



UNIVERSIDAD ESTATAL DE MILAGRO
FACULTAD DE CIENCIAS E INGENIERÍA

TRABAJO DE PROPUESTA TECNOLÓGICA
PREVIO A LA OBTENCIÓN DEL TÍTULO DE INGENIERO EN
SISTEMA

TEMA: PROPUESTA DE DESARROLLO DE UN SISTEMA DE
PREDICCIÓN DE VENTAS USANDO ALGORITMOS DE
INTELIGENCIA ARTIFICIAL.

Autores:

Jiménez Estrada Vicente Manuel

Gómez Herrera George Ezequiel

Tutor:

Ing. León Granizo Oscar Darío

Milagro, Octubre del 2021

ECUADOR

DERECHOS DE AUTOR

Ingeniero.

Fabrizio Guevara Viejó, PhD.

RECTOR

Universidad Estatal de Milagro

Presente.

Yo, Jiménez Estrada Vicente Manuel en calidad de autor y titular de los derechos morales y patrimoniales del trabajo de integración curricular, modalidad virtual, mediante el presente documento, libre y voluntariamente procedo a hacer entrega de la Cesión de Derecho del Autor, como requisito previo para la obtención de mi Título de Grado, como aporte a la Línea de Investigación **Educación, cultura, tecnología en innovación para la sociedad - sub - línea tecnología e innovación educativa / sociedad de la información: gestión, medios y tecnología - sub - línea ingeniería y desarrollo de software / sociedad de la información: gestión, medios y tecnología - sub - línea análisis, procesamiento y modelamiento de datos- ingeniería en sistemas 1S2021**, de conformidad con el Art. 114 del Código Orgánico de la Economía Social de los Conocimientos, Creatividad e Innovación, concedo a favor de la Universidad Estatal de Milagro una licencia gratuita, intransferible y no exclusiva para el uso no comercial de la obra, con fines estrictamente académicos. Conservo a mi favor todos los derechos de autor sobre la obra, establecidos en la normativa citada.

Así mismo, autorizo a la Universidad Estatal de Milagro para que realice la digitalización y publicación de este trabajo de integración curricular en el repositorio virtual, de conformidad a lo dispuesto en el Art. 144 de la Ley Orgánica de Educación Superior.

El autor declara que la obra objeto de la presente autorización es original en su forma de expresión y no infringe el derecho de autor de terceros, asumiendo la responsabilidad por cualquier reclamación que pudiera presentarse por esta causa y liberando a la Universidad de toda responsabilidad.

Milagro, Haga clic aquí para escribir una fecha.

Jiménez Estrada Vicente Manuel

CI: 0926300583

DERECHOS DE AUTOR

Ingeniero.

Fabricio Guevara Viejó, PhD.

RECTOR

Universidad Estatal de Milagro

Presente.

Yo, Gomez Herrera George Ezequiel, en calidad de autor y titular de los derechos morales y patrimoniales del trabajo de integración curricular, modalidad virtual, mediante el presente documento, libre y voluntariamente procedo a hacer entrega de la Cesión de Derecho del Autor, como requisito previo para la obtención de mi Título de Grado, como aporte a la Línea de Investigación **Educación, cultura, tecnología en innovación para la sociedad - sub - línea tecnología e innovación educativa / sociedad de la información: gestión, medios y tecnología - sub - línea ingeniería y desarrollo de software / sociedad de la información: gestión, medios y tecnología - sub - línea análisis, procesamiento y modelamiento de datos- ingeniería en sistemas 1S2021**, de conformidad con el Art. 114 del Código Orgánico de la Economía Social de los Conocimientos, Creatividad e Innovación, concedo a favor de la Universidad Estatal de Milagro una licencia gratuita, intransferible y no exclusiva para el uso no comercial de la obra, con fines estrictamente académicos. Conservo a mi favor todos los derechos de autor sobre la obra, establecidos en la normativa citada.

Así mismo, autorizo a la Universidad Estatal de Milagro para que realice la digitalización y publicación de este trabajo de integración curricular en el repositorio virtual, de conformidad a lo dispuesto en el Art. 144 de la Ley Orgánica de Educación Superior.

El autor declara que la obra objeto de la presente autorización es original en su forma de expresión y no infringe el derecho de autor de terceros, asumiendo la responsabilidad por cualquier reclamación que pudiera presentarse por esta causa y liberando a la Universidad de toda responsabilidad.

Milagro, Haga clic aquí para escribir una fecha.

Gomez Herrera George Ezequiel

CI: 0942491747

APROBACIÓN DEL TUTOR DEL TRABAJO DE TITULACIÓN

Yo, **Ing. León Granizo Oscar Darío** en mi calidad de tutor del trabajo de titulación, elaborado por el estudiante Jiménez Estrada Vicente Manuel. y Gomez Herrera George Ezequiel, cuyo título es **propuesta de desarrollo de un sistema de predicción de ventas usando algoritmos de inteligencia artificial**, que aporta a la Línea de Investigación **Educación, cultura, tecnología en innovación para la sociedad - sub - línea tecnología e innovación educativa / sociedad de la información: gestión, medios y tecnología - sub - línea ingeniería y desarrollo de software / sociedad de la información: gestión, medios y tecnología - sub - línea análisis, procesamiento y modelamiento de datos-ingeniería en sistemas 1S2021** previo a la obtención del Título de Grado ingeniero en sistemas; considero que el mismo reúne los requisitos y méritos necesarios en el campo metodológico y epistemológico, para ser sometido a la evaluación por parte del tribunal calificador que se designe, por lo que lo APRUEBO, a fin de que el trabajo sea habilitado para continuar con el proceso previa culminación de Trabajo de Elija un elemento de la Universidad Estatal de Milagro.

Milagro, Haga clic aquí para escribir una fecha.

Ing. León Granizo Oscar Darío

C.I: 0928368513

APROBACIÓN DEL TRIBUNAL CALIFICADOR

El tribunal calificador constituido por:

Elija un elemento. Haga clic aquí para escribir apellidos y nombres (tutor).

Elija un elemento. Haga clic aquí para escribir apellidos y nombres (Secretario/a).

Elija un elemento. Haga clic aquí para escribir apellidos y nombres (integrante).

Luego de realizar la revisión del Trabajo de Elija un elemento, previo a la obtención del título (o grado académico) de Elija un elemento. presentado por Elija un elemento. Haga clic aquí para escribir apellidos y nombres (estudiante1).

Con el tema de trabajo de Elija un elemento: Haga clic aquí para escribir el tema del Trabajo.

Otorga al presente Trabajo de Elija un elemento, las siguientes calificaciones:

Trabajo de Integración Curricular	[]
Defensa oral	[]
Total	[]

Emite el siguiente veredicto: (aprobado/reprobado) _____

Fecha: Haga clic aquí para escribir una fecha.

Para constancia de lo actuado firman:

	Nombres y Apellidos				Firma
Presidente	Apellidos	y	nombres	de	_____
	Presidente.				
Secretario /a	Apellidos	y	nombres	de	_____
	Secretario				

Integrante Apellidos y nombres de _____
Integrante.

APROBACIÓN DEL TRIBUNAL CALIFICADOR

El tribunal calificador constituido por:

Elija un elemento. Haga clic aquí para escribir apellidos y nombres (tutor).

Elija un elemento. Haga clic aquí para escribir apellidos y nombres (Secretario/a).

Elija un elemento. Haga clic aquí para escribir apellidos y nombres (integrante).

Luego de realizar la revisión del Trabajo de Elija un elemento, previo a la obtención del título (o grado académico) de Elija un elemento. presentado por Elija un elemento. Haga clic aquí para escribir apellidos y nombres (estudiante2).

Con el tema de trabajo de Elija un elemento: Haga clic aquí para escribir el tema del Trabajo.

Otorga al presente Trabajo de Elija un elemento, las siguientes calificaciones:

Trabajo de Integración Curricular	[]
Defensa oral	[]
Total	[]

Emite el siguiente veredicto: (aprobado/reprobado) _____

Fecha: Haga clic aquí para escribir una fecha.

Para constancia de lo actuado firman:

Nombres y Apellidos

Firma

Presidente Apellidos y nombres de _____
Presidente.

Secretario /a Apellidos y nombres de _____
Secretario

Integrante Apellidos y nombres de _____
Integrante.

DEDICATORIA

Dedico esta tesis a Dios quien es el proveedor de toda sabiduría e inteligencia, a mis padres que frente a diversas dificultades siempre me motivaron a seguir adelante, con buenos valores y principios, a mi hermano mayor que es un motivo de superación, a mis compañeros y amigos que a lo largo de mi carrera universitaria me apoyaron y superamos en toda circunstancia, a mi pastor y hermanos en Cristo por sus oraciones y finalmente pero no menos importante a mi esposa Mayra Naranjo, que me ha demostrado su amor y lealtad.

George Ezequiel Gómez Herrera

Dedico esta tesis en primer lugar a Jehová quien me brindó la sabiduría y fortaleza en todo mi proceso de formación profesional.

A mis padres Vicente Jiménez y Lourdes Estrada por ser mi mayor motivación y ejemplo para seguir.

A mis hermanas Fernanda y Eleana Jiménez Estrada que me brindaron su apoyo cuando lo necesité.

En especial a mi novia Ing. Mariela Carrera quien creyó en mí y me inculcó el compromiso, la responsabilidad y las ganas de ser alguien en la vida.

Vicente Manuel Jiménez Estrada

AGRADECIMIENTO

De manera grata elevo mi agradecimiento a la Universidad Estatal de Milagro, a la Facultad de Ciencias e Ingeniería, a mis docentes que nombrarlos uno a uno no terminaría, a todos desde admisión hasta el último semestre que me inculcaron sus enseñanzas y conocimientos, de igual manera extendiendo mi agradecimiento a nuestro tutor el Ingeniero Oscar León que nos acompañó e instruyó en la elaboración y desarrollo de este trabajo, además extender un agradecimiento al MSC. Jhonny Darwin Ortiz Mata por sus observaciones y guía en el trabajo final. También agradezco a mi compañero de tesis que, más que compañero, un amigo.

George Ezequiel Gómez Herrera

Mi agradecimiento a mis familiares quienes me ofrecieron su ayuda cuando lo necesite, a mi tutor el Ing. Oscar León quien nos brindó todos sus conocimientos y acompañamiento en todo el proceso de realización del proyecto propuesto, a mis docentes porque fueron parte fundamental en mi proceso de formación y agradecer a mi compañero de tesis quien trabajó conmigo desde que iniciamos en el camino de ser Ingenieros.

Vicente Manuel Jiménez Estrada

ÍNDICE GENERAL

DERECHOS DE AUTOR	III
APROBACIÓN DEL TUTOR DEL TRABAJO DE TITULACIÓN	IV
APROBACIÓN DEL TRIBUNAL CALIFICADOR	V
DEDICATORIA	VIII
AGRADECIMIENTO	IX
ÍNDICE GENERAL	X
ÍNDICE DE FIGURAS	XII
INDÍCE DE TABLAS	XIII
RESUMEN	1
ABSTRACT	2
CAPÍTULO 1	1
1. INTRODUCCIÓN	1
1.1 Planteamiento del problema	2
1.2 Objetivos	4
1.2.1 Objetivo General	4
1.2.2 Objetivos Específicos	4
1.3 Alcance	4
1.4 Estado del arte	4
1.4.1 Antecedentes del estudio	4
1.4.2 Fundamentación Teórica	7
1.4.2.1 Algoritmos	7
1.4.2.2 Pronóstico de venta	8
1.4.2.3 Algoritmo de predicción	9
1.4.2.4 Tipos de algoritmos de predicción	10
1.4.2.5 Definiciones	16
CAPÍTULO 2	18
2. METODOLOGÍA	18
2.1 Metodología de Investigación	18
2.2 Metodología de Desarrollo	18
2.2.1 SCRUM	18
2.2.2 Fases de la metodología Scrum	19
2.2.3 Especificaciones generales de la Metodología Scrum	19
2.2.4 Equipo Scrum (Roles)	20
2.3 Historias de usuario	21
	X

2.4 SPRINT	27
CAPÍTULO 3	38
3.1 Análisis de Factibilidad	38
3.1.1 Factibilidad Técnica	38
3.2 Propuesta	40
3.2.1 Título	40
3.2.2 Base de modelo de datos	40
3.2.3 Descripción de la data set y los algoritmos	40
3.2.4 Tabla de Resultados	42
3.2.4.1 Resultado de regresión lineal	42
3.2.4.2 Resultados de Árbol de decisión	43
3.2.4.3 Resultados de Redes neuronales	43
3.3 Funcionalidades del Sistema	44
CONCLUSIÓN	51
RECOMENDACIONES	52
BIBLIOGRAFÍA	53
ANEXOS	58

ÍNDICE DE FIGURAS

Figura 1 Pasos para identificar un problema	8
Figura 2 Uso de las aplicaciones de modelos predictivos	10
Figura 3 Tipos de algoritmos	10
Figura 4 Fases de la metodología Scrum	Error! Bookmark not defined.
Figura 5 Reporte Sprint 1	33
Figura 6 Reporte Sprint 2	33
Figura 7 Sprint 3	34
Figura 8 Dataset description	38
Figura 9 Validación del algoritmo de Redes Neuronales	39
Figura 10 Gráfica de Regresión Lineal	40
Figura 11 Modelo final del algoritmo Árbol de Decisión	41
Figura 12 Loguin	Error! Bookmark not defined.
Figura 13 Panel de administración	43
Figura 14 Módulo usuario	43
Figura 15 Nuevo usuario.	44
Figura 16 Módulo ventas	44
Figura 17 Módulo de nueva venta	45
Figura 18 Módulo clientes	45
Figura 19 Nuevo cliente	46
Figura 20 Módulo predicción	46
Figura 21 Carga de datos.	47
Figura 22 Entrenamiento de datos	47
Figura 23 Resultados	48

INDÍCE DE TABLAS

Tabla 1.....	3
Tabla 2.....	20
Tabla 3.....	21
Tabla 4.....	22
Tabla 5.....	23
Tabla 6.....	24
Tabla 7.....	25
Tabla 8.....	26
Tabla 9.....	27
Tabla 10.....	28
Tabla 11.....	29
Tabla 12.....	29
Tabla 13.....	30
Tabla 14.....	31
Tabla 15.....	31
Tabla 16.....	33
Tabla 17.....	34
Tabla 18.....	34
Tabla 19.....	35
Tabla 20.....	38
Tabla 21.....	38
Tabla 22.....	39
Tabla 23.....	39
Tabla 24.....	39
Tabla 25.....	40
Tabla 26.....	42
Tabla 27.....	42
Tabla 28.....	43
Tabla 29.....	43
Tabla 30.....	43
Tabla 31.....	44

Título de trabajo de Integración Curricular: PROPUESTA PARA EL DESARROLLO DE UN SISTEMA DE PREVISIÓN DE VENTAS MEDIANTE ALGORITMOS DE INTELIGENCIA ARTIFICIAL.

RESUMEN

El desarrollo del presente trabajo se centró en proponer el desarrollo de un sistema de predicción de ventas, utilizando algoritmos de inteligencia artificial, para lo cual se recopiló información necesaria para identificar y analizar los datos, se identificó los algoritmos, las técnicas, métodos y herramientas involucradas en el desarrollo del sistema de predicción, luego se seleccionó los algoritmos de inteligencia artificial, para entrenar y predecir los datos, también se determinó el algoritmo que presente mejores resultados y finalmente se propuso un sistema de predicción de ventas. El estado del arte se fundamentó con trabajos de temas similares al propuesto provenientes de citas y referencias bibliográficas de tesis, artículos y revistas científicas. Dentro de la metodología se desarrolló Scrum y todos sus componentes. Finalmente la propuesta de solución estuvo encaminada en primera instancia con el análisis de factibilidad, la selección del algoritmo con el cual se trabajó las predicciones, en este caso, Redes Neuronales, porque presentó los mejores resultados de predicción, utilizando técnicas, métodos y herramientas aplicadas con Scrum, JIRA, HTML, Google, Bootstrap, entre otras, para el desarrollo del sistema de predicción lo cual sirvió para desarrollar el sistema de predicción de ventas que sirvan para una adecuada toma de decisiones.

PALABRAS CLAVE: Algoritmos, Datos, Predicción, Sistema, Ventas.

Curriculum Integration Job Title: PROPOSAL FOR THE DEVELOPMENT OF A SALES FORECASTING SYSTEM USING ARTIFICIAL INTELLIGENCE ALGORITHMS.

ABSTRACT

The development of this work is focused on proposing the development of a sales prediction system, using artificial intelligence algorithms, for which the necessary information was collected to identify and analyze the data, the algorithms, techniques, methods and tools involved in the development of the prediction system were identified, then the artificial intelligence algorithms were selected to train and predict the data, the algorithm with the best results was also determined, and finally a sales prediction system was proposed. The state of the art was based on works on topics similar to the proposed one, coming from citations and bibliographic references from theses, articles and scientific journals. Within the methodology, Scrum and all its components were developed. Finally, the proposed solution was directed in the first instance with the feasibility analysis, the selection of the algorithm with which the predictions were worked, in this case, Neural Networks, because it presented the best prediction results, using techniques, methods and tools applied with Scrum, JIRA, HTML, Google, Bootstrap, among others, for the development of the prediction system which served to develop the sales prediction system that serve for proper decision making.

KEY WORDS: Algorithms, Data, Prediction, Sales, System.

CAPÍTULO 1

1. INTRODUCCIÓN

El desarrollo del trabajo se ha centrado en la propuesta de un sistema de producción de ventas usando algoritmos de inteligencia artificial, la idea del proyecto es brindar a los pequeños y medianos empresarios alternativas para mejorar la gestión de sus actividades comerciales en cuanto a las ventas, para lo cual se consideró una base de datos de empresas para ejecutar el trabajo propuesto.

La predicción de las ventas es una herramienta importante para los empresarios y personal de mantenga un alto nivel de productividad, a través de un adecuado manejo de información precisa obtenido de datos históricos de ventas, de esta manera se predice de manera efectiva los ingresos reales y alcanzables que se obtienen de una constante actualización de datos pasados y presentes.

Dentro de este contexto, el trabajo se ha desarrollado en tres capítulos, los cuales se describirán a continuación: El capítulo I, se centró en desarrollar toda la problemática en base al tema propuesto, para lo cual se estableció los objetivos investigativos, los cuales se cumplieron, también se elaboró el marco teórico fundamentando cada una de las variables con las respectivas citas y referencias bibliográficas.

Capítulo II, se desarrolló la metodología del trabajo, en este caso se empleó la Scrum como parte de desarrollo de un software ágil, la cual cuenta con cuatro fases. Se definieron los Sprints, luego los roles, historias de usuario, estimaciones y finalmente se comenzó el desarrollo del proyecto.

En el capítulo III se definió la propuesta, el diseño de la propuesta, los costos que incurren en el desarrollo, demostrando la viabilidad del trabajo propuesto.

1.1 Planteamiento del problema

Con el paso del tiempo para dar una adecuada administración dentro de las empresas comerciales uno de los aspectos más importantes es la planeación debido a su efectividad en el desarrollo a futuro de las empresas, es decir, son parte fundamental del éxito a corto y largo plazo. Las competencias administrativas y operativas deben ser planificadas y organizadas para anticiparse al futuro con estrategias potenciales que aseguren la permanencia de las organizaciones en el mercado. Conocer plenamente los aspectos económicos permite tomar las mejores decisiones al administrador, puesto que no es una tarea fácil administrar una unidad de negocio para que las ventas proyectadas se cumplan en un espacio de tiempo.

Hace una década las empresas enfocan sus proyecciones en métodos tradicionales, bajo análisis de ventas anuales de periodos anteriores, para las empresas comerciales la planificación es uno de los puntos más importantes a tratar antes de tomar decisiones sobre su futuro. Las grandes empresas dedicadas al comercio de venta masivas o Retails se mueven bajo el histórico de sus ventas, algo que hoy en día puede ser apoyado con diferentes herramientas de análisis de datos y proyecciones, de esta manera establecer un pronóstico de ventas a futuro.

Muchas empresas carecen de tecnología y por ende de un software que procese información histórica para pronosticar ventas que presenten resultados seguros, disminuyendo el nivel de incertidumbre de las empresas al momento de tomar alguna decisión. Los mercados en la actualidad son muy competitivos, el tomar una decisión errónea puede generar pérdidas económicas en las empresas. Motivo por el cual las organizaciones buscan métodos y técnicas eficientes para mantenerse exitosamente dentro del mercado. Es así como la utilización de algoritmos para la proyección de ventas se ha convertido en una de las herramientas dentro de la planificación empresarial.

Cabe mencionar que los métodos utilizados en la predicción de ventas se enfocan en ajustar la información procesada con la realidad. Dentro de este contexto el presente estudio se enfoca en analizar varios algoritmos para realizar un pronóstico de ventas.

En el Ecuador existen muchas empresas que desconocen la importancia de contar con un sistema o herramientas de proyección, las mismas que le dan un estimado más certero para la toma de decisiones, puntos claves para dirigir la empresa a lograr sus objetivos a corto, mediano y largo plazo.

La (Revista Lideres, 2019)menciona que: las ventas en el año 2013 de las empresas de retail llegaron a 17,727 millones de dólares que con respecto al año 2018 lograron obtener un incremento de 1,3% de crecimiento llegando a los 16,863 millones de dólares, lo que conlleva a destacar su impacto de crecimiento y evolución en el país.

Tabla 1

Causas y Consecuencias

Causas	Consecuencias
C1. Proyecciones en métodos tradicionales	E1. Bajo análisis de ventas anuales de periodos anteriores
C2. Desconocimiento de un sistema o herramientas de proyección de ventas	E2. Pérdida a nivel competitivo ante otras empresas
C3. La empresa no cuenta con los recursos económicos	E3. Manejo empírico del historial de las ventas
C4. Tecnología obsoleta	E4. Escaso control de inventario
C5. Falta de planificación	E5. Bajos rendimientos económicos

Nota: Definición de causas y consecuencias, fuente elaboración propia.

1.2 Objetivos

1.2.1 Objetivo General

Proponer el desarrollo de un sistema de predicción de ventas, utilizando algoritmos de inteligencia artificial.

1.2.2 Objetivos Específicos

- Recopilar la información necesaria para identificar y analizar los datos.
- Identificar los algoritmos, las técnicas, métodos y herramientas involucradas en el desarrollo del sistema de predicción.
- Seleccionar los algoritmos de inteligencia artificial, para entrenar y predecir los datos.
- Determinar el algoritmo que presente mejores resultados.
- Proponer un sistema de predicción de ventas.

1.3 Alcance

Se trabajará con un Dataset de una empresa dedicada a ventas de productos masivos obtenida de Kaggle (<https://www.kaggle.com/>), de las cuales se obtendrá información sobre una base de datos de las ventas históricas que ha tenido la empresa, y se realizará la comparación de tres algoritmos de predicción de inteligencia artificial, seleccionando el algoritmo que presente los mejores resultados para poder realizar una propuesta de desarrollo de un sistema de predicción de ventas.

1.4 Estado del arte

1.4.1 Antecedentes del estudio

De acuerdo con un estudio realizado por (Molina, 2020) realizó un trabajo enfocado a la implementación de un modelo analítico, para gestionar la predicción de ventas de un grupo de productos OTC de un establecimiento farmacéutico. El trabajo se

centró en utilizar el mejor modelo utilizando la metodología de investigación conocida como Design Science Research, además de la aplicación de la metodología KDD. Dentro del proceso de construcción de la metodología se consideró oportuno los modelos ARIMAX, Holt Winters y Redes Neuronales, empleando el método de regresión y series temporales, debido a que se utilizó como información datos históricos de las ventas de las subcategorías de los productos. La comparación de los pronósticos de los modelos aplicados demostró una precisión menor al 10%, resultados que fue considerado aceptable debido a su contribución en la reducción del error que el laboratorio maneja, es decir, un 15% en promedio de las unidades proyectadas, por lo tanto, la construcción de los modelos se ajusta a las ventas reales dentro de un tiempo pronosticado, dicho en otras el proyecto aplicado fue factible.

De la misma manera un trabajo realizado por (Robles & Valverde, 2021), el cual se enfocó en la predicción de las ventas, pues la finalidad era incrementarlas dentro del mercado de accesorios y repuestos automotrices, el objetivo se centró en obtener un stock estratégico para potencializar las ventas y a su vez satisfacer las necesidades de los clientes, por lo tanto, aplicaron el Machine Learning y las conocidas Redes Neuronales Recurrentes. También integraron en el trabajo investigativo un sistema web que fue desarrollado con un software externo de logística de Open Source llamado Tryton la selección de este se debió a las diversas funcionalidades de logística que manejaba. Con todo lo realizado pudieron incrementar las ventas de forma eficiente manteniendo una adecuada gestión en la logística de operaciones.

(Bozanic, 2020) centró su conocimiento en establecer un sistema de predicción de precios de venta para una empresa inmersa en el sector inmobiliario en Chile, utilizando algoritmos de Machine Learning. Ante las bajas en las ventas con este estudio buscó reducir los errores de pronósticos de predicción de precios de venta en cada uno de sus

portafolios de inmuebles y operaciones inmobiliarias que se encuentran en competencia en el mercado. Se efectuó una selección de los mejores algoritmos y de esta manera manejar una información que sirva de gran ayuda en el momento de abordar a los clientes interesados en algún inmueble. Esto ha sido posible gracias a la utilización del algoritmo Rendón Forest, debido a la efectividad que ha demostrado ante el que manejaba la empresa en aquel entonces. La información analizada demuestra que las empresas en la actualidad sin importar a la actividad económica que se dedican están buscando métodos para mejorar sus ventas y de esta manera mantener un posicionamiento en el mercado, siendo así el uso de algoritmos como medio para brindarle a las empresas los mejores resultados de predicción.

De la misma manera (Fosca, 2020) también se enfocó en la predicción del precio utilizando las herramientas de Machine Learning de inteligencia artificial a beneficio del grupo de investigación en finanzas aplicadas de la PUCP. El objetivo parte de los conocimientos interdisciplinarios para explorar metodologías de inteligencia artificial dentro del campo de la inversión financiera. Para un mejor entendimiento de los procesos a llevar se presenta un flujograma como pilar para la elaboración de un algoritmo, con la finalidad de estructuras modelos de regresión lineal y SVR, considerando dentro de la categorización el problema central hasta la validación, para finalmente hacer una evaluación de la importancia de la selección en cuanto a los atributos del modelo y así usar algoritmos eficiencias que permitan la comparación de resultados efectivos que ayuden al grupo empresarial tomar las mejores decisiones en cuanto a su actividad económica.

De acuerdo con (Jaramillo, 2020) realizó un estudio sobre la gestión de pedidos para puntos de ventas con aprendizaje automático, con la finalidad de automatizar los pedidos basados en las demandas de los productos que se ofrece en el mercado y así estimar una eficiente rotación de inventarios en los puntos de ventas, con el objetivo de

mantener el control de los inventarios y así evitar el exceso de productos. Para lograr resultados concretos se empleó modelos estadísticos conocidos como Regresión Lineal, Seasonat Arima y Holt-Winters además de los modelos Redes Neuronales Recurrentes y Deep Learning Perceptrón Multicapa, en cambio para la evaluación de los resultados se utilizó las métricas MAE, MSE y R2. La utilización de todos los modelos mencionados hizo posible mantener el control de los inventarios.

Un estudio realizado por (Yosvani, Rivas, Pérez, & Marrero, 2017) tuvo como objetivo realizar un pronóstico de la demanda utilizando Redes Neuronales Artificiales, cabe mencionar que dentro de la planificación de toda empresa los pronósticos son una herramienta importante para establecer los pasos correctos que conlleven al logro de los objetivos. La base de este estudio fue en crear un procedimiento para realizar el pronóstico de la demanda, para lo cual se puso en análisis el procedimiento que maneja la empresa Holguín, la cual se dedica a la comercialización y distribución de medicamentos. Dentro del proceso se escogió a cuatro medicamentos, de esta manera se pudo predecir cuales son los que mayor demanda tienen, lo cual fue de gran aporte dentro del proceso de planificación de la empresa, puesto que podrán realizar con mayor seguridad compras exitosas, evitando así el exceso de productos.

1.4.2 Fundamentación Teórica

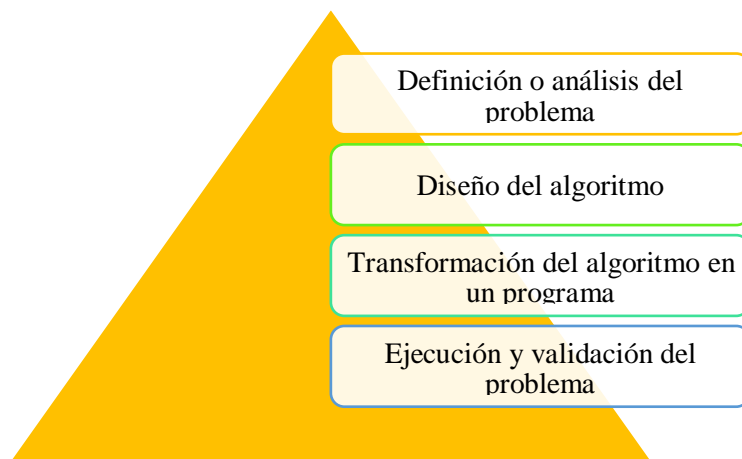
1.4.2.1 Algoritmos

Dentro de la rama informática los algoritmos se los define como una serie de pasos sistemáticos que sirven para dar soluciones inmediatas a un problema específico. Por lo general son muy utilizados por las grandes empresas que manejan enormes y variadas cantidades de productos, y que buscan determinar un stock efectivo para evitar el exceso de estos (Robledano, 2019).

Dentro del sector empresarial en los últimos tiempos los algoritmos se han convertido en una de las herramientas necesarios en el proceso de toma de decisiones para la ejecución de la predicción de ventas o demandas, ya que estas son las que mueven los mercados y la necesidad de las empresas mantener una participación sostenible en el mercado hace que los empresarios visionarios utilicen estos algoritmos.

Dentro del proceso de resolución de un problema es importante considerar los siguientes pasos:

Figura 1 Pasos para identificar un problema



Nota: La figura 1 demuestra los pasos que se deben seguir para identificar un problema a través de algoritmos (Molina, 2020).

1.4.2.2 Pronóstico de venta

Es una técnica efectiva que hace posible el cálculo de las proyecciones de ventas de una forma eficiente y eficaz, utilizando para ello, una fuente de datos que puede ser obtenida de transacciones de inventarios o de las facturas de las ventas realizadas en un periodo de tiempo. Otro aspecto importante del pronóstico de venta es que le permite al empresario visionar el futuro basándose en datos históricos generados por la comercialización de uno o varios productos (Azañero & Ramírez, 2019).

Dentro del proceso de predicción de ventas existen categorías que intervienen, tales como, encuestas, criterio ejecutivo, análisis de regresión y pruebas de mercado y análisis de series de tiempo. La elección del método o métodos varía según los costos, la confiabilidad, propósito del pronóstico, la seguridad de los datos históricos de ventas, el tipo de producto, el tiempo, características del mercado y pericias de los responsables de realizar el pronóstico.

1.4.2.3 Algoritmo de predicción

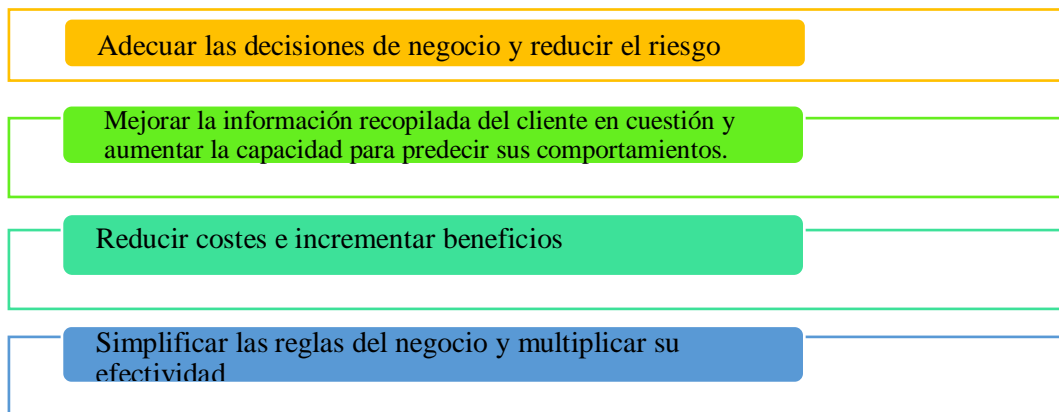
Son un conjunto de técnicas que, a través del entorno del aprendizaje automático, el Big Data, recolección de datos relevantes e identificación de patrones, se pretende presentar una predicción de resultados a futuro de productos que se encuentran dentro de un stock de venta en un mercado específico, para ser de soporte en la toma de decisiones en las etapas de planificación de una empresa. Desde hace algunos años el campo de la predicción automática ha sido de gran aporte en el crecimiento de los negocios sin importar la actividad económica a la cual se dedican.

De acuerdo con los que sostiene (García, y otros, 2017) Los algoritmos son un grupo de procesos ejecutados por medio de técnicas computacionales de análisis de datos que sirven de aporte en la deducción de eventos que determinen escenarios previos a su ejecución, además se brindan nuevas alternativas en la puesta en marcha de un negocio.

Los expertos en el manejo de datos o Data Scientists arriman esfuerzos en poner en funcionamiento dentro del campo de la Big Data a un nivel potencial el tratamiento de datos que permita a las empresas grandes, medianas o pequeñas a generar un valor agregado a través de la predicción en la toma de decisiones. Dentro de este contexto, los modelos predictivos son una herramienta potencial en cualquier sector empresarial. Uno de los ejemplos de mejor uso de estos modelos son el área del marketing en donde se busca

predecir el comportamiento de los clientes para evaluar de acuerdo con un margen de aciertos acciones efectivas en el mercado, por ello, en la actualidad los empresarios buscar asesoría de los entendidos en la materia, para utilizar los mejores métodos y algoritmos que les permita contar con una base de datos confiables para adentrarse más en el mercado. El uso de los modelos predictivos en el mercado empresarial se utiliza para:

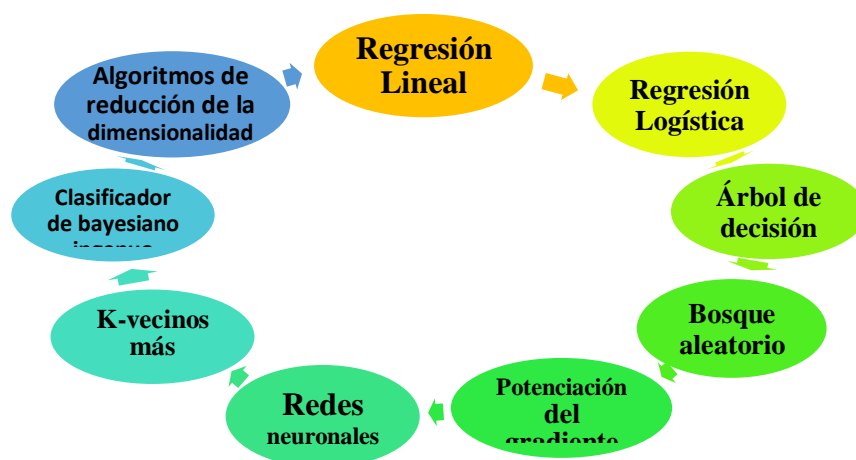
Figura 2 Uso de las aplicaciones de modelos predicción



Nota: En la Figura 2 se muestran las aplicaciones de los modelos predictivos utilizados en el mundo empresarial, (Castillo, 2017).

1.4.2.4 Tipos de algoritmos de predicción

Figura 3 Tipos de algoritmos



Nota: En la Figura 3 se presentan los diferentes tipos de algoritmos que se pueden utilizar en la predicción de ventas, (Sandoval, 2018).

Regresión Lineal

Específicamente se la conoce como una línea recta, en la cual demuestra claramente el ajuste de cada uno de los enlaces numéricos, así como el método de los mínimos cuadrados que se centra en calcular la suma total de los trayectos al cuadrado en puntos específicos que representan datos además de los puntos que se visualizan en la recta que brinda el modelo, con todo ello, el mejor resultado será el brinde los mínimos errores (Salmerón & Rodríguez, 2017).

Las ventajas de la regresión lineal se dan por su facilidad de entendimiento al visualizar la línea claramente, sin embargo, cuando se maneja valores extremos se debe aplicar el sobre ajuste, debido a su sensibilidad en la relación de datos complejos de las variables que se tomen para el análisis.

Regresión logística

Este tipo de regresión sirve como adaptación a regresión lineal de problemas específicos, para lo cual se hace necesario el uso del método de máxima probabilidad para identificar cual es el punto d mayor ocurrencia. Al igual que la regresión lineal es fácil de entender, pero a su vez presenta inconvenientes por su simplicidad, por lo tanto, están sujetos a sobreajuste (Pérez, 2017).

El árbol de decisión

Se muestra como un gráfico con ramificaciones centradas en construcciones lógicas, es decir, estos parten de un grupo compacto de datos y se dividen en varias ramas de acuerdo con una serie de condiciones que a medida que se analiza se seleccionan los aspectos importantes para resolver problemas. Son fáciles de entender y por ende de implementar, aunque no son fuertes para el manejo de datos complejos (Timarán, Caicedo, & Hidalgo, 2018).

Bosque aleatorio (Random Forest)

Este funciona por tomar la medida de varios árboles de decisión tomando en cuenta las muestras de datos. La separación de muestras hace posible analizar la suma de todos los árboles y así obtener buenos resultados, pero si se hace uno a uno pues su análisis sería débil. Entre sus características es que brindan modelos de alta calidad, pero se consideran de difícil entendimiento para las predicciones (Sandoval, 2018).

Potenciación del gradiente (Gradient Boosting)

Tiene similitud con el modelo bosque aleatorio, pero en este se toman árboles de decisión de mayor debilidad, a partir de ellos, optimiza cada muestra de datos en los procesos que va aplicando. Tiene un alto grado de rendimiento, pero hay que tener mucho cuidado con los cambios que se vayan a hacer porque pueden generar cambios radicales en el modelo, es decir, sería demasiado difícil entender las predicciones (Data Science, 2018).

Redes neuronales

Son conocidas porque imitan el comportamiento del cerebro humano, es decir, son módulos de neuronas que se encuentran interconectadas en diferentes capas que permiten el paso de mensajes de una neurona a otra. Estas redes neuronales son muy utilizadas cuando existe desconocimiento del origen de los datos de entrada como los de salida (Macea & Martelo, 2016).

Entre sus características más sobresalientes es que tienen la capacidad de reconocer imágenes, aunque con mucha lentitud, requieren que sean potencializados para mayor efectividad y en cuanto a sus resultados en materia de predicción con casi imposible de entender, por ello, es importante analizar bien los problemas que se desea resolver y dar un adecuado manejo de los datos.

K-vecinos más cercanos (k-NN o Nearest Neighbor)

Este tipo de algoritmo se caracteriza por funcionar de forma agrupada sin ninguna jerarquización, siendo uno de los más usados en el mundo empresarial. Su forma de procesamiento estadístico se centra en reconocer patrones calculando el trayecto de un dato específico con los vecinos que se encuentran más cerca al grupo objeto de estudio. Entre las dificultades que presenta, es cuando maneja muestras pequeñas, es decir, su mayor efectividad es con muestras grandes, por decir, pueden ser utilizadas y útiles para las grandes empresas multinacionales (Tripp, 2020).

Clasificador de bayesiano ingenuo (Naïve Bayes)

Este algoritmo tiene como característica dentro de su procedimiento la presencia, así como la ausencia de una particularidad que no se relaciona con la presencia o ausencia de otras características que forma parte de un grupo que favorece de manera autónoma a la posibilidad de una parte concreta del análisis. Dentro de sus ventajas es que solo requieren de poca cantidad de datos, pero frecuentemente no producen una acertada estimación de probabilidades correctas (Gutiérrez, Margain, Ramírez, & Canul, 2017).

Algoritmos de reducción de dimensionalidad

Esta clase de algoritmos son considerados como los más predictivos, debido a que pueden reducir el número de variables utilizadas en el análisis de los datos, por ello, pueden trabajar sin ningún problema con los algoritmos antes mencionados, sobre todo en conjuntos de datos de grandes dimensiones, dicho en otras palabras, puede ayudar alcanzar buenos resultados (Arroyo, 2016).

Sistema Web

Los sistemas web son conocidos por que pueden utilizarse a través de un servidor web teniendo como nexo necesario al internet a través de un navegador, el cual puede ser escogido por el usuario. En cuanto a las aplicaciones web en la actualidad son muy requeridas por usuarios de la web, por su practicidad y las facilidades del sistema operativo. El uso de estas aplicaciones les resulta a los usuarios un ahorro de dinero y tiempo, puesto que en él pueden encontrar y adquirir productos, servicios o hasta realizar transacciones diarias, sean estas personales como empresariales.

Además, son de gran entretenimiento, pues permite el manejo de programas, plataformas, entre otros, sobre todo lo más importante que el usuario puede ubicarse en cualquier lugar para trabajar desde él, siempre y cuando tenga conexión a internet (Maldonado, 2016).

Los sistemas web son tan flexibles que pueden utilizar el navegador de la preferencia del usuario, siendo estos Chrome, Firefox, Internet, Explorer, entre otros. Otro punto importante es que el usuario puede conectarse de cualquier servidor para acceder a las aplicaciones web, esto es posible porque estas trabajan con bases de datos que hacen posibles procesos y muestren cualquier variedad de información que sea de utilidad del usuario.

En cuanto al uso que el sector empresarial les da a las aplicaciones web, desde que la tecnología vino a evolucionar ha sido un aliado estratégico en las alianzas comerciales entre países, comercialización de productos y servicios en todo el mundo, en fin, las aplicaciones web traen muchos beneficios a las empresas como las personas en general (Rodríguez, 2018).

Arquitectura de Sistema Web

La arquitectura web dentro del campo de la tecnología se la conoce por servir como facilitador en la agrupación visual de cada uno de los contenidos de los sitios web, sea para las empresas y usuarios. Otro punto relevante es que permite la publicidad en la promoción de ventas y demás aspectos que facilitan la navegación del usuario, eh así la importancia de la arquitectura de sistemas en el diseño de páginas web en la cual se pueden observar un sin número de categorías que hacen posible el entendimiento en los usuarios. La arquitectura actual como la tradicional se han trabajado bajo una estructura de construcción, para brindarle al usuario un fácil manejo de las aplicaciones utilizando el navegador de su preferencia sin ninguna dificultad (Traña & Silva, 2016).

Desde el punto de vista empresarial las arquitecturas web tiene como característica el resolver necesidades de un negocio online, el cual puede ofertar un producto o servicio para satisfacer las necesidades, exigencias o expectativas de los clientes, garantizando la calidad, precio y entrega segura a través de la web. Otro aspecto importante es que se encuentran relacionados con el estudio de palabras claves, así como la búsqueda de una variedad de contenidos que están en web.

Arquitectura Cliente Servidor

Dentro de las aplicaciones web uno de los aspectos en los cuales se destacan es en la arquitectura cliente/servidor, el cual se lo conoce como; explorador, navegador o visualizador y por la otra parte está el servidor; que corresponde al servidor web. Existen un sin número de variables que se utilizan en la arquitectura básica de acuerdo con las funcionalidades que contenga el servidor (Lizama, Kindley, & Jeria, 2016)

Entre las funciones de los servidores web es que estos clasifican información formateada para aquellos clientes que los solicitan, tales requerimientos se los realiza por

medio de una conexión de red, utilizando el protocolo http. Cuando se hace la solicitud al protocolo http el servidor web lo recibe, éste localiza la página web en una base de archivos y envía los resultados al navegador donde se realizó la solicitud (Vivar & Quishpe, 2019).

Navegadores Web

- Por medio de un análisis interpreta los requerimientos del usuario y envía los resultados encontrados.

Servidor Web

- Visualización de las aplicaciones web
- Control de la presentación
- Relación recíproca con los servidores de datos

Servidores de datos

- Base de datos
- Ficheros
- Correo, entre otros

1.4.2.5 Definiciones

Dataset. – Es un conjunto de datos que pasan por un proceso de tabulación en cualquier sistema de almacenamiento en donde los datos se encuentran estructurados, es decir, existe una única base central de datos, la misma que puede relacionarse con otras bases de información. Dataset tiene en cada una de sus columnas unas variables y en cada fila de los datos que se estén analizando (Balagueró, 2018).

IA. – La inteligencia artificial no es otra cosa que la mixtura de algoritmos establecidos con la finalidad de crear máquinas que se asemejen a las capacidades que tiene

el ser humano. Esta es una teoría que hasta la actualidad es considerada por los expertos como lejana y sobre todo misteriosa, pero que desde hace varios años se ha venido estudiando para conseguir resultados que puedan beneficiar a las personas en general (Martín & Carbonell, 2016).

Machine Learning. – Considerada una disciplina dentro del campo de la inteligencia artificial y que por medio de algoritmos brinda a los ordenadores la capacidad de identificar patrones dentro de una gran cantidad de datos que pueden ser de gran utilidad en la predicción. Este análisis hace posible que las computadoras realicen tareas Medidas de una manera autónoma sin tener la necesidad de programarlas (Fontalvo, 2019).

Python. – Su lenguaje de programación hace posible la interpretación y orientación que el PERL, el cual cuenta con un alto grado de confiabilidad por su claridad y legibilidad. Expertos sostienen que es fácil de aprender y es portátil, es decir, sus enunciaciones pueden ser interpretadas en diversos sistemas operativos, así como en sistemas basados en Mac, Unix, OS/2, MS-DOS y muchas más versiones de Microsoft Windows 98 (TechTarget, 2018).

Lenguaje de programación. – Lenguaje que se puede interpretar a través de un sistema informático utilizando un programa colector o a su vez por un traductor de instrucciones del procesador; que no es otra cosa que el lenguaje de la máquina que se utiliza en el proceso (Fondo Social Europeo, 2019).

Http. – Son siglas en inglés que son muy conocidas por los usuarios: Hypertext Transfer Protocol, este es el nombre del protocolo con el cual se hace posible la realización de las solicitudes de datos e informaciones y recursos, como documentos de html (Developer, 2021).

CAPÍTULO 2

2. METODOLOGÍA

2.1 Metodología de Investigación

La metodología de la investigación facilita el trabajo del investigador en el desarrollo de las diversas actividades que conllevan al logro de los objetivos establecidos. Por lo tanto, se considera un instrumento metodológico de gran aporte en la identificación de las necesidades que conciernen en este trabajo. Es así como se efectuó un estudio descriptivo sobre los sistemas de predicción de ventas y el uso de algoritmos para la comparación de datos más eficientes que se utilizarán en la propuesta.

Cabe mencionar que en la descripción se detallan las características fundamentales del tema planteado en relación con estudios realizados por otros autores, en los cuales se analizarán los criterios manejados sobre los sistemas de predicción.

La importancia de los estudios descriptivos se centra en especificar las propiedades, características y lineamientos del tema que es sujeto de análisis, es decir, se enfocan en la recolección de información de conceptualizaciones de las variables que intervienen en este trabajo y la relación de estas para que el investigador haga un análisis profundo y concreto de lo que medirá.

2.2 Metodología de Desarrollo

2.2.1 SCRUM

Entre las metodologías de desarrollo ágiles encontramos a Scrum, una herramienta que ha trascendido a lo largo de su aparición, una publicación de Ken Schwaber y Mike Beedle en el año 2001 menciona a Scrum como parte de desarrollo de software ágil. La base principal del método se basa en la cohesión del trabajo en equipo para un eficaz desarrollo.

El equipo dentro de Scrum se ocupan de alcanzar una misma meta, es un equipo autosuficiente, ofrece resultados de calidad a través de interacciones mediante todos los involucrados visualizan el avance, además de estar preparados para los nuevos requisitos que se presenten, esta es una enorme ventaja ante otros métodos de desarrollo de software ya que, dentro de la vida real existen parámetro imprevistos y variables externas que ajustan los requerimientos iniciales, Scrum en cada iteración logra adaptarse y presentar un avance (Mariño & Alfonzo, 2018).

Scrum se lleva a cabo con los llamados sprint, que son actividades en un determinado orden que se ejecutan según su prioridad. Como ya se mencionó en el concepto de metodología ágiles, Scrum realiza entregables funcionales de manera regular del producto, en cada uno de ellos el cliente participa para mejorar, aceptar y/o corregir según sus requerimientos cada sprint terminado.

2.2.2 Fases de la metodología Scrum

- Planificación y estimación
- Implementación
- Revisión y retrospectiva
- Lanzamiento

2.2.3 Especificaciones generales de la Metodología Scrum

Dentro de este punto se realizará la descripción de las especificaciones que formarán parte del presente proyecto, para efectivizar la aplicación de la metodología mencionada, la misma que tiene como objetivo entregar valores en cortos periodos de tiempo, basándose en tres aspectos importantes, como la transparencia, inspección y adaptación. En cuanto a las definiciones que corresponden a la metodología mencionada están las siguientes:

- Distribución adecuada del equipo Scrum según sus roles.

- Elaboración de los Sprint.

Las especificaciones para la aplicación SCRUM nos ayudará definir los sprint, roles y conformación de equipos, tomando decisiones que permitirán definir tareas que serán realizadas por cada uno de los miembros con el fin de proyectar los mejores resultados y un equipo consolidado.

2.2.4 Equipo Scrum (Roles)

Dentro del proceso de aplicación de la metodología Scrum se hace necesario establecer los roles competentes para la intervención y realización del proyecto propuesto. Dentro de los roles establecidos se detallarán a las personas responsables dentro del equipo Scrum: El equipo SCRUM estará definido por tres roles, los cuales se establecen en la siguiente tabla. Este permite tener una mayor comunicación de las tareas y los requerimientos que desea el usuario.

Tabla 2

Roles

ROL	NOMBRES
PRODUCT OWNER	Empresas Retails
SCRUM MASTER	Ing. Óscar León
DEVELOPMENT TEAM	George Gómez Vicente Jiménez

Nota: Definición de Roles, elaboración propia

2.3 Historias de usuario

El análisis realizado para cada requerimiento del Product Backlog, así como las reuniones realizadas con los usuarios permitieron elaborar las historias de usuarios que se presentan a continuación.

Tabla 3

Historia de Usuario N° 1

HISTORIA DE USUARIO			
Número	1	Usuario	Empresas Retails
Nombre Historia	Levantamiento de requerimientos		
Programador Responsable	Vicente Jiménez, George Gómez		
Prioridad en Negocio (Alta, Media, Baja)	Alta	Riesgo en Desarrollo (Alta, Media, Baja)	Media
Estimación	15	N° Sprint asignado	1
Descripción			
<ul style="list-style-type: none"> • Como usuario deseo que se realicen consultas e investigaciones para levantar los requerimientos necesarios para el proyecto. 			
Criterios de aceptación			
<ul style="list-style-type: none"> • Realización de reuniones e investigaciones acerca del proyecto. • Análisis de procesos de previsión de ventas existentes. • Se deben haber levantado todos los requerimientos del proyecto. 			

Nota: Historia de usuario 1, descripción y criterios de aceptación, elaboración propia.

Tabla 4

Historia de Usuario N° 2

HISTORIA DE USUARIO			
Número	2	Usuario	Empresas Retails
Nombre Historia	Búsqueda de los algoritmos		
Programador Responsable	Vicente Jiménez, George Gómez		
Prioridad en Negocio (Alta, Media, Baja)	Media	Riesgo en Desarrollo (Alta, Media, Baja)	Alta
Estimación	15	N° Sprint asignado	1
Descripción			
<ul style="list-style-type: none"> • Como usuario deseo que se realicen consultas sobre algoritmos orientados a la predicción de ventas. 			
Criterios de aceptación			
<ul style="list-style-type: none"> • Identificar los algoritmos utilizados para las predicciones. 			

Nota: Historia de usuario 2, descripción y criterios de aceptación, elaboración propia.

Tabla 5

Historia de Usuario N° 3

HISTORIA DE USUARIO			
Número	3	Usuario	Empresas Retails
Nombre Historia	Selección de los algoritmos		
Programador Responsable	Vicente Jiménez, George Gómez		
Prioridad en Negocio (Alta, Media, Baja)	Alta	Riesgo en Desarrollo (Alta, Media, Baja)	Alta
Estimación	30	Sprint asignado	2
Descripción			
<ul style="list-style-type: none"> • Como usuario deseo que se seleccionen los algoritmos a utilizar para realizar las pruebas con el Data Set. 			
Criterios de aceptación			
<ul style="list-style-type: none"> • Análisis de los mejores algoritmos. • Ventajas y Desventajas de los algoritmos seleccionados. 			

Nota: Historia de usuario 3, descripción y criterios de aceptación, elaboración propia.

Tabla 6

Historia de Usuario N°4

HISTORIA DE USUARIO			
Número	4	Usuario	Empresas Retails
Nombre Historia	Carga y Prueba de los algoritmos seleccionados con el Data Set		
Programador Responsable	Vicente Jiménez, George Gómez		
Prioridad en Negocio (Alta, Media, Baja)	Alta	Riesgo en Desarrollo (Alta, Media, Baja)	Media
Estimación	50	Sprint asignado	2
Descripción			
<ul style="list-style-type: none"> • Como usuario deseo que se carguen los datos del Dataset y se realice el entrenamiento con cada algoritmo seleccionado para generar los resultados. 			
Criterios de aceptación			
<ul style="list-style-type: none"> • Gráficos del entrenamiento de cada algoritmo. 			

Nota: Historia de usuario 4, descripción y criterios de aceptación, elaboración propia.

Tabla 7

Historia de Usuario N°5

HISTORIA DE USUARIO			
Número	5	Usuario	Empresas Retails
Nombre Historia	Análisis y Comparación de resultados		
Programador Responsable	Vicente Jiménez, George Gómez		
Prioridad en Negocio (Alta, Media, Baja)	Alta	Riesgo en Desarrollo (Alta, Media, Baja)	Alta
Estimación	40	Sprint asignado	3
Descripción			
<ul style="list-style-type: none"> ● Como usuario deseo que se realice un análisis de resultados obtenidos y se comparen para seleccionar el algoritmo más apto para la propuesta de nuestro proyecto. 			
Criterios de aceptación			
<ul style="list-style-type: none"> ● Resultado del análisis de resultados. ● Tabla comparativa. 			

Nota: Historia de usuario 5, descripción y criterios de aceptación, elaboración propia.

Tabla 8

Historia de Usuario N°6

HISTORIA DE USUARIO			
Número	6	Usuario	Empresas Retails
Nombre Historia	Propuesta del sistema de predicción de ventas		
Programador Responsable	Vicente Jiménez, George Gómez		
Prioridad en Negocio (Alta, Media, Baja)	Alta	Riesgo en Desarrollo (Alta, Media, Baja)	Media
Estimación	50	Sprint asignado	3
Descripción			
<ul style="list-style-type: none"> • Como usuario deseo que se desarrolle la propuesta del sistema de predicción de ventas para determinar y mostrar el resultado final del proyecto. 			
Criterios de aceptación			
<ul style="list-style-type: none"> • Implementación de la propuesta. 			

Nota: Historia de usuario 6, descripción y criterios de aceptación, elaboración propia.

PRODUCT BACKLOG

Tabla 9

Product Backlog

N°	REQUERIMIENTO	SPRINT	ESTIMACIÓN	INCREMENTO
1	Levantamiento de requerimientos.	1	30	30
2	Búsqueda de los algoritmos			
3	Selección de los algoritmos	2	80	110
4	Carga y Prueba de los algoritmos seleccionados con el Data Set			
5	Análisis y Comparación de resultados	3	90	200
6	Propuesta del sistema de predicción de ventas			

Nota: Requerimientos necesarios para la implementación del proyecto, elaboración propia

2.4 SPRINT

Sprints a ejecutar

Para llegar al punto de ejecución inicialmente se debe establecer las fases que intervendrán en la realización de la metodología Scrum, para el proyecto el equipo

SCRUM decidió implementar tres sprint con una duración de 11 semanas, el cual se encuentra detallado en la tabla Sprints a ejecutar. Se presenta la descripción de las fases que son parte de esta propuesta:

Tabla 10

Sprints a ejecutar

N° Sprint	Sprint Goal	Duración	Fecha Inicio	Fecha Fin
1	Recolección de datos	2 semanas	13/07/2021	25/09/2021
2	Entrenamiento y análisis de algoritmos	4 semanas	27/09/2021	22/10/2021
3	Aplicativo Web	5 semanas	25/10/2021	26/11/2021

Nota: Duración de los sprint, fecha de inicio y fecha fin, elaboración propia.

Sprint Planning

La Planificación de los String son reuniones que se elaboran al comienzo de cada sprint para definir el alcance que tendrá. Estas reuniones sirven para decidir y derivar las tareas que formarán el sprint a realizarse.

De las reuniones del Sprint Planning nace el Sprint Backlog que es la composición del listado de tareas que se va a llevar a cabo en el sprint a iniciar, y estas tareas contienen el detalle de la siguiente tabla:

Tabla 11

Detalle de Sprint

Responsable:	El responsable del equipo de desarrollo en llevarla a cabo.
Estimación:	Puntos de complejidad estimados a la tarea.
Nº Días:	Tiempo estimado de conclusión de la tarea.
Historia de Usuario:	Número de historia de usuario de la tarea.

Nota: Detalles que se encuentran dentro del Sprint, elaboración propia.

Los siete primeros números representan el peso o apreciación de las tareas en la serie Fibonacci, en donde puede visualizar el porcentaje de tamaño o complicación de la tarea analizada, determinando la detección con la tarea 1 que presente la complejidad de menor rango, siendo esta la referencia para la estimación de las tareas que se trataran. Los resultados son presentados a continuación en la tabla 12:

Tabla 12

Puntos de estimación de tarea

Estimación (Fibonacci)	Porcentaje de Complejidad
1	10%
2	25%
3	40%
5	55%
8	70%
13	85%
21	100%

Nota: Se muestra la estimación de cada uno de los requerimientos, elaboración propia

SPRINT PLANNING N° 1

Tomando en cuenta las estimaciones, el equipo de desarrollo considero tomar los dos primeros requerimientos del Product Backlog para el primer sprint:

- Levantamiento de requerimientos
- Búsqueda de los algoritmos

Dando como resultado de la reunión a continuación el Sprint Backlog.

Tabla 13

Sprint Backlog N°1

N°	Tarea	Responsable	Estimación	N° Días	Historia de Usuario
1	Levantamiento de requerimientos.	George Gómez, Vicente Jiménez	15	7	1
2	Búsqueda de los algoritmos.	George Gómez, Vicente Jiménez	15	7	2
TOTAL, INCREMENTO DEL SPRINT			30	14	

Nota: Se muestra la realización del primer Sprint, elaboración propia.

SPRINT PLANNING N° 2

Tomando en cuenta las estimaciones, el equipo de desarrollo considero tomar los dos requerimientos del Product Backlog para el segundo sprint:

- Selección de los algoritmos
- Carga y prueba de los algoritmos seleccionados con el Data Set.

Dando como resultado de la reunión a continuación el Sprint Backlog.

Tabla 14**Sprint Backlog N°2**

N°	Tarea	Responsable	Estimación	N° Días	Historia de Usuario
1	Selección de los algoritmos	George Gómez, Vicente Jiménez	20	10	3
2	Carga de los algoritmos seleccionados con el Data Set.	George Gómez	25	6	4
3	Prueba de los algoritmos seleccionados con el Data Set	Vicente Jiménez	35	12	
TOTAL, INCREMENTO DEL SPRINT			80	28	

Nota: Se muestran detalles del segundo Sprint, elaboración propia.

SPRINT PLANNING N° 3

Tomando en cuenta las estimaciones, el equipo de desarrollo considero tomar los dos requerimientos del Product Backlog para el tercer sprint:

- Análisis y Comparación de resultados
- Propuesta del sistema de predicción de ventas.

Dando como resultado de la reunión a continuación el Sprint Backlog.

Tabla 15**Sprint Backlog N°3**

N°	Tarea	Responsable	Estimación	N° Días	Historia de Usuario
1	Comparación de resultados.	George Gómez	20	7	5
2	Análisis de resultados.	Vicente Jiménez	20	7	
3	Propuesta del sistema de predicción de ventas.	George Gómez Vicente Jiménez	50	21	6
TOTAL, INCREMENTO DEL SPRINT			90	35	

Nota: Se muestran detalles del Sprint 3, elaboración propia.

Sprints

En este apartado se puede visualizar la descripción de cada sprint.

Sprint N° 1.

Las actividades que se realizaron en este sprint son las siguientes:

- Levantamiento de requerimientos, para lo cual se procedió a obtener una lista de los requerimientos funcionales y no funcionales priorizados por su valor para el mencionado Product Owner que tiene el trabajo de gestionar el Product Backlog, para lo cual se introdujo los requisitos nuevos de acuerdo con su descubrimiento y priorización conforme las necesidades del cliente, así como del retorno de la inversión que el cliente obtendrá de esta implementación.
- Búsqueda de los algoritmos, se procedió a buscar en sitios web tales como artículos científicos, tesis de pregrado, posgrado, revistas científicas.

Sprint N° 2.

Las actividades que se realizaron en este sprint son las siguientes:

- Selección de los algoritmos, luego de la búsqueda de algoritmos se procedió a la selección de 3 algoritmos lo cual se utilizará para la comparación de datos.
- Carga de los algoritmos seleccionados con el Dataset, se cargarán los algoritmos y los datos en una suite de código abierto que contiene varias aplicaciones, conceptos que sirven para el desarrollo y librerías.

- Prueba de los algoritmos seleccionados con el Dataset, una vez cargado los datos se entrenan y se pondrán a prueba los algoritmos para comprobar que no exista algún inconveniente y presentar los resultados.

Sprint N° 3.

Las actividades que se realizaron en este sprint son las siguientes:

- Comparación de resultados, los resultados que muestran los algoritmos seleccionados serán comparados luego de varias pruebas.
- Análisis de resultados, esto nos permitirá tomar las mejores conclusiones de cuál algoritmo mostró mejores resultados.
- Propuesta del sistema de predicción de ventas, luego del análisis de resultados para ver qué algoritmo mostró mejores resultados, se propuso diseñar un sistema de predicción de ventas que permita al usuario utilizar el mejor algoritmo para predecir sus ventas a futuro.

SPRINT DAILY

El sprint daily se realiza al terminar un sprint y se muestra las fechas en la que se revisó el sprint terminado junto al equipo de trabajo.

Tabla 16

Sprint Daily N°1

N°	Tarea	Responsable	Fecha inicio	Fecha fin	Fecha de revisión
1	Levantamiento de requerimientos.	George Gómez Vicente Jiménez	13/08/2021	02/09/2021	03/09/2021
2	Búsqueda de los algoritmos.	George Gómez Vicente Jiménez	04/09/2021	24/09/2021	25/09/2021

Nota: Se muestra las fechas sprint terminado y su revisión, elaboración propia.

Tabla 17**Sprint Daily N°2**

N°	Tarea	Responsable	Fecha inicio	Fecha fin	Fecha de revisión
1	Selección de los algoritmos	George Gómez Vicente Jiménez	27/09/2021	03/10/2021	04/10/2021
2	Carga de los algoritmos seleccionados con el Data Set.	George Gómez	05/10/2021	12/10/2021	13/10/2021
3	Prueba de los algoritmos seleccionados con el Data Set	Vicente Jiménez	14/10/2021	21/10/2021	22/10/2021

Nota: Se muestra las fechas sprint terminado y su revisión, elaboración propia.

Tabla 18**Sprint Daily N° 3.**

N°	Tarea	Responsable	Fecha inicio	Fecha fin	Fecha de revisión
1	Comparación de resultados.	George Gómez Vicente Jiménez	25/10/2021	30/10/2021	31/10/2021
2	Análisis de resultados.	George Gómez	01/11/2021	08/11/2021	09/10/2021
3	Propuesta del sistema de predicción de ventas.	Vicente Jiménez	10/11/2021	25/11/2021	26/11/2021

Nota: Se muestra las fechas sprint terminado y su revisión, elaboración propia.

INCREMENTO/SPRINT

Tabla 19

Incrementos

N°	REQUERIMIENTO	POR HACER	EN CURSO	LISTO
1	Levantamiento de requerimientos.			X
2	Búsqueda de los algoritmos			X
3	Selección de los algoritmos			X
4	Carga y Prueba de los algoritmos seleccionados con el Data Set		X	
5	Análisis y Comparación de resultados	X		
6	Propuesta del sistema de predicción de ventas	X		

Nota: Se muestran los incrementos de los requerimientos, elaboración propia.

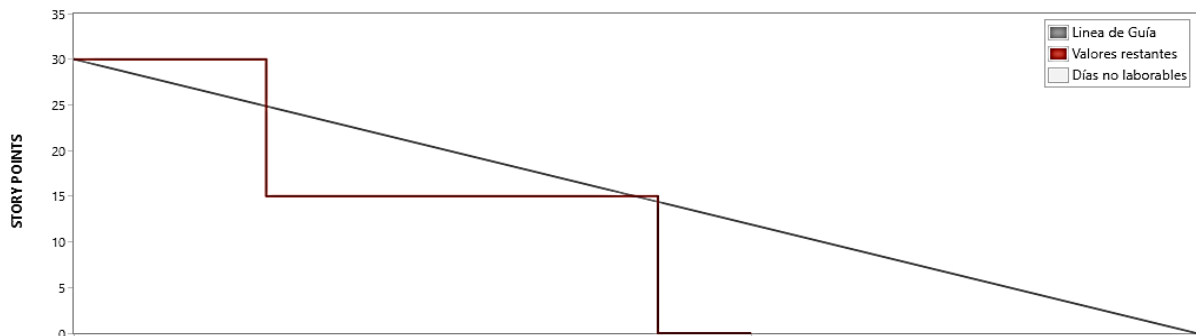
BURNDOWN CHART

En el burndown chart mostrado:

- El equipo inicia con un proyecto que fue planeado para 3 sprints.
- Se inició con 340 story points de trabajo.
- Para finalizar los 3 sprints, se planificó realizar el primer sprint en 30 puntos, el segundo en 110 y el tercero en 200

El primer sprint fue bueno y se completaron 30, dejando 310.

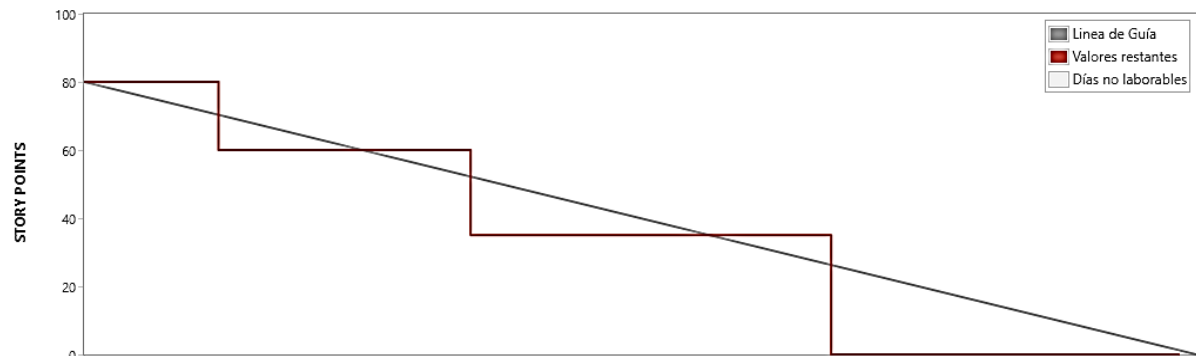
Figura 4 Reporte Sprint 1



Nota: Gráfica del primer sprint listo con las versiones completadas, elaboración propia.

El segundo sprint cumplió con todos los requerimientos completando los 110 puntos, dejando 200 puntos para finalizar.

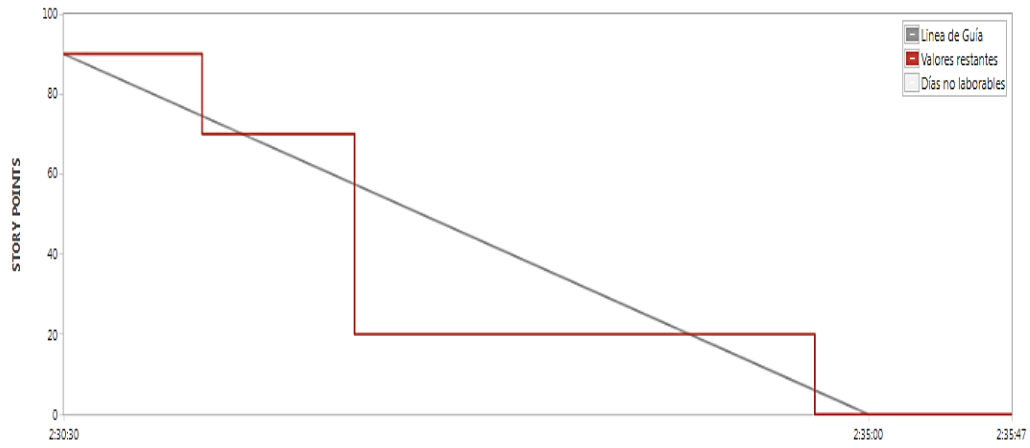
Figura 5 Reporte Sprint 2



Nota: Gráfica del segundo sprint listo con las versiones completadas, elaboración propia.

El tercer sprint se completó 200 puntos, el cual permitió finalizar los sprint sin inconvenientes como se muestra en la gráfica 3.

Figura 6 Sprint 3



Nota: Gráfica del tercer sprint listo terminado, elaboración propia.

CAPÍTULO 3

3. PROPUESTA DE SOLUCIÓN

3.1 Análisis de Factibilidad

Para el análisis de la factibilidad se realizó la elección del algoritmo, el diseño de sistema, levantamiento de información, obteniendo los siguientes resultados:

3.1.1 Factibilidad Técnica

Tabla 20

Características del Hardware

Cantidad	Hardware	Descripción
2	HP Laptop 15	Procesador (11th Gen Intel(R) Core (TM) i5-1135G7 @ 2.40GHz 2.42 GHz) Disco duro (256GB SSD + 16GB Intel Optane) RAM (8,00 GB (7,65 GB usable)

Nota: Se muestran las características del Hardware, elaboración propia

Tabla 21

Características del Software

Software	Versión
Python	Python 3.10.0
Plantilla Bootstrap	Bootstrap 4.3.1
xampp-windows-x64	7.1.33-1-VC14
Sublime text	Sublime Text 4121
Windows 10 Home	Sistema operativo 64 bits, procesador en x64

Nota: Se muestran las características del Software, elaboración propia

3.1.2 Factibilidad Económica

Tabla 22

Descripción costo del Hardware

Cantidad	Hardware	Costo Unitario	Costo Total
2	HP Laptop 15	\$ 750,00	\$ 1.500,00
			\$
	Total		1.500,00

Nota: Se muestran los costos del Hardware, elaboración propia

Tabla 23

Descripción costo del software

Software	Costo
Python	\$ -
Plantilla Bootstrap	\$ -
xampp-windows-x64	\$ -
Sublime text 3	\$ -
Windows 10	\$ 16,75
Total	\$ 16,75

Nota: Se muestran los costos del Software, elaboración propia.

Tabla 24

Descripción costo de talento humano

Cantidad	Talento humano	Costo Unitario	Costo Total
2	Programadores	\$ 400,00	\$ 800,00
	Total		\$ 800,00

Nota: Se muestran los costos del talento humano, elaboración propia

Tabla 25

Costos totales

Descripción	Costo
Hardware	\$ 1.500,00
Software	\$ 16,75
Talento Humano	\$ 800,00
Total	\$ 2.316,75

Nota: Se muestra la suma total de los elementos de la propuesta del sistema, elaboración propia

La descripción de cada uno de los componentes que forman parte de la propuesta ha demostrado que es viable, puesto que el costo de inversión asciende a \$2339.85 estando dentro de los parámetros del mercado, por lo tanto, se considera que este proyecto es viable.

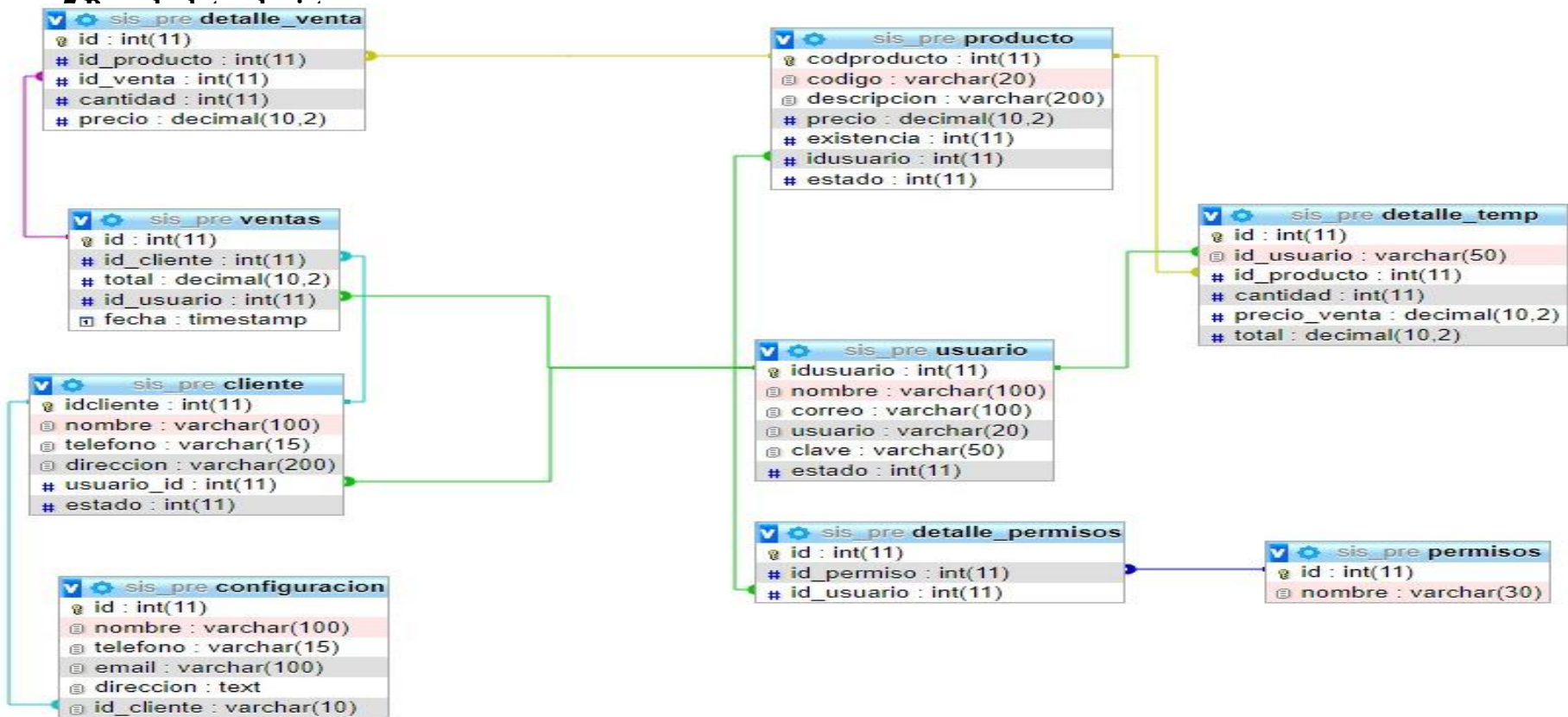
3.2 Propuesta

3.2.1 Título

“Propuesta de desarrollo de un sistema de predicción de ventas usando algoritmos de inteligencia artificial”

Esta propuesta tiene como finalidad proponer un sistema de proyección de ventas, para lo cual se ha estudiado tres algoritmos de inteligencia artificial orientados a la predicción.

3.2.2 Base de modelo de datos



Nota: Base de datos del sistema de predicción de ventas, elaboración propia.

3.2.3 Descripción de la data set y los algoritmos

Dataset

El Dataset que se utilizó fue obtenido de la plataforma Web Kaggle, que contiene las ventas desde el año 2013 al 2018 de la empresa de consumo masivo Corporación Favorita del Ecuador. El mismo que contiene información sobre ventas, fechas, productos, información de la tienda, entre otras. De los cuales para el efecto de análisis del Dataset se tomó en consideración las ventas unitarias de un periodo de tiempo (2015-2017), dando un total de 49,442 transacciones.

Las variables que se utilizaron son fechas y transacciones y se las representan con (x) para la entrada y la salida con (y).

Justificación

Para la implementación de los algoritmos seleccionados se utilizó la biblioteca Scikit-learn, que es una librería de código libre para Python, esta biblioteca contiene algoritmos de clasificación, clustering, regresión, entre otras. Es compatible con otras librerías de Python como NumPy, SciPy y matplotlib, útiles para visualizar resultados. Scikit Learn es una herramienta útil, simple y eficiente dentro del análisis predictivo de datos.

En el modelo de regresión lineal simple se dividió el Dataset en cuatro columnas para el manejo de datos, se separaron las variables que contienen las fechas y las transacciones respectivas de cada año. Se realizó la separación del 80% para train y el 20% para el test.

Para el entrenamiento de algoritmo se utilizó el modelo de regresión lineal simple que brinda la librería de Scikit-learn, además de detectar y eliminar los valores atípicos,

se utilizó el enfoque IQR (Inter Quartile Range). También se empleó el coeficiente de determinación o coeficiente de correlación de Pearson, una prueba que mide la relación estadística entre dos variables continuas.

En el algoritmo se muestra las gráficas de valores iniciales, entrenamientos, y predicción, así como la presentación del error medio cuadrático.

Árbol de decisión de regresión

El algoritmo realizó la implementación de los códigos de acceso abierto que ofrece Scikit-Learn, el mismo que procede a dividir los datos en train y test, además del entrenamiento de los datos. Para el modelo inicial del árbol se dio un máximo de crecimiento de 5 nodos y una semilla de 30000 transacciones, valores que pueden ser ajustados.

El modelo utiliza también un algoritmo de pruning (poda) por validación cruzada que hace que el árbol crezca a su mayor capacidad para luego ser podado, obteniendo un árbol más optimo. La estructura final del árbol arrojó una profundidad de 12 nodos con 332 nodos terminales. El modelo también presenta el error cuadrático medio (rmse).

Redes Neuronales

Se usó una Red Neuronal MLP (Multilayered Perceptron - perceptrón multicapa) feedforward de pocas capas. Para la división de los datos de entrenamiento y prueba, se lo dividió en el 80% para el train y el 20% para el test.

Se usaron como entrada 7 inputs, 1 capa oculta con 7 neuronas; este valor se lo escogió al azar, pero se puede variar y ser ajustado, la salida será una sola neurona. Como función de activación se usó tangente hiperbólica puesto que utilizó valores entre -1 y 1.

El algoritmo empleó el error cuadrático medio (rmse) de nuestro modelo, este utilizó el optimizador Adam y como métrica de perdida error absoluto media, y presentó los resultados de predicción.

3.2.4 Tabla de Resultados

3.2.4.1 Resultado de regresión lineal

Tabla 26

Descripción del dataset

	Transacciones2015	Transacciones2016
Count	353.000000	353.000000
Mean	84710.441926	84365.362606
Std	9644.520795	8719.193831
Min	2202.000000	61574.000000
25%	78165.000000	77696.000000
50%	82462.000000	82331.000000
75%	92717.000000	90428.000000
Max	108362.000000	118222.000000

Nota: Descripción del Dataset como cantidad de transacciones, la media, desviación estándar, etc., elaboración propia

Tabla 27

Comparación de Datos

	Actual	Predicción
0	69628	84734.424256
2	89018	84734.424256
3	88846	84734.424256
4	80412	84734.424256
5	83344	84734.424256

Nota: Se realiza una comparación con los datos reales y predichos, elaboración propia.

3.2.4.2 Resultados de Árbol de decisión

Tabla 28

Error cuadrático medio (RMSE)

El error (rmse) del test	
Modelo Inicial	400.1507845
Modelo Final	70.75338127

Nota: Error del test del modelo inicial y final tras aplicar pruning.

3.2.4.3 Resultados de Redes neuronales

Tabla 29

Comparación de datos

	Real	Predicción	Diferencia
0	1174.000000	1257.331055	-83.331055
1	967.999878	1026.339233	-58.339355
2	1254.999878	1187.371826	67.628052
3	757.000000	1155.676147	-398.676147
4	644.000122	1090.068970	-446.068848

Nota: Se comparan los datos reales con los de la predicción, elaboración propia

Tabla 30

	Real	Predicción	Diferencia
Count	21650.000000	21650.000000	21650.000000
Mean	1656.723145	1576.418091	80.304504
Std	956.998718	736.134644	673.749634
Min	5.000036	799.373962	-2776.952637
25%	1019.000061	1199.243317	-321.855865
50%	1362.000122	1310.365662	-26.368774
75%	1981.750092	1558.613068	271.909912
máx.	8307.000000	7670.610352	6638.695312

Nota: Se, comparan los datos reales con los de la predicción, elaboración propia.

Tabla 31

Comparación del Error medio Cuadrático

EL ERROR MEDIO CUADRATICO	
REGRESION LINEAL SIMPLE	91.5
ARBOL DE DECISION (REGRESION)	70.68
REDES NEURONALES	25.1

Nota: Se compara los resultados de los tres algoritmos, elaboración propia.

En conclusión, el algoritmo de Redes neuronales se comportó de mejor manera en el tratamiento de datos y su entrenamiento. Unos de los resultados más evidentes fue el Error Cuadrático Medio (rmse) que se obtuvo, el cual se comparó el resultado de cada algoritmo, regresión lineal con 91.5, árbol de decisión con 70.68 y redes neuronales con 25,1.

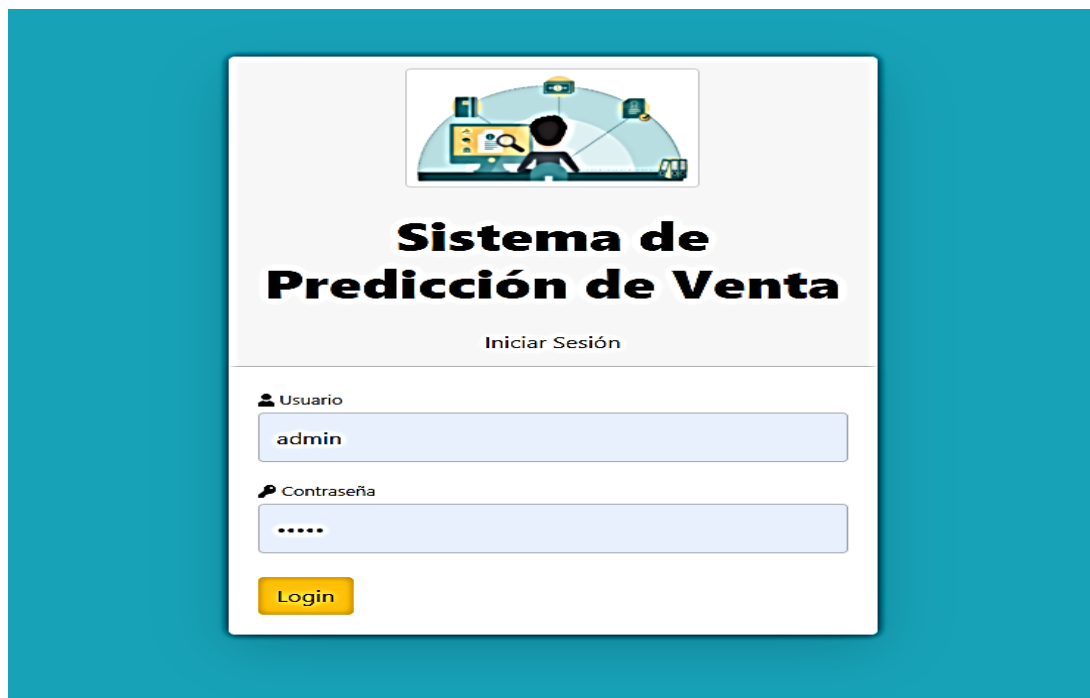
3.3 Funcionalidades del Sistema

Proponemos desarrollar el sistema de proyección de ventas en el lenguaje Php, ya que permite una conexión con la base de datos que fue realizada en phpMyAdmin, para el diseño utilizamos una plantilla de Bootstrap el cual cuenta con diferentes librerías como son CSS, JS library entre otras, utilizando además HTML, JAVASCRIPT y AJAX para el diseño de nuestro sistema.

A continuación, se presenta el diseño de la propuesta para el diseño de un sistema de predicción de ventas:

La Figura 13 muestra el Login el cual será la pantalla inicial donde las personas autorizadas podrán tener acceso al sistema.

Figura 8 Login



Nota: Pantalla inicial Login, elaboración propia.

La Figura 14 se muestra el panel de administración donde se podrá acceder a cada uno de los formularios del sistema, además de presentar las gráficas de los productos más vendidos y que se encuentran en stock.

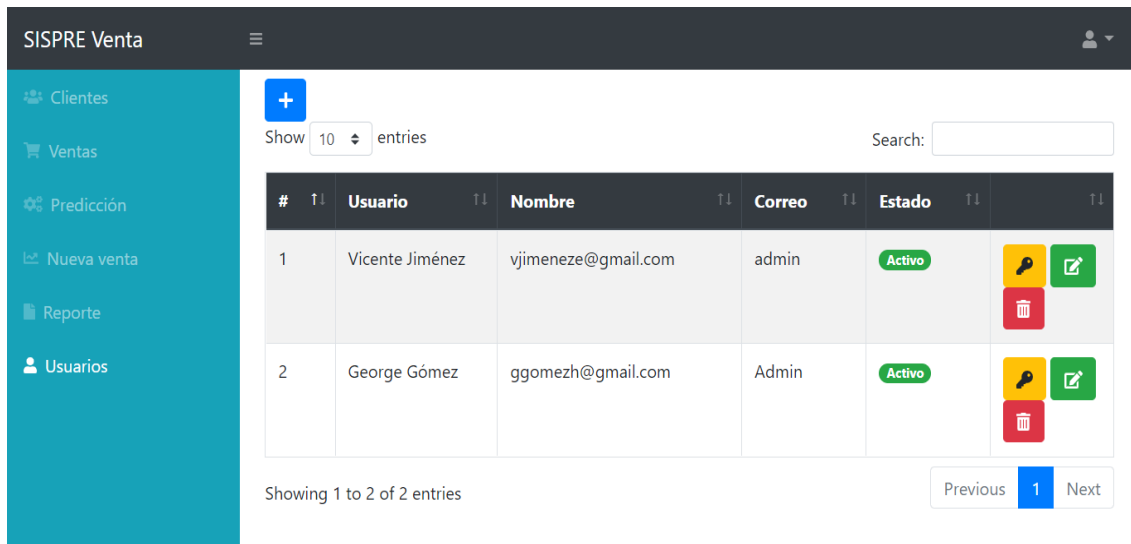
Figura 9 Panel de administración



Nota: Página principal de panel de administración, elaboración propia.

La Figura 15, se muestra el módulo de usuarios, en el cual se puede observar a las personas autorizadas al sistema, donde se podrá dar permisos a los usuarios además de podrá editar y eliminar un usuario.

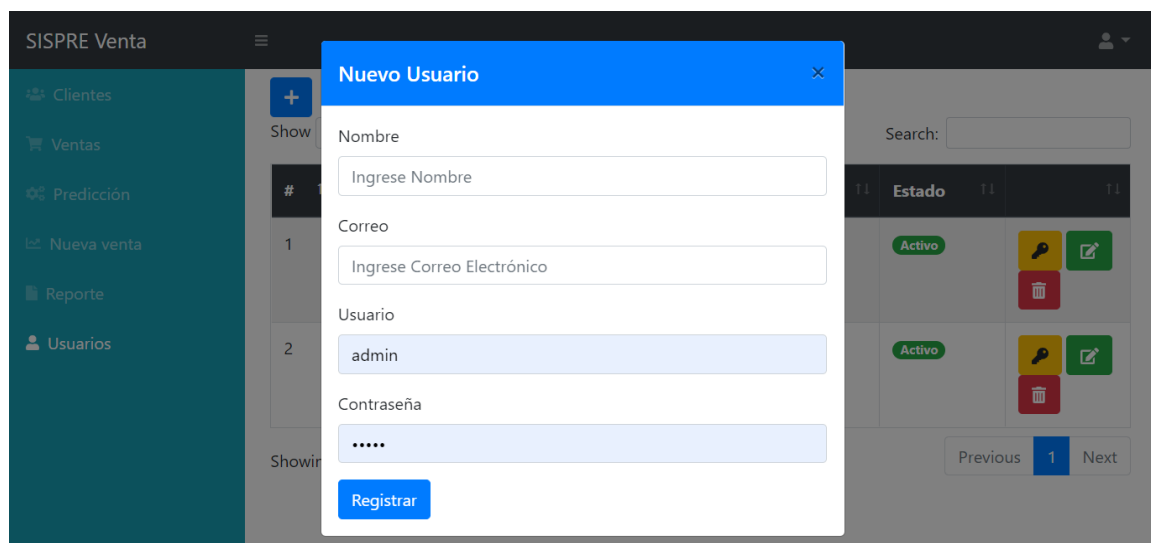
Figura 10 Módulo usuario



Nota: Modulo Usuario, el cual permite dar permisos, editar y eliminar un usuario, elaboración propia.

La Figura 16 muestra el módulo usuario cuenta con la opción de registrar un nuevo usuario, el cual tendrá permiso para poder ingresar al sistema.

Figura 11 Nuevo usuario.



Nota: Registro de nuevo usuario en el módulo Usuario, elaboración propia.

La Figura 17 presenta el módulo de ventas donde se puede visualizar los datos del cliente y la transacción que realizó, además de poder obtener los datos a través de archivos PDF y CSV.

Figura 12 Módulo ventas

The screenshot shows the 'SISPRE Venta' interface. On the left is a teal sidebar with menu items: Clientes, Ventas, Predicción, Nueva venta, Reporte, and Usuarios. The main area displays a table of sales with columns: #, Cliente, Producto, Total, and Fecha. There are two entries, both for 'Juan Pérez' with a total of 0.00. Each entry has PDF and CSV export icons. At the bottom right, there are 'Previous', '1', and 'Next' navigation buttons.

#	Cliente	Producto	Total	Fecha
1	Juan Pérez	Juan Pérez	0.00	0000-00-00 00:00:00
2	Juan Pérez	Juan Pérez	0.00	0000-00-00 00:00:00

Nota: Descripción de ventas, elaboración propia.

La Figura 18 presenta como generar una nueva venta y de esta manera se ingresará los datos del cliente, los datos de la venta y del vendedor, además de generar el total a pagar guardando los datos.

Figura 13 Módulo de nueva venta





The screenshot shows the 'SISPRE Venta' interface for creating a new sale. The sidebar is the same as in Figure 12. The main area is titled 'Datos del Cliente' and contains form fields for 'Nombre', 'Teléfono', and 'Dirección'. Below this is a blue header for 'Datos Venta'. Underneath, it shows the 'VENDEDOR' as 'VICENTE JIMÉNEZ' and a 'Buscar Producto' section with a search input field. At the bottom, there is a table with columns: Id, Descripción, Cantidad, Precio, Precio Total, and Acción. The 'Total Pagar' is shown as 0.00. A 'Generar Venta' button is located at the bottom left.

Id	Descripción	Cantidad	Precio	Precio Total	Acción
Total Pagar			0.00		

Nota: Venta nueva con datos, elaboración propia.

La Figura 19 muestra el módulo cliente, los datos del cliente se podrán visualizar, el cual permitirá editar y eliminar datos de algún cliente.

Figura 14 Módulo clientes

#	Nombre	Teléfono	Dirección	Estado	
1	Juan Pérez	8296826298	Quito y Jaime Roldós	Activo	 
2	Carlos Piguave	2147483647	Guayas - Milagro	Activo	 

Nota: Lista de clientes, elaboración propia.

La Figura 20 muestra el módulo de clientes, también cuenta con la opción de crear un nuevo cliente, el cual permitirá obtener los datos para futuras ventas.

Figura 15 Nuevo cliente

Nuevo Cliente

Nombre

Teléfono

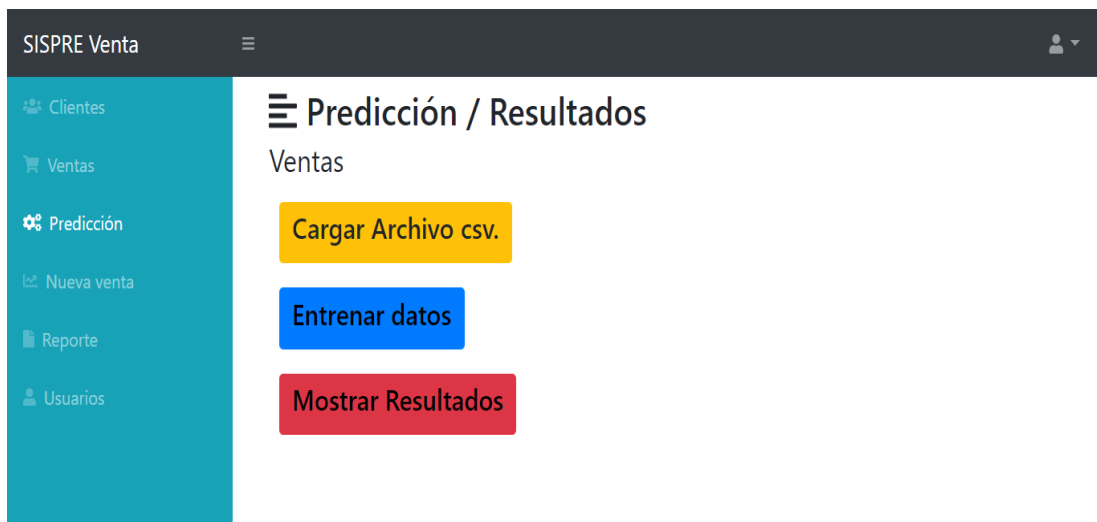
Dirección

Guardar Cliente

Nota: Lista de clientes, elaboración propia.

La Figura 21 presenta la pantalla principal del módulo de predicción, el mismo que cuenta con tres botones, los cuales permitirán cargar los datos, entrenar y presentar los resultados de la predicción.

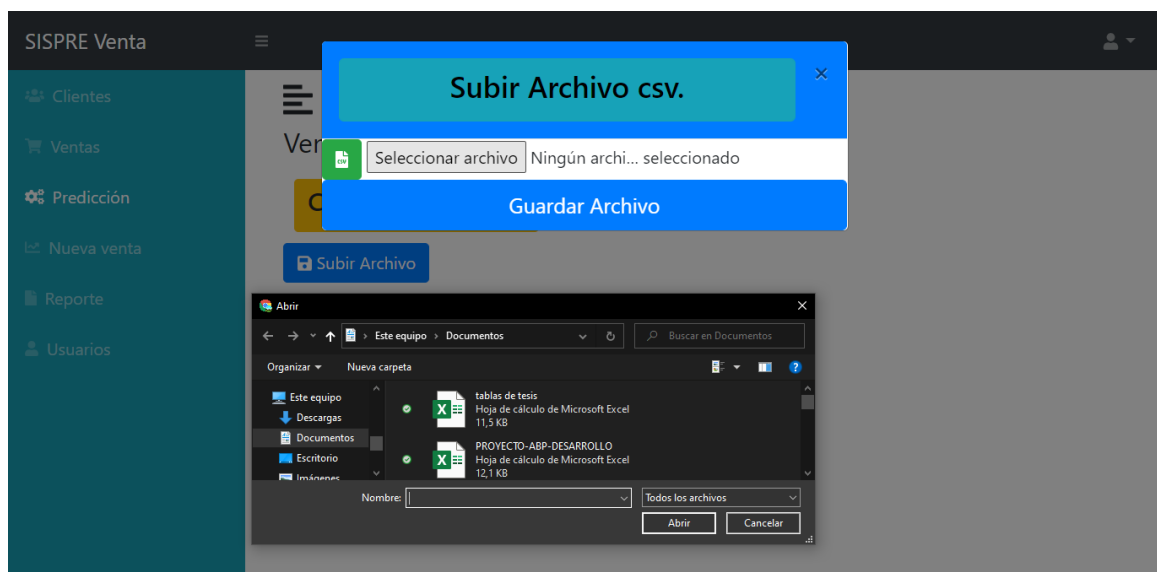
Figura 16 Módulo predicción



Nota: Pantalla principal del módulo de Predicción, elaboración propia.

La Figura 22 presenta el módulo Predicción el cual nos permitirá cargar el archivo de los datos en formato CSV permitiendo elegir el archivo desde un PC.

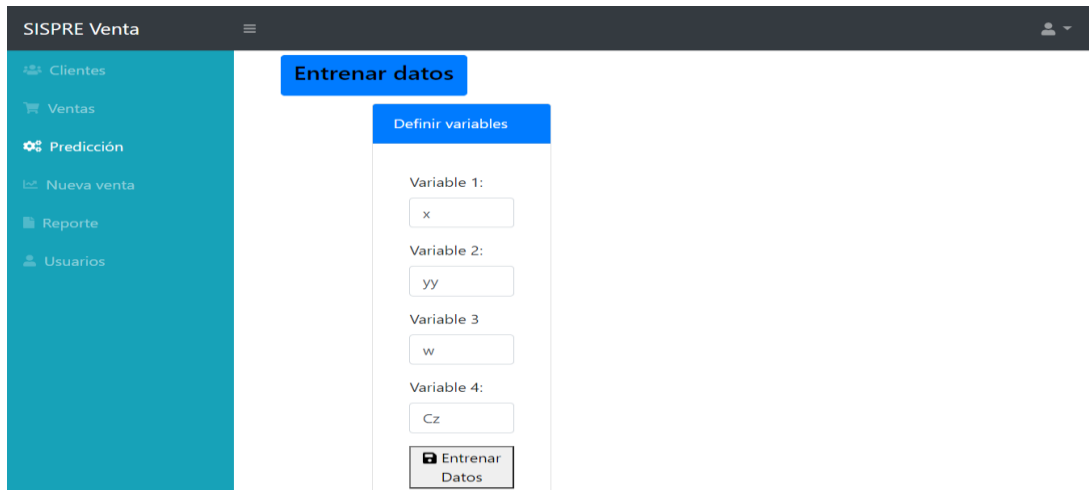
Figura 17 Carga de datos.



Nota: Carga de datos en formato csv, elaboración propia.

La Figura 23 muestra un formulario donde se definirá las variables para el entrenamiento de los datos.

Figura 18 Entrenamiento de datos



Nota: Entrenamiento de datos, elaboración propia.

La Figura 24 muestra los resultados, el cual permitirá visualizar los resultados, luego de haber entrenado y realizada la predicción con el sistema, además poder obtener los datos en archivos de formato PDF y CSV de Excel.

Figura 19 . Resultados



Nota: Resultado de predicciones, fuente elaboración propia.

CONCLUSIÓN

El desarrollo del trabajo propuesto permitió hacer un levantamiento de información exitoso, tanto para el manejo de datos, utilización de metodologías, diseño entre otros aspectos, concluyendo con los siguientes:

- La recopilación de información se la realizó en diferentes fuentes bibliográficas, tales como tesis de pregrado, posgrado, artículos y revistas científicas, que permitieron recabar datos relevantes que fueron de gran aporte en el proceso de identificación y análisis de datos que sirvieron para realizar la propuesta de solución ante la problemática planteada.
- Para la identificación de los tres algoritmos se realizó una investigación teórica sobre los algoritmos que han presentado mejores resultados en los estudios realizados por otros investigadores. En cuanto a las técnicas, métodos y herramientas aplicadas se utilizó Scrum, JIRA, HTML, Google, Bootstrap, entre otras, para el desarrollo del sistema de predicción.
- Los algoritmos utilizados para la comparación entre los mismos fueron redes neuronales, regresión lineal y árbol de decisión, de los cuales se seleccionó para la propuesta el algoritmo redes neuronales, porque presentó los mejores resultados de predicción, lo cual sirvió para proponer el sistema de predicción de ventas.

RECOMENDACIONES

Las recomendaciones se las realizó de acuerdo con el conocimiento y experiencias en el desarrollo de esta propuesta, por parte de los autores:

- Se recomienda que al momento de desarrollar el sistema este sea fácil de utilizar y amigable, para que el usuario se sienta cómodo al momento de realizar el entrenamiento y las pruebas.
- Para futuras investigaciones se recomienda realizar pruebas con otros algoritmos y comparar y obtener los resultados en tiempo real, para las pruebas de los algoritmos se recomienda usar datos de entrenamientos y de prueba, donde se suele utilizar el 80% de los datos para entrenamiento y el 20% para las pruebas.
- Se recomienda que al momento de cargar los datos se definan bien las variables, porque al predecir se presentarán los resultados de acuerdo con los datos presentados (históricos-actuales), y al no definir adecuadamente se podrán presentar datos erróneos que afectarían al desarrollo de las empresas y a las metas que deseen cumplir.

BIBLIOGRAFÍA

- Alania, P. (2018). *Aplicación de técnicas de minería de datos para predecir la deserción estudiantil de la facultad de ingeniería de la universidad nacional Daniel Alcides Carrión*. Pasco. Obtenido de http://repositorio.undac.edu.pe/bitstream/undac/829/1/T026_40573846_M.pdf
- Arroyo, J. (2016). Métodos de reducción de dimensionalidad: Análisis comparativo de los métodos APC, ACPP y ACPK. *Revista Uniciencia*, 30(1), 10. Obtenido de <https://www.redalyc.org/journal/4759/475948285007/475948285007.pdf>
- Azañero, J., & Ramírez, R. (2019). *Análisis comparativo de algoritmos de pronóstico de ventas para su implementación en Mypes del sector ferretería de Chiclayo*. Pimentel, Perú. Obtenido de <https://repositorio.uss.edu.pe/bitstream/handle/20.500.12802/6995/AZA%C3%91ERO%20BURGA%20JHYNNO%20%26%20RAMIREZ%20SIPION%20RICHARD.pdf?sequence=1&isAllowed=y>
- Balagueró, T. (2018). *¿Qué son los datasets y los dataframes en el Big Data?* Obtenido de <https://www.deustoformacion.com/blog/programacion-tic/que-son-datasets-dataframes-big-data>
- Benites, L. (2018). *Introducción, estructura de algoritmos, conceptos, instrucciones con algoritmos, instrumentos que utilizan, aplicaciones prácticas*. Lima, Perú. Obtenido de <https://repositorio.une.edu.pe/bitstream/handle/UNE/4110/Algor%C3%ADmica.pdf?sequence=1&isAllowed=y>
- Bozanic, M. (2020). *Sistema de predicción de precios-venta de inmuebles en el mercado del sector inmobiliario de la región metropolitana de la república de Chile con el uso de algoritmos de Machine Learning*. Santiago de Chile. Obtenido de <http://repositorio.uchile.cl/bitstream/handle/2250/176909/Sistema-de-predicci%C3%B3n-de-precios-venta-de-inmuebles-en-el-mercado-del-sector-inmobiliario-de-la-Regi%C3%B3n-Metropolitana-de.pdf?sequence=1&isAllowed=y>
- Carrión, C. (2018). *Aplicación de redes neuronales artificiales para la predicción de la recuperación de planta concentradora en Minsur S.A.* Arequipa. Obtenido de <http://repositorio.unsa.edu.pe/bitstream/handle/UNSA/6629/IMcaoscj.pdf?sequence=1&isAllowed=y>
- Castillo, P. (2017). *Aplicación de aprendizaje automático para la predicción de clientes potenciales en procesos de mercadotecnia*. Obtenido de

<https://cimat.repositorioinstitucional.mx/jspui/bitstream/1008/755/1/ZAC%20TE%2061.pdf>

- Conde, D. (2018). *Inteligencia Artificial con Tensor Flow para predicción de comportamientos*. Sevilla. Obtenido de <http://bibing.us.es/proyectos/abreproy/91796/fichero/TFG-1796-CONDE.pdf>
- Data Science. (2018). *Impulso de gradiente – Lo que necesitas saber*. Obtenido de <https://datascience.eu/es/aprendizaje-automatico/impulso-de-gradiente-lo-que-necesitas-saber/>
- Developer. (2021). *Generalidades del protocolo HTTP*. Obtenido de <https://developer.mozilla.org/es/docs/Web/HTTP/Overview>
- Dorado, S., & Velásquez, V. (2020). *Pronóstico de demanda utilizando inteligencia artificial*. Cali. Obtenido de https://repository.icesi.edu.co/biblioteca_digital/bitstream/10906/87548/1/TG03014.pdf
- Fondo Social Europeo. (2019). *Glosario de términos utilizados en sistemas de gestión de información*. España. Obtenido de http://incual.mecd.es/documents/20195/1873855/IFC304_3+-+A_GL_Documento+publicado/2bdf96b1-a185-4da5-94ea-d4b6e8cea8ee
- Fontalvo, H. (2019). Metodología de Aprendizaje Automático para la Clasificación y Predicción de Usuarios en Ambientes Virtuales de Educación. *Revista Información tecnológica*, 30(1). Obtenido de https://www.scielo.cl/scielo.php?script=sci_arttext&pid=S0718-07642019000100247&lng=en&nrm=iso&tlng=en
- Fosca, A. (2020). *Desarrollo de un modelo para la predicción del precio del cobre empleando*. Lima. Obtenido de https://tesis.pucp.edu.pe/repositorio/bitstream/handle/20.500.12404/17730/FOSCA_GAMARRA_ALMUDENA_DESARROLLO_MODELO_PREDICCION.pdf?sequence=1&isAllowed=y
- García, E., Sandoval, M., Carrillo, J. A., Pérez, P., Etchevers, J. D., & Macedo, A. (2017). Modelos de predicción y clasificación de deficiencias de hierro en el cultivo de frijol (*Phaseolus Vulgaris* L.) con redes neuronales bayesianas regularizadas y árboles de clasificación. *Revista Fitotecnia Mexicana*, 40(1), 19. Obtenido de <https://www.redalyc.org/journal/610/61051194003/61051194003.pdf>
- Gutiérrez, G., Margain, M. d., Ramírez, T. A., & Canul, J. (2017). Un modelo basado en el Clasificador Naïve Bayes para la evaluación del desempeño docente. *Revista*

- Iberoamericana de Educación a Distancia*, 20(2), 293-313. Obtenido de <https://www.redalyc.org/pdf/3314/331453132014.pdf>
- Jacinto, R. (2019). “*Redes neuronales para predicción de contaminación del aire en Carabayllo-Lima*”. Lima. Obtenido de http://repositorio.unfv.edu.pe/bitstream/handle/UNFV/3428/UNFV_JACINTO_HERRERA_RAUL_TRINIDAD_MAESTRIA_2019.pdf?sequence=1&isAllowed=y
- Jaramillo, J. (2020). *Gestión de pedidos para puntos de venta con aprendizaje automático*. Bogotá. Obtenido de <https://repositorio.uniandes.edu.co/bitstream/handle/1992/49180/u833927.pdf?sequence=1>
- Lizama, O., Kindley, G., & Jeria, J. (2016). *Redes de computadores. Arquitectura Cliente - Servidor*. Obtenido de <http://profesores.elo.utfsm.cl/~agv/elo322/1s16/projects/reports/Proyecto%20Cliente%20-%20Servidor.pdf>
- Macea, A., & Martelo, D. (2016). *Predicción con redes neuronales de un proceso de secado de ñame (dioscorea rotundata) en horno microondas*. Berastegui. Obtenido de https://repositorio.unicordoba.edu.co/bitstream/handle/ucordoba/1030/PREDIC_1.PDF?sequence=1&isAllowed=y
- Maldonado, J. (2016). *Desarrollo e implementación de un sistema web de seguimiento y evaluación de las prácticas pre-profesionales para la facultad de ingeniería escuela civil de la PUGE*. Quito. Obtenido de http://repositorio.puce.edu.ec/bitstream/handle/22000/12562/Tesis_Teoria.pdf?sequence=1&isAllowed=y
- Mariño, S. I., & Alfonzo, P. L. (Julio de 2018). *Implementación de SCRUM en el diseño del proyecto del Trabajo Final de Aplicación*. Obtenido de <https://www.redalyc.org/pdf/849/84933912009.pdf>
- Martín, A., & Carbonell, S. (2016). La producción científica en Inteligencia Artificial: revistas del primer cuartil indexadas en Scopus Sciverse. *Revista Ciencias de la Información*, 47(2), 25- 32. Obtenido de <https://biblat.unam.mx/hevila/Cienciasdelainformacion/2016/vol47/no2/3.pdf>
- Molina, K. (2020). *Implementación de un modelo analítico para la predicción de la venta del portafolio de*. Sangolquí, Ecuador. Obtenido de <http://repositorio.espe.edu.ec/xmlui/bitstream/handle/21000/22561/T-ESPE-043875.pdf?sequence=1&isAllowed=y>

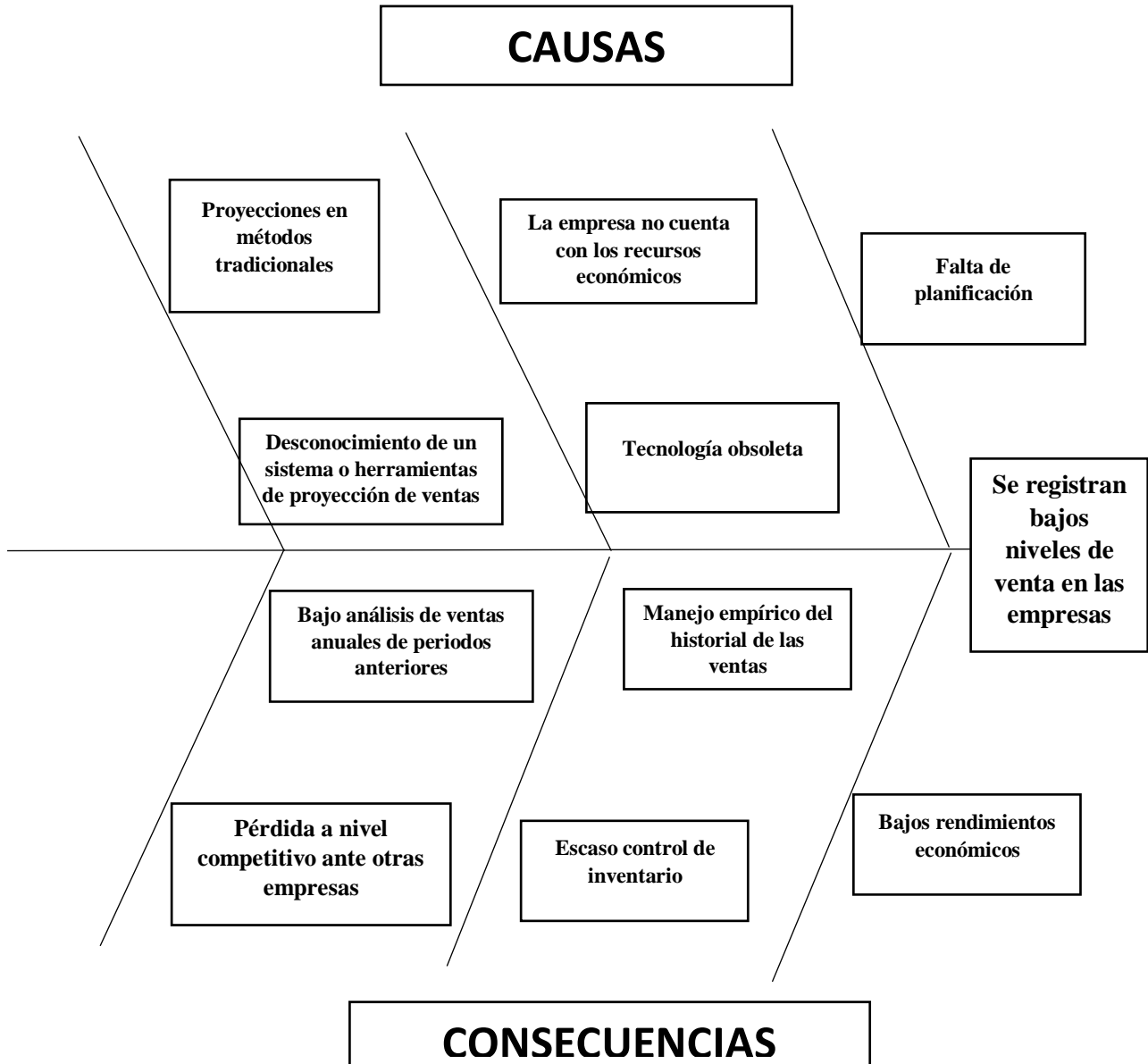
- Mustapha, M. (2015). *Aplicación de Redes Neuronales para la Resolución de Problemas Complejos en Confiabilidad y Riesgo*. Las Palmas. Obtenido de https://accedacris.ulpgc.es/bitstream/10553/18929/4/0728036_00000_0000.pdf
- Palacios, C. (2020). *Análisis y predicción de las tendencias de venta en el*. Quito. Obtenido de <https://repositorio.usfq.edu.ec/bitstream/23000/8799/1/146132.pdf>
- Pamies, B. (2017). *Predicción de la probabilidad de éxito en la adquisición de clientes*. Obtenido de https://rua.ua.es/dspace/bitstream/10045/69432/1/Prediccion_de_la_probabilidad_de_exito_en_la_adqui_PAMIES_CARTAGENA_BENJAMIN.pdf
- Pérez, J. (2017). *La regresión logística como modelo de predicción del riesgo crediticio en las organizaciones de la economía social y solidaria*. Universidad Internacional del Ecuador, Quito. Obtenido de <https://www.uv.mx/iiesca/files/2018/03/23CA201702.pdf>
- Revista Lideres. (11 de Noviembre de 2019). *El 'retail' mueve en Ecuador más de USD 16 800 millones anuales*. Obtenido de <https://www.revistalideres.ec/lideres/retail-ecuador-ventas-centros-comerciales.html>
- Robles, M., & Valverde, M. (2021). *Sistema de predicción para incrementar las ventas de accesorios y respuestos automotrices en la empresa GGP automotriz*. Lima. Obtenido de https://repositorio.usmp.edu.pe/bitstream/handle/20.500.12727/7424/robles_rma-valverde_cmy.pdf?sequence=3&isAllowed=y
- Rodríguez, J. (2018). *Desarrollo e implementación web del formulario mensual de establecimientos de hospedaje en la Dircetur – Puno - 2017*. Puno-Perú. Obtenido de http://repositorio.unap.edu.pe/bitstream/handle/UNAP/9734/Rodriguez_Alcos_Jair_Anthony.pdf?sequence=1&isAllowed=y
- Salmerón, R., & Rodríguez, E. (2017). *Métodos cuantitativos para un modelo de regresión lineal con multicolinealidad*. *Revista de Métodos Cuantitativos para la Economía y la Empresa*, 169-189. Obtenido de <https://www.redalyc.org/pdf/2331/233154079006.pdf>
- Sandoval, L. (2018). *Algoritmos de aprendizaje automático para análisis y predicción de datos*. Obtenido de http://www.redicces.org.sv/jspui/bitstream/10972/3626/1/Art6_RT2018.pdf
- TechTarget. (2018). *Python*. Obtenido de <https://searchdatacenter.techtarget.com/es/definicion/Python>

- Timarán, R., Caicedo, J., & Hidalgo, A. (2018). Árboles de decisión para predecir factores asociados al desempeño académico de. *Revista Investigación Desarrollo Innovador*, 363-378. Obtenido de <https://dialnet.unirioja.es/descarga/articulo/6976577.pdf>
- Traña, M., & Silva, E. (2016). *Desarrollo de un Sistema de Información, con Arquitectura Orientada a Servicios, que permita la gestión de los Procesos de la Biblioteca “Alejandro Sequeira Hernández” de la UNAN FAREM – Chontales en el año 2015*. Juigalpa. Obtenido de <https://repositorio.unan.edu.ni/3848/1/11083.pdf>
- Tripp, C. (2020). Revisión de literatura de modelos computacionales para la predicción de la velocidad del viento de 2004 a 2016. *Revista de Investigación en Tecnologías de la*, 8(15), 27-40. Obtenido de <http://portal.amelica.org/ameli/jatsRepo/368/3681483004/3681483004.pdf>
- Velasquez, E. (2018). *Predicción del consumo de gas natural aplicando redes neuronales*. La Paz, Bolivia. Obtenido de <https://repositorio.umsa.bo/xmlui/bitstream/handle/123456789/7765/T.2705.pdf?sequence=1&isAllowed=y>
- Vivar, L., & Quishpe, F. (2019). *Gestión del contenido multimedia de la cartelera virtual vía Streaming de la escuela de formación de tecnólogos*. Quito. Obtenido de <https://bibdigital.epn.edu.ec/bitstream/15000/20370/1/CD%209841.pdf>
- Yosvani, O., Rivas, A., Pérez, M., & Marrero, F. (2017). Procedimiento para el pronóstico de la demanda mediante redes neuronales artificiales. *Ciencias Holguín*, 23(1), 1-18. Obtenido de <https://www.redalyc.org/pdf/1815/181549596004.pdf>
- Zapata, J. (2019). *Aplicación de las redes neuronales para determinar los pronósticos mensuales de la inflación para el año 2018*. Piura. Obtenido de <https://repositorio.unp.edu.pe/bitstream/handle/UNP/2014/EST-ZAP-LAU-2019.pdf?sequence=1&isAllowed=y>

ANEXOS

Anexo 1

Diagrama de espina de pescado (Ishikawa)



Anexo 2

Creación de base de datos y tablas

```
-- phpMyAdmin SQL Dump
-- version 5.1.1
-- https://www.phpmyadmin.net/
--
-- Servidor: 127.0.0.1
-- Tiempo de generación: 31-03-2022 a las 06:15:19
-- Versión del servidor: 10.4.22-MariaDB
-- Versión de PHP: 8.1.0

SET SQL_MODE = "NO_AUTO_VALUE_ON_ZERO";
START TRANSACTION;
SET time_zone = "+00:00";

/*!40101 SET
@OLD_CHARACTER_SET_CLIENT=@@CHARACTER_SET_CLIENT */;

/*!40101 SET
@OLD_CHARACTER_SET_RESULTS=@@CHARACTER_SET_RESULTS */;

/*!40101 SET
@OLD_COLLATION_CONNECTION=@@COLLATION_CONNECTION */;
/*!40101 SET NAMES utf8mb4 */;

--
-- Base de datos: `sis_venta`
--
-- Estructura de tabla para la tabla `cliente`
--
CREATE TABLE `cliente` (
  `idcliente` int(11) NOT NULL,
  `nombre` varchar(100) COLLATE utf8_spanish_ci NOT NULL,
```

```

`telefono` varchar(15) COLLATE utf8_spanish_ci NOT NULL,
`direccion` varchar(200) COLLATE utf8_spanish_ci NOT NULL,
`usuario_id` int(11) NOT NULL,
`estado` int(11) NOT NULL DEFAULT 1
) ENGINE=InnoDB DEFAULT CHARSET=utf8 COLLATE=utf8_spanish_ci;
--
-- Volcado de datos para la tabla `cliente`
--
INSERT INTO `cliente` (`idcliente`, `nombre`, `telefono`, `direccion`, `usuario_id`,
`estado`) VALUES
(1, 'UNEMI', '8296826298', 'S/D', 1, 1),
(2, 'Ezequiel Gomez', '2147483647', 'Ecuador - Milagro', 1, 1);
--
-- Estructura de tabla para la tabla `configuracion`
--
CREATE TABLE `configuracion` (
  `id` int(11) NOT NULL,
  `nombre` varchar(100) COLLATE utf8_spanish_ci NOT NULL,
  `telefono` varchar(15) COLLATE utf8_spanish_ci NOT NULL,
  `email` varchar(100) COLLATE utf8_spanish_ci NOT NULL,
  `direccion` text COLLATE utf8_spanish_ci NOT NULL
) ENGINE=InnoDB DEFAULT CHARSET=utf8 COLLATE=utf8_spanish_ci;
--
-- Volcado de datos para la tabla `configuracion`
--
INSERT INTO `configuracion` (`id`, `nombre`, `telefono`, `email`, `direccion`) VALUES
(1, 'Ezequiel Gomez', '925491523', 'egomezhl@unemi.edu.ec', 'Ecuador-Milagro');
-----
--
-- Estructura de tabla para la tabla `detalle_permisos`

```

```

-
CREATE TABLE `detalle_permisos` (
  `id` int(11) NOT NULL,
  `id_permiso` int(11) NOT NULL,
  `id_usuario` int(11) NOT NULL
) ENGINE=InnoDB DEFAULT CHARSET=utf8mb4;
--
-- Volcado de datos para la tabla `detalle_permisos`
--
INSERT INTO `detalle_permisos` (`id`, `id_permiso`, `id_usuario`) VALUES
(8, 3, 2),
(9, 4, 2),
(10, 5, 2),
(11, 6, 2);
-----
--
-- Estructura de tabla para la tabla `detalle_temp`
--
CREATE TABLE `detalle_temp` (
  `id` int(11) NOT NULL,
  `id_usuario` varchar(50) COLLATE utf8_spanish_ci NOT NULL,
  `id_producto` int(11) NOT NULL,
  `cantidad` int(11) NOT NULL,
  `precio_venta` decimal(10,2) NOT NULL,
  `total` decimal(10,2) NOT NULL
) ENGINE=InnoDB DEFAULT CHARSET=utf8 COLLATE=utf8_spanish_ci;
-----
--
-- Estructura de tabla para la tabla `detalle_venta`
--

```

```

CREATE TABLE `detalle_venta` (
  `id` int(11) NOT NULL,
  `id_producto` int(11) NOT NULL,
  `id_venta` int(11) NOT NULL,
  `cantidad` int(11) NOT NULL,
  `precio` decimal(10,2) NOT NULL
) ENGINE=InnoDB DEFAULT CHARSET=utf8mb4;

--
-- Volcado de datos para la tabla `detalle_venta`
--

INSERT INTO `detalle_venta` (`id`, `id_producto`, `id_venta`, `cantidad`, `precio`)
VALUES
(1, 4, 1, 2, '13.00'),
(2, 3, 1, 1, '16.00'),
(3, 4, 2, 3, '13.00');

-----

--
-- Estructura de tabla para la tabla `permisos`
--

CREATE TABLE `permisos` (
  `id` int(11) NOT NULL,
  `nombre` varchar(30) NOT NULL
) ENGINE=InnoDB DEFAULT CHARSET=utf8mb4;

--
-- Volcado de datos para la tabla `permisos`
--

INSERT INTO `permisos` (`id`, `nombre`) VALUES
(1, 'configuración'),
(2, 'usuarios'),
(3, 'clientes'),
(4, 'productos'),

```

```

(5, 'ventas'),
(6, 'nueva_venta');
-----
--
-- Estructura de tabla para la tabla `prediccion`
--

CREATE TABLE `prediccion` (
  `predicc` int(11) DEFAULT NULL
) ENGINE=InnoDB DEFAULT CHARSET=utf8mb4;
-----
--
-- Estructura de tabla para la tabla `producto`
--

CREATE TABLE `producto` (
  `codproducto` int(11) NOT NULL,
  `codigo` varchar(20) COLLATE utf8_spanish_ci NOT NULL,
  `descripcion` varchar(200) COLLATE utf8_spanish_ci NOT NULL,
  `precio` decimal(10,2) NOT NULL,
  `existencia` int(11) NOT NULL,
  `usuario_id` int(11) NOT NULL,
  `estado` int(11) NOT NULL DEFAULT 1
) ENGINE=InnoDB DEFAULT CHARSET=utf8 COLLATE=utf8_spanish_ci;
--
-- Volcado de datos para la tabla `producto`
--

INSERT INTO `producto` (`codproducto`, `codigo`, `descripcion`, `precio`, `existencia`,
`usuario_id`, `estado`) VALUES
(2, '321', 'Gaseosa kr', '10.00', 180, 1, 1),
(3, '654', 'Galletas', '16.00', 8, 1, 1),
(4, '987', 'Sandia', '13.00', 55, 1, 1),

```



```

(5, '12345', 'prueba', '1503.00', 50, 0, 1),
(6, '12345', 'prueba', '1503.00', 50, 0, 0);
-----

--
-- Estructura de tabla para la tabla `usuario`
--

CREATE TABLE `usuario` (
  `idusuario` int(11) NOT NULL,
  `nombre` varchar(100) COLLATE utf8_spanish_ci NOT NULL,
  `correo` varchar(100) COLLATE utf8_spanish_ci NOT NULL,
  `usuario` varchar(20) COLLATE utf8_spanish_ci NOT NULL,
  `clave` varchar(50) COLLATE utf8_spanish_ci NOT NULL,
  `estado` int(11) NOT NULL DEFAULT 1
) ENGINE=InnoDB DEFAULT CHARSET=utf8 COLLATE=utf8_spanish_ci;

--
-- Volcado de datos para la tabla `usuario`
--

INSERT INTO `usuario` (`idusuario`, `nombre`, `correo`, `usuario`, `clave`, `estado`)
VALUES
(1, 'Ezequiel Gomez', 'egomez1@unemi.edu.ec', 'admin',
'21232f297a57a5a743894a0e4a801fc3', 1),
(2, 'Vicente Jimenez', 'vjimeneze@unemi.edu.ec', 'admin',
'f4f068e71e0d87bf0ad51e6214ab84e9', 1);
-----

--
-- Estructura de tabla para la tabla `ventas`
--

CREATE TABLE `ventas` (
  `id` int(11) NOT NULL,
  `id_cliente` int(11) NOT NULL,

```

```

`producto` int(11) NOT NULL,
`total` decimal(10,2) NOT NULL,
`id_usuario` int(11) NOT NULL,
`fecha` timestamp NOT NULL DEFAULT current_timestamp() ON UPDATE
current_timestamp()
) ENGINE=InnoDB DEFAULT CHARSET=utf8mb4;
--
-- Volcado de datos para la tabla `ventas`
--

INSERT INTO `ventas` (`id`, `id_cliente`, `producto`, `total`, `id_usuario`, `fecha`)
VALUES
(1, 1, 0, '42.00', 1, '2021-05-16 14:35:54'),
(2, 1, 0, '39.00', 1, '2021-05-16 14:39:39');
-- Índices para tablas volcadas
--
--
-- Indices de la tabla `cliente`
--
ALTER TABLE `cliente`
  ADD PRIMARY KEY (`idcliente`);
--
-- Indices de la tabla `configuracion`
--
ALTER TABLE `configuracion`
  ADD PRIMARY KEY (`id`);
--
-- Indices de la tabla `detalle_permisos`
--
ALTER TABLE `detalle_permisos`
  ADD PRIMARY KEY (`id`);

```

```

--
-- Indices de la tabla `detalle_temp`
--
ALTER TABLE `detalle_temp`
  ADD PRIMARY KEY (`id`);
--
-- Indices de la tabla `detalle_venta`
--
ALTER TABLE `detalle_venta`
  ADD PRIMARY KEY (`id`);
--
-- Indices de la tabla `permisos`
--
ALTER TABLE `permisos`
  ADD PRIMARY KEY (`id`);
--
-- Indices de la tabla `producto`
--
ALTER TABLE `producto`
  ADD PRIMARY KEY (`codproducto`);
--
-- Indices de la tabla `usuario`
--
ALTER TABLE `usuario`
  ADD PRIMARY KEY (`idusuario`);
--
-- Indices de la tabla `ventas`
--
ALTER TABLE `ventas`
  ADD PRIMARY KEY (`id`);

```

```

--
-- AUTO_INCREMENT de las tablas volcadas
--
--
-- AUTO_INCREMENT de la tabla `cliente`
--
ALTER TABLE `cliente`
  MODIFY `idcliente` int(11) NOT NULL AUTO_INCREMENT,
  AUTO_INCREMENT=4;
--
-- AUTO_INCREMENT de la tabla `configuracion`
--
ALTER TABLE `configuracion`
  MODIFY `id` int(11) NOT NULL AUTO_INCREMENT, AUTO_INCREMENT=2;
--
-- AUTO_INCREMENT de la tabla `detalle_permisos`
--
ALTER TABLE `detalle_permisos`
  MODIFY `id` int(11) NOT NULL AUTO_INCREMENT, AUTO_INCREMENT=12;
--
-- AUTO_INCREMENT de la tabla `detalle_temp`
--
ALTER TABLE `detalle_temp`
  MODIFY `id` int(11) NOT NULL AUTO_INCREMENT, AUTO_INCREMENT=15;
--
-- AUTO_INCREMENT de la tabla `detalle_venta`
--
ALTER TABLE `detalle_venta`
  MODIFY `id` int(11) NOT NULL AUTO_INCREMENT, AUTO_INCREMENT=4;
--
-- AUTO_INCREMENT de la tabla `permisos`

```

```

--
ALTER TABLE `permisos`
  MODIFY `id` int(11) NOT NULL AUTO_INCREMENT, AUTO_INCREMENT=7;
-- AUTO_INCREMENT de la tabla `producto`
ALTER TABLE `producto`
  MODIFY `codproducto` int(11) NOT NULL AUTO_INCREMENT,
  AUTO_INCREMENT=7;

-- AUTO_INCREMENT de la tabla `usuario`
ALTER TABLE `usuario`
  MODIFY `idusuario` int(11) NOT NULL AUTO_INCREMENT,
  AUTO_INCREMENT=3;

--
-- AUTO_INCREMENT de la tabla `ventas`
--
ALTER TABLE `ventas`
  MODIFY `id` int(11) NOT NULL AUTO_INCREMENT, AUTO_INCREMENT=3;
COMMIT;
/*!40101 SET CHARACTER_SET_CLIENT=@OLD_CHARACTER_SET_CLIENT */;
/*!40101 SET CHARACTER_SET_RESULTS=@OLD_CHARACTER_SET_RESULTS
*/;
/*!40101 SET COLLATION_CONNECTION=@OLD_COLLATION_CONNECTION
*/;

```

Anexo 3

Códigos de algoritmos de predicción.

Regresión lineal

```

#Librerías
import numpy as np
import random
import pandas as pd
from scipy.stats import pearsonr
from sklearn import linear_model
from sklearn.metrics import mean_squared_error, r2_score

```

```

import matplotlib.pyplot as plt
%matplotlib inline

data=pd.read_csv("/content/drive/MyDrive/transactions.csv/transaccionescaba.csv")
data.head()

#Descripcion del Dataset como cantidad de transacciones, la media, desviación estándar, et
c.
data.describe()

''' Detección '''
# IQR RANGO INTERQUARTIL valores atípicos
Q1 = np.percentile(data['transacciones2015'], 25,
                    interpolation = 'midpoint')

Q3 = np.percentile(data['transacciones2016'], 75,
                    interpolation = 'midpoint')
IQR = Q3 - Q1

print("forma antigua: ", data.shape)

# Límite superior
upper = np.where(data['transacciones2015'] >= (Q3+1.5*IQR))
# Límite inferior
lower = np.where(data['transacciones2016'] <= (Q1-1.5*IQR))

''' Eliminar los valores atípicos'''
data.drop(upper[0], inplace = True)
data.drop(lower[0], inplace = True)

print("Nueva forma: ", data.shape)

#Separa las fechas y los datos
data1 = data[["transacciones2015", "transacciones2016"]]
data2 = data[["fecha2015", "fecha2016"]]
print(data.head())

#Presenta en un diagrama la distribución de los datos
plt.scatter(data2["fecha2015"], data1["transacciones2015"])
plt.scatter(data2["fecha2016"], data1["transacciones2016"])
plt.title("2015 VS 2016")
plt.xlabel("FECHA")
plt.ylabel("TRANSACCIONES")
plt.show()

```

```

corr_test = pearsonr(x = data['transacciones2015'], y = data['transacciones2016'])
print("Coeficiente de correlación de Pearson: ", corr_test[0])
#Mide la relacion estadistica entre dos variables continuas. Relación entre dos variables
print("P-value: ", corr_test[1])

#Separa los datos de entrenamiento y prueba
#80% para entrenamiento y el 20% para prueba
num = int(len(data)*0.8)

train= data1[:num]
test = data1[num:]

print("Data: ",len(data1))
print("Train: ",len(train))
print("Test: ",len(test))

#Se llama al modelo de regresión lineal simple de Scikit-learn
regr = linear_model.LinearRegression()

train_x = np.array(train[["transacciones2015"]])
train_y = np.array(train[["transacciones2016"]])

regr.fit(train_x,train_y)

coeficientes = regr.coef_
intercept = regr.intercept_

print("Slope: ",coeficientes[0])
err = 91.5
#Print("Intercept: ",intercept)

#Se presenta los datos del dataset inicial
predited_data = regr.predict(data1[["transacciones2016"]])
predited_data[0:5]

#Se predice con los datos de entrenamiento
predited_train = regr.predict(train[["transacciones2016"]])
predited_train[0:5]

#Se predice con el algoritmo ya entrenado
predited_test = regr.predict(test[["transacciones2016"]])
predited_test[0:5]

#Se presenta los datos iniciales
plt.scatter(data1["transacciones2016"],data1["transacciones2015"])

```

```
plt.plot(data["transacciones2016"],predited_data,color="red")
plt.title("DISTRIBUCION DE DATOS ANTES DEL ENTRENAMIENTO")
plt.xlabel("FECHA")
plt.ylabel("TRANSACCIONES")
plt.show()
```

#Se presentan los datos del entrenamiento

```
plt.scatter(train["transacciones2016"],train["transacciones2015"])
plt.plot(train["transacciones2016"],predited_train,color="red")
plt.title("DISTRIBUCION DE DATOS DESPUES DEL ENTRENAMIENTO")
plt.xlabel("FECHA")
plt.ylabel("TRANSACCIONES")
plt.show()
```

#Se presenta los datos después del entrenamiento

```
plt.scatter(test["transacciones2015"],test["transacciones2016"])
plt.plot(test["transacciones2015"],predited_test,color="red")
plt.title("DISTRIBUCION DE DATOS FINAL")
plt.xlabel("FECHA")
plt.ylabel("TRANSACCIONES")
plt.show()
```

#Se realiza una comparación con los datos reales y predichos

```
A_P_data = pd.DataFrame({"Actual":data1["transacciones2016"],"Prediccion":predited_test[:][0][0]})
print(A_P_data.head())
```

#Se muestra en una grafica de barras dichos datos

```
A_P_data.head(10).plot(kind='bar',figsize=(12,6))
plt.show()
```

#Se presentan los valores residuales y el error medio cuadrático

```
test_x = np.array(test[["transacciones2015"]])
test_y = np.array(test[["transacciones2016"]])
```

```
predited_y = regr.predict(test_x)
```

```
res = (predited_y - test_y)
```

```
rmse = err
```

```
RSS = (res*res).sum()
```

```
print("Redidual: ",RSS)
```

```
print(f"El error (rmse) de test es: {rmse}")
```

Árbol de decisión


```

# Tratamiento de datos
# -----
import numpy as np
import pandas as pd

# Gráficos
# -----
import matplotlib.pyplot as plt

# Preprocesado y modelado
# -----

from sklearn.model_selection import train_test_split
from sklearn.tree import DecisionTreeRegressor
from sklearn.tree import plot_tree
from sklearn.tree import export_graphviz
from sklearn.tree import export_text
from sklearn.model_selection import GridSearchCV
from sklearn.metrics import mean_squared_error

# Configuración warnings
# -----
import warnings
warnings.filterwarnings('once')

# Box Plot
import seaborn as sns
sns.boxplot(data['transacciones'])

''' Detection '''
# IQR RANGO INTERQUARTIL valores atípicos
#
Q1 = np.percentile(data['transacciones'], 25,
                    interpolation = 'midpoint')

Q3 = np.percentile(data['transacciones'], 75,
                    interpolation = 'midpoint')
IQR = Q3 - Q1

print("Old Shape: ", data.shape)

# Upper bound
upper = np.where(data['transacciones'] >= (Q3+1.5*IQR))
# Lower bound
lower = np.where(data['transacciones'] <= (Q1-1.5*IQR))

''' Removing the Outliers '''
data.drop(upper[0], inplace = True)

```

```

data.drop(lower[0], inplace = True)

print("New Shape: ", data.shape)

# División de los datos en train y test
# -----
X_train, X_test, y_train, y_test = train_test_split(
    data.drop(columns = "fecha"),
    data['transacciones'],
    random_state = 30000
)

# Creación del modelo
# -----
modelo = DecisionTreeRegressor(
    max_depth = 5, #profundidad máxima que puede alcanzar el árbol
    random_state = 30000 #semilla para que los resultados sean reproducibles.
    #Tiene que ser un valor entero.
)

# Entrenamiento del modelo
# -----
modelo.fit(X_train, y_train)

# Estructura del árbol creado
# -----
fig, ax = plt.subplots(figsize=(12, 5))

print(f"Profundidad del árbol: {modelo.get_depth()}")
print(f"Número de nodos terminales: {modelo.get_n_leaves()}")

plot = plot_tree(
    decision_tree = modelo,
    feature_names = data.drop(columns = "fecha").columns,
    class_names = 'transacciones',
    filled = True,
    impurity = False,
    fontsize = 10,
    precision = 2,
    ax = ax
)

# Pruning (const complexity pruning) por validación cruzada
# -----
# Valores de ccp_alpha evaluados
param_grid = {'ccp_alpha':np.linspace(0, 80, 20)}

# Búsqueda por validación cruzada
grid = GridSearchCV(

```

```

# El árbol se crece al máximo posible para luego aplicar el pruning
estimator = DecisionTreeRegressor(
    max_depth = None,
    min_samples_split = 3,
# número mínimo de observaciones que debe de tener un nodo para que pueda dividirse
    #
    min_samples_leaf = 1,
    random_state = 10000
),
param_grid = param_grid,
cv = 10,
refit = True,
return_train_score = True
)

grid.fit(X_train, y_train)

fig, ax = plt.subplots(figsize=(6, 3.84))
scores = pd.DataFrame(grid.cv_results_)
scores.plot(x='param_ccp_alpha', y='mean_train_score', yerr='std_train_score', ax=ax)
scores.plot(x='param_ccp_alpha', y='mean_test_score', yerr='std_test_score', ax=ax)
ax.set_title("Error de validacion cruzada vs hiperparámetro ccp_alpha");

# Mejor valor ccp_alpha encontrado
# -----
grid.best_params_

# Estructura del árbol final
modelo_final = grid.best_estimator_
print(f"Profundidad del árbol: {modelo_final.get_depth()}")
print(f"Número de nodos terminales: {modelo_final.get_n_leaves()}")

fig, ax = plt.subplots(figsize=(7, 5))
plot = plot_tree(
    decision_tree = modelo_final,
    feature_names = data.drop(columns = "fecha").columns,
    class_names = 'transacciones',
    filled = True,
    impurity = False,
    ax = ax
)

# Error de test del modelo inicial
predicciones = modelo.predict(X = X_test)
rmse = mean_squared_error(
    y_true = y_test,
    y_pred = predicciones,
    squared = False
)

```

```

    )
print(f"El error (rmse) de test es: {rmse}")

# Error de test del modelo final (tras aplicar pruning)
predicciones = modelo_final.predict(X = X_test)
rmse = mean_squared_error(
    y_true = y_test,
    y_pred = predicciones,
    squared = False
)
print(f"El error (rmse) de test es: {rmse}")

```

Redes neuronales

```

#Librería
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
plt.rcParams['figure.figsize'] = (16, 9)
plt.style.use('fast')
#Estructura y modelo del algoritmo
from keras.models import Sequential
from keras.layers import Dense,Activation,Flatten
from sklearn.preprocessing import MinMaxScaler
#Carga del dataset
dataset = pd.read_csv('/content/drive/MyDrive/transactions.csv/transacciones.csv', parse_
dates=[0], header=None,index_col=0, squeeze=True,names=['fecha','transacciones'])
dataset1.head()

''' Detection '''
# IQR RANGO INTERQUARTIL valores atípicos
Q1 = np.percentile(dataset1['transacciones'], 25,
                    interpolation = 'midpoint')

Q3 = np.percentile(dataset1['transacciones'], 75,
                    interpolation = 'midpoint')
IQR = Q3 - Q1

print("Old Shape: ", dataset1.shape)

# Upper bound
upper = np.where(dataset1['transacciones'] >= (Q3+1.5*IQR))
# Lower bound
lower = np.where(dataset1['transacciones'] <= (Q1-1.5*IQR))

''' Removing the Outliers '''

```

```
dataset1.drop(upper[0], inplace = True)
dataset1.drop(lower[0], inplace = True)
```

```
print("New Shape: ", dataset1.shape)
```

```
#Describe lo que contiene nuestro Dataset.
```

```
#Son un total de 46322 registros, la media de venta de unidades es de 1492 y un desvío de 680
```

```
dataset1.describe()
```

```
#Presentamos los valores mensuales en la grafica
```

```
plt.plot(meses['2015'].values)
```

```
plt.plot(meses['2016'].values)
```

```
plt.plot(meses['2017'].values)
```

```
#Descripción de las variables
```

```
#data: Secuencia de observaciones como una lista o matriz 2D NumPy. Requerido.
```

```
#n_in: Número de observaciones de retraso como entrada ( X ). Los valores pueden estar entre [1..len(datos)] Opcional. El valor predeterminado es 1.
```

```
#n_out: Número de observaciones como salida ( y ). Los valores pueden estar entre [0..len(datos)-1]. Opcional. El valor predeterminado es 1.
```

```
#dropnan : booleano si se eliminan o no las filas con valores NaN. Opcional. El valor predeterminado es Verdadero.
```

```
#La función devuelve un único valor.
```

```
#return:Pandas DataFrame de series encuadradas para aprendizaje supervisado
```

```
PASOS=7
```

```
# convertir series en aprendizaje supervisado
```

```
def series_to_supervised(data, n_in=1, n_out=1, dropnan=True):
```

```
    n_vars = 1 if type(data) is list else data.shape[1]
```

```
    dataset = pd.DataFrame(data)
```

```
    cols, names = list(), list()
```

```
    # secuencia de entrada (t-n, ... t-1)
```

```
    for i in range(n_in, 0, -1):
```

```
        cols.append(dataset.shift(i))
```

```
        names += [('var%d(t-%d)' % (j+1, i)) for j in range(n_vars)]
```

```
    # secuencia de pronóstico (t, t+1, ... t+n)
```

```
    for i in range(0, n_out):
```

```
        cols.append(dataset.shift(-i))
```

```
        if i == 0:
```

```
            names += [('var%d(t)' % (j+1)) for j in range(n_vars)]
```

```
        else:
```

```

        names += [('var%d(t+%d)' % (j+1, i)) for j in range(n_vars)]
# unir todo
agg = pd.concat(cols, axis=1)
agg.columns = names
# soltar filas con valores NaN
if dropnan:
    agg.dropna(inplace=True)
return agg

# Load dataset
values = dataset.values
# Asegurar que todos los datos sean flotantes
values = values.astype('float32')
# Normalizar características
scaler = MinMaxScaler(feature_range=(-1, 1))
values=values.reshape(-1, 1) # esto lo hacemos porque tenemos 1 sola dimension
scaled = scaler.fit_transform(values)
# Marco como aprendizaje supervisado
reframed = series_to_supervised(scaled, PASOS, 1)
reframed.head()

# Dividir en conjuntos de entrenamiento y prueba
values = reframed.values
n_train_days = 37057 - (9265+PASOS)
train = values[:n_train_days, :]
test = values[n_train_days:, :]
# dividir en entrada y salida
x_train, y_train = train[:, :-1], train[:, -1]
x_val, y_val = test[:, :-1], test[:, -1]
# # remodelar la entrada para que sea 3D [muestras, intervalos de tiempo, características]
x_train = x_train.reshape((x_train.shape[0], 1, x_train.shape[1]))
x_val = x_val.reshape((x_val.shape[0], 1, x_val.shape[1]))
print(x_train.shape, y_train.shape, x_val.shape, y_val.shape)

#Modelo

def crear_modeloFF():
    model = Sequential()
    model.add(Dense(PASOS, input_shape=(1,PASOS),activation='tanh'))
    model.add(Flatten())
    model.add(Dense(1, activation='tanh'))
    model.compile(loss='mean_absolute_error',optimizer='Adam',metrics=["mse"])
    model.summary()
    return model

#Entrenamiento
EPOCHS=30

```

```

model = crear_modeloFF()
history=model.fit(x_train,y_train,epochs=EPOCHS,validation_data=(x_val,y_val),batch_size=PASOS)

```

#En la gráfica observamos la comparacion del modelo predicho y el real.

```

results=model.predict(x_val)
plt.scatter(range(len(y_val)),y_val,c='g')
plt.scatter(range(len(results)),results,c='r')
plt.title('validate')
plt.show()

```

#Gráficos de loss

```

plt.plot(history.history['loss'])
plt.title('loss')
plt.plot(history.history['val_loss'])
loss = history.history['val_loss']
plt.title('validate loss')
plt.show()

```

#Grafico de Accuracy

```

plt.title('Accuracy')
plt.plot(history.history['mse'])
plt.show()
acuracyyy = history.history['mse']

```

#Se comparan los datos reales con los de la predicción

```

compara = pd.DataFrame(np.array([y_val, [x[0] for x in results]]).transpose()
compara.columns = ['real', 'prediccion']

```

```

inverted = scaler.inverse_transform(compara.values)

```

```

compara2 = pd.DataFrame(inverted)
rmse = 25.1
compara2.columns = ['real', 'prediccion']
compara2['diferencia'] = compara2['real'] - compara2['prediccion']
compara2.head()

```

#Se comparan los datos reales con los de la predicción

```

compara2['real'].plot()
compara2['prediccion'].plot()
real = compara2['real']
prediccion = compara2['prediccion']
real[50:100].plot()
prediccion[50:100].plot()

```

#Mostramos el error cuadrático medio que nos genera el modelo.

```

print(f"El error (rmse) de test es: {rmse}")

```

