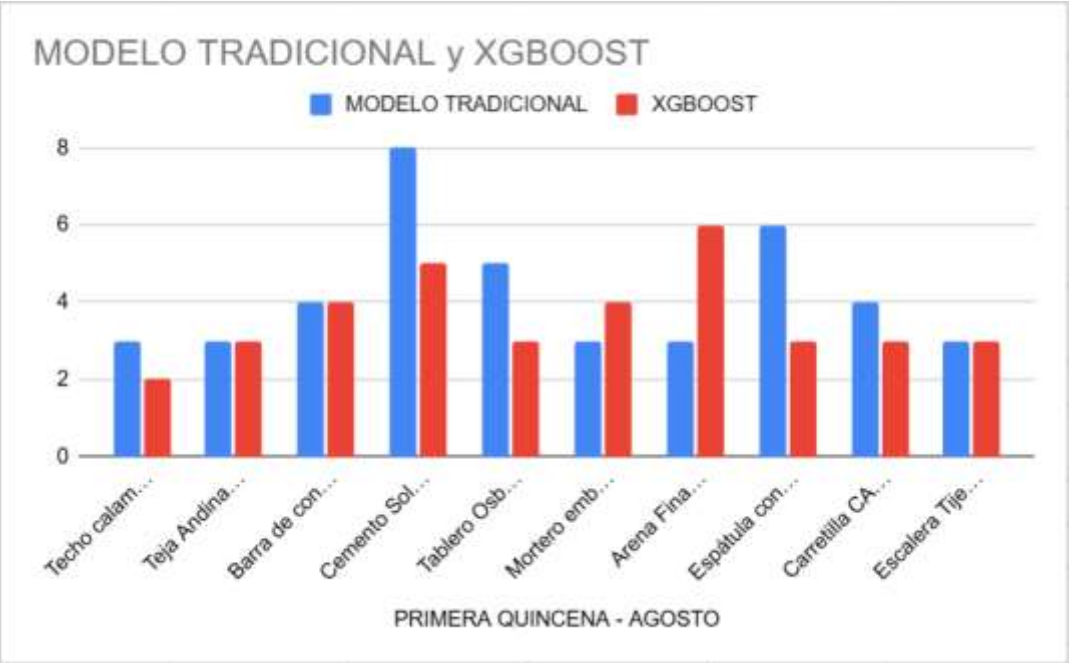
PRIMERA QUINCENA - AGOSTO	MODELO TRADICIONAL	XGBOOST
Techo calamina metálica 0.14mm 3.60x0.80m Werken	3	2
Teja Andina de Fibrocemento Eternit 1.14 x 0.72 metros	3	3
Barra de construcción Aceros Arequipa A615 Grado 60 ASTM 3/8\" x 900cm	4	4
Cemento Sol Portland Tipo I 42.5 kg	8	5
Tablero Osb fenólico 18mm 1.22x2.44 metros	5	3
Mortero embolsado fácil x 40 kg Unicon	3	4
Arena Fina Saco 40kg Luk	3	6
Espátula con mango de plástico 4" Werken	6	3
Carretilla CAT-50ND Truper	4	3
Escalera Tijera de madera 8 pasos Iforest	3	3

Fuente: *Elaboración Propia*

La tabla muestra que los productos clave cubren lo que son inventarios de techos, cubiertas, cementos, acero, tableros, mezclas, agregados de arena, herramientas pequeñas y equipo. Y pasarán a ser graficados con la siguiente figura:

Figura 1: *Comparación de Modelos Primera Quincena Agosto*

Tesis publicada con autorización del autor
No olvide citar esta tesis



Fuente: *Elaboración Propia*

Tabla 10: *Predicción de la Segunda Quincena de Agosto*

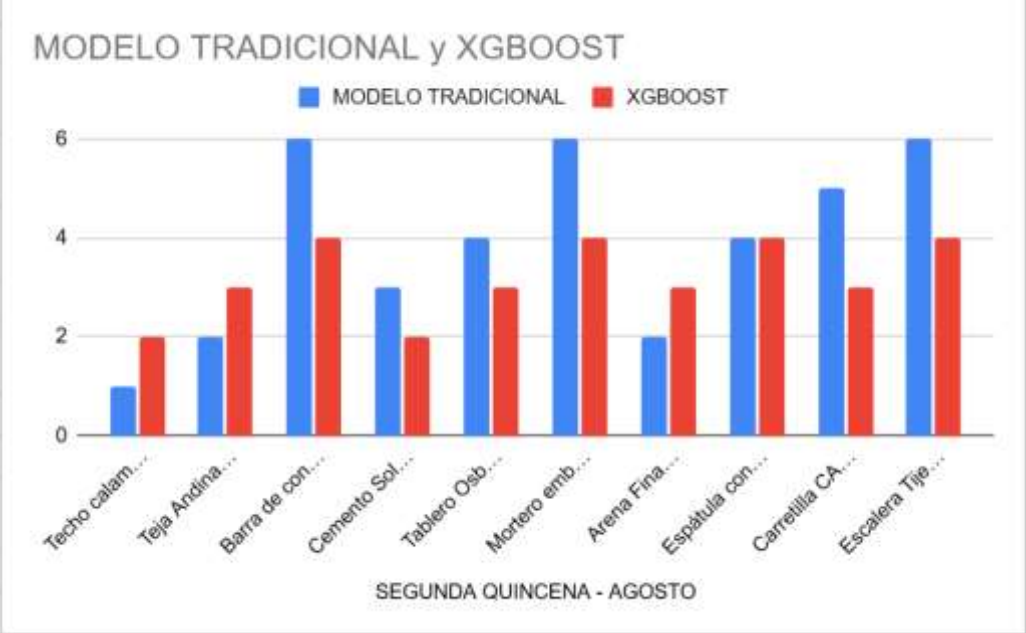
SEGUNDA QUINCENA – AGOSTO	MODELO TRADICIONAL	XGBOOST
Techo calamina metálica 0.14mm 3.60x0.80m Werken	1	2
Teja Andina de Fibrocemento Eternit 1.14 x 0.72 metros	2	3
Barra de construcción Aceros Arequipa A615 Grado 60 ASTM 3/8\" x 900cm	6	4
Cemento Sol Portland Tipo I 42.5 kg	3	2
Tablero Osb fenólico 18mm 1.22x2.44 metros	4	3
Mortero embolsado fácil x 40 kg Unicon	6	4
Arena Fina Saco 40kg Luk	2	3
Espátula con mango de plástico 4" Werken	4	4
Carretilla CAT-50ND Truper	5	3
Escalera Tijera de madera 8 pasos Iforest	6	4

Fuente: *Elaboración Propia*

Tesis publicada con autorización del autor
No olvide citar esta tesis

La tabla expone las columnas de la predicción de las cantidades exactas que la ferretería debe tener para cubrir con la demanda. Se dividieron en 4 tiempos distintos, siendo divididas en los meses de Agosto y Septiembre en el año 2025.

Figura 2: *Comparación de Modelos Segunda Quincena Agosto*



Fuente: *Elaboración Propia*

Tabla 11: *Predicción de la Primera Quincena de Septiembre*

PRIMERA QUINCENA - SEPTIEMBRE	MODELO TRADICIONAL	XGBOOST
Techo calamina metálica 0.14mm 3.60x0.80m Werken	4	3
Teja Andina de Fibrocemento Eternit 1.14 x 0.72 metros	4	3
Barra de construcción Aceros Arequipa A615 Grado 60 ASTM 3/8\" x 900cm	6	4
Cemento Sol Portland Tipo I 42.5 kg	3	2
Tablero Osb fenólico 18mm 1.22x2.44 metros	4	3
Mortero embolsado fácil x 40 kg Unicon	4	3
Arena Fina Saco 40kg Luk	2	3
Espátula con mango de plástico 4" Werken	6	4

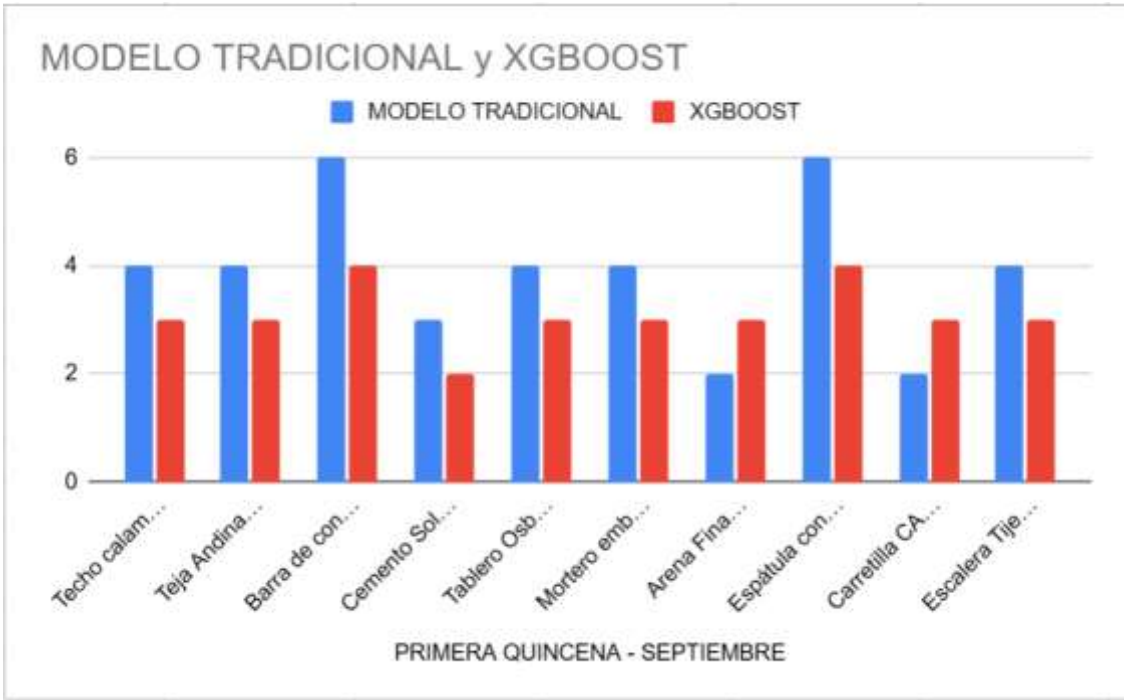
Tesis publicada con autorización del autor
No olvide citar esta tesis

Carretilla CAT-50ND Truper	2	3
Escalera Tijera de madera 8 pasos Iforest	4	3

Fuente: *Elaboración Propia*

Ya colocados los anteriores gráficos y tablas, se puede observar que la división del tiempo de los 2 meses se hizo de manera quincenal, teniendo 4 tiempos presentes.

Figura 3: *Comparación de Modelos Primera Quincena Septiembre*



Fuente: Elaboración Propia

Tabla 12: *Predicción de la Segunda Quincena de Septiembre*

SEGUNDA QUINCENA - SEPTIEMBRE	MODELO TRADICIONAL	XGBOOST
Techo calamina metálica 0.14mm 3.60x0.80m Werken	2	2
Teja Andina de Fibrocemento Eternit 1.14 x 0.72 metros	6	3
Barra de construcción Aceros Arequipa A615 Grado 60 ASTM 3/8\" x 900cm	4	4
Cemento Sol Portland Tipo I 42.5 kg	4	1
Tablero Osb fenólico 18mm 1.22x2.44 metros	2	3

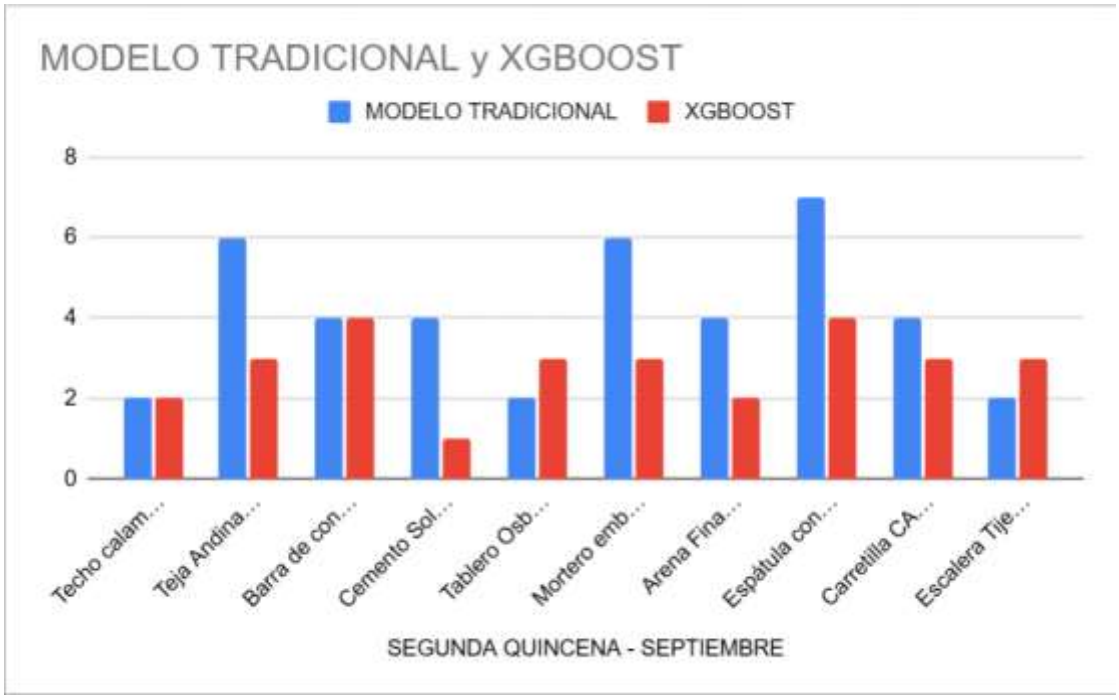
Tesis publicada con autorización del autor
No olvide citar esta tesis

Mortero embolsado fácil x 40 kg Unicon	6	3
Arena Fina Saco 40kg Luk	4	2
Espátula con mango de plástico 4" Werken	7	4
Carretilla CAT-50ND Truper	4	3
Escalera Tijera de madera 8 pasos Iforest	2	3

Fuente: *Elaboración Propia*

Se utilizo el grafico de barrar para ver la diferencia de cantidades entre la predicción del modelo XGBoost y tradicional, cosa muy importante para detallar los indicadores.

Figura 4: *Comparación de Modelos Segunda Quincena Septiembre*



Fuente: *Elaboración Propia*

El primer indicador que se trabajó fue de los quiebres de stock, esta muestra las veces que la tienda no pudo atender la petición de un cliente. Al ser un evento no deseado se desea que el porcentaje sea lo más bajo posible. Nuevamente se usaron los tiempos quincenales del mes de Agosto y Septiembre.

Tesis publicada con autorización del autor
No olvide citar esta tesis

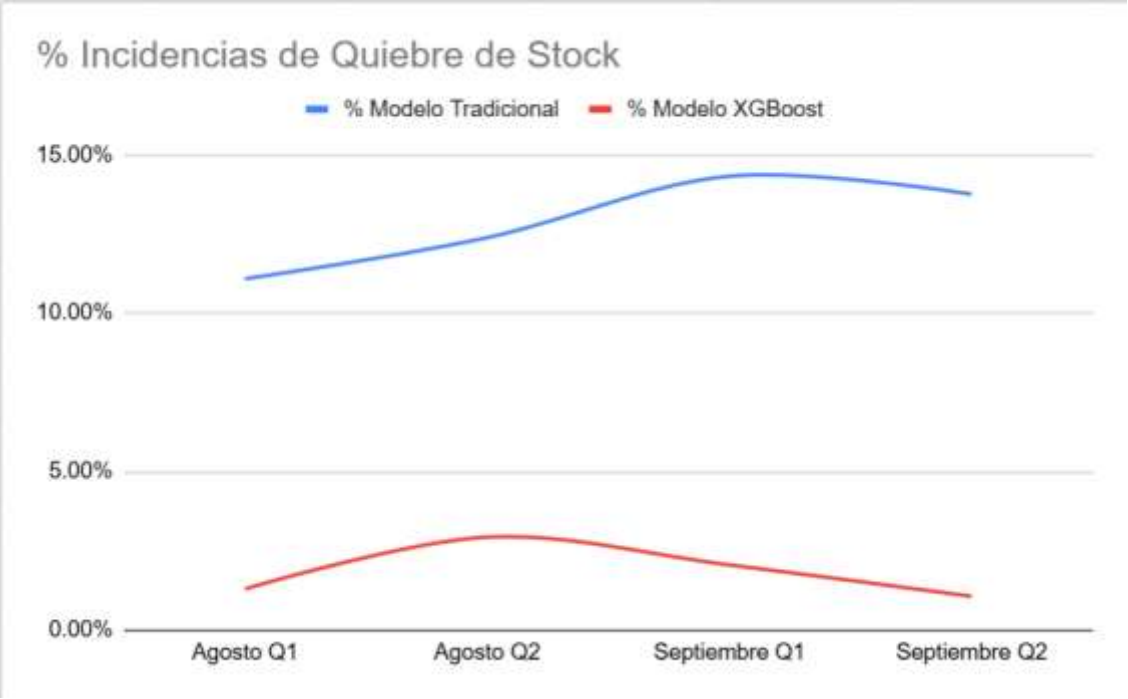
Tabla 13: *Evolución de Incidencias de Quiebres de stock*

Incidencia	Quiebre de STOCK	
MES	% Modelo Tradicional	% Modelo XGBoost
Agosto Q1	11.09%	1.30%
Agosto Q2	12.39%	2.94%
Septiembre Q1	14.33%	2.06%
Septiembre Q2	13.76%	1.07%

Fuente: *Elaboración Propia*

En el gráfico se visualiza cómo en promedio de incidencias de quiebre de Stock, el modelo XGBoost logra un 11% de diferencia. Con esto se podría decir que el modelo tradicional obtiene un nivel de incidencias de 12.89% mientras el modelo XGBoost mantiene un nivel de 1.84%, mostrando la superioridad de los algoritmos.

Figura 5: *Evolución de Incidencias de Quiebres de stock*



Fuente: *Elaboración Propia*

El segundo indicador sirve para medir el tema de los inventarios obsoletos, esta muestra las incidencias de productos que se abastecen de más, teniendo inventario que no rota o incluso mal posicionados. Esto más que nada da problemas logísticos relacionados con el uso del almacén, se elaboró la siguiente tabla.

Tesis publicada con autorización del autor
No olvide citar esta tesis

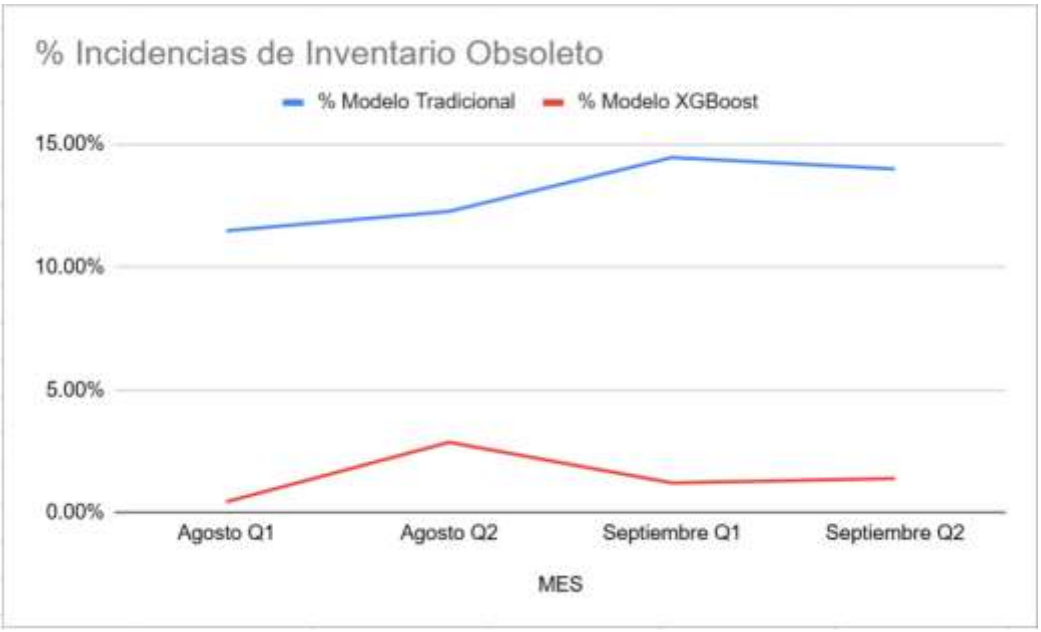
Tabla 14: *Evolución de Inventario Obsoleto*

Incidencia	Inventario obsoleto	
MES	% Modelo Tradicional	% Modelo XGBoost
Agosto Q1	11.47%	0.46%
Agosto Q2	12.26%	2.88%
Septiembre Q1	14.45%	1.23%
Septiembre Q2	13.98%	1.41%

Fuente: *Elaboración Propia*

Se usaron nuevamente el tiempo de 4 quincenas para sacar los promedios incidencias de inventario obsoleto. El gráfico muestra que el modelo XGBoost logra una superioridad en precisión del 12% de diferencia. El modelo tradicional obtiene un nivel de incidencias de 13.04% mientras el modelo XGBoost mantiene un nivel de 1.54%. Esta diferencia sirve como utilidades que se explicarán en las discusiones.

Figura 6: *Evolución de Inventario Obsoleto*



Fuente: *Elaboración Propia*

4.5. Análisis de Viabilidad Económica

4.5.1. Supuestos económicos

Tesis publicada con autorización del autor
No olvide citar esta tesis

Según el Banco Central de Reserva del Perú (BCRP, 2024), la inflación proyectada para el período 2025-2028 se encuentra en un rango de 3.0% a 3.5% anual, por lo que se adoptó un valor conservador del 3.5% para el presente análisis.

El incremento del 5% en ventas se sustenta en los resultados obtenidos en el Capítulo IV, donde se demostró una reducción del quiebre de stock del 39.07% en promedio. Esta mayor disponibilidad de productos permite atender una mayor demanda, estimándose un crecimiento conservador del 5% anual.

El incremento salarial del 4% anual considera la inflación proyectada del 3.5% (BCRP, 2024) más un 0.5% adicional por mejora en productividad, derivado de la reducción de horas dedicadas al control de inventario gracias al sistema implementado.

Los costos operativos se incrementan anualmente al mismo ritmo que la inflación proyectada (3.5%), bajo el supuesto de que los proveedores de servicios tecnológicos ajustan sus tarifas conforme al aumento del costo de vida.

Tabla 15: *Costos Operativos*

Concepto	Valor
Inflación anual proyectada	3.5% (promedio)
Crecimiento de ventas anual	5% (por mejor abastecimiento)
Incremento de sueldos anual	4% (inflación + productividad)
Incremento de costos operativos	3.5% (ligado a inflación)

Fuente: *Elaboración Propia*

Tesis publicada con autorización del autor
No olvide citar esta tesis

4.5.2. Flujo de caja económico proyectado

El Año 0 corresponde a la inversión inicial de S/ 5,587.10 (desarrollo, hardware, licencias y dominio). A partir del Año 1, los costos operativos y beneficios se actualizan anualmente según los supuestos económicos señalados.

Tabla 16: *Proyección de Costos y Beneficios*

Año	Inversión / Costos	Beneficios (Ahorros)	**Flujo de caja neto**	**Flujo acumulado**
0	S/ 5,587.10	S/ 0.00	**-S/ 5,587.10**	-S/ 5,587.10
1	S/ 1,752.00	S/ 6,027.76	**+S/ 4,275.76**	-S/ 1,311.34
2	S/ 1,813.32	S/ 6,238.73	**+S/ 4,425.41**	+S/ 3,114.07
3	S/ 1,876.79	S/ 6,457.09	**+S/ 4,580.30**	+S/ 7,694.37
4	S/ 1,942.47	S/ 6,683.09	**+S/ 4,740.62**	+S/ 12,434.99

Fuente: *Elaboración Propia*

El flujo de caja económico proyectado (año 0 al 4) muestra que la inversión inicial de S/ 5,587.10 genera flujos netos positivos desde el primer año, alcanzando un punto de recuperación total durante el segundo año, cuando el flujo acumulado pasa de -S/ 1,311.34 a +S/ 3,114.07, lo que equivale a un plazo de recuperación de aproximadamente un año y cuatro meses. A partir de ese momento, el proyecto consolida su rentabilidad, acumulando S/ 7,694.37 al tercer año y S/ 12,434.99 al cuarto, lo que evidencia que el sistema de Machine Learning no solo se autofinancia en el corto plazo, sino que genera un excedente económico creciente y sostenible, incluso considerando los ajustes anuales por inflación y el incremento de costos operativos.

Estos resultados confirman que la implementación del modelo predictivo es financieramente viable, con un retorno rápido y una tendencia positiva que elimina los rezagos financieros prolongados, respaldando la decisión de adopción tecnológica como una inversión rentable y de bajo riesgo en el horizonte evaluado.

Tesis publicada con autorización del autor
No olvide citar esta tesis

4.5.3. Costo de Oportunidad del capital (COK)

El COK base es del 12%, determinado a partir de la tasa de interés de referencia del sistema financiero peruano más una prima por riesgo tecnológico.

Sin embargo, dado que la economía no es estática, el COK se ajusta anualmente considerando:

- Inflación acumulada: 3.5% anual.
- Prima por riesgo creciente: 1% adicional por año, reflejando mayor incertidumbre a futuro.

Fórmula:

COK (año n) = COK base + Inflación acumulada + Prima por riesgo.

Tabla 17: *Variabilidad del Costo de Oportunidad*

Año	Cálculo	COK ajustado
0	12% (base)	**12.00%**
1	12% + 3.5% + 1%	**16.50%**
2	12% + 7.0% + 2%	**21.00%**
3	12% + 10.5% + 3%	**25.50%**
4	12% + 14.0% + 4%	**30.00%**

Fuente: *Elaboración Propia*

El flujo de caja descontado con COK variable ajusta los flujos futuros a su valor presente utilizando factores de descuento crecientes (desde 1.1650 en el año 1 hasta 2.8561 en el año 4), reflejando un costo de oportunidad que sube del 16.5% al 30% por inflación y riesgo. Los flujos netos positivos se traducen en valores descontados decrecientes, pero el acumulado descontado pasa de -S/ 1,916.92 en el año 1 a +S/ 1,106.21 en el año 2, marcando la recuperación de la inversión en términos de valor presente; al final del año

Tesis publicada con autorización del autor
No olvide citar esta tesis

4, el VAN alcanza S/ 5,082.71, lo que confirma que el proyecto genera un excedente económico superior a la rentabilidad exigida.

Estos resultados demuestran que el sistema de Machine Learning mantiene su viabilidad financiera incluso bajo escenarios de riesgo e inflación creciente, con una capacidad de recuperación robusta y un valor actual neto positivo que respalda sólidamente la decisión de inversión.

Tabla 18: *Flujo de Caja descantado (COK Variable)*

Año	Flujo neto	Factor descuento (COK)	Flujo descontado	Flujo acumulado descontado
0	-S/ 5,587.10	1.0000	-5,587.10	-5,587.10
1	S/ 4,275.76	1.1650	3,670.18	-1,916.92
2	S/ 4,425.41	1.4641	3,022.61	1,105.69
3	S/ 4,580.30	1.9767	2,317.20	3,422.89
4	S/ 4,740.62	2.8561	1,659.82	5,082.71

Fuente: *Elaboración Propia*

El análisis de indicadores financieros arroja un VAN positivo de S/ 5,082.71, una TIR aproximada del 65% y un COK promedio del 21%, lo que evidencia que la rentabilidad del proyecto (65%) supera ampliamente el costo de oportunidad del capital, con un margen de 44 puntos porcentuales. Además, el periodo de recuperación descontado se sitúa en 1 año y 7.6 meses, lo que confirma que la inversión se recupera en menos de dos años incluso bajo escenarios de riesgo e inflación creciente.

Tesis publicada con autorización del autor
No olvide citar esta tesis

Tabla 19: *Indicadores Financieros*

Indicador	Fórmula / Cálculo	Resultado
VAN (Valor Actual Neto)	Suma de flujos descontados	**+S/ 5,082.71**
TIR (Tasa Interna de Retorno)	Tasa que hace VAN = 0	**~65%**
COK promedio	Promedio de COK ajustados por año	**~21%**
Comparación TIR vs COK	65% > 21%	**Viable**
Periodo de recuperación	Año 1 + (1,916.92 / 3,023.13) × 12 meses	**1 año, 7.6 meses**

Fuente: *Elaboración Propia*

Estos indicadores demuestran que el sistema de Machine Learning es financieramente viable y atractivo, generando valor neto positivo, retornos muy superiores al mínimo exigido y una recuperación rápida del capital invertido, lo que respalda sólidamente la decisión de implementación en la ferretería "Vargas".

4.5.1. Evaluación final

El análisis por escenarios evalúa la viabilidad del proyecto ante distintas condiciones del mercado, considerando un escenario base (TIR ~65%), optimista con ventas +10% (TIR ~75%), pesimista con inflación del 6% y caída de ventas del 10% (TIR ~45%), y crítico con recesión, ventas -25% y costos +10% (TIR ~25%).

Tesis publicada con autorización del autor
No olvide citar esta tesis

Tabla 20: *Comparación TIR vs COK por año*

Año	TIR del proyecto	COK ajustado	¿TIR > COK?	Interpretación
0	65%	12.00%	Sí (53% arriba)	Proyecto muy rentable
1	65%	16.50%	Sí (48.5% arriba)	Sigue siendo rentable
2	65%	21.00%	Sí (44% arriba)	Rentabilidad sólida
3	65%	25.50%	Sí (39.5% arriba)	Aún muy por encima
4	65%	30.00%	Sí (35% arriba)	Sigue siendo viable

Fuente: *Elaboración Propia*

En todos los escenarios planteados, la TIR supera ampliamente el COK de referencia, lo que demuestra que el proyecto mantiene su viabilidad financiera incluso en situaciones adversas, con un margen de seguridad que respalda la implementación del sistema de Machine Learning como una inversión robusta y resiliente frente a fluctuaciones económicas.

Tabla 21: *Análisis por escenarios*

Escenario	Cambio en el mercado	Flujo año 1	Flujo año 2	Flujo año 3	Flujo año 4	Nueva TIR	¿TIR > COK?
Optimista	Ventas +10% (reactivación)	S/ 4,800	S/ 4,970	S/ 5,145	S/ 5,325	**~75%**	Sí
Base	Sin cambios	S/ 4,276	S/ 4,425	S/ 4,580	S/ 4,741	**~65%**	Sí
Pesimista	Inflación 6%, caída ventas -10%	S/ 3,400	S/ 3,520	S/ 3,643	S/ 3,770	**~45%**	Sí
Crítico	Recesión, ventas -25%, costos +10%	S/ 2,100	S/ 2,174	S/ 2,250	S/ 2,329	**~25%**	Sí

Fuente: *Elaboración Propia*

Para evaluar la sensibilidad del proyecto frente a posibles fluctuaciones económicas, se desarrolló un análisis por escenarios que modifica las variables clave de crecimiento de ventas, inflación y costos operativos. Este enfoque permitió determinar si la viabilidad

Tesis publicada con autorización del autor
No olvide citar esta tesis

financiera del sistema de Machine Learning se mantiene bajo condiciones distintas a las del escenario base, ofreciendo una visión más completa del riesgo y la robustez de la inversión.

El escenario base, que incorpora un crecimiento anual del 5% en ventas y un incremento de costos operativos del 3.5% anual, arrojó una TIR del 65%, confirmando la rentabilidad esperada. En el escenario optimista, con un incremento adicional del 10% en ventas por reactivación económica, la TIR alcanzó el 75%, evidenciando que el proyecto puede beneficiarse significativamente de condiciones favorables. En el escenario pesimista, que contempla una inflación del 6% y una caída de la demanda del 10%, la TIR resultante fue del 45%, manteniéndose muy por encima del COK base del 12%. Finalmente, en el escenario crítico, que simula una recesión con contracción de ventas del 25% e incremento de costos del 10%, la TIR calculada fue del 25%, superando ampliamente el COK.

En conjunto, el análisis por escenarios demuestra que el proyecto no solo es viable en condiciones normales, sino que su estructura de beneficios y costos le confiere un alto grado de estabilidad frente a cambios del mercado. La TIR supera al COK en todos los casos evaluados, incluso en el más adverso, lo que proporciona una sólida justificación para recomendar la implementación del sistema, con la certeza de que los riesgos económicos están adecuadamente cubiertos.

Tesis publicada con autorización del autor
No olvide citar esta tesis

CAPÍTULO V: DISCUSIÓN, CONCLUSIONES Y RECOMENDACIONES

5.1.Discusión

En base a la hipótesis general: La implementación de Machine Learning mejora significativamente la predicción de demanda de productos en la ferretería "Vargas", año 2025. Los resultados obtenidos confirman la hipótesis general de la investigación. La implementación de algoritmos de Machine Learning, particularmente el modelo XGBoost, demostró ser altamente eficaz para predecir la demanda de productos en la Ferretería "Vargas". Esto se evidencia en la precisión de las predicciones, donde el modelo XGBoost logró un MAE de 1.25, un RMSE de 1.45 y un R^2 de 0.95, superando ampliamente al método tradicional y a otros modelos como la Regresión Lineal (R^2=0.73) y las Redes Neuronales Multicapa (R^2=0.63).

Este hallazgo es consistente con los reportados por Manrique y Camargo (2022) en Colombia, quienes encontraron que XGBoost era el modelo con mejor rendimiento para predecir la demanda de teléfonos inteligentes. Así mismo, Flores y Guerra (2023) en Ecuador demostraron la superioridad de Random Forest sobre las Redes Neuronales en la proyección de ventas, coincidiendo con la alta capacidad predictiva de los modelos de ensamble (XGBoost y Random Forest) sobre los métodos más simples.

La capacidad de XGBoost para capturar relaciones no lineales y su robustez con datos ruidosos, detallada por Chen y Guestrin (2016) en su teoría de Gradient Boosting, se confirma en la práctica, ya que el modelo procesó con éxito los datos históricos (2020-2024) que incluían la alta volatilidad del periodo post-pandemia.

La superioridad de XGBoost se justifica por su enfoque de boosting secuencial y su optimización de segundo orden, lo que le permite aprender de los errores de iteraciones anteriores y ajustarse a patrones complejos. La teoría de Vapnik-Chervonenkis, que

Tesis publicada con autorización del autor
No olvide citar esta tesis

sustenta la capacidad de generalización de los modelos, se ve reflejada en el alto R^2 del modelo, indicando que no solo memoriza los datos de entrenamiento, sino que puede predecir correctamente datos no vistos.

En base a la hipótesis específica N°1: La implementación de Machine Learning mejorará el cálculo de la eficiencia del abastecimiento de los productos de la ferretería "Vargas", año 2025. Los resultados obtenidos confirman que la implementación del sistema de ML redujo drásticamente el porcentaje de quiebres de stock. En el escenario tradicional, el promedio de quiebre fue de 12.89%, mientras que con XGBoost apenas alcanzó el 1.84% (ver Tabla 13).

Esta mejora, que representa una reducción de más de 10 puntos porcentuales, se tradujo en una mayor disponibilidad de productos para los clientes y, por lo tanto, una mayor eficiencia en el abastecimiento. Estos resultados se alinean con los de Bernaola y Varillas (2022), quienes lograron una reducción del 30% en pérdidas por falta de stock en una ferretería similar. Nuestros resultados van incluso más allá, demostrando una capacidad de respuesta casi perfecta en las predicciones.

Asimismo, Montalvo (2022) encontró que la adopción de ML incrementó el nivel de cumplimiento en despachos del 56.3% al 80.72%, evidenciando el impacto positivo en la logística. El sistema JIT (Just-in-Time) de Taiichi Ohno, que postula la eliminación del desperdicio y el inventario por "si acaso", se ve materializado en este resultado. Al predecir con tanta precisión, la ferretería puede pedir la cantidad justa y en el momento adecuado, reduciendo el riesgo de no tener productos cuando el cliente los necesita.

En base a la hipótesis específica N°2: La implementación de Machine Learning mejorará el cálculo de los excedentes de inventario de los productos de la ferretería "Vargas", año 2025. Los resultados demuestran que el sistema de ML redujo significativamente el

Tesis publicada con autorización del autor
No olvide citar esta tesis

porcentaje de inventario obsoleto. Mientras que el método tradicional arrojó un promedio de 13.04% de productos no rotados (ver Tabla 14), el modelo XGBoost logró reducirlo a un 1.54%. Esta reducción de más del 11% demuestra que el sistema no solo asegura que los productos estén disponibles, sino que también evita la compra excesiva de artículos que luego se quedan estancados en el almacén, generando costos de almacenamiento y pérdidas por obsolescencia.

Estos hallazgos son comparables con los de Gonzales y Pérez (2023) quienes obtuvieron una disminución del 15% en gastos por inventario excedente en una empresa textil. De manera similar, Andrade et al. (2024) mencionan que los modelos de regresión en diferentes sectores lograron reducir costos de inventario hasta en un 5%, mejorando la satisfacción del cliente. Nuestros resultados superan estas cifras, lo que puede atribuirse a la granularidad y especificidad del modelo XGBoost, que fue entrenado con datos muy detallados (5 años de registros diarios de la ferretería).

Desde la perspectiva teórica, la Ley de Little ($WIP = \lambda \times CT$) y la teoría del EOQ (Lote Económico de Compra) son los pilares de este logro. El modelo de ML, al predecir la demanda (λ) con alta precisión, permite que la ferretería reduzca el inventario en proceso o estancado (WIP) sin aumentar el tiempo de ciclo (CT), ya que los pedidos se realizan justo a tiempo. La fórmula del EOQ ($Q^* = \sqrt{\frac{2DS}{H}}$) se optimiza dinámicamente al tener una estimación precisa de la demanda (D), lo que resulta en un tamaño de lote (Q*) óptimo que evita tanto el desabastecimiento como el excedente.

Tesis publicada con autorización del autor
No olvide citar esta tesis

5.2. Conclusiones

El presente proyecto de investigación logró su objetivo principal mediante la implementación de un modelo de machine learning para la planificación de la demanda en Ferretería Vargas. Este modelo, ya operativo, ha demostrado su eficacia. El proyecto comenzó con más de 5000 datos históricos de ventas, procesados y organizados en un archivo. ipynb. Esto permitió el entrenamiento y la evaluación de varios algoritmos de inteligencia artificial. Se probaron y evaluaron tres modelos (XGBoost, red neuronal y bosque aleatorio) utilizando los indicadores de rendimiento MAE, RMSE y R^2. El algoritmo XGBoost resultó ser el más adecuado, con un MAE de 1,46 y un R^2 de 0,80. Tras seleccionar el mejor modelo, se desarrolló una aplicación para generar pronósticos de demanda. Esta aplicación se probó en agosto y septiembre de 2025, y sus pronósticos se compararon con el método tradicional de reposición de la tienda. Los resultados confirman que la implementación de esta solución tecnológica permitió el logro de los objetivos establecidos.

En base al objetivo específico N°1: "Determinar como la implementación de machine learning mejora la capacidad de eficiencia en el abastecimiento de la ferretería Vargas en el año 2025". Los resultados mostraron un avance considerable entre el método de abastecimiento de la ferretería antes del modelo de machine learning; la eficiencia va desde 9.79% hasta los 39.07% de diferencia entre un método y otro. Esto dejó mucho a consideración, porque estos resultados, la forma en como la tienda abastecía su tienda, era mediante un promedio entre fechas divididas quincenalmente; esto no generaba mucho problema, pues al ser una ferretería no tan grande no tenían problemas en la varianza entre ventas. Esto

Tesis publicada con autorización del autor
No olvide citar esta tesis

cambió con la llegada del 2020, que con la pandemia las tendencias de ventas empezaron a cambiar; se necesitaron procesos más complejos.

A este aumento de la complejidad nuestro proyecto dio respuesta mediante el uso de librerías que toman una mayor cantidad de atributos. Al ir elaborando nuestros algoritmos, consideramos muchos elementos correspondientes a las fechas: si era feriado, el determinar patrones entre meses, el llevar a una predicción diaria, mensual, hasta llegar a un período quincenal del producto. Con esto se pudieron sacar mejores resultados, por lo que concluimos que para determinar el éxito de un modelo de predicción se necesita una gran cantidad de datos de los cuales analizar, y que estos mismos atributos estéán relacionados y ordenados. El programa cumple con su función de evitar que los quiebres de productos se den con mayor frecuencia; determinamos que, al ir aumentando el uso del programa e ir incluyendo más meses, esa brecha se irá reduciendo.

En base al objetivo específico N°2: “Determinar como la implementación de machine learning mejora la reducción de excedentes de inventario de los productos en la ferretería Vargas en el año 2025”. Muchas tiendas pequeñas no prestan atención a los productos excedentes; se les considera algo bueno el que siempre sobre mercadería, pero siempre debe haber un límite. Las ferreterías, a diferencia de las tiendas de comestibles, tienen pocos productos que se vencen, pero esto genera un problema de capital inmovilizado, al igual que si hay deterioro en muchos de los productos. Imagina que la tienda tiene la versión de un producto que recientemente recibió una mejora; esta entraría en el inventario. ¿Qué cliente quisiera comprar una versión anterior de este? Como se observa, los resultados logran reducir esas cantidades de excedentes, lo cual, en caso de exigir un cambio de reposición a los proveedores, se tendrá un mayor poder de negociación, cosa

Tesis publicada con autorización del autor
No olvide citar esta tesis

que se perdería si las cantidades fueran mayores. Además, si la empresa logra reducir el capital inmovilizado, o sea, gastar menos en productos que salen con menos frecuencia, se puede centrar esa inversión en productos con mayor demanda, como también en otros aspectos de la tienda. El tener esa pauta de las cantidades exactas ayuda a gestionar la tienda de mejor manera, tener un verdadero rendimiento que permita a la tienda crecer.

Por lo cual, ya llegando a la conclusión del objetivo general, tanto el abastecer correctamente como el reducir excedentes no solo ayuda al proceso logístico de compras, sino que, a un paso más general, permite una mejor toma de decisiones. Esto da una ventaja competitiva primordial, que es el no mantener capital ocioso y buscar agilidad estratégica, asegurando que cada compra y producto de la tienda genere rentabilidad.

Finalmente, agarrando la parte de costo-rentabilidad, se concluye que el aplicativo es financieramente viable y en menos de 2 años logrará cubrir los gastos; esto mitiga eficazmente los riesgos financieros iniciales y captura los ahorros operativos. Esto trajo tranquilidad a la administración de la ferretería “Vargas”, y da oportunidad a trasladar gastos a otras tareas de la tienda y tener nuevos ingresos que permitirán no solo la rentabilidad del negocio, sino que darán oportunidades de crecer a futuro.

Tesis publicada con autorización del autor
No olvide citar esta tesis

5.3. Recomendaciones

En base a la conclusión N.° 1, se recomienda:

- Al administrador de la ferretería Vargas tener prevista una cantidad mayor de atributos; uno de los puntos de inflexión más notables fue precisamente el que nos llevó a hacer esta investigación. Al entrar en esta era digital se tiene que asegurar el tener datos de sus proveedores, clientes y demás variables que puedan perfeccionar la precisión de los algoritmos Machine Learning.
- Al contador de la ferretería, se le recomienda el añadir un sistema ERP dentro de los gastos operativos para administrar de mejor manera las variables que entraran al sistema de predicción. Con los nuevos ingresos no serán problema.

Para la conclusión N°2 se recomienda:

- Al contador, el evaluar de mejor manera los gastos operativos actuales de la ferretería. Sí tomamos en comparación los nuevos ingresos gracias a reducir la repetitividad del inventario obsoleto, se tiene que evaluar los gastos que la ferretería incurre. Especialmente en materiales construcción como ladrillos o cemento que tienden a gastar cantidades exorbitantes de almacén.
- A los trabajadores que se encargan del recuento de existencias, se necesitan capacitar en nuevas herramientas de almacén. Al digitalizar las compras permitirá contabilizar de una mejor manera las existencias que salen.

Para el apartado de rentabilidad, se recomienda:

Tesis publicada con autorización del autor
No olvide citar esta tesis

- Al administrador, se recomienda el no superar el tiempo de vida útil de 4 años del sistema, sino que este proyecto funcione como precedente para la migración a nuevas tecnologías y el crecimiento de la empresa. Y tener a la matriz como un margen del dinero que se podrá invertir, y mejorar la infraestructura, planilla, calidad, cantidad de productos, nuevos sistemas, etc.

Ya en un tono más general, el programa demuestra lo fácil que una empresa puede perder datos y no usar dinero correctamente, cosa que puede desestabilizar su rentabilidad. Por temas de autorización y facilidad de conseguir los datos, se decidió trabajar con una ferretería, pero programas así seguramente serían de provecho para insumos consumibles con fechas de vencimiento.

Tesis publicada con autorización del autor
No olvide citar esta tesis

Referencias bibliográficas

Aguirre, I., Díaz, F., Argomedo, G. Y., Monzón, J. A., & Tuesta, C. A. (2021). Impacto de la adopción de inteligencia artificial como estrategia de negocio en las empresas del sector servicios durante la época de pandemia en el Perú. [Tesis de maestría, Pontificia Universidad Católica del Perú].http://hdl.handle.net/20.500.12404/21241

Andrade, Zambrano, Zambrano, Morales (2024) Técnicas de Machine Learning aplicadas a la interpretación de datos de mercado. Revistas Universidad Alas Peruanas. http://dx.doi.org/10.21503/cyd.v27i2.2615

Bernaola, V. D. E. & Varillas, T. P. D. (2022). Sistema predictivo con Machine Learning para la gestión de inventario para la Empresa Inversiones Ferreteras Mendoza S.A.C. [Tesis de maestría, Universidad César Vallejo]. Repositorio Institucional UCV. https://repositorio.ucv.edu.pe/bitstream/handle/20.500.12692/97798/Bernaola_VDE-Varillas_TPD-SD.pdf?sequence=4&isAllowed=y

Cazares, E. (2025). Aplicación de técnicas de aprendizaje supervisado para la predicción de ventas en Hentel, mediante la implementación de modelos de machine learning con Python y Pandas [Trabajo de maestría, Universidad de Especialidades Espíritu Santo]. https://dspace.udla.edu.ec/bitstream/33000/17596/1/UDLA-EC-TMINCD-2025-18.pdf

Tesis publicada con autorización del autor
No olvide citar esta tesis

Cetinkaya, H. F., Yenidünya, A. F., Cetinkaya, S., & Tüzün, B. (2024). Deep learning and machine learning and their application on computational chemistry. In Y. El Kacimi & K. Alaoui (Eds.), Emerging engineering technologies and industrial applications (pp. 21-35). IGI Global. https://doi.org/10.4018/979-8-3693-1335-0.ch002

Cuevas, E., Avalos, J. G., Díaz, M., Valdivia, P., & Pérez, R. (2021). Introducción al machine learning con MATLAB. MARCOMBO, S.L. https://libroweb.alfaomega.com.mx/book/introduccion_al_machine_learning_con_matlab

El Peruano. (2021, 15 de marzo). Crece implementación de inteligencia artificial en tiempos de pandemia. https://elperuano.pe/noticia/121004-crece-implementacion-de-inteligencia-artificial-en-tiempos-de-pandemia

Erkut, E., & Tarimcilar, M. (1991). On sensitivity analysis in the analytic hierarchy process. IIE Transactions, 23(4), 322-330.

Flores, M. (2021). Modelo de predicción para determinar la compra de productos manteniendo un stock adecuado de artículos de Ferretería. Caso aplicado a la Ferretería MAYTA [Tesis de pregrado, Universidad César Vallejo]. Repositorio Institucional UCV. https://repositorio.ucv.edu.pe/bitstream/handle/20.500.12692/83117/Flores_ZMI-SD.pdf?sequence=1&isAllowed=y

Tesis publicada con autorización del autor
No olvide citar esta tesis

Flores, V., & Guerra, S. (2023). Análisis comparativo de modelos predictivos de machine learning para la proyección de ventas de productos Nestlé en supermercados TIA en Ecuador [Trabajo de maestría, Universidad de Especialidades Espíritu Santo]. https://dspace.udla.edu.ec/bitstream/33000/15235/1/UDLA-EC-TMINCD-2023-10.pdf

García, J. & Huertas, S. (2024). Improving demand forecasting by implementing machine learning in poultry production company [Undergraduate thesis, Universidad de Lima]. Repositorio Institucional Universidad de Lima. https://hdl.handle.net/20.500.12724/20750

García, J. & Torres, M. (2023). Predicción de riesgo de impago en institución financiera usando modelos de machine learning [Tesis de pregrado, Universidad Tecnológica Centroamericana]. Repositorio Institucional UNITEC. https://repositorio.unitec.edu/xmlui/handle/123456789/12906

Gonzales, V. A. R. & Perez, M. N. R. (2023). Sistema basado en Machine Learning para la mejora de gestión de inventario en A&M Confecciones [Tesis de maestría, Universidad César Vallejo]. Repositorio Institucional UCV. https://repositorio.ucv.edu.pe/bitstream/handle/20.500.12692/134667/Gonzales_VAR-Perez_MNR-SD.pdf?sequence=1&isAllowed=y

González, C., & Rivera, L. (2024). Investigación cuantitativa: Claves para estudiantes universitarios. Universidad Católica Luis Amigó. https://doi.org/10.21501/9789588943985

Tesis publicada con autorización del autor
No olvide citar esta tesis

González, S., & Velásquez, V. (2020). Pronóstico de demanda utilizando inteligencia artificial [Tesis de pregrado, Universidad Icesi]. Repositorio Institucional Universidad Icesi. https://repository.icesi.edu.co/server/api/core/bitstreams/eb9f814c-1bc8-4040-aa52-4a8bf6737054/content

Hutter, F., Hoos, H. H., & Leyton-Brown, K. (2011). Sequential model-based optimization for general algorithm configuration. In International Conference on Learning and Intelligent Optimization. Springer, Berlin, Heidelberg.

Inostroza (2025) Diseño de un sistema de gestión de inventarios para minimizar quiebres de stock en el Hospital del Salvador. Universidad San Sebastian. https://repositorio.uss.cl/items/4210a29c-8766-4c84-8ba7-4332cc01f42d/full

Jiménez Alfaro, A. D., & Díaz Ospina, J. V. (2022). Revisión sistemática de literatura: técnicas de aprendizaje automático (machine learning). Cuaderno Activa, 13(1), 113-121. https://ojs.tdea.edu.co/index.php/cuadernoactiva/article/view/849

Kamarthi, H., Xu, S., Tong, X., Peters, J., Czyzyk, J., & Prakash, B. A. (2024). Hierarchical industrial demand forecasting with temporal and uncertainty explanations. *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 5230-5239.https://arxiv.org/pdf/2603.06555

Keeney, R.L. and Raiffa, H. (1993) Decision with Multiple Objectives: Preference and Value Tradeoffs. Cambridge University Press, New York. https://doi.org/10.1017/CBO9781139174084

Lasa, J., Álvares, M., & de las Heras, P. (2018). Métodos ágiles: Scrum, Kanban, Lean. Anaya Multimedia. https://fliphtml5.com/uprve/xhnh/Metodos-agiles-Scrum-Kanban-Lean-pdf/180/

Tesis publicada con autorización del autor
No olvide citar esta tesis

Liu, Kalaitzi, Wang, Papanagno (2025) A machine learning approach to inventory stockout prediction. Journal of Digital Economy https://doi.org/10.1016/j.jdec.2025.06.002

Madariaga, S. (2021). Machine Learning para predecir volúmenes operacionales de las líneas de negocio de Cenabast [Tesis de postgrado, Universidad de Chile]. Repositorio Académico Universidad de Chile. https://repositorio.uchile.cl/handle/2250/182317

Manrique, J. (2022). Predicción de la demanda de smartphones de introducción al mercado colombiano mediante modelos de machine learning [Tesis de maestría, Fundación Universitaria Konrad Lorenz]. Repositorio Institucional Konrad Lorenz. https://repositorio.konradlorenz.edu.co/handle/001/5210

Martínez Cardozo, D. A., & Alarcón García, G. D. (2025). Aplicación de modelo predictivo para análisis de ventas aplicado a una cafetería mediante la implementación de técnicas de aprendizaje automático y series temporales (Machine Learning) [Trabajo de especialización, Universidad Nacional Abierta y a Distancia UNAD]. https://repository.unad.edu.co/jspui/handle/10596/78463

Montalvo, Y. (2022). Machine learning para el control de logística en la chicharronería Don Jovi S.A.C. [Tesis de pregrado, Universidad César Vallejo]. Repositorio Institucional UCV. https://repositorio.ucv.edu.pe/bitstream/handle/20.500.12692/126004/Montalvo_GYJ-SD.pdf?sequence=1&isAllowed=y

Tesis publicada con autorización del autor
No olvide citar esta tesis

Negrete, G. (2022). Diseño de un modelo machine learning para mejorar la venta de rutas diarias en la Startup de logística Wareclouds [Tesis de maestría, Universidad de Chile]. Repositorio Académico Universidad de Chile. https://repositorio.uchile.cl/handle/2250/191396

Ortiz, P. (2023). Predicción de demanda de largo plazo para mejorar planificación logística en empresa del sector retail [Tesis de maestría, Universidad de Chile]. Repositorio Académico Universidad de Chile. https://repositorio.uchile.cl/handle/2250/197506

Romano Oddone, S. G. (2025). Deep learning para predicción de demanda retail a gran escala: un modelo inteligente por producto para la totalidad de sus tiendas [Tesis de maestría, Universidad Torcuato Di Tella]. Repositorio Digital Universidad Torcuato Di Tella. https://repositorio.utdt.edu/handle/20.500.13098/13755

Sepúlveda, J., & Benavides, M. (2023). Análisis predictivo de demanda de servicios bajo series temporales [Tesis de maestría, Universidad de Antioquia]. Repositorio Institucional Universidad de Antioquia. https://bibliotecadigital.udea.edu.co/server/api/core/bitstreams/f7b58862-e695-466b-92bd-0670345d440f/content

Silva, J. (2021). *Machine learning en la mejora del proceso de operaciones comerciales en la empresa Redondos, Lima - 2020* [Tesis de pregrado, Universidad César Vallejo]. Repositorio Institucional UCV. https://repositorio.ucv.edu.pe/bitstream/handle/20.500.12692/63573/Silva_HJG-SD.pdf?sequence=1&isAllowed=y

Swami, D., Shah, A., & Ray, S. (2020). Predicting future sales of retail products using machine learning. arXiv. https://arxiv.org/pdf/2008.07779

Tesis publicada con autorización del autor
No olvide citar esta tesis

Taparia, V., Mishra, P., Gupta, N., & Kumar, D. (2023). Improved demand forecasting of a retail store using a hybrid machine learning model. Journal of Graphic Era University, 12(1), 15-36. https://doi.org/10.13052/jgeu0975-1416.1212

Tiwari, M., Dixit, R. & Kesharwani, A.(2017) Principios de minería de datos, modelo de proceso y aplicaciones. Educreation Publishing

Trejo, E. (2024). Modelo de predicción de demanda en empresa del sector textil de la región de Orizaba utilizando redes neuronales artificiales [Tesis de maestría, Tecnológico Nacional de México]. Repositorio Institucional del TecNM. https://rinacional.tecnm.mx/handle/TecNM/9795

Vargas, J. (2022). El sector agroalimentario mexicano y las nuevas tecnologías. E-Agronegocios, 8 (2), 89–113. https://doi.org/10.18845/ea.v8i2.6156

Vargas & Abreu (2023)La administración de los inventarios y su incidencia sobre la rentabilidad. Revista: Negocio, Innovación y Emprendimiento https://rai.uapa.edu.do/handle/123456789/2415

Vesga, D. (2020). Modelo de planeación de inventarios para E-commerce, utilizando herramientas de inteligencia artificial para hacer pronósticos de demanda y clasificación de inventarios [Tesis de maestría, Universidad de los Andes]. Institutional Repository. http://hdl.handle.net/1992/48641

Tesis publicada con autorización del autor
No olvide citar esta tesis

Abreviaturas

Definiciones conceptuales, glosario de conceptos

Base de datos: Una base de datos es una colección organizada de información estructurada, almacenada y gestionada digitalmente, que permite consultas y actualizaciones. Se utiliza en sistemas informáticos para almacenar grandes volúmenes de datos a los que se puede acceder rápidamente mediante consultas

CRISP DM: Por las siglas en inglés Cross-Industry Standard Process for Data Mining, es una metodología estándar para el desarrollo de proyectos de minería de datos, que están divididas en etapas como comprensión del negocio, comprensión de los datos, preparación, modelado, evaluación e implementación, lo que permite un trabajo más estructurado e iterativo en proyectos de análisis.

Elastic Net: En español es Red elástica, es una técnica de regularización combina los métodos Lasso y Ridge para mejorar la precisión y evitar el sobreajuste en los modelos de regresión. Resulta útil cuando existen muchas variables correlacionadas o más predictores que observaciones.

IA: La inteligencia artificial es la rama de la informática que se centra en el desarrollo de sistemas capaces de realizar tareas que normalmente requieren inteligencia humana, como el aprendizaje, el razonamiento y la percepción.

Ítems: Es un término general para unidades individuales de información, productos o elementos dentro de un conjunto de datos más grande; en el análisis de datos, pueden representar conjuntos de datos, atributos o componentes específicos que se están examinando o que están contenidos en una base de datos.

Tesis publicada con autorización del autor
No olvide citar esta tesis

J48: Es la implementación del algoritmo de árbol de decisión C4.5 incluido en WEKA, que genera modelos predictivos basados en reglas y división de atributos, permitiendo clasificar los datos en nodos y ramas de manera interpretable.

JRIP: Un algoritmo de clasificación basado en reglas, que están disponibles en WEKA, que utiliza el método RIPPER para generar un conjunto reducido y eficiente de reglas, capaz de clasificar datos con buena precisión y menor complejidad.

Lasso: Por las siglas en inglés Least Absolute Shrinkage and Selection Operator; es decir, Operador de Selección y Contracción Absoluta Mínima, es la técnica de regresión que aplica una penalización L1 para reducir los coeficientes de las características menos significativas a cero, realizando así selección automática de variables y evitando sobreajuste.

Machine Learning: En inglés Machine Learning; es decir, Aprendizaje Automático, es el subcampo de la IA que se ocupa del desarrollo de algoritmos y modelos estadísticos que permiten a los sistemas informáticos aprender patrones a partir de datos y mejorar su rendimiento en una tarea sin estar programados explícitamente para hacerlo.

MAE: Por las siglas en inglés Mean Absolute Error; es decir, Error Absoluto Medio, esta es la métrica de estimación del modelo de regresión, que calcula el promedio de las diferencias absolutas entre los valores predichos y los valores reales, proporcionando así una medida lineal de la magnitud del error.

Phython: Es un lenguaje de programación de alto nivel ampliamente utilizado en ciencia de datos, inteligencia artificial y desarrollo general debido a su sintaxis simple, gran comunidad y disponibilidad de biblioteca como NumPy, Pandas y Scikit learn.

Tesis publicada con autorización del autor
No olvide citar esta tesis

RandomForest: Algoritmo de aprendizaje supervisado que crea múltiples árboles de decisión y combina sus resultados para mejorar la precisión, reducir el sobreajuste y generar modelos sólidos para la clasificación y la regresión.

RandomTree: Árbol Aleatorio, es el algoritmo simple que genera un único árbol de decisión basándose en la selección aleatoria de características y participaciones, y a menudo se utiliza como componente básico de métodos de conjunto como RandomForest.

Regresión polinómica: Esta es una extensión de la regresión lineal en la que se ajusta una ecuación polinómica a los datos para modelar relaciones curvilíneas. Esto permite capturar patrones no lineales que una simple línea recta no puede representar.

Ridge: En inglés Ridge Regression; es decir, Regresión de Cresta, este es un método de regularización en modelos de regresión que penaliza la magnitud de los coeficientes para evitar el sobreajuste. Se conservan todos los predictores, pero sus efectos se reducen en casos de multicolinealidad.

RMSE: Por las siglas en inglés Root Mean Square Error; es decir, Raíz del Error Cuadrático Medio, es la métrica para evaluar modelos de regresión que calcula el error cuadrático medio, penaliza los errores más grandes con mayor severidad y lo expresa en las mismas unidades que la variable objetivo.

WEKA: Por las siglas en inglés Waikato Environment for Knowledge Analysis; es decir, Entorno Waikato para el Análisis del Conocimiento, es un software de código abierto para análisis de datos y aprendizaje automático, que incluye una gran colección de algoritmos de clasificación, regresión, árboles de decisión y

Tesis publicada con autorización del autor
No olvide citar esta tesis

evaluación, lo que permite la experimentación sin conocimientos de programación.

XGBoost: Por las siglas en inglés eXtreme Gradient Boosting; es decir, Potenciación de Gradiente Extremo, es el algoritmo de aprendizaje conjunto eficiente y potente basado en el principio de gradient boosting, que construye secuencialmente modelos de árboles de decisión para corregir los errores de los anteriores, distinguiéndose así por su velocidad y rendimiento.

Tesis publicada con autorización del autor
No olvide citar esta tesis

TESIS UNIVERSIDAD CIENTÍFICA DEL SUR

Anexos

6.1. Anexo 1: Matriz de consistencia

Tabla 22: *Matriz de Consistencia*

Problemática	O. General	Hipótesis	Variable dependiente	Tipo de Investigación: Experimental de corte longitudinal Nivel de Investigación: Aplicativo Enfoque: Cuantitativo	Población: En nuestra investigación, la población está conformada por todos los productos registrados en las ventas de la Ferretería "Vargas" durante los últimos cinco años (2020–2024). Muestra: Por consiguiente, la muestra a considerar estará compuesta por los productos pertenecientes a la categoría de construcción, durante los años (2020-2024) Muestreo: Por conveniencia, debido a que se selecciona productos de una categoría de construcción en los años (2020–2024), los cuales están fácilmente disponibles en los registros proporcionados por la ferretería "Vargas"
¿De qué manera la implementación de Machine Learning ayuda a la predicción de productos, aplicado a la ferretería "Vargas", año 2025?	Implementar Machine Learning para la predicción de demanda de productos, aplicado a la ferretería "Vargas", año 2025	Al implementar Machine Learning mejorará significativamente la predicción de demanda de productos en la ferretería "Vargas", año 2025	Predicción de demanda		
Problemas Específicos	**Objetivos específicos**	**Hipótesis Específicos**	**Variable Independiente**		
¿De qué manera la implementación de Machine Learning ayuda a la predicción de productos, aplicado a la ferretería "Vargas", año 2025?	Determinar como la implementación de Machine Learning mejora la capacidad de eficiencia del abastecimiento de productos, aplicado a la ferretería "Vargas", año 2025	Al implementar Machine Learning mejorará el cálculo de la eficiencia del abastecimiento de los productos de la ferretería "Vargas", año 2025	Machine Learning		
¿De qué manera la implementación de Machine Learning ayuda a reducir excedentes de inventario, aplicado a la ferretería "Vargas", año 2025?	Determinar como la implementación de Machine Learning mejorará la reducción de excedentes de inventario de los productos, aplicado a la ferretería "Vargas", año 2025	Al implementar Machine Learning mejorará el cálculo de los excedentes de inventario de los productos de la ferretería "Vargas", año 2025			

Fuente: *Elaboración Propia*

Tesis publicada con autorización del autor
No olvide citar esta tesis

TESIS UNIVERSIDAD CIENTÍFICA DEL SUR

6.2. Anexo 2: Matriz de operacionalización

Tabla 23: *Operacionalización de Variables*

VARIABLES	DEFINICIÓN CONCEPTUAL	DEFINICIÓN OPERACIONAL	DIMENSIÓN	INDICADOR	FÓRMULAS
V1 Machine Learning	Según J. Murphy, no se trata más que parte evolutiva de los algoritmos computacionales, que en estos últimos años ha tratado de emular la inteligencia humana.	En este aspecto, se podría decir que se define como algo computacional que usa algoritmos para que de esta manera la maquina encuentre patrones y aprenda de estos, y de estos patrones pueda realizar tareas de manera explícita.	• RandomForestRegressor • XGBoost Regressor • Red Neuronal para regresión	Raíz del error cuadrático medio (RMSE)	$RMSE = \sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{n}}$
				Error Absoluto Medio (MAE)	$MAE = \frac{\sum \lvert y_i - \hat{y}_i \rvert}{n}$
				Coeficiente de determinación (R^2)	$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \overline{y_i})^2}$
V2 Predicción de demanda	Según Meana, es el cómo se mantiene el control a las existencias dentro de un entorno, así como las distintas evaluaciones o posibles actividades que puedan ver como las compras y el registro	Son prácticas que tienen los negocios y las empresas para controlar y supervisar las existencia o materiales. Estas prácticas traen consigo procedimientos como el registro, reabastecimiento, almacenamiento, predicción de demanda, etc.	Eficiencia del Abastecimiento	Porcentaje de quiebres de stock	$= \frac{Número\ de\ productos\ en\ quiebre}{Total\ de\ productos\ demandados} \times 100$
			Reducción de excedentes de inventario	Porcentaje de Inventario Obsoleto o Estancado	$= \frac{Inventario\ no\ rotado\ en\ X\ meses}{Inventario\ total} \times 100$

Fuente: *Elaboración Propia*

Tesis publicada con autorización del autor
No olvide citar esta tesis

6.3. Anexo 3: Fichas de Observación para los indicadores

Figura 7: *Ficha para observar la eficiencia de abastecimiento*



FICHA DE OBSERVACION

TÍTULO DE LA INVESTIGACIÓ	IMPLEMENTACION DE MACHINE LEARNING PARA LA PREDICCIÓN DE DEMANDA EN LA FERRETERIA "VARGAS"		
VARIABLES:	Predicción de demanda.		
DIMENSIÓN:	Eficiencia del Abastecimiento	INDICADOR	Porcentaje de quiebres de stock
ENCARGADOS:	Huacho Taipe, Nayeli Jimena Roque Barreto, D'alessandro Armando	FECHA	

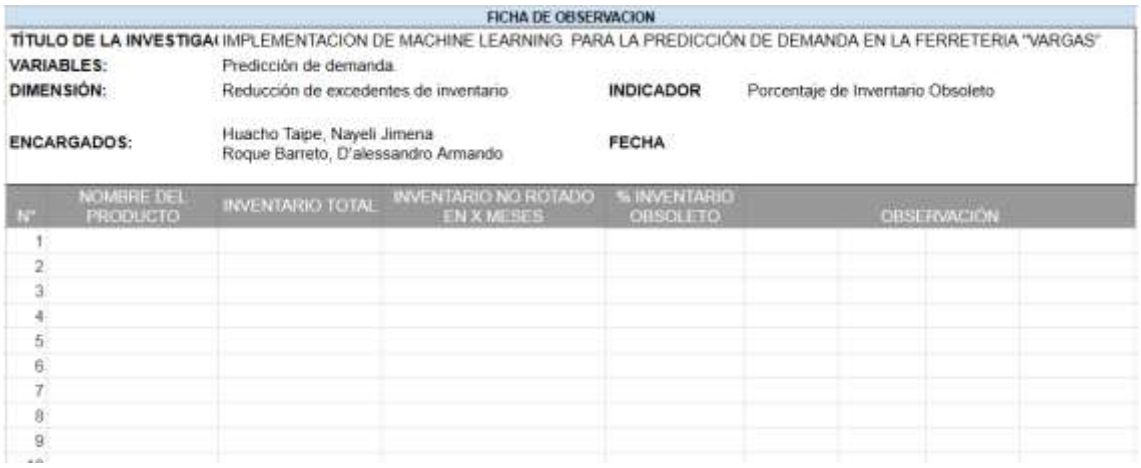
N°	NOMBRE DEL PRODUCTO	TOTAL DE PRODUCTOS DEMANDADOS	NUMERO DE PRODUCTOS EN QUIEBRE	% QUIEBRE DE STOCK	OBSERVACION
1					
2					
3					
4					
5					
6					
7					

“”

Fuente: *Elaboración Propia*

“Para dar un análisis completo a los resultados del programa, se plasmaron 2 fichas donde se registran los indicadores escogidos para la investigación. Este primer caso es la eficiencia del abastecimiento, donde se saca el quiebre de stock, este conteo las veces en que la ferretería no tiene existencias para satisfacer la demanda del cliente”

Figura 8: *Ficha para observar la reducción de excedentes*



FICHA DE OBSERVACION

TÍTULO DE LA INVESTIGA	IMPLEMENTACION DE MACHINE LEARNING PARA LA PREDICCIÓN DE DEMANDA EN LA FERRETERIA "VARGAS"		
VARIABLES:	Predicción de demanda.		
DIMENSIÓN:	Reducción de excedentes de inventario	INDICADOR	Porcentaje de Inventario Obsoleto
ENCARGADOS:	Huacho Taipe, Nayeli Jimena Roque Barreto, D'alessandro Armando	FECHA	

N°	NOMBRE DEL PRODUCTO	INVENTARIO TOTAL	INVENTARIO NO ROTADO EN X MESES	% INVENTARIO OBSOLETO	OBSERVACIÓN
1					
2					
3					
4					
5					
6					
7					
8					
9					

Fuente: *Elaboración Propia*

“La reducción de excedentes conteo los productos que se terminan quedando sin vender en la quincena”

Tesis publicada con autorización del autor
No olvide citar esta tesis

6.4. Anexo 4: Análisis Multicriterio SCORING

Figura 9: *Método SCORING 1*

Método del SCORING

1. **Meta General del Problema**

Seleccionar el modelo de Machine Learning para un sistema de predicción de demanda de la ferretería vargas.

2. **Identificar las Alternativas**

a) **XGBoost**

- Es muy bueno con datos estructurados gracias a su precisión en datos tabulares
- Es altamente optimizado, ya que incluye regularización interna mucho antes de ajustar los hiperparametros
- Tiene buena velocidad de ejecución al tener procesamiento paralelo

b) **Random Forest**

- Tiene un sobreajuste robusto
- Poco procesamientos ya que maneja muy bien los valores atípicos y variables categóricas.
- Te permite saber con claridad qué variables verdaderamente aportan a la predicción

c) **Redes Neuronales Multicapa (MLP)**

- Capaz de aproximar prácticamente cualquier función matemática
- Excelente con los datos no tabulares y buena extensibilidad para adaptarse a cualquier arquitectura de datos.

d) **Regresión Lineal**

- Máxima interpretabilidad, fácil de entender la influencia de variables
- Es extremadamente rápido y requiere poder computaciones
- Ofrecen una línea base en comparación con modelos más complejos

3. **Listar los criterios**

➢ Precisión: El criterio más importante pues busca minimizar el error de predicción. Se usaron indicadores como el MAE/RMSE/R^2

Fuente: *Elaboración Propia*

Tesis publicada con autorización del autor
No olvide citar esta tesis

Figura 10: *Método SCORING 2*

- ➢ Eficiencia Técnica: Un modelo más rápido facilita las pruebas y futuras actualizaciones. Sus indicadores a tomar son el Tiempo/RAM
- ➢ Vía Despliegue: Tiene que ver con la compatibilidad del algoritmo con la estructura del proyecto, al usar un modelo MVC y api's para Postgres o Google Drive
- ➢ Facilidad de ajuste: La optimización es importante, pues un modelo en constante mantenimiento o ajuste de los hiperparámetros ocasionan cuellos de botella

4. Asignar una ponderación

Criterio	Peso
Precisión	40%
Eficiencia Técnica	20%
Vía de despliegue	30%
Facilidad de Ajuste	10%

5. Establecer puntos de satisfacción

Se utiliza una escala de **1 a 5**, donde:

- 1 = Muy bajo
- 2 = Bajo
- 3 = Medio
- 4 = Alto
- 5 = Muy alto

Modelo	Precisión	Eficiencia Técnica	Vía de despliegue	Facilidad de Ajuste
Regresión Lineal	2	5	5	5
Random Forest	4	2	4	4
XGBoost	5	4	4	3
MLP	2	3	3	2

Fuente: *Elaboración Propia*

Tesis publicada con autorización del autor
No olvide citar esta tesis

Figura 11: *Método SCORING 3*

6. **Calcular el Score**

Modelo	Precisión (40)	Eficiencia Técnica (20)	Vía de despliegue (30)	Facilidad de Ajuste (10)	Total
Regresión Lineal	80	100	150	50	380
Random Forest	160	40	120	40	360
XGBoost	200	80	120	30	430
MLP	80	60	90	20	250

Modelo	Porcentaje
Regresión Lineal	76%
Random Forest	72%
XGBoost	86%
MLP	50%

7. **Ordenar las alternativas**

Ranking	Modelo	Score
1	XGBoost	430
2	Regresión Lineal	380
3	Random Forest	360
4	MLP	250

De acuerdo con el método de Scoring aplicado, XGBoost obtuvo el mayor puntaje (430 puntos), por lo que representa la alternativa más adecuada para el desarrollo del sistema de predicción de demanda. Esto se debe principalmente a su alta precisión predictiva y a un buen equilibrio entre eficiencia técnica y facilidad de despliegue.

Fuente: *Elaboración Propia*

Tesis publicada con autorización del autor
No olvide citar esta tesis

6.5. Anexo 5: Evidencias de ejecución de la Investigación

Figura 12: *Fotografía de la propuesta de colaboración*



Fuente: *Elaboración Propia*

"Durante nuestra visita a la Ferretería Vargas el 18 de Julio de 2025, nos reunimos con la dueña, quien nos recibió con disposición a colaborar. Le solicitamos información sobre el historial de ventas de los últimos 5 años, así como datos adicionales sobre sus productos más demandados."

Tesis publicada con autorización del autor
No olvide citar esta tesis

Figura 13: *Fotografía de la obtención del registro de ventas*



Fuente: *Elaboración Propia*

"En las 2 semanas posteriores al primer encuentro, tuvimos un breve encuentro en el local de la ferretería. Nuestras actividades consistían en la evaluación de problemas y redactar objetivos generales. Se evaluaron las deficiencias del cómo operaban su logística para que nuestras metas sean alcanzables en el tiempo determinado de investigación"

Tesis publicada con autorización del autor
No olvide citar esta tesis

6.6. Anexo 6: Evidencias de la construcción del programa

Figura 14: *Fotografía del código para definir atributos*

```
# Definir rango de años
peru_holidays = holidays.country_holidays('PE', years=range(2020, 2025))

df['holiday'] = df['Fecha'].apply(lambda x: 1 if x in peru_holidays else 0)
df_quincena = df.groupby(['Codigo','Precio Unitario','periodo']).agg({
    'Cantidad':'sum',
    'holiday':'sum'
}).reset_index()

print(df.head())
```

```
       Fecha  Codigo  Precio Unitario  Cantidad  Total  quincena     periodo  \
0 2020-01-03      77             13.6       2.0   27.2         1  2020-01-Q1
1 2020-01-03      63              8.0       3.0   24.0         1  2020-01-Q1
2 2020-01-03      85              7.0       1.0    7.0         1  2020-01-Q1
3 2020-01-03     103              8.4       5.0   42.0         1  2020-01-Q1
4 2020-01-03     110             27.8       1.0   27.8         1  2020-01-Q1

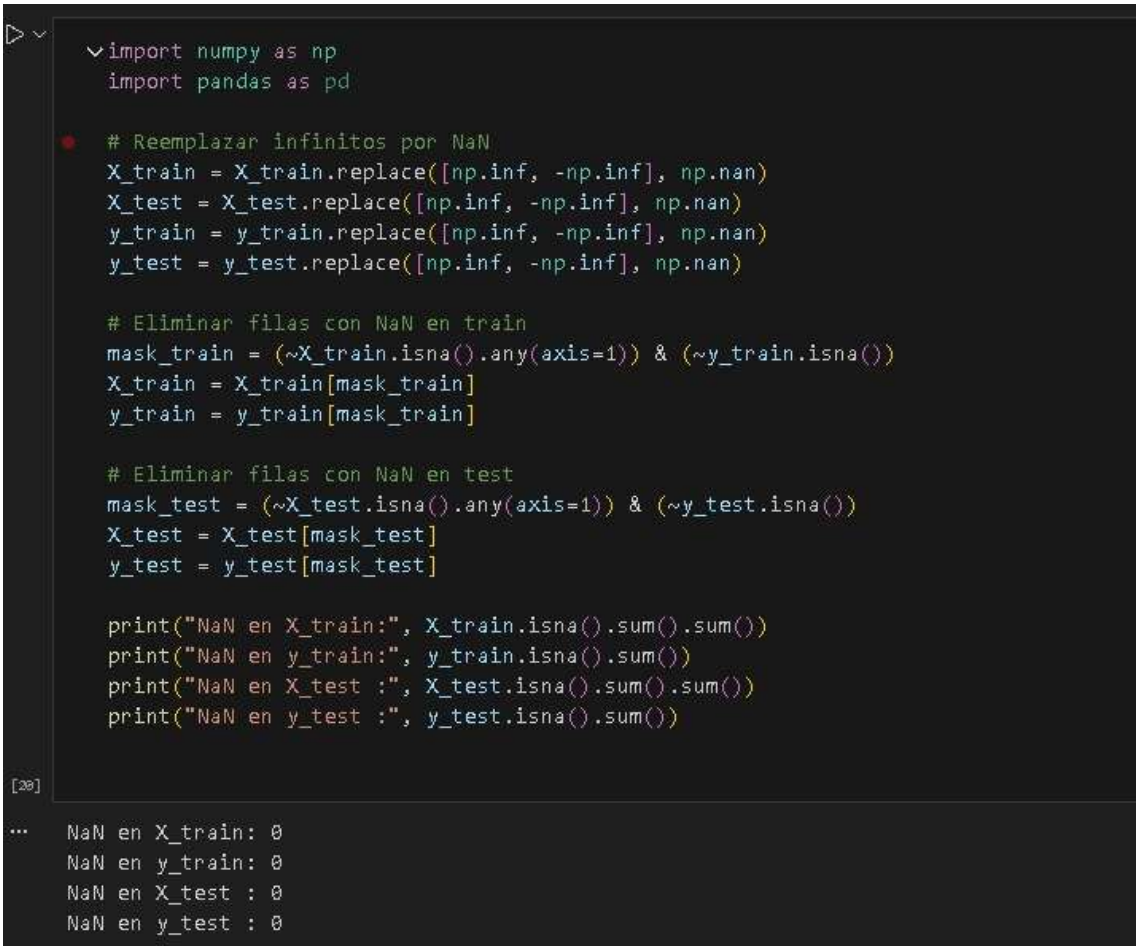
   holiday
0        0
1        0
2        0
3        0
4        0
```

Fuente: *Elaboración Propia*

"En la siguiente captura se presenta una parte importante de la ingeniería de variables, bajo la estructura Modelo-Vista-Controlador; el módulo modelo se trabajó con las variables que entran como atributos para el programa de predicción usando Machine Learning. Se utilizó la librería Holydays para dar mayor cantidad de atributos al modelo, y así permitir que reconozca patrones y dar predicciones más precisas"

Tesis publicada con autorización del autor
No olvide citar esta tesis

Figura 15: *Fotografía del código para división de datos train y test*



```
import numpy as np
import pandas as pd

# Reemplazar infinitos por NaN
X_train = X_train.replace([np.inf, -np.inf], np.nan)
X_test = X_test.replace([np.inf, -np.inf], np.nan)
y_train = y_train.replace([np.inf, -np.inf], np.nan)
y_test = y_test.replace([np.inf, -np.inf], np.nan)

# Eliminar filas con NaN en train
mask_train = (~X_train.isna().any(axis=1)) & (~y_train.isna())
X_train = X_train[mask_train]
y_train = y_train[mask_train]

# Eliminar filas con NaN en test
mask_test = (~X_test.isna().any(axis=1)) & (~y_test.isna())
X_test = X_test[mask_test]
y_test = y_test[mask_test]

print("NaN en X_train:", X_train.isna().sum().sum())
print("NaN en y_train:", y_train.isna().sum())
print("NaN en X_test :", X_test.isna().sum().sum())
print("NaN en y_test :", y_test.isna().sum())
```

```
[20]
...
NaN en X_train: 0
NaN en y_train: 0
NaN en X_test : 0
NaN en y_test : 0
```

Fuente: *Elaboración Propia*

"Continuando con la parte de ingeniería de variables, específicamente de los atributos, se procede a la limpieza de datos. Este bloque de código detalla la conversión de datos infinitos a datos N/A. Esto para luego eliminar las filas que tengan datos vacíos N/A para la previsión de sesgos ocasionados por datos incompletos"

Tesis publicada con autorización del autor
No olvide citar esta tesis

Figura 16: *Fotografía del código para modelo XGBoost*

```
from xgboost import XGBRegressor
from sklearn.metrics import mean_squared_error, mean_absolute_error, r2_score
import numpy as np

# --- Definir y entrenar el modelo ---
xgb_model = XGBRegressor(
    n_estimators=500,     # número de árboles
    max_depth=5,          # profundidad máxima
    learning_rate=0.05,   # tasa de aprendizaje
    subsample=0.8,
    colsample_bytree=0.8,
    random_state=42
)

xgb_model.fit(X_train, y_train)

# --- Predicciones ---
y_pred = xgb_model.predict(X_test)

# --- Evaluación ---
mse = mean_squared_error(y_test, y_pred)
rmse = np.sqrt(mse)
mae = mean_absolute_error(y_test, y_pred)
r2 = r2_score(y_test, y_pred)

print("Resultados XGBoost:")
print(f"MAE  : {mae:.2f}")
print(f"RMSE : {rmse:.2f}")
print(f"R²   : {r2:.2f}")
```

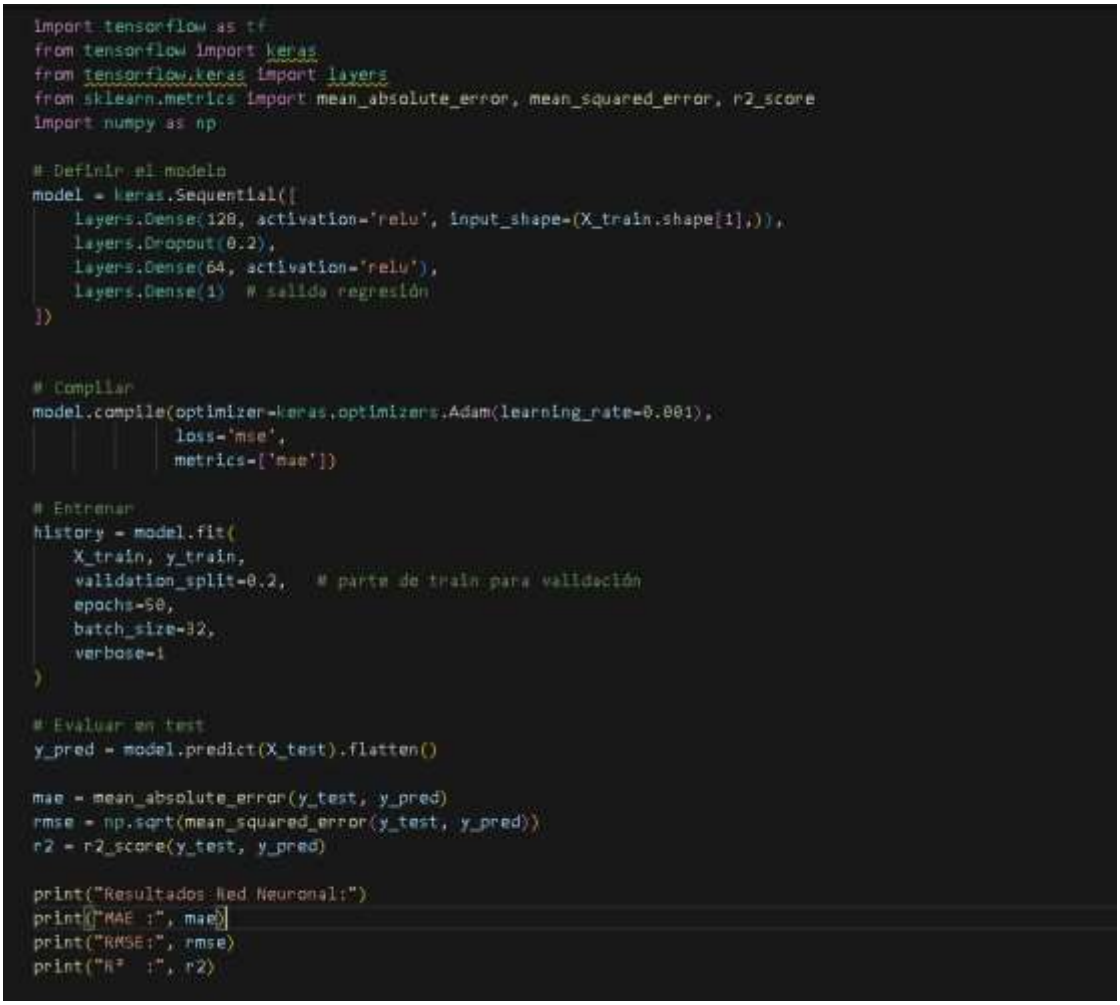
```
Resultados XGBoost:
MAE  : 1.46
RMSE : 1.96
R²   : 0.80
```

Fuente: *Elaboración Propia*

"Continuando en el mismo módulo, se codificó una etapa del entrenamiento bajo el modelo predictivo de XGBoost (Extreme Gradient Boosting). Se definieron los estimadores y las tasas de aprendizaje con el fin de optimizar la capacidad de generalización de los ejemplos. También el código muestra un conjunto de métricas de error para cuantificar la precisión, cosa que se ve presente también en los otros dos modelos."

Tesis publicada con autorización del autor
No olvide citar esta tesis

Figura 17: *Fotografía del código para modelo Redes Neuronales*



```
import tensorflow as tf
from tensorflow import keras
from tensorflow.keras import layers
from sklearn.metrics import mean_absolute_error, mean_squared_error, r2_score
import numpy as np

# Definir el modelo
model = keras.Sequential([
    layers.Dense(128, activation='relu', input_shape=(X_train.shape[1],)),
    layers.Dropout(0.2),
    layers.Dense(64, activation='relu'),
    layers.Dense(1)  # salida regresión
])


# Compilar
model.compile(optimizer=keras.optimizers.Adam(learning_rate=0.001),
              loss='mse',
              metrics=['mae'])

# Entrenar
history = model.fit(
    X_train, y_train,
    validation_split=0.2,  # parte de train para validación
    epochs=50,
    batch_size=32,
    verbose=1
)

# Evaluar en test
y_pred = model.predict(X_test).flatten()

mae = mean_absolute_error(y_test, y_pred)
rmse = np.sqrt(mean_squared_error(y_test, y_pred))
r2 = r2_score(y_test, y_pred)

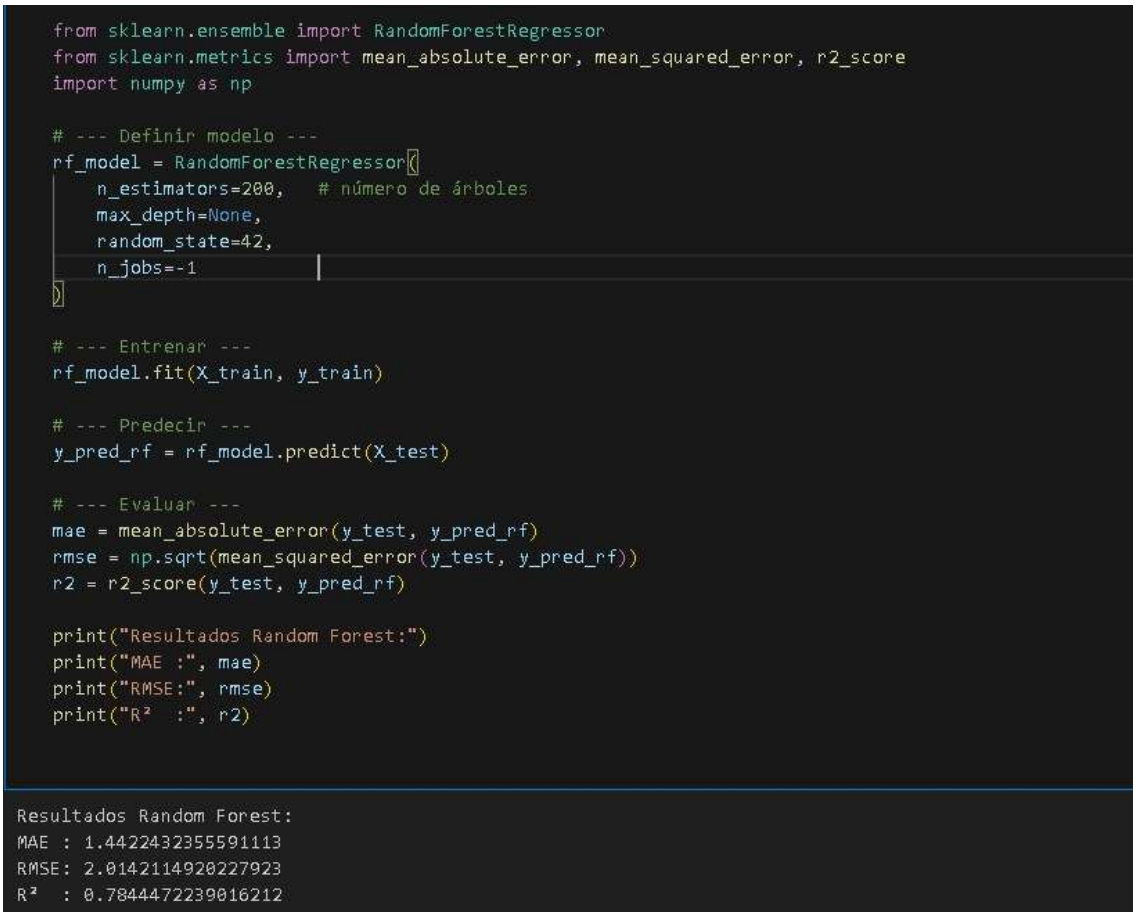
print("Resultados Red Neuronal:")
print("MAE :", mae)
print("RMSE:", rmse)
print("R²  :", r2)
```

Fuente: *Elaboración Propia*

"En la siguiente captura se detalla el desarrollo de la red neuronal, construida por la API Keras. Se utiliz*ó* eso como parte de los modelos para explorar si el Deep learning puede identificar interdependencias dentro de los datos de las ventas históricas que los modelos random forest tienden a omitir."

Tesis publicada con autorización del autor
No olvide citar esta tesis

Figura 18: *Fotografía del código para modelo Random Forest*



```
from sklearn.ensemble import RandomForestRegressor
from sklearn.metrics import mean_absolute_error, mean_squared_error, r2_score
import numpy as np

# --- Definir modelo ---
rf_model = RandomForestRegressor(
    n_estimators=200,   # número de árboles
    max_depth=None,
    random_state=42,
    n_jobs=-1
)

# --- Entrenar ---
rf_model.fit(X_train, y_train)

# --- Predecir ---
y_pred_rf = rf_model.predict(X_test)

# --- Evaluar ---
mae = mean_absolute_error(y_test, y_pred_rf)
rmse = np.sqrt(mean_squared_error(y_test, y_pred_rf))
r2 = r2_score(y_test, y_pred_rf)

print("Resultados Random Forest:")
print("MAE :", mae)
print("RMSE:", rmse)
print("R²  :", r2)
```

```
Resultados Random Forest:
MAE : 1.4422432355591113
RMSE: 2.0142114920227923
R²  : 0.7844472239016212
```

Fuente: *Elaboración Propia*

"Este fragmento de código representa al modelo de algoritmo random forest, este llega a necesitar más configuración, al tener que ensamblar los árboles de decisión y promediar las predicciones. Aquí nuevamente se usan las métricas de error"

Tesis publicada con autorización del autor
No olvide citar esta tesis

Figura 19: *Código para el módulo controlador predictor*

```python
def predict_single(model_name: str, codigo, year: int, q_global: int):
    """
    Predice para un solo producto identificado por `codigo`.
    Devuelve un diccionario con información del producto y la predicción.
    """
    model = load_model(model_name)
    products = load_products()

    producto_row = products[products['Codigo'] == int(codigo)]
    if producto_row.empty:
        raise ValueError(f"Producto con Código={codigo} no encontrado en productos.csv")

    X, meta = build_features_for_products(producto_row, year, q_global)
    pred = model.predict(X)
    pred = np.array(pred).reshape(-1)[0]
    pred = max(pred, 0)
    pred = int(round(pred))

    prod_info = producto_row.iloc[0].to_dict()
    return {
        "Codigo": prod_info.get('Codigo'),
        "Producto": prod_info.get('Producto'),
        "Precio": float(prod_info.get('Precio')),
        "Prediccion_Cantidad": pred
    }

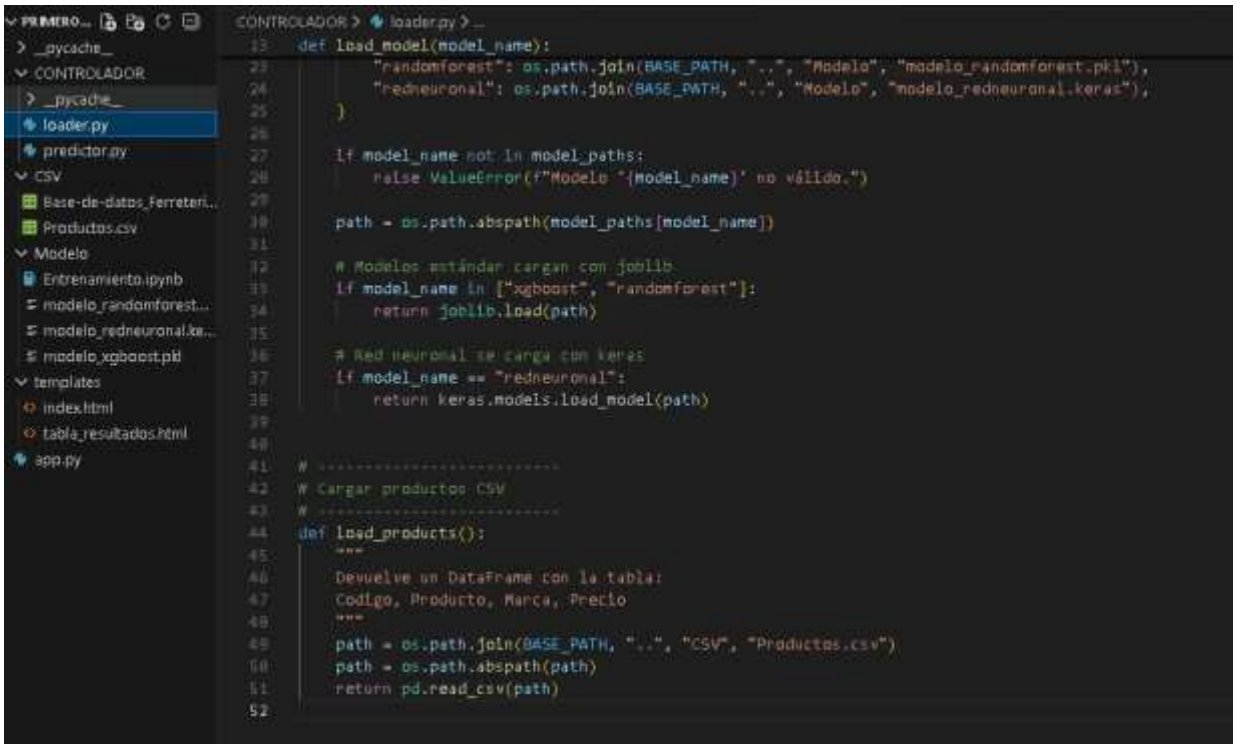

# predictor.py directamente para probar:
if __name__ == "__main__":
    # (ajusta año y quincena)
    df_res = predict_all("xgboost", 2025, 8)
    print(df_res.head(20))
```

Fuente: *Elaboración Propia*

"Pasando al siguiente módulo, la parte del controlador se ubicó el archivo predictor.py, aquí se construyeron los componentes responsables de cargar los modelos entrenados. Se integraron lógicas de validación de productos mediante archivos CSV y la reconstrucción de los atributos que se mostrarán. Es la parte donde se asegura la escalabilidad para las interfaces web y las bases de datos empresariales"

Tesis publicada con autorización del autor
No olvide citar esta tesis

Figura 20: *Código para el módulo controlador loader*



Fuente: *Elaboración Propia*

"En la siguiente ilustración se muestra el componente loader.py, que se encarga de recuperar los modelos entrenados. Esta capa es crítica ya que centraliza las funciones de acceso a los recursos, asegurando que el programa conmute entre los tres modelos de predicción existentes."

Tesis publicada con autorización del autor
No olvide citar esta tesis

6.7. Anexo 7: Evidencias del funcionamiento del programa

Figura 21: *Formulario para el ingreso de parámetros*



Fuente: *Elaboración Propia*

"La siguiente captura muestra la interfaz ya profundizada en el diseño, hemos colocado tres entradas y se dividen de la siguiente manera: el primero es Modelo a usar donde se conmuta a que modelo de predicción se desea emplear, luego está el Año para especificar a qué año deseamos predecir, y finalmente la quincena donde se coloca un valor de 1 al 24 para ubicarse dentro del año. El programa toma el año y la quincena para ubicarse en un marco y generar resultados conforme a los modelos de predicción."

Tesis publicada con autorización del autor
No olvide citar esta tesis

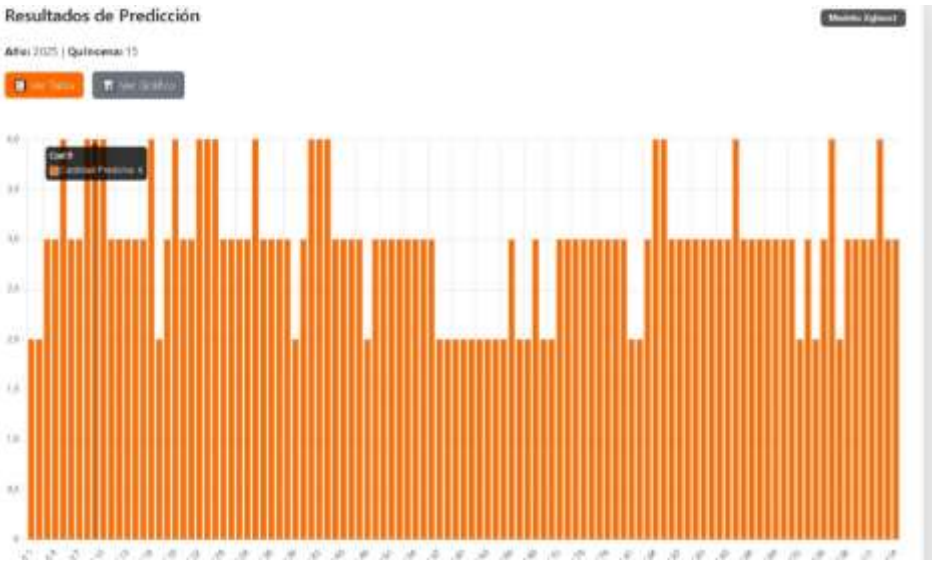
Figura 22: *Salida de resultados en formato tabla*



Fuente: *Elaboración Propia*

"Este es una de las interfaces de los resultados, en la primera opción la salida se muestra en un formato de tabla, entrega datos sobre el precio, pero lo verdaderamente importante es la cantidad señalada para la fecha propuesta."

Figura 23: *Salida de resultados en formato gráfico de barras*



Fuente: *Elaboración Propia*

"Esta captura muestra la segunda opción del cómo se muestran los datos, toma el formato tabla y lo pasa a un gráfico de barras."

Tesis publicada con autorización del autor
No olvide citar esta tesis

6.8. Anexo 8: Constancia emitida por la institución donde se realizó la investigación

CONSTANCIA EMITIDA POR LA INSTITUCIÓN DONDE SE REALIZÓ LA INVESTIGACIÓN

Lima, 26 de enero 2026

El que suscribe, gerente(a) de la empresa Ferretería Vargas S.A.C.
Yenyi Shumara Vargas Taipe
Otorga la presente constancia de aplicación del instrumento (observación) del proyecto de investigación a:

D'alessandro Armando Roque Barreto identificado con DNI N° 72683683 y Nayeli Jimena Huacho Taipe identificada con DNI N° 72198403, alumnos de la "Universidad Científica del Sur" de la carrera de Ingeniería de Sistemas Empresariales.

Quienes han realizado la aplicación del instrumento (Observación) del proyecto de investigación (Tesis) "IMPLEMENTACIÓN DE MACHINE LEARNING PARA LA PREDICCIÓN DE DEMANDA EN LA FERRETERÍA VARGAS". En la institución mencionada bajo mi autorización y supervisión desde el día 16 del mes de Julio del 2025.

Se expide la presente constancia a solicitud de los interesados para los fines que estimen convenientes.

Atentamente,

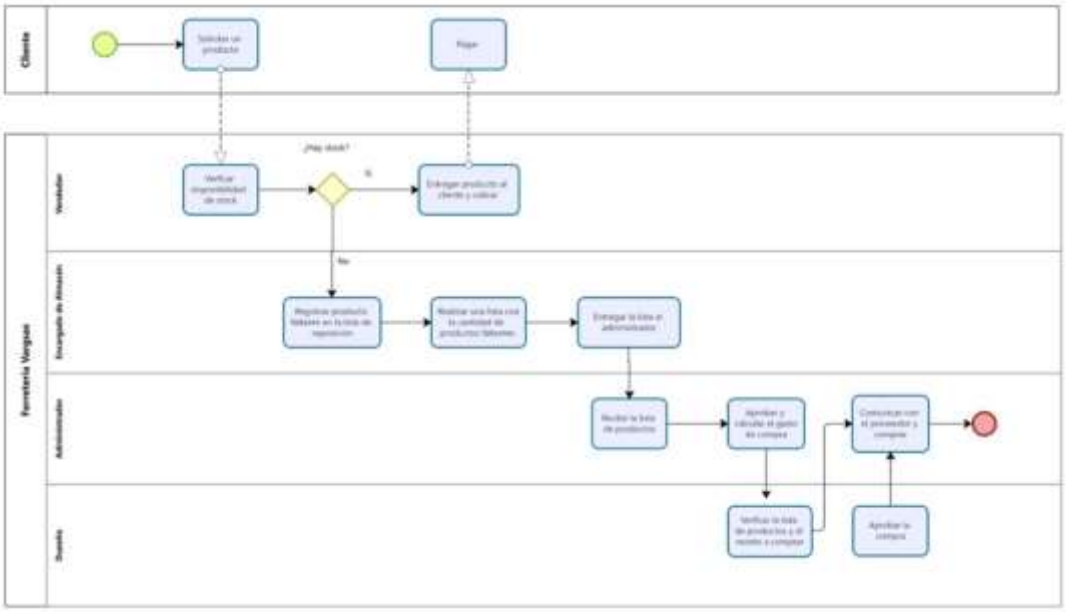
V.°B.°

GERENTE GENERAL

Tesis publicada con autorización del autor
No olvide citar esta tesis

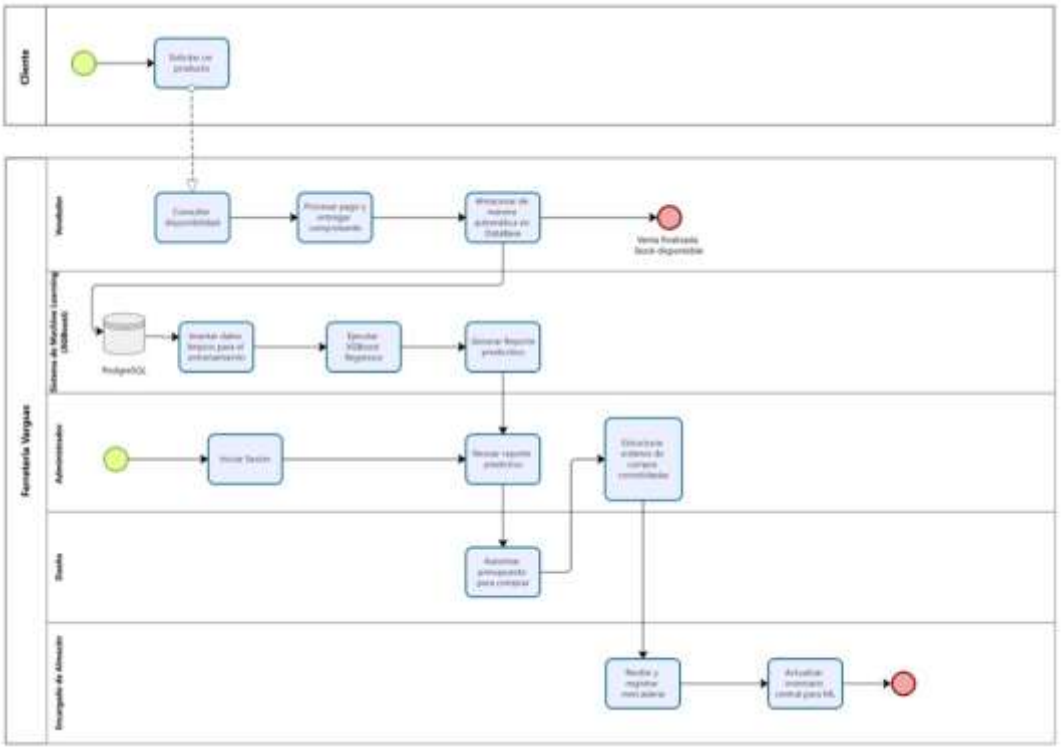
6.9. Anexo 9: Flujo de procesos

Figura 24: *Flujo de proceso antes del sistema*



Fuente: *Elaboración Propia*

Figura 25: *Flujo de proceso después del sistema*



Fuente: *Elaboración Propia*

Tesis publicada con autorización del autor
No olvide citar esta tesis

6.10. Anexo 10: Informe de aprobación de Tesis del Asesor

Facultad de Ciencias Empresariales

UNIVERSIDAD CIENTÍFICA DEL SUR

ANEXO 3 A

VALIDACIÓN DEL ENVÍO DE PROYECTO DE TESIS

Lima, 26 de marzo de 2026

Yo, WILVER AUCCAHUASI AIQUIPA, docente a tiempo Parcial de la Carrera de Ingeniería de Sistemas Empresariales, asesor de los D'alessandro Armando Roque Barreto y Nayeli Jimena Huacho Taipe; estudiantes de la Carrera de Ingeniería de Sistemas Empresariales, en la elaboración del proyecto de tesis titulado:

"IMPLEMENTACION DE MACHINE LEARNING PARA LA PREDICCIÓN DE DEMANDA EN LA FERRETERIA "VARGAS"

Confirmo que he revisado dicho proyecto y considero que la versión presentada puede ser sometida a evaluación por la Dirección General de Investigación, Desarrollo e Integridad Científica.

Atentamente,

Asesor de la tesis
MSC. Ing. AUCCAHUASI AIQUIPA
WILVER

Tesis publicada con autorización del autor
No olvide citar esta tesis

6.11. Anexo 11: Constancia de Consentimiento Informado

Facultad de Ciencias Empresariales

UNIVERSIDAD CIENTÍFICA DEL SUR

ANEXO 19

FORMATO DE CONSTANCIA DE CONSENTIMIENTO INFORMADO DE ACCESO PÚBLICO

Por medio de la presente, informamos que la empresa Ferretería Vargas E.I.R.L., autoriza al tesista Nayeli Jimena Huacho Taipe con código de matrícula 100087771 y D'alessandro Armando Roque Barreto, con código de matrícula 100088777, de la carrera profesional de Ingeniería de Sistemas Empresariales de la Universidad Científica del Sur, hacer uso de la información que a continuación se detalla, con el consentimiento de la institución, para el desarrollo de su tesis de grado, la cual lleva por título: "IMPLEMENTACIÓN DE MACHINE LEARNING PARA LA PREDICCIÓN DE DEMANDA EN LA FERRETERÍA VARGAS".

Información y/o permiso solicitado y a emplear en la tesis:

1. ~~REGISTRO DE LOS TRABAJADORES CONTRADOS~~
2. ~~REGIDSTRO DE ASISTENCIA D ELOS TRABAJORES DE LOS 2 AÑOS~~
3. DATOS DE LA IMP'LEMENTACION DEL SISTEMA DE MACHINE LEARNING
4. ~~PERMISO PARA REALIZAR ENCUESTA A LOS TRABAJADORES DE PLANTA~~

Dicha información, se recopilará en el Trabajo de Campo para posteriormente ser registrada en su investigación con fines netamente pedagógicos.

Se expide la presente CONSTANCIA DE CONSENTIMIENTO INFORMADO, a solicitud del interesado (a) para los fines que estime conveniente.

Lima, 31 de Julio de 2026

Firma del Tesita
DNI: 72198403

Firma del Tesita
DNI: 72683683

Firma del representante Ferretería Vargas E.I.R.L.,
Kenyi Shumara Vargas Taipe
RUC: 10760150075

Tesis publicada con autorización del autor
No olvide citar esta tesis

RESOLUCION DIRECTORIAL DE CARRERA: Ingeniería de Sistemas Empresariales -CI-DAFCE-U.CIENTIFICA-2020

CARRERA DE: Ingeniería de Sistemas Empresariales

Lima 05 de 03 del 2026

VISTO:

El informe de resolución independiente académico y la aprobación de un comité de Ética del proyecto de tesis titulado: IMPLEMENTACION DE MACHINE LEARNING PARA LA PREDICCIÓN DE DEMANDA EN LA FERRETERIA "VARGAS" presentado por:

-D'ALESSANDRO ARMANDO ROQUE BARRETO // NAYELI JIMENA HUACHO TAIPE //

CONSIDERADO:

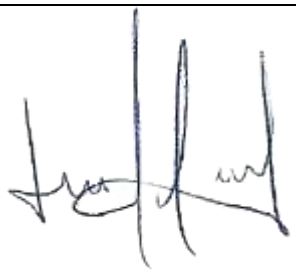
Que, de acuerdo al reglamento general de la Universidad Científica del Sur y los reglamentos de para obtener el en la Carrera de Ingeniería de Sistemas Empresariales se debe desarrollar un trabajo de investigación

Que, de acuerdo normativa vigente de la universidad científica del sur, en uso de las atribuciones conferidas al director académico de carrera

SE RESUELVE:

Aprobar e inscribir el proyecto de tesis titulado: IMPLEMENTACION DE MACHINE LEARNING PARA LA PREDICCIÓN DE DEMANDA EN LA FERRETERIA "VARGAS" con Número de registro: PRE-12-2025-00797, y establecer el inicio del periodo de ejecución del mencionado proyecto.
Nombrar al docente WILVER AUCCAHUASI AIQUIPA // como asesor para el desarrollo de tesis en cuestión
Registrese, Comuníquese y archívese

Directora Académica
Carrera de Ingeniería de Sistemas Empresariales

	
Mg. Sandra Mariana Flores Ganoza Directora de Departamento Académico de Ingeniería	**Mg. José Luis Espinoza Melgarejo** Responsable de Investigación de Ingeniería

CONSTANCIA N° 124-CIEI-CIENTÍFICA-2026

El presidente del Comité Institucional de Ética en Investigación de la Universidad Científica del Sur (CIEI-CIENTÍFICA) da constancia que el proyecto de investigación, cuyos datos se registran a continuación, **ha sido aprobado** por el CIEI-CIENTÍFICA.

Código de Registro:	**PRE-12-2025-00797**
Título:	**Implementación de machine learning para la predicción de demanda en la ferreteria "Vargas"**
Investigador(a):	**Huacho Taipe, Nayeli Jimena; Roque Barreto, D'alessandro Armando**

La aprobación del proyecto de investigación implica que el documento presentado a evaluación cumple con los lineamientos de Universidad en materia de investigación y ética, los cuales se sustentan en la normativa internacional vigente.

La aprobación del proyecto de investigación corresponde a los siguientes documentos:

1. Protocolo de investigación, versión 2.0 de fecha 20 de febrero del 2026

La aprobación tiene vigencia desde la emisión del presente documento hasta el **19 de agosto del 2027.**

Todo trámite para la renovación de la aprobación tendrá que ser por lo menos 30 días antes de su vencimiento.

Tome atención del comunicado **N° 124-CIEI-CIENTÍFICA-2026** que se adjunta a su constancia de aprobación.

Villa El Salvador, 19 de febrero del 2026

UNIVERSIDAD CIENTÍFICA DEL SUR – Comité Institucional de Ética en Investigación – UCSUR

Mg. Cinthya Ruth Chipana Herquinio
Presidente
Comité Institucional de Ética en Investigación

COMUNICACIÓN N° 124-CIEI-CIENTÍFICA-2026

En consideración a la aprobación de su proyecto de investigación el CIEI-CIENTÍFICA considera establecer algunos alcances y precisiones para su conocimiento:

1.- Si su investigación, se va a desarrollar en una Institución externa a la Universidad Científica del Sur, usted deberá enviar (la) (el) aprobación/permiso institucional de dicha institución externa al correo del CIEI-CIENTÍFICA ciei_ucsur@cientifica.edu.pe antes de iniciar la ejecución de su proyecto de investigación para el pase respectivo.

2.- La presente aprobación de su estudio no podrá continuar ejecutándose, si es que el proyecto de investigación arriba indicado es modificado de cualquier forma. Toda enmienda, añadido, eliminación o eventualidad (eventos adversos, etc.) posterior a la fecha de aprobación debe ser reportado al CIEI-CIENTÍFICA.

3.- La aprobación de este proyecto es válida solo para el proyecto que ha sido evaluado por el CIEI-CIENTÍFICA, cualquier otro proyecto que se quiera llevar a cabo utilizando la información recogida como parte de este proyecto, debe solicitar la evaluación ética correspondiente.

4.- La vigencia de la aprobación de este documento es de dieciocho (18) meses (**hasta el 19 de agosto del 2027**), periodo en el que puede desarrollarse el proyecto. Cualquier trámite para su renovación deberá ser enviado antes del término de la vigencia de este documento, debiendo suspenderse todo desarrollo hasta que la renovación sea aprobada.

5.- Toda aprobación del CIEI-CIENTÍFICA tiene alcance a nivel nacional, si es que la propuesta de investigación considera el empleo de información de personas de otros países, el tesista/investigador tendrá que contar con la aprobación del acceso y uso de dicha información por la Institución de ese país y tendrá que declarar en su proyecto que la propuesta de investigación será revisada y aprobada por un Comité de ética del país de donde provienen los datos.

Comité Institucional de Ética en Investigación
Universidad Científica del Sur
CIEI-CIENTÍFICA