



UNIVERSIDAD CÉSAR VALLEJO

**FACULTAD DE INGENIERÍA Y ARQUITECTURA
ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS**

**Machine Learning para predecir el riesgo crediticio de un cliente
en la Empresa FUTECH PERU S.A.C, 2022**

TESIS PARA OBTENER EL TÍTULO PROFESIONAL DE:

Ingeniero de Sistemas

AUTOR:

Yaranga Vite, Italo Paul (orcid.org/0000-0001-7380-135X)

ASESOR:

Mg. Quiñones Nieto, Yamil Alexander (orcid.org/0000-0003-4474-0556)

LÍNEA DE INVESTIGACIÓN:

Sistemas de Información y Comunicaciones

LÍNEA DE RESPONSABILIDAD SOCIAL UNIVERSITARIA:

Desarrollo económico, empleo y emprendimiento

LIMA – PERÚ
2022

Dedicatoria

Dedico este trabajo con todo mi amor y cariño a mi Familia:

Mi esposa Zarú Franzu, mis hijos Italo Matias y Gianfranco Ariel.

Asimismo, también se lo dedico de manera especial a mi Madre Anita Vite, a mis hermanas Victoria, Patricia y mi hermano Elmo y a mis tías Rosa, Cecilia.

Agradecimiento

Realizar un trabajo de investigación requiere el apoyo de muchas personas, agradecido de forma muy especial a mi familia por toda la paciencia y comprensión, a mi Madre por todo el apoyo brindado y no puedo dejar de mencionar a mi amigo Gustavo Coronel y Alfonso Barrera que sin su apoyo hubiera sido muy dificultoso terminar este proyecto.

Índice de contenidos

	Página
Carátula	
Dedicatoria	ii
Agradecimiento	iii
Índice de contenidos	iv
Índice de tablas	v
Índice de gráficos y figuras	vii
Resumen	ix
Abstract	x
I. INTRODUCCIÓN	1
II. MARCO TEÓRICO	5
III. METODOLOGÍA	20
3.1. Tipo y diseño de investigación	21
3.2. Variables y operacionalización	21
3.3. Población, muestra y muestreo	22
3.4. Técnicas e instrumentos de recolección de datos	22
3.5. Procedimientos	22
3.6. Método de análisis de datos	23
3.7. Aspectos éticos	23
IV. RESULTADOS	24
V. DISCUSIÓN	40
VI. CONCLUSIONES	42
VII. RECOMENDACIONES	44
REFERENCIAS	46
ANEXOS	57

Índice de tablas

Tabla 1	Matriz de confusión – Support Vector Machine	35
Tabla 2	Matriz de confusión – Random Forest	35
Tabla 3	Matriz de confusión – Naibes Bayes	36
Tabla 4	Matriz de confusión – K Nearest Neighbor	36
Tabla 5	Matriz de confusión – Decision Tree	37
Tabla 6	Tabla cruzada cálculo de precisión con algoritmo Support Vector Machine	38
Tabla 7	Tabla cruzada cálculo de precisión con algoritmo Random Forest	38
Tabla 8	Tabla cruzada – cálculo de precisión con algoritmo Naibes Bayes	39
Tabla 9	Tabla cruzada – cálculo de precisión con algoritmo K Nearest Neighbor	39
Tabla 10	Tabla cruzada – cálculo de precisión con algoritmo Decision Tree	40
Tabla 11	Resumen de cuadro comparativo de algoritmos	40
Tabla 12	Tabla cruzada – cálculo de sensibilidad con algoritmo Support Vector Machine	41
Tabla 13	Tabla cruzada – cálculo de sensibilidad con algoritmo Random Forest	41
Tabla 14	Tabla cruzada – cálculo de sensibilidad con algoritmo Naibes Bayes	42
Tabla 15	Tabla cruzada – cálculo de sensibilidad con algoritmo K Nearest Neighbor	42
Tabla 16	Tabla cruzada – cálculo de sensibilidad con algoritmo Decision Tree	43
Tabla 17	Comparación de los resultados mediante el indicador sensibilidad	43
Tabla 18	Tabla cruzada – cálculo de especificidad con algoritmo Support Vector Machine	44
Tabla 19	Tabla cruzada – cálculo de especificidad con algoritmo Random Forest	44
Tabla 20	Tabla cruzada – cálculo de especificidad con algoritmo Naibes Bayes	45
Tabla 21	Tabla cruzada – cálculo de especificidad con algoritmo K Nearest Neighbor	45
Tabla 22	Tabla cruzada – cálculo de sensibilidad con algoritmo Decision Tree	46
Tabla 23	Comparación de los resultados mediante el indicador especificidad	46
Tabla 24	Resumen de cuadro comparativo de algoritmos	47
Tabla 25	Medida de Kappa de Cohen – Support Vector Machine	47
Tabla 26	Medida de Kappa de Cohen – Random Forest	48

Tabla 27	Medida de Kappa de Cohen – Neibes Bayes	48
Tabla 28	Medida de Kappa de Cohen – K Nearest Neighbor	48
Tabla 29	Medida de Kappa de Cohen – Decision Tree	48
Tabla 30	Resumen de cuadro comparativo de los algoritmos de las medidas del nivel de concordancia de utilizando la medida de acuerdo de Kappa Cohen	49

Índice de figuras

Figura 1	Entrenamiento de modelos de Machine Learning	20
Figura 2	Matriz de confusión	20
Figura 3	Modelo predictivo Decision Tree	23
Figura 4	Modelo predictivo Árboles de decisión RandomForest	24
Figura 5	Modelo predictivo Maquina de Vectores de Soporte	24
Figura 6	Modelo predictivo del Algoritmo Naive Bayes	25
Figura 7	Representación del Algoritmo k vecinos más cercanos.	25
Figura 8	Etapas de la metodología KDD	27
Figura 9	El diagrama del diseño experimental tipo pre-experimental	31
Figura 10	Nivel de Kappa de Cohen	37
Figura 11	Cuestionario utilizando Google Forms	85
Figura 12	Definición de variables – SPSS Statistics	86
Figura 13	Datos importados de Google Forms a Excel	89
Figura 14	Datos importados de Excel a SPSS Statistics	89
Figura 15	Cuantificación de ítems no contestados parte 1 - SPSS Statistics	90
Figura 16	Cuantificación de ítems no contestados parte 2 - SPSS Statistics	90
Figura 17	No se evidencio que hubo ítems nulos en el total de registros (500) parte 2 - SPSS Statistics	91
Figura 18	Nodo origen de datos	92
Figura 19	Transformación de datos con nodo derivar - SPSS Modeler	92
Figura 20	Ejemplo de la transformación de datos - SPSS Modeler	93
Figura 21	Aplicación del nodo tipo - SPSS Modeler	93
Figura 22	Ejemplo del nodo tipo - SPSS Modeler	94
Figura 23	Aplicación del nodo filtro - SPSS Modeler	94
Figura 24	Algoritmos de aprendizaje automático - SPSS Modeler	95
Figura 26	Precisión de modelo utilizando Decision Tree – SPSS Modeler	96

Figura 27	Precisión de modelo utilizando Random Forest - SPSS Modeler	96
Figura 28	Precisión de modelo utilizando Support Vector Machine	97
Figura 29	Precisión de modelo utilizando Naibes Bayes - SPSS Modeler	97
Figura 30	Precisión de modelo utilizando K Vecino más Cercano-SPSS Modeler	98
Figura 31	Entorno de la Suite de Anaconda	99
Figura 32	Entorno de Jupiter Notebook	100
Figura 33	Entorno de Servidor local de Jupiter Notebook	101
Figura 34	Diseño del modelo Físico de la base de datos	102
Figura 35	Arquitectura del sistema o aplicaciones web	103
Figura 36	Verificación de credenciales para el acceso al sistema	104
Figura 37	Interfaz de principal del uso de la Solución web	105
Figura 38	Interfaz de la aplicación para la predicción del riesgo crediticio del cliente	106
Figura 39	Resultados de la ejecución de la aplicación web para la predicción del riesgo crediticio del cliente	107

Resumen

En el presente trabajo de investigación se utilizó los modelos de Machine Learning para predecir el riesgo crediticio de los clientes de una empresa en estudio, se utilizó la metodología KDD, así mismo herramientas como SPSS statistic y SPSS Modeler para el uso de los modelos predictivo.

El objetivo de esta investigación es determinar en qué porcentaje Machine Learning permite predecir el riesgo crediticio de los clientes con precisión, sensibilidad y especificidad, con el fin de poder identificar a los clientes con probabilidad de alto o bajo riesgo crediticio.

En esta investigación se utilizó una población de 500 clientes, así mismo se usó la totalidad de la población como muestra. Por otro lado, el estudio es de tipo aplicada, con un diseño de investigación experimental de tipo pre-experimental de un solo grupo, ya que luego de aplicar Machine Learning se podrá observar los resultados y realizar la medición.

Como resultado en relación a los indicadores de precisión, sensibilidad y especificidad para los algoritmos Support Vector Machine, Random Forest, Naïves Bayes, K Nearest Neighbor, Decision Tree, se valida que Machine Learning permite predecir el riesgo crediticio de los clientes de la empresa de estudio, así mismo el algoritmo con mejores resultados para esta casuística fue Support Vector Machine con un valor de 99,8%.

Palabras clave: Machine Learning, riesgo crediticio, métricas de precisión, inteligencia artificial.

Abstract

In the present research work, Machine Learning models were used to predict the credit risk of the clients of a company under study, the KDD methodology was used, as well as tools such as SPSS statistic and SPSS Modeler for the use of predictive models.

The objective of this research is to determine to what percentage Machine Learning allows predicting the credit risk of clients with precision, sensitivity and specificity, in order to be able to identify clients with a probability of high or low credit risk.

In this investigation, a population of 500 clients was used, likewise the entire population was used as a sample. On the other hand, the study is of an applied type, with a pre-experimental experimental research design of a single group, since after applying Machine Learning the results can be observed and the measurement carried out.

As a result, in relation to the indicators of precision, sensitivity and specificity for the Support Vector Machine, Random Forest, Naives Bayes, K Nearest Neighbor, Decision Tree algorithms, it is validated that Machine Learning allows predicting the credit risk of the company's clients. study, likewise the algorithm with the best results for this casuistry was Support Vector Machine with a value of 99.8%.

Keywords: Machine Learning, credit risk, precision metrics, artificial intelligence.

I. INTRODUCCIÓN

En el ámbito internacional, nos encontramos con el constante cambio de la economía hace que cada vez las empresas sean más complejas en cada actividad que realizan, Las empresas que se encuentra en rubro financiero exige un alto manejo de predicción y calificación de los clientes, con la finalidad de prevenir perjuicios futuros que afecten a las organizaciones. Según García (2017). La información es indispensable cuando se realiza un pronóstico. De tal manera, Que el pronóstico contribuye a la planificación que permite producir un modelo. Un pronóstico se puede realizar para saber que sucederá con los datos finales que se utilicen en diferente ámbito. La necesidad de utilizar pronósticos en el ámbito de riesgos crediticio es de vital importancia.

A nivel nacional actualmente, las empresas están reportando índices de morosidad de sus clientes muy altos por encima del promedio en la economía nacional, por tal motivo se han implementado diferentes formas tecnológicas de predicción de clientes lo cual va a permitir reducir el índice de morosidad de los clientes.

Las empresas aprovechan la tecnología actual para seguir creciendo económicamente, los sistemas de información permiten automatizar los procesos de una organización, la compra de una solución informática permite gestionar las operaciones crediticias desarrollada en diferentes plataformas de trabajo. los sistemas expertos vinculados en un entorno empresarial utilizando Inteligencia Artificial (IA). Esta especialidad permite automatizar el trabajo intelectual en diferentes ámbitos (Russell y Norvig, 2017).

Los modelos predictivos que utilizan la Inteligencia Artificial, permiten “emular los procesos de asimilación de conocimiento y reproducen el comportamiento que realiza el órgano humano del cerebro: Las Redes Neuronales Artificiales” (Hilera y Martínez, 2015). Arredondo Vidal (2016) determina mediante modelos basados en los complejos sistemas nerviosos de seres vivos como animales y personas con abundantes interconexiones.

En el ámbito local la empresa FUTECH PERU S.A.C, ofrece soluciones tecnológicas y la venta de equipos de procesamiento de datos, lo cual ha permitido que la empresa utilice tecnología de vanguardia como Machine Learning para realizar la predicción de riesgos crediticio, en este contexto

Navarro y Ríos (2015). Realizaron un estudio en el desarrollo de una solución de Business Intelligence con la finalidad de brindar un mejor soporte a la toma de decisiones.

La problemática principal que se presenta en la empresa FUTECH PERU S.A.C en la actualidad, se basa por la falta de un mecanismo de verificación de la información actualizada del cliente en un proceso de admisión comercial de operaciones donde no estamos seguros si debemos autorizar las o no, por consiguiente, la inadecuada calificación de un posible cliente con lleva a varios factores negativos que afectan el patrimonio de la organización, por otro lado la decisión de otorgar un crédito a un cliente está sujeta a la subjetividad del comité encargado conocido también como sesgo cognitivo, en ese sentido es indispensable el uso de tecnologías modernas actuales como Machine Learning que mediante un conjunto de algoritmos predictivos permiten que el proceso de verificación sea automatizado y permitan procesar, analizar y aprender de los datos con el fin de obtener una mejor precisión, sensibilidad y especificidad al momento de predecir el riesgo crediticio de un cliente.

Por ello se plantea el problema general ¿En qué medida Machine Learning permite predecir el riesgo crediticio de un cliente en la empresa FUTECH PERÚ SAC, 2022?

Y como problemas específicos, ¿En qué medida Machine Learning permite predecir con precisión el riesgo crediticio de un cliente en la empresa FUTECH PERÚ SAC, 2022?, ¿En qué medida Machine Learning permite predecir con sensibilidad el riesgo crediticio de un cliente en la empresa FUTECH PERÚ SAC, 2022?, ¿En qué medida Machine Learning permite predecir con especificidad el riesgo crediticio de un cliente en la empresa FUTECH PERÚ SAC, 2022?

La justificación teórica, se centra básicamente en aplicar de diferentes conceptos relacionados en la tecnología de software y modelo predictivos que van ayudar a las empresas a mejorar sus operaciones comerciales.

La justificación práctica, está orientada en Machine Learning donde se fundamenta en modelos predictivos con diferentes características que nos van a

permitir aplicar los en los modelos predictivos para poder obtener mejores resultados en riesgos crediticios.

La justificación económica, se centra en uso de la tecnología y los modelos predictivos los cuales van a permitir mejorar los riesgos crediticios para cual les permitirá alcanzar una economía sostenible.

En esta investigación se planteó como objetivo general: Aplicar Machine Learning para predecir el riesgo crediticio de un cliente en la empresa FUTECH PERÚ SAC, 2022.

Y como objetivos específicos: Determinar en qué porcentaje Machine Learning permite predecir con precisión el riesgo crediticio de un cliente en la empresa FUTECH PERÚ SAC, 2022., Determinar en qué porcentaje Machine Learning permite predecir con sensibilidad el riesgo crediticio de un cliente en la empresa FUTECH PERÚ SAC, 2022., Determinar en qué porcentaje Machine Learning permite predecir con especificidad el riesgo crediticio de un cliente en la empresa FUTECH PERÚ SAC, 2022.

Para demostrar el objetivo general se planteó como hipótesis general:

Machine Learning predice significativamente el riesgo crediticio de un cliente en la empresa FUTECH PERÚ SAC, 2022.

Y como hipótesis específicas: Machine Learning predice significativamente con precisión el riesgo crediticio de un cliente en la empresa FUTECH PERÚ SAC, 2022., Machine Learning predice significativamente con sensibilidad el riesgo crediticio de un cliente en la empresa FUTECH PERÚ SAC, 2022., Machine Learning predice significativamente con especificidad el riesgo crediticio de un cliente en la empresa FUTECH PERÚ SAC, 2022.

II. MARCO TEÓRICO

En la coyuntura actual se están desarrollando investigaciones que utilizan el Machine Learning aplicado a diversas directrices en el ámbito nacional e internacional, en ese sentido a continuación se detallan investigaciones encontradas.

Ossa & Jaramillo (2021), realizaron una investigación sobre *“Machine Learning para la evaluación del riesgo crediticio”*, tiene como objetivo comparar un modelo predictivo de regresión logística con algunos de las técnicas de predicción de aprendizaje automático, la metodología predictiva se centró en identificar factores de riesgo, permite realizar una evaluación compacta de posible incumplimiento, señala los factores de riesgo que indican las probabilidades. La recolección de los datos fue suministrada en una base de datos histórica que contine información de los pagos realizados por los clientes correspondiente a una cantidad de 15,060 registros y 18 variables, se obtuvo como resultado que el modelo Random Forest brinda una mejor exactitud en 90% de acuerdo a las métricas evaluadas.

Francés (2020), realizo una investigación referente al *“Impacto de Machine Learning en el Sistema Financiero, mediante el método inductivo”*, que sostiene que los métodos no supervisados permiten realizar un análisis y un procesamiento de datos, a diferencia de los métodos supervisados detectan las posibles predicciones de riesgos que pueden desencadenar el incumplimiento. Mientras los algoritmos sean más sofisticados y tengan una mejor performance van a permitir mejores predicciones; la recolección de datos que se utilizó fue mediante fuentes de aplicaciones comerciales financiera que brindan información, El estudio llego a la conclusión, donde indican que es uso de aplicaciones con algoritmos de aprendizaje automático, aumentan la productividad en un 60% de los colaboradores, reduciendo el 20% del margen de esfuerzo de trabajo, reduciendo los de tareas que se realizan con frecuencia para realizar las operaciones más importantes, e incrementará la importancia de las habilidades humanas.

Ruiz (2020) realiza un estudio de *“Desarrollo una aplicación utilizando Machine Learning que permite realizar predicciones de consumo de suministros eléctricos”*, se desarrolló mediante una metodología predictiva, la recopilación de datos se realizó mediante datos históricos de una serie de consumos

proporcionados por las comercializadoras. La investigación realizada resume indicando que el uso de machine learning va a cambiar el sector de forma positiva, permitiendo a los asalariados reducir su tiempo en un 20% y utilizar las máquinas con la finalidad de realizar las operaciones rutinarias. El sistema también detecta inconsistencia de datos de un conjunto de correlaciones momentáneas de consumo para su posterior verificación y medicación. El sistema se sustenta de datos de histórico donde se realiza un proceso de entrenamiento con cada un suministro eléctrico proporcionado por sus fuentes.

Aceituno (2019), realizó una investigación de un *“Modelo de predicción para la evaluación crediticia usando Machine Learning en un sector micro financiero”*, la investigación tuvo como objetivo utilizar un modelo de predicción de riesgo crediticio mediante un modelo de Machine Learning que permita brindar una aproximación de asertividad al momento de otorgar financiamientos. utilizó una metodología predictiva, con respecto a la recopilación de los datos se utilizaron una base de datos con datos históricos de 15,015 de clientes. Los datos finales con respecto a la precisión del modelo de red neuronal son de 93.7% y la regresión logística del 86.1%.

Ortiz & Haro (2017), realizaron un estudio denominado *“Modelo tecnológico de análisis predictivo basado en Machine Learning para evaluación de riesgo crediticio”*, tiene como objetivo aplicar un modelo predictivo sobre una plataforma Cloud aplicado en la evaluación de riesgo crediticio. es un modelo predictivo basado en aprendizaje automático para el uso de proceso de verificación de estado crediticio, tomaron en cuenta un proceso donde se le agrega valor a los datos de entrada que fueron obtenidos desde una base de datos histórica hasta encontrar la predicción en tanto se optimiza el modelo predictivo, se definieron las siguientes etapas como la extracción, ingestión, preparación, análisis, transformación, y consumo de los 5 del análisis predictivo se establecieron en gestión, cumplimiento la autenticación y la autorización. Los resultados obtenidos tienen una precisión del 93.3% mediante el modelo de árboles de decisión de doble clasificación a diferencia del modelo de regresión logística.

Rodríguez & Miñano (2017), realizaron estudio del *“Desarrollo de una aplicación informática basada en un modelo de Machine Learning para mejorar la evaluación de préstamos crediticios”*, como objetivo se plantearon mejorar la

verificación de los préstamos crediticios utilizando un software que utiliza modelo de Machine Learning, la metodología que utilizaron en la Investigación fue pre experimental, donde se fundamentada en un modelo aprendizaje automático, para mejorar las evaluaciones crediticias, utilizaron una base de datos histórica de 1460 registro de préstamos realizados en el año 2015, Los resultados obtenidos se centraron en un 98.8% de sensibilidad, 25% de exactitud, 87% de eficacia, inherentes a los indicadores de la variable causal.

En relación a Machine Learning, Hurwitz y Kirsch (2018, p. 20) sostiene que la preparación con un conjunto de datos, antes de la implementación del modelo, debe centrarse en el aprendizaje iterativo. Un proceso de modelado iterativo desarrollado dentro de la cobertura en línea incrementa la relación entre las entidades de datos. La complejidad de este contexto, es posible omitir los vínculos y conducta basados en la competencia humana. Las competencias que han completado la fase de formación se pueden utilizar en tiempo real. En el aspecto del aprendizaje supervisado de la inteligencia artificial, la caracterización de los datos de entrada que ayudan a evaluar sus competencias en un tiempo limitado, a diferencia de los desarrolladores humanos que se enfocan en determinar qué hacer. el análisis de la inteligencia artificial seleccionada para el análisis cuántico y los resultados se incluyen en la decisión tomadas.

Izarra (2016), quién realizó un estudio de “*Riesgo Crediticio Y La Morosidad En La Cooperativa De Ahorro Y Crédito*”, que tiene como objetivo de analizar de la relación entre el riesgo crediticio y la relación directa con la morosidad en una entidad financiera el año 2014. En la investigación es tipo aplicada donde se obtuvo el 57% de los ejecutivos señalan con respecto al riesgo crediticio que está en un nivel alto, asimismo, hay una relación negativa con la morosidad en un -88.53%, Se evidencio la relación entre las variables de estudio; llegando a determinar que la morosidad explica en un 77.57% de la evaluación del riesgo crediticio.

Borrero y Bedolla (2020), realizaron un estudio de “Predicción de riesgo crediticio en Colombia usando técnicas de inteligencia artificial”, su objetivo reconocer posibles clientes de alto riesgo crediticio, Se utiliza modelos utilizando tres técnicas de predicción de Machine Learning, se utilizaron datos históricos

para utilizar los modelos de predicciones, el rendimiento observado indica que los árboles de decisión son más precisos para el uso riesgo crediticio con un valor de una métrica de área bajo de la curva ROC de 88.29%.

Con respecto a las teorías utilizadas en el estudio podemos mencionar como: Según Russell y Norving, 2004. La inteligencia artificial sintetiza y automatiza tareas intelectuales y es, por lo tanto, relevante para cualquier ámbito de la actividad intelectual humana.

Según Torra, 2011. El aprendizaje automático es el rendimiento de un algoritmo para incrementarse si el algoritmo aprende de la actividad ejecutada y de sus errores, debe existir un control de las etapas de entrenamiento y desarrollo del modelo.

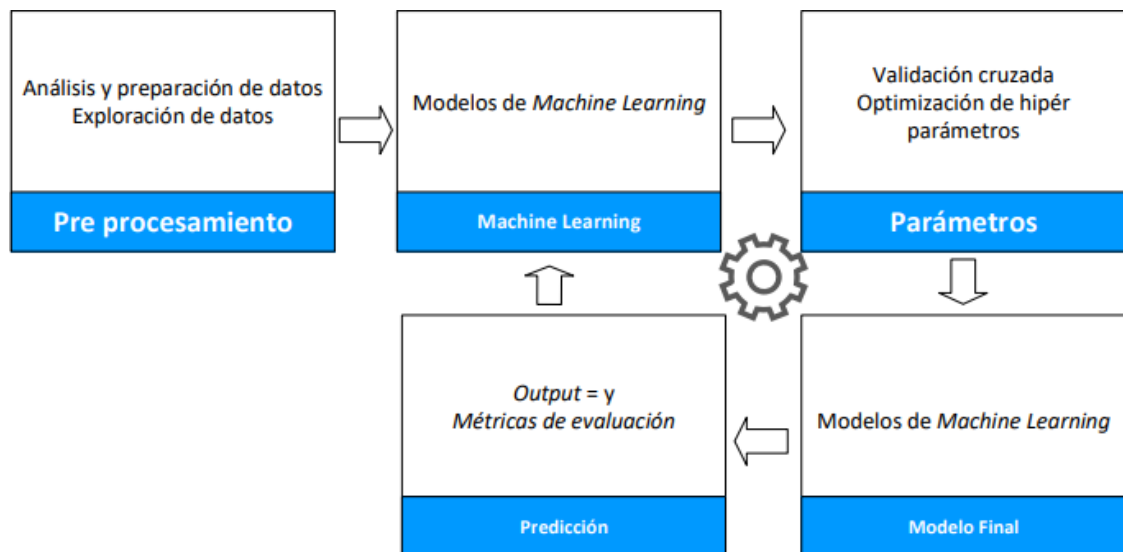
Según Romero Rojas, B. (2020, p.3). Machine Learning es una tecnología de software moderna que se basa es un método del análisis de datos que permiten automatizar la elaboración de modelos predictivos basados estadísticamente, que permite que un computador la capacidad de aprender de los datos, permiten, identificar patrones que permitan que ayuden a la toma de decisiones con mínima participación del ser humano.

Según Klaine, Imra, Onireti, y Souza, (2017). Machine Learning se clasifica de acuerdo al tipo de supervisión que tienen durante su entrenamiento, las cuatro categorías principales son, Aprendizaje Supervisado, No Supervisado, Semi Supervisado y Aprendizaje De Reforzamiento.

El aprendizaje supervisado dependiendo de la naturaleza de la variable que queremos predecir (variable dependiente) nos permite utilizar una clasificación o de regresión. Cuando la variable dependiente es categórica utilizamos clasificación, si la variable es continua utilizamos regresión.

Así mismo, Al entrenamiento de los algoritmos de Machine Learning, Según Raschka & Vahid (2017). Todo empieza con el pre procesamiento de datos, y posteriormente se configura el dataset en los algoritmos de Machine Learning, seguidamente se fijan los hiperparámetros para conseguir los modelos finales y las métricas de evaluación para conocer el rendimiento del modelo.

Figura 1. Entrenamiento de modelos de Machine Learning



Fuente: Adaptado de Raschka & Vahid (2017)

Por otro lado, los modelos predictivos serán medibles mediante sus métricas de evaluación de precisión, sensibilidad y exactitud. Según AGGARWAL (2019, p. 497), la evaluación se realiza mediante una matriz de confusión, la curva ROC.

Figura. 2 matriz de confusión

Matriz de confusión

		Predicted Class	
		Positive	Negative
Actual Class	Positive	(TP) True Positive	(FN) False Negative
	Negative	(FP) False Positive	(TN) True Negative

Fuente: Segura (2018)

Según Ortega (2021, p. 27). La precisión es la relación entre las predicciones correctas y el número total de predicciones correctas previstas.

$$\text{Precisión} = \frac{TP+TN}{TP+TN+FP+FN} * 100$$

Según Ortega (2021, p. 28). la sensibilidad evalúa la exactitud de las muestras positivas y las negativas respectivamente.

$$\text{sensibilidad} = \frac{TP}{TP+FN} * 100$$

Según Ortega (2021, p. 28), La especificidad de la clasificación es la unión de los pronósticos acertados y la cantidad total de pronósticos, con qué repetición es acertado el clasificador.

$$\text{Especificidad} = (\text{TN} / (\text{TN} + \text{FP})) \times 100$$

Según Gonzáles, 2014. La importancia de utilizar Machine Learning en las organizaciones que cada día generan una gran cantidad de datos exponencialmente, el uso de estos datos mediante el uso de máquinas de aprendizaje de forma adecuada representa una gran ventaja competitiva para la organización, permite tener un mayor grado de precisión en sus predicciones realizadas en operaciones comerciales.

Según Gonzáles (2014), La minería de datos es un proceso que permite descubrir patrones datos mediante un conjunto de modelo de predicción de Machine Learning, se puede realizar predicciones con nuevos datos, la data histórica de adquisición de vehículos vía online se puede usar para analizar el comportamiento de los clientes en sus procesos de compra y poder determinar patrones de comportamiento.

Según Castro, A. M., & Castro, J. A. M. (2014). El crédito es una de forma prestar un capital, donde el cliente se realiza un compromiso de retornar el capital solicitado más los intereses generados en un periodo de tiempo acordado según las condiciones establecidas.

Según Romero Rojas, B. (2020, p.13). El riesgo crediticio es la posibilidad de contraer una pérdida provocada por el incumplimiento de pago por la parte del cliente al cual se le facilito un capital para un determinado periodo de devolución.

Kendall (2011, p.31). Los sistemas web son aplicaciones donde los usuarios los utilizan ingresando vía Internet desde un explorador web. Es una aplicación de software desarrollada en un lenguaje interpretado por los navegadores web.

Montoro, A. F. (2012, p.3). Python es un lenguaje de programación de alto nivel soportado por varios sistemas operativos y utilizado en diferentes plataformas de

trabajo, en estos últimos años venido creciendo la demanda del uso de Python en aplicaciones de inteligencia de negocio y ciencia de datos.

Lan Gilfillan.(2003 p.40). MySQL es un gestor de base de datos con gran demanda en el mercado. Es un programa capaz de albergar una gran cantidad de datos que permiten cubrir las necesidades de las organizaciones con respecto al almacenamiento de sus datos. MySQL es un motor de base de datos que permite interpretar el lenguaje de consulta estructurado de base de datos (SQL).

Toro (2017), Anaconda es una plataforma que está conformada por un conjunto de programas para Data Science con Python y R. Se usa principalmente para el aprendizaje automático, el procesamiento de datos a gran escala, la informática científica y el análisis predictivo.

Como es de uso general, se tiene que seleccionar un conjunto de herramientas para implementar el análisis, conocidas como librerías, de las cuales para este proyecto se destacan:

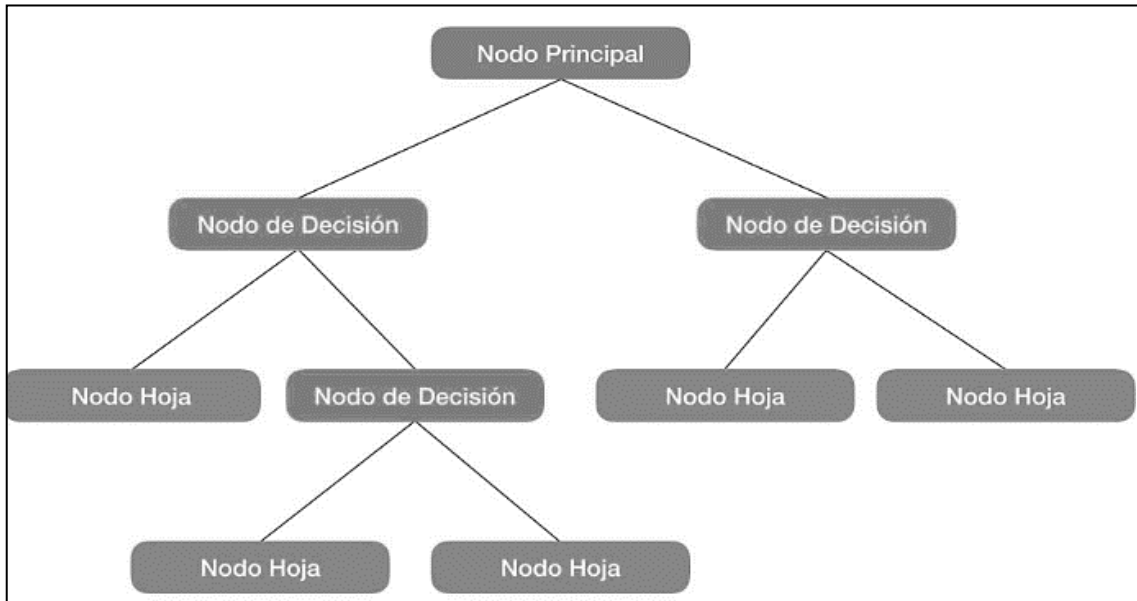
- Scikit-learn: es una biblioteca de aprendizaje automático de software libre permiten realizar selección de características, como de modelados de alto nivel de modelos de clasificación y regresión, incluyendo su validación. Según (Jiménez, 2019).

Así, mismo podemos indicar que los Modelos de Machine Learning como: Árboles de decisión, Random Forest, Support Vector Machine y Naive Bayes presentan las siguientes características.

Según Charris (2018, p. 1), el algoritmo de árbol de decisión “tiene como finalidad el aprendizaje inductivo en base a construcciones lógicas”. Un modelo de árbol de decisión se caracteriza por un nodo principal; el cual desencadena en nodos hijos que representan las reglas de decisión.

Esta técnica se sustentó en el modelo cuya definición se determinó en la siguiente expresión.

Figura 3: Decision Tree



Fuente: Rodríguez Vicente (2018)

Según Breiman (2001), el algoritmo Random Forest utiliza una mezcla de predictores de árboles, donde cada árbol necesita de los datos de un vector aleatorio muestreado de forma aislada y con la misma repartición para los elementos de árboles en el bosque.

$h(x,$

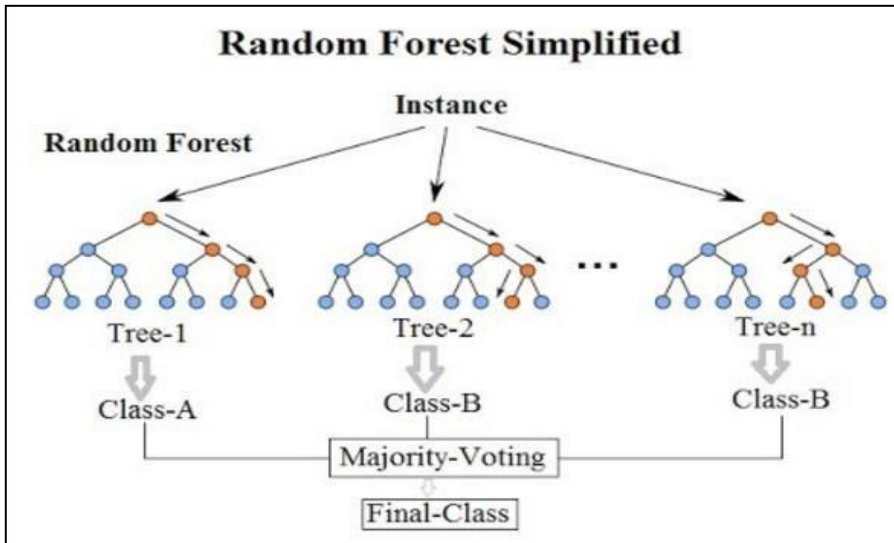
$\Theta_k)$

Donde:

h: Clasificador estructurado en árbol

k: k - ésimo árbol x: Vector de atributos de entrada

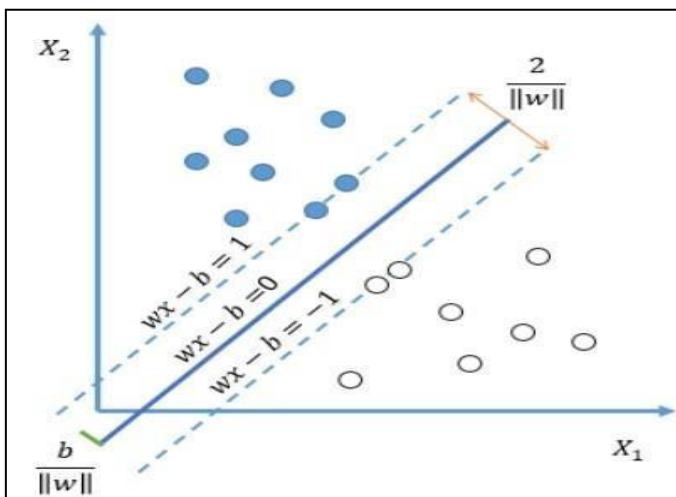
Figura4: Arboles de decisión RandomForest



Fuente: Prinzie A, Poel D (2007).

Según Deisenroth et al., (2019), el algoritmo Support Vector Machine se clasifica como un algoritmo de aprendizaje supervisado, se identifica como clasificador, este modelo tiene un enfoque que permite razonar de forma geométrica basado en productos internos y proyecciones.

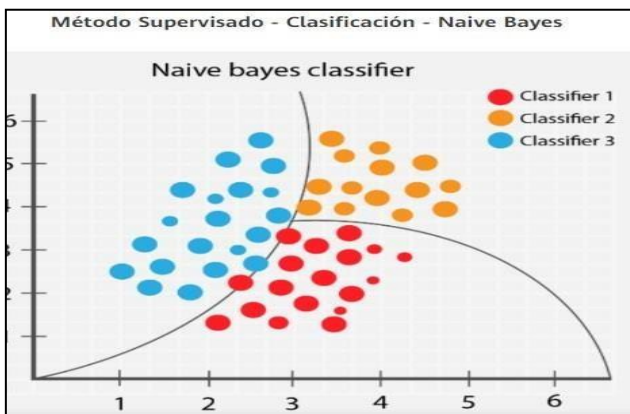
Figura 5. Representación del Algoritmo Máquina de Vectores de Soporte



Fuente: Nguyen (2016)

Según Barrientos (2009, p 22), El modelo Naive Bayes, representa un método que utiliza la estadística para definir que las características son aisladas, es decir que los atributos en un conjunto de datos no tienen relación con otros atributos presentes.

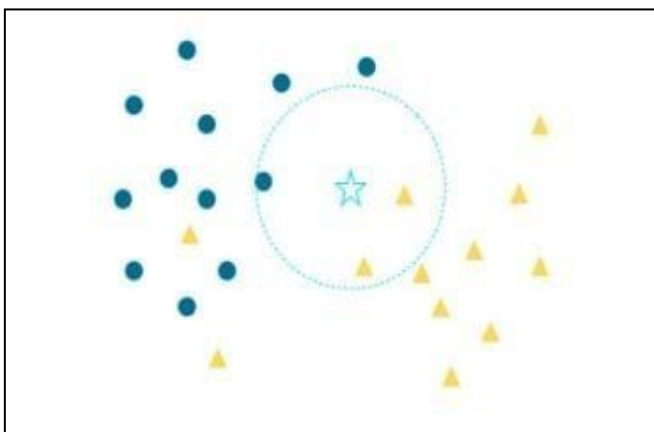
Figura 6. Representación del Algoritmo Naive Bayes



Fuente: Nguyen (2016)

Según Morales (2009), k vecinos más cercanos es una técnica de Machine Learnin que se basa en instancias, el proceso que realiza es muy sencillo: se guardan las operaciones de entrenamiento de datos históricos y cuando se necesita clasificar un nuevo objeto, se recuperan los objetos que tiene mayor similitud y se usa su separación para clasificar al nuevo objeto.

Figura 7. Representación del Algoritmo k vecinos más cercanos.



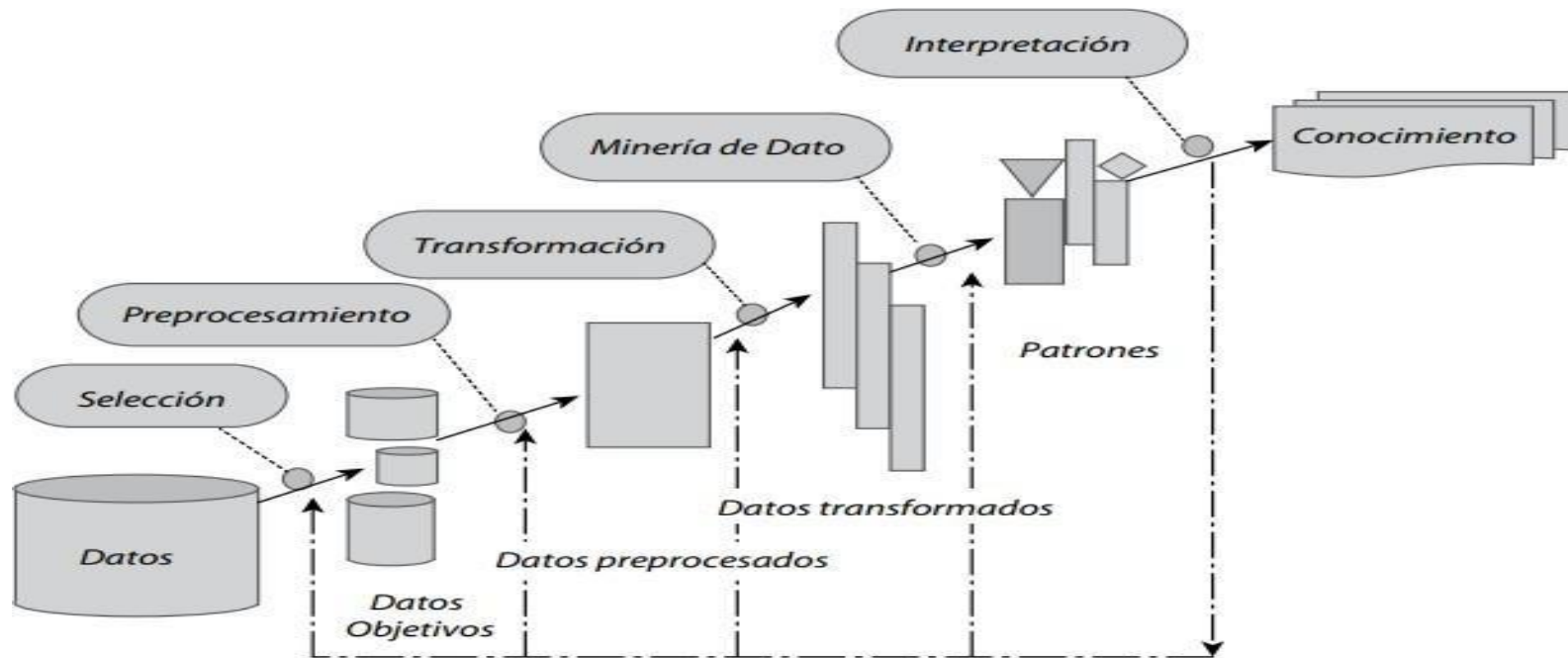
Fuente: Morales (2019)

En relación a la metodología tecnológica se utilizó KDD (Knowledge Discovery in Database), Según el autor Fayyad (1996, p. 82), “son un conjunto de procesos secuenciales que permite reconocer patrones apropiados dentro de un conjunto de datos”, la iteración que se realiza entre cada una de las etapas permite obtener un nuevo conocimiento beneficioso para toma de decisiones.

La Metodología KDD consta de 5 etapas:

Selección, Pre-procesamiento, Transformación, Minería de datos, Interpretación / Evaluación

Figura 8. Etapas de la metodología KDD



Fuente: Timarán-Pereira (2016, p.65)

En esta etapa inicial de selección radica en el estudio y el reconocimiento de los objetivos de la metodología KDD desde el punto de vista del cliente, Según Fayyad (1996). Esta etapa es muy importante porque se debe conocer anticipadamente los objetivos, procesos y actividades que realiza la organización, un conocimiento anticipado permitirá dirigir el proceso de minería de datos de forma óptima para obtener resultados de calidad.

En esta segunda etapa de pre-procesamiento / limpieza se identifican datos faltantes, ruidosos o nulos, se usan diferentes prácticas para uniformizar los datos que no se conocen, datos vacíos, duplicidad, etc (Moine, 2013, p. 11).

En esta tercera etapa es transformación, su finalidad es quitar los factores irrelevantes según lo que se desea conseguir en el proceso, para ellos se hace uso de técnicas para disminuir el número de factores (Fayyad 1996).

En esta cuarta etapa minería de datos se seleccionan los algoritmos apropiados que se utilizaran para realizar la elección de patrones mediante los datos y también se selecciona los modelos y parámetros apropiados según sea necesidad de la problemática. Según Thuraisingham, (1999). En la fase de la minería de datos se va utilizar técnicas de Machine Learning mediante algoritmos de clasificación y de regresión entre los cuales podemos mencionar: Decision Tree, Random Forest, Support Vector Machine, K Nearest Neighbor y Naive Bayes que se detallan seguidamente.

En quinta etapa la evaluación se analizan los patrones encontrados y probablemente se vuelven a revisar los procesos anteriores para siguientes iteraciones. En esta etapa se puede visualizar los patrones encontrados y también se realiza la eliminación de los patrones repetidos e irrelevantes y selección de patrones útiles que son entendibles para los usuarios. Según Timarán (2016).

Según, Torres Quezada (2020). En esta etapa se realiza un análisis de los resultados que se obtuvieron de los algoritmos utilizados. El estudio se realizó mediante métricas de rendimiento como la matriz de confusión producidas por los algoritmos lo cual permite determinar elección de un cliente apto aun crédito.

Otros de los aspectos de gran relevancia es la apreciación de los algoritmos de Machine Learning, que han permitido que más personas hagan uso de estos modelos de predicción en toma de decisiones. En especial donde se tiene que tomar decisiones de alto riesgo. Según Ortega (2021)

Por tanto, de lo indicado anteriormente por varios autores, se puede mencionar que la predicción del riesgo crediticio es un factor determinante y existe una variedad de líneas de investigación para concluir cómo influyen, teniendo en consideración e importancia en un ámbito de alto riesgo.

III. METODOLOGÍA

3.1. Tipo y diseño de investigación

Esta investigación de tipo aplicada, debido a que se aplica el conocimiento de Machine Learning mediante el uso de un conjunto de algoritmos de aprendizaje que van a permitir la predicción del riesgo crediticio de un cliente, Según Thomas y Manz (2017), mantiene que una investigación aplicada nos ayuda a entender el uso del conocimiento para realizar un sistema que permita apoyar en la solución de problemas y pronosticar resultados.

El diseño de investigación es experimental de tipo pre-experimental, se utilizó un grupo de clientes que fueron registrados mediante un cuestionario digital y posteriormente se aplicó Machine Learning para conseguir una observación para poder efectuar mediciones. Según Hernández (2014, p 141) “los pre-experimentos radican en aplicar un tratamiento a un grupo y posteriormente llevar a cabo una medición de variables”

Figura 9: El diagrama del diseño experimental tipo pre-experimental



Fuente: Bernal (2010)

3.2. Variables y Operacionalización

Las variables identificadas en el proyecto de investigación son: Machine Learning variable independiente y predecir el riesgo crediticio de un cliente la variable dependiente, a su vez la variable dependiente cuenta con una dimensión de riesgo crediticio; asimismo, cuenta con tres indicadores como precisión, sensibilidad y la especificidad; de tal manera que la operacionalización de variables se detalla en el Anexo N° 01 y 02.

Definición conceptual de la variable dependiente predicción riesgo crediticio de un cliente.

La variable dependiente predicción del riesgo crediticio de un cliente representa la existencia de pérdida potencial debido al incumplimiento de un compromiso de la persona a quien se dio un crédito (o en algunos términos y condiciones de la

transacción)", la predicción de riesgo de pago en las pequeñas empresas financieras debe ser tratada de forma diferente a la banca comercial por motivo que no cuenta con el software y la tecnología adecuada para medir el riesgo de microcréditos. Según Lara (2016. p.90).

Definición operacional de la variable dependiente predicción del riesgo crediticio de un cliente

Para predecir el riesgo crediticio de un cliente se ha tomado en cuenta una dimensión riesgo crediticio y tres indicadores, (a) Precisión, (b) Sensibilidad, y la (c) especificidad, la escala de medición que se estableció es la razón.

La operacionalización a detalle de estas variables se encuentra en el ANEXO N° 2.

3.3. Población

En este estudio se utilizó una población de 500 registro de clientes que fueron obtenidos mediante un cuestionario, la muestra está compuesta por la población. El muestreo es de tipo no probabilístico. Según Hernández Fernández y Baptista (2018), señalan que una población es una agrupación de elementos, con atributos idénticos, que se utilizan en un estudio.

3.6. Técnicas e instrumentos de recolección de datos

Técnicas de recolección de datos

Según García (2003, p.2), El cuestionario es utilizado para obtener registro de datos. La técnica para obtener los datos se realizó mediante un cuestionario de 26 preguntas que fueron registradas en un software de administrador de encuestas y estuvieron relacionadas con preguntas de carácter personal (6), factores económicos (13), factores financieros (3), factores socio-culturales (3), variable objetiva (1), también se hizo uso de una ficha de observación donde se tomó los valores de los indicadores de precisión, sensibilidad y especificidad de los modelos predictivos.

Validez

La autenticidad del instrumento de la ficha de registro de datos, del estudio se decidió mediante un juicio de expertos. Hernández, Fernández y Baptista (2014), mencionan que la autenticidad es el grado en que un instrumento realiza la medición de una variable que intenta demostrar.

3.7. Procedimientos

Los datos de los clientes empresa fueron obtenidos mediante un cuestionario realizado en software de administración de encuestas, que se encuentra en el (ANEXO 9) donde se visualiza las preguntas consideradas para este estudio, Se utilizo un conjunto de fases mediante una metodología para Machine Learning como KDD lo cual nos permitió realizar una minería de datos y haces uso de un conjunto de modelos predictivos como Support Vector Machine, Random Forest, Naives Bayes, Decision Tree.

La validación de hipótesis del estudio se realizó mediante una matriz de confusión proporcionada por un software para la minería de datos que mide el desempeño del algoritmo, el nivel de concordancia y el grado de significancia de los datos se utilizó el coeficiente de Kappa de Cohen.

3.8. Método de análisis de datos

Los datos se obtuvieron mediante un cuestionario virtual (ANEXO 8) que permitió que los clientes de la empresa en estudio registren sus datos, luego se utilizó un análisis predictivo utilizando el SPSS Modeler, donde se obtuvo información referente con la precisión, sensibilidad y especificidad, Así mismo indicar que se aplica la estadística predictiva, luego se aplicó las fases de la metodología KDD, (ANEXO 13).

3.8. Aspectos éticos

La investigación desarrollada tiene un grado de responsabilidad, sinceridad en los resultados obtenidos como de la confiabilidad de los datos brindados por la en la empresa FUTECH PERU S.A.C, donde se acabó el estudio. Es conveniente indicar que el desarrollo de la tesis utilizando los aspectos solicitados por la Universidad Cesar Vallejo, Sé reconoce las fuentes de información de los autores citándolos con el formato APA.

IV. RESULTADOS

Los resultados del estudio se basan en cada indicador de precisión, sensibilidad y especificidad que son obtenidos mediante la comparación de cinco algoritmos de Machine Learning para determinar el mejor modelo predictivo, la hipótesis fue validada mediante el índice de Kappa de Cohen que nos permite conocer el nivel de concordancia y el nivel de significancia.

Así mismo se muestra la matriz de confusión obtenida mediante el software SPSS Modeler de cada una de las técnicas de Machine Learning:

- Support Vector Machine

Tabla 1: Matriz de confusión - Support Vector Machine

Rendimiento (Riesgo crediticio)	Predicción		
	Bajo	Alto	Porcentaje correcto
Bajo	138	0	100%
Alto	1	361	99,7%
	27,8%	72,2%	99,80

Origen: Realización del investigador

Se utilizaron una cantidad de 500 clientes registrados, se identificó las casillas 1 y 4 de la matriz que tienen los valores 138 y 361 que son los clientes de bajo riesgo crediticio y un 0,2% de alto riesgo que han sido predichos por el modelo predictivo con una precisión de asertividad del 99,80%.

- Random Forest

Tabla 2: Matriz de confusión - Random Forest

Rendimiento (Riesgo crediticio)	Predicción		
	Bajo	Alto	Porcentaje correcto
Bajo	136	2	98,5%
Alto	17	345	95,1%
	30,6%	69,4%	96,20

Origen: Realización del investigador

Se utilizaron una cantidad de 500 clientes registrados, se identificó las casillas 1 y 4 de la matriz que tienen los valores 136 y 345 que son los clientes de bajo riesgo crediticio y un 3,8% de alto riesgo que han sido predichos por el modelo predictivo con una precisión de asertividad del 96,20%.

- Naibes Bayes

Tabla 3: Matriz de confusión - Naibes Bayes

	Predicción		
Rendimiento (Riesgo crediticio)	Bajo	Alto	Porcentaje correcto
Bajo	120	18	85%
Alto	9	353	97,5%
	25,8%	74,2%	94,60

Origen: Realización del investigador

Se utilizaron una cantidad de 500 clientes registrados, se identificó las casillas 1 y 4 de la matriz que tienen los valores 120 y 353 que son los clientes de bajo riesgo crediticio y un 5,4% de alto riesgo que han sido predichos por el modelo predictivo con una precisión de asertividad del 94,60%.

- K Nearest Neighbor

Tabla 4: Matriz de confusión - K Nearest Neighbor

	Predicción		
Rendimiento (Riesgo crediticio)	Bajo	Alto	Porcentaje correcto
Bajo	93	45	51,6%
Alto	21	341	93,8%
	22,8%	77,2%	86,8%

Origen: Realización del autor

Se utilizaron una cantidad de 500 clientes registrados, se identificó las casillas 1 y 4 de la matriz que tienen los valores 93 y 341 que son los clientes de bajo riesgo crediticio y un 13,2% de alto riesgo que han sido predichos por el modelo predictivo con una precisión de asertividad del 86,8%.

- Decision Tree

Tabla 5: Matriz de confusión - Decision Tree

Rendimiento (Riesgo crediticio)	Predicción		
	Bajo	Alto	Porcentaje correcto
Bajo	76	62	18,4%
Alto	25	337	92,6%
	20,2%	79,8%	82,6%

Origen: Realización del investigador

Se utilizaron una cantidad de 500 clientes registrados, se identificó las casillas 1 y 4 de la matriz que tienen los valores 53 y 344 que son los clientes de bajo riesgo crediticio y un 17,4% de alto riesgo que han sido predichos por el modelo predictivo con una precisión de asertividad del 82,6%.

Así mismo, se muestran los resultados para las hipótesis específicas, utilizando el software estadístico SPSS se obtuvo el índice de Kappa de Cohen que indica el grado de acuerdo entre dos mediciones. Según Manterola (2018, p.261) “el índice de Kappa de Cohen es una correlación que puede variar de -1 a +1, donde 1 representa la concordancia.

Figura 10: Nivel de Kappa de Cohen

Valores	Interpretación
< 0,01	No acuerdo
0,01 - 0,20	Ninguna a escaso
0,21 - 0,40	Regular o razonable
0,41 - 0,60	Moderado
0,61 - 0,80	Substancial
0,81 - 1,00	Casi perfecto

Fuente: Manterola (2018)

HE1: El Machine Learning permite predecir con precisión el riesgo crediticio de un cliente.

- Support Vector Machine

Tabla 6: Tabla cruzada - cálculo de precisión con algoritmo Support Vector Machine

			Resumen		Total
			Bajo	Alto	
predicción	Bajo	Recuento	138	1	139
		% del total	27,6%	0,2%	27,8%
	Alto	Recuento	0	361	361
		% del total	0,0%	72,2%	72,2%
Total		Recuento	138	362	500
		% del total	27,6%	72,4%	100,0%

Origen: Realización del investigador

$\text{Precisión} = (\text{TP} + \text{TN} / \text{TP} + \text{TN} + \text{FP} + \text{FN}) * 100$	$= (138 + 361) / (138 + 361 + 1 + 0) * 100$ $= 99,8\%$
--	--

Apreciación: Machine Learning permite predecir el riesgo crediticio de un cliente con una precisión del 99.8% mediante el modelo Support Vector Machine.

- Random Forest

Tabla 7: Tabla cruzada - cálculo de precisión con algoritmo Random Forest

			Resumen		Total
			Bajo	Alto	
predicción	Bajo	Recuento	136	17	153
		% del total	27,2%	3,4%	30,6%
	Alto	Recuento	2	345	347
		% del total	0,4%	69,0%	69,4%
Total		Recuento	138	362	500
		% del total	27,6%	72,4%	100,0%

Origen: Realización del investigador

$\text{Precisión} = (\text{TP} + \text{TN} / \text{TP} + \text{TN} + \text{FP} + \text{FN}) * 100$	$= (136 + 345) / (136 + 345 + 17 + 2) * 100$ $= 96,2\%$
--	---

Apreciación: Machine Learning permite predecir el riesgo crediticio de un cliente con una precisión de 96.2% mediante el modelo de Random Forest.

- Naibes Bayes

Tabla 8: Tabla cruzada - cálculo de precisión con algoritmo Naibes Bayes

			Resumen		Total
			Bajo	Alto	
predicción	Bajo	Recuento	120	9	129
		% del total	24,0%	1,8%	25,8%
	Alto	Recuento	18	353	371
		% del total	3,6%	70,6%	74,2%
Total		Recuento	138	362	500
		% del total	27,6%	72,4%	100,0%

Origen: Realización del investigador

$\text{Precisión} = (\text{TP} + \text{TN} / \text{TP} + \text{TN} + \text{FP} + \text{FN}) * 100$	$= (120 + 353) / (120 + 353 + 9 + 18) * 100$ $= 94,6\%$
--	---

Apreciación: Machine Learning permite predecir el riesgo crediticio de un cliente con una precisión de 94,6% mediante el modelo de Naibes Bayes.

- K Nearest Neighbor

Tabla 9: Tabla cruzada - cálculo de precisión con algoritmo K Nearest Neighbor

			Resumen		Total
			Bajo	Alto	
Predicción	Bajo	Recuento	93	21	114
		% del total	18,6%	4,2%	22,8%
	Alto	Recuento	45	341	386
		% del total	9,0%	68,2%	77,2%
Total		Recuento	138	362	500
		% del total	27,6%	72,4%	100,0%

Origen: Realización del investigador

$\text{Precisión} = (\text{TP} + \text{TN} / \text{TP} + \text{TN} + \text{FP} + \text{FN}) * 100$	$= (93 + 341) / (93 + 341 + 21 + 45) * 100$ $= 86,8\%$
--	--

- Decision Tree

Tabla 10: Tabla cruzada - cálculo de precisión con algoritmo Decision Tree

			Resumen		Total
			Bajo	Bajo	
predicción	Bajo	Recuento	76	25	101
		% del total	15,2%	5,0%	20,2%
	Alto	Recuento	62	337	399
		% del total	12,4%	67,4%	79,8%
Total		Recuento	138	362	500
		% del total	27,6%	72,4%	100,0%

Origen: Realización del investigador

$\text{Precisión} = (\text{TP} + \text{TN} / \text{TP} + \text{TN} + \text{FP} + \text{FN}) * 100$	$= (76 + 337) / (76 + 337 + 62 + 25) * 100$ $= 82,6\%$
--	--

Apreciación: Machine Learning permite predecir el riesgo crediticio de un cliente con una precisión de 82,6% mediante el modelo de Decision Tree.

Tabla 11: Comparación de los resultados mediante el indicador precisión

ALGORITMO	RESULTADO (%)
Support Vector Machine	99,8%
Random Forest	96,2%
Naibes Bayes	94,6%
K Nearest Neighbor	86,8%
Decision Tree	82,6%

Origen: Realización del investigador

Apreciación: En la tabla 11 se puede observar que el algoritmo con mejor resultado que es el Support Vector Machine con el 99,8% de presión al predecir el riesgo crediticio de un cliente.

HE2: El Machine Learning permite predecir con sensibilidad el riesgo crediticio de los clientes.

- Support Vector Machine

Tabla 12: Tabla cruzada - cálculo de sensibilidad con algoritmo Support Vector Machine

			Resumen		Total
			Bajo	Alto	
predicción	Bajo	Recuento	138	1	139
		% del total	27,6%	0,2%	27,8%
	Alto	Recuento	0	361	361
		% del total	0,0%	72,2%	72,2%
Total		Recuento	138	362	500
		% del total	27,6%	72,4%	100,0%

Origen: Realización del investigador

Sensibilidad=(TP/TP+FN)) x 100	=(138/(138+0))*100 =100%
--------------------------------	-----------------------------

Apreciación: Machine Learning permite predecir el riesgo crediticio de un cliente con una sensibilidad de 100% mediante el modelo de Support Vector Machine.

- Random Forest

Tabla 13: Tabla cruzada - cálculo de sensibilidad con algoritmo Random Forest

			Resumen		Total
			Bajo	Alto	
predicción	Bajo	Recuento	136	17	153
		% del total	27,2%	3,4%	30,6%
	Alto	Recuento	2	345	347
		% del total	0,4%	69,0%	69,4%
Total		Recuento	138	362	500
		% del total	27,6%	72,4%	100,0%

Origen: Realización del investigador

Sensibilidad=(TP/TP+FN)) x 100	=(136/(136+2))*100 =98,6%
--------------------------------	------------------------------

Apreciación: Machine Learning permite predecir el riesgo crediticio de un cliente con una sensibilidad de 98.6% mediante el modelo de Random Forest.

- Naibes Bayes

Tabla 14: Tabla cruzada - cálculo de sensibilidad con algoritmo Naibes Bayes

			Resumen		Total
			Bajo	Alto	
predicción	Bajo	Recuento	120	9	129
		% del total	24,0%	1,8%	25,8%
	Alto	Recuento	18	353	371
		% del total	3,6%	70,6%	74,2%
Total		Recuento	138	362	500
		% del total	27,6%	72,4%	100,0%

Origen: Realización del investigador

$Sensibilidad = (TP / (TP + FN)) \times 100$	$= (120 / (120 + 18)) \times 100$ $= 87\%$
--	---

Apreciación: Machine Learning permite predecir el riesgo crediticio de un cliente con una sensibilidad de 87% mediante el modelo de Naibes Bayes.

- K Nearest Neighbor

Tabla 15: Tabla cruzada - cálculo de sensibilidad con algoritmo K Nearest Neighbor

			Resumen		Total
			Bajo	Alto	
Predicción	Bajo	Recuento	93	21	114
		% del total	18,6%	4,2%	22,8%
	Alto	Recuento	45	341	386
		% del total	9,0%	68,2%	77,2%
Total		Recuento	138	362	500
		% del total	27,6%	72,4%	100,0%

Origen: Realización del investigador

$Sensibilidad = (TP / (TP + FN)) \times 100$	$= (93 / (93 + 45)) \times 100$ $= 67,4\%$
--	---

Apreciación: Machine Learning permite predecir el riesgo crediticio de un cliente con una sensibilidad de 67,4% mediante el modelo de K Nearest Neighbor.

- Decision Tree

Tabla 16: Tabla cruzada - cálculo de sensibilidad con algoritmo Decision Tree

			Resumen		Total
			Bajo	Bajo	
predicción	Bajo	Recuento	76	25	101
		% del total	15,2%	5,0%	20,2%
	Alto	Recuento	62	337	399
		% del total	12,4%	67,4%	79,8%
Total		Recuento	138	362	500
		% del total	27,6%	72,4%	100,0%

Origen: Realización del investigador

$\text{Sensibilidad} = (\text{TP} / (\text{TP} + \text{FN})) \times 100$	$= (76 / (76 + 62)) \times 100$ $= 55,1\%$
--	--

Apreciación: Machine Learning permite predecir el riesgo crediticio de un cliente con una sensibilidad de 55,1% mediante el modelo de Decision Tree.

Tabla 17: Comparación de los resultados mediante el indicador sensibilidad

ALGORITMO	RESULTADO (%)
Support Vector Machine	100%
Random Forest	98,6%
Naibes Bayes	87%
K Nearest Neighbor	67,4%
Decision Tree	55,1%

Origen: Realización del investigador

Apreciación: En la tabla 17 se puede observar que el algoritmo con mejor resultado que es el Support Vector Machine con el 100% de sensibilidad al predecir el riesgo crediticio de un cliente.

HE3: El Machine Learning permite predecir con especificidad el riesgo crediticio de los clientes.

- Support Vector Machine.

Tabla 18: Tabla cruzada - cálculo de especificidad con algoritmo Support Vector Machine

			Resumen		Total
			Bajo	Alto	
predicción	Bajo	Recuento	138	1	139
		% del total	27,6%	0,2%	27,8%
	Alto	Recuento	0	361	361
		% del total	0,0%	72,2%	72,2%
Total		Recuento	138	362	500
		% del total	27,6%	72,4%	100,0%

Origen: Realización del investigador

$\text{Especificidad} = \frac{TN}{(TN+FP)} \times 100$	$= \frac{361}{(361+1)} \times 100$ $= 99,7\%$
--	---

Apreciación: Machine Learning permite predecir el riesgo crediticio de un cliente con una especificidad de 99,7% mediante el modelo de Support Vector Machine.

- Random Forest

Tabla 19: Tabla cruzada - cálculo de especificidad con algoritmo Random Forest

			Resumen		Total
			Bajo	Alto	
predicción	Bajo	Recuento	136	17	153
		% del total	27,2%	3,4%	30,6%
	Alto	Recuento	2	345	347
		% del total	0,4%	69,0%	69,4%
Total		Recuento	138	362	500
		% del total	27,6%	72,4%	100,0%

Origen: Realización del investigador

$\text{Especificidad} = \frac{TN}{(TN+FP)} \times 100$	$= \frac{345}{(345+17)} \times 100$ $= 95,3\%$
--	--

Apreciación: Machine Learning permite predecir el riesgo crediticio de un cliente con una especificidad de 95.3% mediante el modelo de Random Forest.

- Naibes Bayes

Tabla 20: Tabla cruzada - cálculo de especificidad con algoritmo Naibes Bayes

			Resumen		Total
			Bajo	Alto	
predicción	Bajo	Recuento	120	9	129
		% del total	24,0%	1,8%	25,8%
	Alto	Recuento	18	353	371
		% del total	3,6%	70,6%	74,2%
Total		Recuento	138	362	500
		% del total	27,6%	72,4%	100,0%

Origen: Realización del investigador

$\text{Especificidad} = (\text{TN} / (\text{TN} + \text{FP})) \times 100$	$= (353 / (353 + 9)) \times 100$ $= 97,5\%$
---	---

Apreciación: Machine Learning permite predecir el riesgo crediticio de un cliente con una sensibilidad de 97,5% mediante el modelo Naibes Bayes.

- K Nearest Neighbor

Tabla 21: Tabla cruzada - cálculo de especificidad con algoritmo K Nearest Neighbor

			Resumen		Total
			Bajo	Alto	
Predicción	Bajo	Recuento	93	21	114
		% del total	18,6%	4,2%	22,8%
	Alto	Recuento	45	341	386
		% del total	9,0%	68,2%	77,2%
Total		Recuento	138	362	500
		% del total	27,6%	72,4%	100,0%

Origen: Realización del investigador

$\text{Especificidad} = (\text{TN} / (\text{TN} + \text{FP})) \times 100$	$= (341 / (341 + 21)) \times 100$ $= 94,2\%$
---	--

Apreciación: Machine Learning permite predecir el riesgo crediticio de un cliente con una especificidad de 94,2% mediante el modelo de K Vecino más Cercano.

- Decision Tree

Tabla 22: Tabla cruzada - cálculo de sensibilidad con algoritmo Decision Tree

			Resumen		Total
			Bajo	Bajo	
predicción	Bajo	Recuento	76	25	101
		% del total	15,2%	5,0%	20,2%
	Alto	Recuento	62	337	399
		% del total	12,4%	67,4%	79,8%
Total		Recuento	138	362	500
		% del total	27,6%	72,4%	100,0%

Origen: Realización del investigador

$\text{Especificidad} = (\text{TN} / (\text{TN} + \text{FP})) \times 100$	$= (337 / (337 + 25)) \times 100$ $= 93,1\%$
---	--

Apreciación: Machine Learning permite predecir el riesgo crediticio de un cliente con una especificidad de 93,1% mediante el modelo de Decision Tree.

Apreciación: Machine Learning permite predecir el riesgo crediticio de un cliente con una especificidad de 94,2% mediante el modelo de K Vecino más Cercano.

Tabla 23: Comparación de los resultados mediante el indicador especificidad

ALGORITMO	RESULTADO (%)
Support Vector Machine	99,7%
Naibes Bayes	97,5%
Random Forest	95,3%
K Nearest Neighbor	94,2%
Decision Tree	93,1%

Origen: Realización del investigador

Apreciación: En la tabla 9 se puede observar que el algoritmo con mejor resultado que es el Support Vector Machine con el 99,7% de especificidad al predecir el riesgo crediticio de un cliente.

- Hipótesis General:

Ho: Machine Learning no predice significativamente el riesgo crediticio de un cliente en la empresa FUTECH PERÚ SAC, 2022.

H1: Machine Learning predice significativamente el riesgo crediticio de un cliente en la empresa FUTECH PERÚ SAC, 2022.

Tabla 24: Cuadro comparativo de los resultados de los algoritmos

INDICADORES			
ALGORITMO	PRESICIÓN	SENSIBILIDAD	ESPECIFICIDAD
Support Vector Machine	99,8%	100%	99,7%
Random Forest	96,2%	98,6%	95,3%
Naibes Bayes	94,6%	87%	97,5%
K Nearest Neighbor	86,8%	67,4%	94,2%
Decision Tree	82,6%	55,1%	93,1%

Origen: Realización del investigador

Mediante los indicadores de precisión, sensibilidad y especificidad de modelo de árbol de decisión, Support Vector Machine, Random Forest, Naibes Bayes, K Nearest Neighbor, Decision Tree, se verifica que el modelo con mejor capacidad para predecir el riesgo crediticio Support Vector Machine con un valor de 99,8% en todos los indicadores.

Asu vez se hace uso del índice de Kappa de Cohen para obtener el nivel de concordancia y el nivel de significancia que hay entre dos observaciones.

Tabla 25: Medida de Kappa de Cohen - Support Vector Machine

Medidas simétricas					
		Valor	Error estándar asintótico ^a	T aproximada ^b	Significación aproximada
Medida de acuerdo	Kappa	,995	,005	22,249	,000
N de casos válidos		500			

Origen: Realización del investigador

Tabla 26: Medida de Kappa de Cohen - Random Forest

Medidas simétricas					
		Valor	Error estándar asintótico ^a	T aproximada ^b	Significación aproximada
Medida de acuerdo	Kappa	,908	,021	20,357	,000
N de casos válidos		500			

Origen: Realización del investigador

Tabla 27: Medida de Kappa de Cohen - Neibes Bayes

Medidas simétricas					
		Valor	Error estándar asintótico ^a	T aproximada ^b	Significación aproximada
Medida de acuerdo	Kappa	,862	,026	19,298	,000
N de casos válidos		500			

Origen: Realización del investigador

Tabla 28: Medida de Kappa de Cohen - K Nearest Neighbor

Medidas simétricas					
		Valor	Error estándar asintótico ^a	T aproximada ^b	Significación aproximada
Medida de acuerdo	Kappa	,651	,039	14,674	,000
N de casos válidos		500			

Origen: Realización del investigador

Tabla 29: Medida de Kappa de Cohen - Decision Tree

Medidas simétricas					
		Valor	Error estándar asintótico ^a	T aproximada ^b	Significación aproximada
Medida de acuerdo	Kappa	,525	,044	11,992	,000
N de casos válidos		500			

Origen: Realización del investigador

Tabla 30: Resumen de cuadro comparativo de los algoritmos de las medidas del nivel de concordancia de utilizando la medida de acuerdo de Kappa Cohen y el nivel de significancia que hay entre dos observaciones.

ALGORITMO	Medida de Kappa de Cohen	
	Medida de acuerdo Kappa	Significación aproximada
Support Vector Machine	0,995	,000
Random Forest	0,98	,000
Naibes Bayes	0,862	,000
K Nearest Neighbor	0,651	,000
Decision Tree	0,525	,000

Origen: Realización del investigador

Interpretación: En el índice de Kappa que se observa en todos los algoritmos de Machine Learning cuenta con un valor que representa que hay una buena concordancia entre dos observaciones y el nivel de significancia es menor a 5% por tanto, se acepta la hipótesis alterna.

En conclusión, el algoritmo con mejor nivel de concordancia después de haber evaluado a los algoritmos de clasificación como Random Forest, Naibes Bayes, K Nearest Neighbor, Decision Tree es Support Vector Machine con un 99,5% según la medida de acuerdo a Kappa de Cohen y un grado de significancia menor a 5% por tanto, se acepta la hipótesis alterna.

V. DISCUSIÓN

Con los resultados de este estudio, se desarrolló una aplicación para predecir el riesgo crediticio de un cliente en una empresa en estudio.

La verosimilitud de la predicción de Machine Learning sobre el riesgo crediticio se cuantificó mediante los indicadores precisión, sensibilidad y especificidad. La precisión mediante Support Vector Machine se cuantificó en 99,8%, con el modelo de Random Forest un 96,2%, Naïves Bayes 94,6%, K Vecino más Cercano 86,8%, y mediante Decision Tree un 82,6%. Con el indicador sensibilidad Support Vector Machine se cuantificó en 100%, con el modelo Random Forest un 98%, Naïves Bayes un 87%, K Vecino más Cercano 67,4% y mediante Decision Tree 55,1%. El indicador especificidad se valoró Support Vector Machine se cuantificó en 99,7%, con el modelo Random Forest un 95,3%, Naïves Bayes un 97,5%, K Nearest Neighbor 94,2% y con el modelo Decision Tree 93,1%. Respecto al riesgo crediticio, se evidenció que el modelo Support Vector Machine alcanzó una predicción del 99,8% en la fase final de predicción.

Los resultados de esta investigación se condicen con lo que obtuvieron Ortiz et al. (2017) quienes mediante arboles de decisión de doble clasificación alcanzaron una precisión del 93.3%. De manera similar, Rodríguez et al. (2017) en la aplicación informática basada en Machine Learning, para mejorar la evaluación de los préstamos crediticios alcanzaron un 98.8% de sensibilidad, 25% de especificidad, 87% de eficacia. Otro modelo relevante y acorde a los que se obtuvieron en este estudio desarrollo Aceituno (2019), cuya metodología se orientó al nivel estructural cuyos resultados obtenidos en relación a la precisión con el modelo de red neuronal fueron de 93.7% y la regresión logística del 86.1%. En este contexto, se condice el estudio de Ossa & Jaramillo (2021) que estimó el riesgo crediticio en una cartera de consumo, que identificó factores de riesgo basada en la noción de (Dominguez & Aldana, 2001), Los resultados estimaron el riesgo crediticio en probabilidades de por lo menos 80%.

Estos resultados están incluidos en el marco conceptual relacionados a Machine Learning en definición de Hurwitz y Kirsch (2018), y el riesgo crediticio en la noción de Jiménez (2016), Izarra (2016).

VI. CONCLUSIÓN

Primera: De acuerdo a la información obtenida de la presente investigación efectuada en la empresa FUTECH SAC, se determina que utilizando Machine Learning se puede implementar una aplicación sustentada en un modelo Support Vector Machine para predecir el riesgo crediticio de los clientes, que fue confrontado con algoritmos de Random Forest se cuantificó, modelo Naibes Bayes, K Nearest Neighbor y los Decision Tree, cuyo efecto conjunto en la predicción se valoró en 99.8% para Random Forest, 96.2% para el modelo de Naibes Bayes, 94.6%, árboles de decisión, 79.04%.

Segunda: Luego de usar varios algoritmos como: Support Vector Machine, Random Forest, Naibes Bayes, K Nearest Neighbor y Decision Tree se determinó que el modelo de Machine Learning que brinda la mejor precisión al momento de predecir el riesgo crediticio de un cliente en la empresa en estudio es el modelo de Support Vector Machine con un 99,8%.

Tercera: Luego de usar varios algoritmos como: Support Vector Machine, Random Forest, Naibes Bayes, K Nearest Neighbor y Decision Tree se determinó que el modelo de Machine Learning que brinda la mejor sensibilidad al momento de predecir el riesgo crediticio de un cliente en la empresa en estudio es el modelo de Support Vector Machine con un 100%.

Cuarta: Luego de usar varios algoritmos como: Support Vector Machine, Random Forest, Naibes Bayes, K Nearest Neighbor y Decision Tree se determinó que el modelo de Machine Learning que brinda la mejor especificidad al momento de predecir en el riesgo crediticio de un cliente en la empresa en estudio es el modelo de Support Vector Machine con un 99,7%.

VII. RECOMENDACIONES

Con los resultados obtenidos en esta investigación, se sugiere que el Gerente General de la empresa FUTECH SAC, implementar la aplicación del producto de esta investigación, dada su precisión del 99,8% al momento de predecir el riesgo crediticio de un cliente.

- Se recomienda utilizar la mayor cantidad de técnicas y métricas de Machine Learning para comparar resultados y elegir el modelo predictivo que tenga mayor capacidad predictiva al momento de predecir el riesgo crediticio de un cliente.
- Se recomienda utilizar un gran volumen de datos históricos y utilizar técnicas de minería de datos para seleccionar los datos más relevantes de los clientes con la finalidad de mejorar los resultados de precisión del modelo predictivo para identificar a los clientes de alto riesgo crediticio.
- Se recomienda fortalecer los resultados obtenidos en la presente investigación mediante el uso de servicios de proveedores de información clasificada crediticia que brinden información actualizada del comportamiento crediticio de los clientes que permita comparar con los resultados obtenidos del modelo de clasificación Support Vector Machine.
- Se recomienda que el administrador del área de préstamos realice una revisión de los indicadores de precisión de los clientes que fueron calificados utilizando su método de calificación y lo contraste con el indicador de precisión del modelo predictivo Support Vector Machine que obtuvo una precisión de 99,8%

REFERENCIAS

- Abirami, S., & Chitra, P. (2020). Energy-efficient edge based real-time healthcare support system. In *Advances in Computers* (1st ed., Vol. 117, Issue 1). Elsevier Inc. <https://doi.org/10.1016/bs.adcom.2019.09.007>
- Abolhassan, F. (2017). *The Drivers of Digital Transformation. Why There's No Way Around the Cloud*. Bonn, Germany: Springer International.
- Aceltuno Rojo.(2017). modelo predictivo de análisis de riesgo crediticio usando machine learning en una entidad del sector microfinanciero. http://repositorio.unap.edu.pe/bitstream/handle/UNAP/14481/Miguel_Romilio_Aceituno_Rojo.pdf?sequence=1&isAllowed=y
- Alencar, R. (2018). Python notebook using data from Porto Seguro's Safe Driver Prediction. Kaggle. <https://www.kaggle.com/rafjaa/resampling-strategies-for-imbalanced-datasets>
- Altman, E. I. (2003). *Medición integral del riesgo de crédito*. books.google.com.
- Avram, M. G. (2014). Advantages and Challenges of Adopting Cloud Computing from an Enterprise Perspective. *Procedia Technology*.
- Támara Ayús, A. L., Vargas Ramírez, H., Joaquín Cuartas, J., & Chica Arrieta, I. E. (2019). Regresión logística y redes neuronales como herramientas para realizar un modelo Scoring. *Revista Lasallista de Investigación*, 197. <http://hdl.handle.net/10567/2645>
- Bakshi, C. (2020). Random Forest Regression. Levelup.Gitconnected. <https://levelup.gitconnected.com/random-forest-regression-209c0f354c84>
- Bazarbash, M. (2019). FinTech in Financial Inclusion: Machine Learning Applications in Assessing Credit Risk. *IMF Working Papers*, 19(109), 1. <https://doi.org/10.5089/9781498314428.001>
- BARRIENTOS, R., CRUZ, N., ACOSTA, H., RABATTE, I., GOGESCOECHEA,

M., PAVÓN, P. y BLÁZQUEZ, S., 2009. Árboles De Decisión Como Herramienta En El Diagnóstico Médico. Artículo Original [en línea], pp. 20-24. Disponible en:
https://www.uv.mx/rm/num_antteriores/revmedica_vol9_num2/articulos/arboles.pdf

Bermingham, M. L.-W. (2015). Application of high-dimensional feature selection: evaluation for genomic prediction in man. *Scientific reports*. nature.com.

Betancourt, G. A. (2005). Las máquinas de soporte vectorial (SVMs). *revistas.utp.edu.co*, 67-69.

Bonilla, M. O. (2003). Modelos Paramétricos y no Paramétricos en Problemas de Credit Scoring. *Revista Española de Financiación Y Contabilidad*, 833-869.

Bourel, M. (2012). Métodos de agregación de modelos y aplicaciones. *researchgate.net*, 1-15.

Borrero-Tigreros, D., & Bedoya-Leiva, O. F. (2020). Predicción de riesgo crediticio en Colombia usando técnicas de inteligencia artificial. *Revista UIS Ingenierías*, 19(4), 37-52.

Breiman, L. (2001). Random forests. *Machine learning*, 45(1), 5-32.
<https://doi.org/10.1023/A:1010933404324>

Cambroner, C. G. (2006). Algoritmos de aprendizaje: knn & kmeans. *Inteligencia en Redes de Comunicación*.

Carhuancho Mendoza, I. M., Sicheri Monteverde, L., Nolzco Labajos, F. A., Guerrero Bejarano, M. A., & Casana Jara, K. M. (2019). Metodología de la investigación holística. GUAYAQUIL/UIDE/2019.
<https://repositorio.uide.edu.ec/handle/37000/3893>

Castro, A. M., & Castro, J. A. M. (2014). Crédito y cobranza. Grupo Editorial Patria.
<https://books.google.es/books?hl=es&lr=&id=gtXhBAAQBAJ&oi=f>

nd&pg=PP1&dq=Jos%C3%A9+Antonio+Morales+Castro&ots=p1gULdll-c&sig=Aicy0F_Q4nXGszoeI4Nay7eJ5XQ#v=onepage&q=Jos%C3%A9%20Antonio%20Morales%20Castro&f=false, pag.36

Campos, G. y Lule, N. (2012). La observación, un método para el estudio de la realidad. *Xihmai*, ISSN-e 1870-6703, Vol. 7, Nº. 13, 2012, págs. 45-60. <https://dialnet.unirioja.es/servlet/articulo?codigo=3979972>

Chan, C. (2018). What is a ROC Curve and How to Interpret It. *Display R Blog*. <https://www.displayr.com/what-is-a-roc-curve-how-to-interpret-it/>

Chandra, B., M. Gupta y M.P. Gupta, Robust Approach for Estimating Probabilities in Naive-Bayes Classifier. In *International Conference on Pattern Recognition and Machine Intelligence* (pp. 11-16), Springer, Berlin, Heidelberg, December (2007)

Chappell & Associates. (2015). *Introducing Azure Machine Learning*. San Francisco, California.

Charris, L., Henriquez, C., Hernandez, S., Jimeno, L., Guillen, O., & Moreno, S. (2018). Análisis comparativo de algoritmos de árboles de decisión en el procesamiento de datos biológicos. *Investigación y desarrollo en TIC*, 9(1), 26-34., <http://revistas.unisimon.edu.co/index.php/identific/article/view/3158>

Corso, C. L. (2009). Aplicación de algoritmos de clasificación supervisada usando Weka. *investigacion.frc.utn.edu.ar*, 1-11.

Dha, V. (2013). Data science and prediction. *Communications of the ACM*. *Communications of the ACM*, 64-65,66,67,68,69,70,71,72,73.

Domínguez, E., & Aldana, D. (2001). Logistic regression: An example of its use in Endocrinology. *Revista Cubana de Endocrinología*, 21(1).

Deisenroth, M., Faisal, A., & Soon, C. (2019). *Mathematics for Machine Learning*. Cambridge, United Kingdom: Cambridge University Press

- Dunkel y Weber, 2007; Ramaswamy, 2005; redes neuronales artificiales (Atiya, 2001). Predicción mediante Redes Neuronales Artificiales (RNA). <https://doi.org/10.17268/sci.agropecu.2012.03.02>
- El Naqa, I. (2015). Machine Learning in Radiation Oncology. ScienceDirect, 3-11.
- Fayyad, U. M., Piatetsky-Shapiro, G., & Smyth, P. (1996, August). Knowledge Discovery and Data Mining: Towards a Unifying Framework. In KDD (Vol. 96, pp. 82-88). https://www.aaai.org/Papers/KDD/1996/KDD96-014.pdf?utm_campaign=ml4devs-newsletter&utm_medium=email&utm_source=Revue%20newsletter
- Frawley, W. J., Piatetsky-Shapiro, G., & Matheus, C. J. (1992). Knowledge discovery in databases: An overview. AI magazine, 13(3), 57-57. DOI: <https://doi.org/10.1609/aimag.v13i3.1011>
- Francés Monedero. (2020). impacto del machine learning en el sistema financiero, <http://hdl.handle.net/11531/42692>
- García, M. L. (2017). Modelos para medir el riesgo de crédito de la banca. Cuadernos de administración, 1-33.
- GARCÍA, T., 2003. El cuestionario como instrumento de investigación/evaluación. Página del proyecto de apoyo para profesionales de la formación (PROMETEO) de la Junta de Andalucía [En línea], pp. 28. Disponible en:http://www.univsantana.com/sociologia/El_Cuestionario.pdf.
- Gilley, S. (2019). Entrenamiento de modelos de riesgo crediticio: Azure Machine Learning Studio. Obtenido de Entrenamiento de modelos de riesgo crediticio: Azure Machine Learning Studio: <https://docs.microsoft.com/es-es/azure/machine-learning/studio/tutorial-part2-credit-risk-train>.
- Gonzales, Andres (2014). ¿Qué es Machine Learning? Recuperado de <http://cleverdata.io/quees-machine-learning-big-data/>

- Guillén Uyen, J. (2001). Morosidad crediticia y tamaño: Un análisis de la crisis bancaria peruana *. Revista Concurso de Investigación para Jóvenes Economistas 2001-2002.
- Gutiérrez, J. F. (2011). Pronóstico de incumplimientos de pago mediante máquinas de vectores de soporte: una aproximación inicial a la gestión del riesgo de crédito. researchgate.net.
- Hassanien, S. A. Vorobyov and M. W. Morency, "Efficient Transmit Beamspace Design for Search-Free Based DOA Estimation in MIMO Radar," in *IEEE Transactions on Signal Processing*, vol. 62, no. 6, pp. 1490-1500, March 15, 2014, doi: 10.1109/TSP.2014.2299513.
- Hand, D. J., & Henley, W. E. (1997). Statistical classification methods in consumer credit scoring: a review. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 160(3), 523-541.
- Hernández-Sampieri, R., Fernández Collado, C., & Baptista Lucio, P. (2018). Metodología de la investigación (Vol. 4). México: McGraw-Hill Interamericana.
- Huaman.(2020): "Implementación de un sistema de gestión de seguridad electrónica con Machine Learning dirigido a Prosegur Perú para gestión de seguridad en viviendas de Lima Metropolitana", <https://hdl.handle.net/20.500.12867/2842>
- Hurwitz, J., & Kirsch, D. (2018). Machine Learning. (I. John Wiley & Sons, Ed.) New York: IBM Limited Edition.
- Hurtado, M. J. R (2010)., Baños,. La investigación formativa como metodología de aprendizaje en la mejora de competencias transversales. *Procedia-Social and Behavioral Sciences*, 196, 177-182.
- Hurtado de Barrera, J. (2010). Metodología de la investigación. Guía para la comprensión Holística de la ciencia. 4a edición. Bogotá, Colombia: Quirón

Ediciones SA Cooperativa Editorial Magisterio. Caracas, Venezuela: Ciea-Sypal.(p. 539 – 542)

Jiménez López, N. R. (2016). La gestión de la calidad crediticia como alternativa de solución a los problemas de morosidad de la cartera de la micro y pequeña empresa y su efecto en los resultados económicos y financieros de la caja municipal de ahorro y crédito del santa-años 2014 y 2015, <https://hdl.handle.net/20.500.13032/741>

Jiménez Cárdenas, E. A. (2019). Análisis de la Red social Twitter para la identificación de patrones que generan oportunidades de negocio en la ciudad de Guayaquil utilizando el entorno de trabajo Jupyter Notebook y el Lenguaje de Programación Python (Bachelor's thesis, Universidad de Guayaquil. Facultad de Ciencias Matemáticas y Físicas. Carrera de Ingeniería en Sistemas Computacionales.).

Izarra C. (2016). Riesgo Crediticio y la Morosidad en la cooperativa de ahorro y crédito Huancavelica LTDA. N° 582 - Huancavelica - Periodo 2014. Huancavelica: Universidad Nacional de Huancavelica. Retrieved from <https://repositorio.unh.edu.pe/bitstream/handle/UNH/1358/TP%20-%20UNH.%20CONT.%200088.pdf?sequence=1&isAllowed=y>

Kendall, K. & Kendall, J. (2011) Análisis y Diseño de Sistemas. Octava Edición. PEARSON EDUCACIÓN, México.

Klaine, P., Imra, M., Onireti, O., & Souza, R. (2017). A survey of machine learning techniques applied to selforganizing cellular networks. IEEE Communications Surveys & Tutorials, 19(4), 23922431.

Lara Rubio, Iván; Rayo Cantón, Salvador; y Camino Blasco, David (2016). “Un modelo credit scoring para instituciones de microfinanzas en el marco de Basilea II”. Journal of economics, finance and administrative science, pp. 90. Junio de 2016.

- Lan Gilfillan(2003). La Biblia de MySQL, Editorial Anaya Multimedia, S.A., 2003., Mexico.
- Microsoft Corporation. (2016). Entendiendo Azure. Guía para Desarrolladores. Redmond, Washington, E.E.U.U: Microsoft Corporation.
- Miklos, T., & Tello, M. (2007). Planeación prospectiva y estratégica. Documento procedente del V Encuentro de estudios Prospectivos. (pag. 37)
- Moine, J. M. (2013). Metodologías para el descubrimiento de conocimiento en bases de datos: un estudio comparativo (Doctoral dissertation, Universidad Nacional de La Plata). <https://doi.org/10.35537/10915/29582>
- Montoro, A. F. (2012). Python 3 al descubierto. RC libros.
- Morales, E., 2009, “Descubrimiento de conocimiento en bases de datos”, Obtenido el día 11 de Julio de 2009, desde la World Wide Web en el sitio <http://ccc.inaoep.mx/~emorales/Cursos/KDD/principal.html>
- Nguyen, K. (2015). Introduction to Machine Learning. nguyenducminhkhoy.com, 1-59.
- Nguyen, L. (2016). Tutorial on Support Vector Machine. Some Novel Algorithms for Global Optimization and Relevant Subjects, (6).
- Otzen, Tamara, & Manterola, Carlos. (2017). Técnicas de Muestreo sobre una Población a Estudio. International Journal of Morphology, 35(1), 227-232. <https://dx.doi.org/10.4067/S0717-95022017000100037>
- Ortiz, C., & Haro, B. (2017). Modelo tecnológico de análisis predictivo basado en machine learning para evaluación de riesgo crediticio. Universidad Peruana de Ciencias Aplicadas (UPC). <http://hdl.handle.net/10757/656207>

- Ortega-Martín, S. (2021). Comparativa de modelos de Machine Learning interpretables para la predicción de riesgo crediticio (Master's thesis), <https://reunir.unir.net/handle/123456789/12210>
- Ossa, W. & Jaramillo, V. (2021). Machine Learning para la estimación del riesgo de crédito en una cartera de consumo. Universidad Eafit, <http://hdl.handle.net/10784/29589>
- Pacheco, V. G. (Enero de 2019). Big Data & Data Science Blog: Una Breve Historia del Machine Learning. Obtenido de LUCA AI Powered Decisions: <https://data-speaks.luca-d3.com/2018/11/una-brevehistoria-del-machine-learning.html>
- Paola A, C. H. (2004). Aplicacion de Arboles de decision en modelos de riesgo crediticio. *Revsita Colombiana de Estadística*, 13.
- Planas, M. R. (2004). La importancia de los datos. *Nutricion hospitalaria*. SciELO Espana, 11-13.
- Raschka, S., & Vahid, M. (2017). *Python Machine Learning Second Edition*. In Packt Publishing (2). United Kingdom: Cambridge University Press. doi: 10.1017/CBO9781107415324.004
- Reed, D. A. (2005). Computational science: ensuring America's competitiveness. *apps.dtic.mil*, 1-55.
- Rendón, V., & Landman, J. K. (2016). Uso de la hoja de cálculo para analizar datos cualitativos. *Magis: Revista Internacional de Investigación en Educación*, 9(18), 29-48.
- Rodríguez, , J. J., & Miñano, , M. M. (2017). Desarrollo de una aplicación informática basada en un modelo de mahine learning para mejorar la evaluación de préstamos crediticios (Tesis de licenciatura). Repositorio de la Universidad Privada del Norte.

- Romero Rojas, B. (2020). Una introducción a los modelos de Machine Learning (Bachelor's thesis).
- Russo, C., Ramón, H., Alonso, N., Cicerchia, B., Esnaola, L., & Tessore, J. P. (2016). Tratamiento masivo de datos utilizando técnicas de Machine Learning.
- Ruiz b.(2020) El propósito del proyecto es el desarrollo de un sistema de Aprendizaje Automático que permita realizar predicciones de consumo de suministros eléctricos. <http://hdl.handle.net/10609/106646>
- Russell S. & Norvig P. (2004). Inteligencia Artificial un enfoque moderno (2da Edición). España: Pearson Educación.
- SAGNER T, A. (2012). EL INFLUJO DE CARTERA VENCIDA COMO MEDIDA DE RIESGO DE CREDITO: ANALISIS Y APLICACION AL CASO DE CHILE. *Revista de análisis económico*, 27-53.
- Saju, J., & Chacko, J. (2017). Sustainable development of microfinance customers: An empirical investigation based on India. *Journal of Enterprise Information Management*, (30), 1 - 30. doi: 10.1108/EL-01-2017-0019
- Salim, J. U. H. A. N. A., Yahya, Y. A. R. I. N. A., Othman, M., & Rashid, N. (2007, December). The Use of Holistic Approach to Knowledge Management Initiative in Managing Information in Higher Learning Institution: A Perspective. In 6th WSEAS International Conference on E-ACTIVITIES (pp. 347-352). https://www.researchgate.net/profile/Juhana-Salim-3/publication/237571770_The_Use_of_Holistic_Approach_to_Knowledge_Management_Initiative_in_Managing_Information_in_Higher_Learning_Institution_A_Perspective/links/549792b50cf29b94482624a8/The-Use-of-Holistic-Approach-to-Knowledge-Management-Initiative-in-Managing-Information-in-Higher-Learning-Institution-A-Perspective.pdf

- Schuyten, G. (1991). Statistical thinking in psychology and education. biblio.ugent.be, 486-490.
- Sánchez (2021). El riesgo crediticio y la morosidad en la caja municipal de ahorro y crédito Huancayo S.A.- Wanchaq, 2021.
<https://hdl.handle.net/20.500.12692/85938>
- Takeyas, B. (2007). Introducción a la inteligencia artificial. itnuevolaredo.edu.mx, 1-3.
- Technica, G. B.-S. (2005). LAS MÁQUINAS DE SOPORTE VECTORIAL (SVMs). revistas.utp.edu.co, 47-101.
- Timarán-Pereira, S. R., Hernández-Arteaga, I., Caicedo-Zambrano, S. J., Hidalgo-Troya, A. y AlvaradoPérez, J. C. (2016). El proceso de descubrimiento de conocimiento en bases de datos. En Descubrimiento de patrones de desempeño académico con árboles de decisión en las competencias genéricas de la formación profesional (pp. 63-86). Bogotá: Ediciones Universidad Cooperativa de Colombia. doi: <http://dx.doi.org/10.16925/9789587600490>
- Torres-Quezada, Y. (2021). Minería de datos para determinar los factores más influyentes en la ocurrencia de siniestros de tránsito en Ecuador en el año 2020. CEDAMAZ, 11(2), 124-132., DOI: 10.54753/cedamaz.v11i2.1181
- Toro, L. (2017). Anaconda Distribution: La Suite más completa para la Ciencia de datos con Python. línea]. Available: <https://blog.desdelinux.net/ciencia-de-datos-con-python>
- Torra, V. (2011). La inteligencia artificial. *Lychnos*, (7), 14-19.
- Thomas, E. y Manz, D. (2017). Research Methods for Cyber Security. Synggress, 269. Recuperado de: <https://doi.org/10.1016/B978-0-12-805349-2.00035-2>

- Thuraisingham, B., & Maning, D. (1999). technologies, techniques, tools, and Trends. CRC press.
- Trujillo Fernández, D. (2017). Aplicación de metodologías Machine Learning en la gestión de riesgo de crédito. oa.upm.es, 69-70
- USKOV, V, 2019. Machine Learning – based Predictive Analytics of Student Academic Performance in STEM Education. IEEE Global Engineering Education Conference (EDUCON), 2019. pp. 1370-1376.
- Vazques,(2019): sistema predictivo basado en un modelo credit scoring de aprendizaje automático para la medición del riesgo crediticio en los créditos pyme de la edpyme alternativa s.a.
- Vela, Z. S., & Caro, A. A. (2015). Herramientas financieras en la evaluación del riesgo de crédito (1 era. ed.). Lima: Fondo Editorial de la UIGV.
- Villada, F., Muñoz, N., & García-Quintero, E. (2016). Redes neuronales artificiales aplicadas a la predicción del precio del oro. Información tecnológica, 27(5), 143-150. <http://dx.doi.org/10.4067/S0718-07642016000500016>
- WHITNEY, F. L. (2016). Elementos de investigación. Barcelona. Omega.

Anexo 1: Matriz de Consistencia

TÍTULO: Machine Learning para predecir el riesgo crediticio de un cliente en la empresa FUTECH PERU S.A.C, 2022

AUTOR: Italo Paul Yaranga Vite

PROBLEMAS	OBJETIVOS	HIPÓTESIS	VARIABLES	DIMENSIONES	INDICADOR	Metodología
¿En qué medida Machine Learning permite predecir el riesgo crediticio de un cliente en la empresa FUTECH PERÚ SAC, 2022?	Aplicar Machine Learning para predecir el riesgo crediticio de un cliente en la empresa FUTECH PERÚ SAC, 2022.	Machine Learning predice significativamente el riesgo crediticio de un cliente en la empresa FUTECH PERÚ SAC, 2022.	Variable Independiente: Machine Learning es una técnica moderna en el desarrollo software el cual nos permite aprender mediante datos de entrenamiento. Según Hurwitz y Kirsch (2018, P. 20)			Tipo De Investigación: Aplicada Diseño de Investigación: Experimental de tipo Pre Experimental
¿En qué medida Machine Learning permite predecir con precisión el riesgo crediticio de un cliente en la empresa FUTECH PERÚ SAC, 2022?	Determinar en qué porcentaje Machine Learning permite predecir con precisión el riesgo crediticio de un cliente en la empresa FUTECH PERÚ SAC, 2022.	Machine Learning predice significativamente con precisión el riesgo crediticio de un cliente en la empresa FUTECH PERÚ SAC, 2022.	Variable dependiente: predecir el riesgo crediticio de un cliente. Según. Lara (2016, P.91)	Métricas de precisión Según Ortega, (2021)	Precisión $(TP+TN/TP+TN+FP+FN) * 100$ Sensibilidad $(TP/TP+FN) x 100$ Especificidad $(TN/TN+FP) x 100$ Ortega,2021, P..27	

<p>¿En qué medida Machine Learning permite predecir con sensibilidad el riesgo crediticio de un cliente en la empresa FUTECH PERÚ SAC, 2022?</p>	<p>Determinar en qué porcentaje Machine Learning permite predecir con sensibilidad el riesgo crediticio de un cliente en la empresa FUTECH PERÚ SAC, 2022.</p>	<p>Machine Learning predice significativamente con sensibilidad el riesgo crediticio de un cliente en la empresa FUTECH PERÚ SAC, 2022.</p>				
<p>¿En qué medida Machine Learning permite predecir con exactitud el riesgo especificidad de un cliente en la empresa FUTECH PERÚ SAC, 2022?</p>	<p>Determinar en qué porcentaje Machine Learning permite predecir con especificidad el riesgo crediticio de un cliente en la empresa FUTECH PERÚ SAC, 2022.</p>	<p>Machine Learning predice significativamente con especificidad el riesgo crediticio de un cliente en la empresa FUTECH PERÚ SAC, 2022.</p>				

Anexo 2: Operacionalización de la Variable

TÍTULO: Machine Learning para predecir el riesgo crediticio de un cliente en la empresa FUTECH PERU S.A.C, 2022

AUTOR: Italo Paul Yaranga Vite

Variable	Definición Conceptual	Definición Operacional	Dimensión	Indicadores	Escala de medición
Independiente: Machine Learning	Machine Learning es una técnica moderna en el desarrollo software el cual nos permite aprender mediante datos de entrenamiento con la finalidad de predecir datos. Según Hurwitz y Kirsch (2018, P. 20)	Machinne Learning permite predecir el riesgo crediticio de un cliente mediante un modelo de regresión logística mediante data histórica y dato actuales.			
Variable dependiente: predecir el riesgo crediticio de un cliente	Predecir el riesgo crediticio de un cliente son procedimientos estadísticos que se usan para clasificar a aquellos que solicitan crédito en tipo de riesgo bueno o malo. Lara (2016, P.91)	Para medir la predicción del riesgo crediticio de un cliente se usa las métricas de precisión, sensibilidad y especificidad, la cual será obtenida a través de Machine Learning.	Métricas de precisión Según (Ortega, 2021)	Precisión: Es la relación entre las predicciones correctas y el número total de predicciones correctas. Sensibilidad: evalúa la exactitud de las muestras positivas y las negativas respectivamente. Especificidad: son pronósticos acertados y la cantidad total de pronósticos, con qué repetición es acertado el clasificador Ortega, (2021, P..27)	Razón

ANEXO N° 3: Comparación de algoritmos de clasificación de aprendizaje automático

ML Algorithm	Main focus on task type	Parametric	Training speed	Prediction speed	Automatically learns features
Linear Regression	Regression	Yes	Fast	Fast	No
Logistic Regression	Classification	Yes	Fast	Fast	No
Decision Tree Classification	Either	No	Fast	Depends on value of "n"	Yes
K Nearest Neighbor	Classification	No	Fast	Fast	No
ANN	Either	No	Slow	Fast	Yes
Naïve Bayes Classification	Classification	Yes	Fast (except feature extraction)	Fast	No
Random Forest Classification	Classification	No	Slow	Mode-rate	Yes
SVM Classification	Either	No	Slow	Fast	Yes

Fuente: Uskov (2019)

ANEXO N° 4: INSTRUMENTO DE OBSERVACIÓN PARA EL ALGORITMO
SUPPORT VECTOR MACHINE
FICHA DE REGISTRO

Tipo de Prueba	Post Test
Investigador	Italo Paul Yaranga Vite
Fecha de inicio	

Algoritmo	Support Vector Machine
-----------	------------------------

Matriz de confusión:

		Predicción	
		Positive	Negative
Observación	Positive	(TP) True Positive	(FN) False Negative
	Negative	(FP) False Positive	(TN) True Negative

Métricas a Evaluar:

Ítem	Indicador	Medida	Fórmula	Precisión
1	Precisión	Razón	$Precisión = (TP+TN/TP+TN+FP+FN) * 100$	99,8%
2	Sensibilidad	Razón	$Sensibilidad = (TP/TP+FN) * 100$	100%
3	Especificidad	Razón	$Especificidad = (TN/TN+FP) * 100$	99,7%

ANEXO N° 5: INSTRUMENTO DE OBSERVACIÓN PARA EL ALGORITMO
RANDOM FOREST
FICHA DE REGISTRO

Tipo de Prueba	Post Test
Investigador	Italo Paul Yaranga Vite
Fecha de inicio	

Algoritmo	Random Forest
-----------	---------------

Matriz de confusión:

		Predicción	
		Positive	Negative
Observación	Positive	(TP) True Positive	(FN) False Negative
	Negative	(FP) False Positive	(TN) True Negative

Métricas a Evaluar:

Ítem	Indicador	Medida	Fórmula	Precisión
1	Precisión	Razón	$Precisión = (TP+TN / TP+TN+FP+FN) * 100$	96,2%
2	Sensibilidad	Razón	$Sensibilidad = (TP / TP+FN) * 100$	98,6%
3	Especificidad	Razón	$Especificidad = (TN / TN+FP) * 100$	95,3%

ANEXO N° 6: INSTRUMENTO DE OBSERVACIÓN PARA EL ALGORITMO
 NAIBES BAYES
 FICHA DE REGISTRO

Tipo de Prueba	Post Test
Investigador	Italo Paul Yaranga Vite
Fecha de inicio	

Algoritmo	Naibes Bayes
-----------	--------------

Matriz de confusión:

		Predicción	
		Positive	Negative
Observación	Positive	(TP) True Positive	(FN) False Negative
	Negative	(FP) False Positive	(TN) True Negative

Métricas a Evaluar:

Ítem	Indicador	Medida	Fórmula	Precisión
1	Precisión	Razón	$Precisión = (TP+TN / TP+TN+FP+FN) * 100$	94,6%
2	Sensibilidad	Razón	$Sensibilidad = (TP / TP+FN) * 100$	87%
3	Especificidad	Razón	$Especificidad = (TN / TN+FP) * 100$	97,5%

ANEXO N° 7: INSTRUMENTO DE OBSERVACIÓN PARA EL ALGORITMO
K VECINO MÁS CERCANOS
FICHA DE REGISTRO

Tipo de Prueba	Post Test
Investigador	Italo Paul Yaranga Vite
Fecha de inicio	

Algoritmo	K Nearest Neighbor
-----------	--------------------

Matriz de confusión:

		Predicción	
		Positive	Negative
Observación	Positive	(TP) True Positive	(FN) False Negative
	Negative	(FP) False Positive	(TN) True Negative

Métricas a Evaluar:

Ítem	Indicador	Medida	Fórmula	Precisión
1	Precisión	Razón	$Precisión = \frac{TP+TN}{TP+TN+FP+FN} * 100$	86,8%
2	Sensibilidad	Razón	$Sensibilidad = \frac{TP}{TP+FN} * 100$	67,4%
3	Especificidad	Razón	$Especificidad = \frac{TN}{TN+FP} * 100$	94,2%

ANEXO N° 8: INSTRUMENTO DE OBSERVACIÓN PARA EL ALGORITMO
DECISION TREE
FICHA DE REGISTRO

Tipo de Prueba	Post Test
Investigador	Italo Paul Yaranga Vite
Fecha de inicio	

Algoritmo	Decision Tree
-----------	---------------

Matriz de confusión:

		Predicción	
		Positive	Negative
Observación	Positive	(TP) True Positive	(FN) False Negative
	Negative	(FP) False Positive	(TN) True Negative

Métricas a Evaluar:

Ítem	Indicador	Medida	Fórmula	Precisión
1	Precisión	Razón	$Precisión = (TP+TN / TP+TN+FP+FN) * 100$	82,6%
2	Sensibilidad	Razón	$Sensibilidad = (TP / TP+FN) * 100$	55,1%
3	Especificidad	Razón	$Especificidad = (TN / TN+FP) * 100$	93,1%

Anexo 9: Instrumento para la Recolección de Datos Cuestionario

Item	Factor	Aspecto	Valor
1	Datos Personales	¿Género biológico?	Masculino / Femenino
2		¿Cuál es tu edad?	Dato Numérico entre 18 - 70
3		¿A qué distrito perteneces?	
4		¿Cuál es tu estado civil?	Soltero/Casado/Divorciado/Viudo
5		¿Tipo de vivienda?	Alquilada/Familiar/Propia
6		¿Cuál es su nivel instrucción?	Primarios, Secundarios, Técnica, Universitaria
7	Datos Laborales	¿Cuál es su cargo actual?	Empleado/Ejecutivo/funcionario/Gerente General
8		¿Cuánto tiempo lleva laborando?	Dato Numérico
9		¿Cuál es su modalidad de contrato laboral?	Tiempo parcial/Tiempo completo
10		¿Tipo de sector socioeconómico donde labora?	Industrial/Agricultura/Ganadería/Transporte/Pesca/Telecomunicaciones/Construcción/Comercio
11		¿Cuál es su situación laboral?	Dependiente/Independiente/Casa/Jubilado/Estudiante
12		¿Cuál es su ingreso bruto mensual?	Dato Numérico
13		¿Cuál es su monto de otros ingresos?	Dato Numérico
14	Datos Laborales del Conyugue	¿Cuál es su situación laboral?	Dependiente/Independiente/Casa/Jubilado/Estudiante
15		¿Cuál es su cargo actual?	Empleado/Ejecutivo/funcionario/Gerente General
16		¿Cuánto tiempo lleva laborando?	Dato Numérico
17		¿Cuál es su modalidad de contrato laboral?	Tiempo parcial/Tiempo completo
18		¿Cuál es su ingreso bruto mensual?	Dato Numérico
19		¿Cuál es su monto de otros ingresos?	Dato Numérico
20	Datos Financieros	¿Tiene prestamos con otras instituciones financieras?	SI / NO
21		¿Cuenta con Vehículo, Casa u otros bienes?	SI / NO
22		¿Cuenta con depósito a plazo fijo?	SI / NO
23	Socio cultural	¿Cuántos integrantes hay en su grupo familiar?	Dato Numérico
24		¿Cuántos integrantes de su grupo familiar trabajan?	Dato Numérico
25		¿Cuántos integrantes de su grupo familiar estudian?	Dato Numérico
26	Variable objetivo	¿Cuál es el tipo de riesgo crediticio?	Alto//Bajo

ANEXO N° 10: VALIDACIÓN DE INSTRUMENTO - EXPERTO 1

Apellidos y Nombre del experto: Coronel Castillo Erick Gustavo

Título y/o grado: Maestro en Ingeniero de Sistemas e Informática

Fecha: 24/10/2022

Nombre del instrumento: Cuestionario

Autor: Italo Paul Yaranga Vite

Título de investigación: Machine Learning para predecir el riesgo crediticio de un cliente en la empresa FUTECH PERU S.A.C, 2022.

INDICADORES	CRITERIOS	Deficiente 0-20%	Regular 21-50%	Bueno 51- 70%	Muy Bueno 71-80%	Excelente 81-100%
Claridad	Promedio de Validación					95
Objetividad	Esta expresado en conducta observable.					95
Actualidad	Es adecuado al avance de la ciencia.					95
Organización	Existe una organización lógica					95
Suficiencia	Comprende los aspectos de cantidad y calidad.					95
Intencionalidad	Adecuado para valorar aspectos del sistema metodológico y científico					95
Consistencia	Está basado en aspectos teóricos, científicos acordes a la tecnología.					95
Coherencia	Está basado en aspectos teóricos, científicos acordes a la tecnología.					95
Metodología	Responde al propósito del trabajo bajo los objetivos a lograr					95
Pertinencia	El instrumento es adecuado al tipo de investigación.					95
Promedio de Validación						95

Promedio de valoración: 95

Observaciones:



FIRMA

ANEXO N° 11: VALIDACIÓN DE INSTRUMENTO - EXPERTO 2

Apellidos y Nombre del experto: JOSE LUIS TRELLES SUCA

Título y/o grado: MAESTRO EN EDUCACIÓN CON MENCIÓN EN DOCENCIA E INVESTIGACIÓN UNIVERSITARIA

Fecha: 24/10/2022

Nombre del instrumento: Cuestionario

Autor: Italo Paul Yaranga Vite

Título de investigación: Machine Learning para predecir el riesgo crediticio de un cliente en la empresa FUTECH PERU S.A.C, 2022.

INDICADORES	CRITERIOS	Deficiente 0-20%	Regular 21-50%	Bueno 51- 70%	Muy Bueno 71-80%	Excelente 81-100%
Claridad	Promedio de Validación					95
Objetividad	Esta expresado en conducta observable.					95
Actualidad	Es adecuado al avance de la ciencia.					95
Organización	Existe una organización lógica					95
Suficiencia	Comprende los aspectos de cantidad y calidad.					95
Intencionalidad	Adecuado para valorar aspectos del sistema metodológico y científico					95
Consistencia	Está basado en aspectos teóricos, científicos acordes a la tecnología.					95
Coherencia	Está basado en aspectos teóricos, científicos acordes a la tecnología.					95
Metodología	Responde al propósito del trabajo bajo los objetivos a lograr					95
Pertinencia	El instrumento es adecuado al tipo de investigación.					95
Promedio de Validación						95

Promedio de valoración: 95

Observaciones:



FIRMA

ANEXO N° 12: VALIDACIÓN DE INSTRUMENTO - EXPERTO 3

Apellidos y Nombre del experto: Juscamaita Rico Elvis Harol

Título y/o grado: Maestro en Ingeniero de Sistemas e Informática

Fecha: 24/10/2022

Nombre del instrumento: Cuestionario

Autor: Italo Paul Yaranga Vite

Título de investigación: Machine Learning para predecir el riesgo crediticio de un cliente en la empresa FUTECH PERU S.A.C, 2022.

INDICADORES	CRITERIOS	Deficiente 0-20%	Regular 21-50%	Bueno 51- 70%	Muy Bueno 71-80%	Excelente 81-100%
Claridad	Promedio de Validación					99
Objetividad	Esta expresado en conducta observable.					99
Actualidad	Es adecuado al avance de la ciencia.					97
Organización	Existe una organización lógica					99
Suficiencia	Comprende los aspectos de cantidad y calidad.					95
Intencionalidad	Adecuado para valorar aspectos del sistema metodológico y científico					98
Consistencia	Está basado en aspectos teóricos, científicos acordes a la tecnología.					95
Coherencia	Está basado en aspectos teóricos, científicos acordes a la tecnología.					98
Metodología	Responde al propósito del trabajo bajo los objetivos a lograr					95
Pertinencia	El instrumento es adecuado al tipo de investigación.					97
Promedio de Validación						97.2

Promedio de valoración: 97.2

Observaciones:



FIRMA

ANEXO N° 13: Autorización de la empresa en estudio Futech

Solicitud de Carta de Presentación para Investigación.



FACULTAD DE
INGENIERÍA Y ARQUITECTURA

Fecha		
Día	Mes	Año
01	08	2022

FICHA PARA SOLICITAR CARTA DE PRESENTACIÓN PARA INVESTIGACIÓN

DATOS DEL ALUMNO:

Apellido Paterno	Apellido Materno	Nombres
Yaranga	Vite	Italo Paul
Teléfono Fijo / Celular	DNI	E-mail
952834106	40245171	ipyv@hotmail.com

DATOS DE LA EMPRESA:

PUBLICA <input type="checkbox"/>		PRIVADA <input checked="" type="checkbox"/>	
Razón Social		Teléfono	
FUTECH PERU S.A.C.		(511) 428.5888	
Dirección		E-mail	
JR. MOQUEGUA NRO. 157 INT. 211		main@futech.com.pe	
Distrito	Provincia	Departamento	
LIMA	LIMA	LIMA	
Apellidos y Nombres del Representante		Cargo	
CASTRO LUNA, ANTONIO FLORENTINO		GERENTE GENERAL	

Observaciones:

Firma de Director(a) Escuela

ANTONIO CASTRO LUNA
GERENTE GENERAL
FUTECH PERU S.A.C






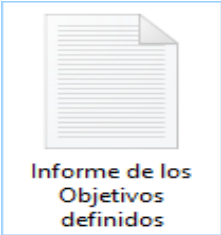
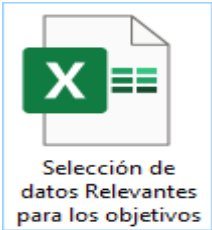
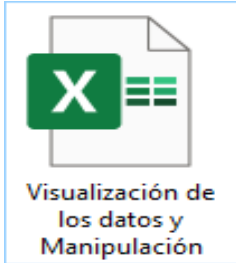
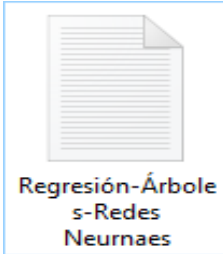
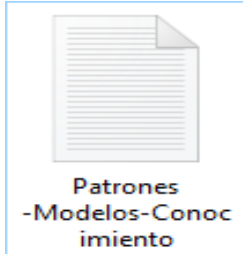
Firma del Solicitante

CAMPUS: Lima norte



Firmado digitalmente por:
CASTRO LUNA ANTONIO
FLORENTINO FIR 09373851 hard
Motivo: En señal de
conformidad
Fecha:01/08/2022 12:34:41-0500

Anexo 16: Gráfico de la Metodología de la tecnológica.

METODOLOGÍA		Basada en Knowledge Discovery in Databases o KDD				
SELECCIÓN		PRE-PROCESAMIENTO	TRANSFORMACIÓN	MINERIA DE DATOS	INTERPRETACION	
<ul style="list-style-type: none"> • Se definen los objetivos específicos a lograr. • Selección de fuente de datos. 		Manipular los datos para eliminación de ruido y datos aislados o atípicos.	Continuar reduciendo la cantidad de datos para el modelo, Visualizar los datos de forma gráfica para tener una mejor perspectiva de las Variables.	Seleccionar la técnica de Machine Learning como: Modelo regresión, Árboles de decisión y Redes Neuronales	Verificación del logro de los objetivos establecido con respecto a la predicción de riesgo de un cliente	
						
Entregable	 <p>Informe de los Objetivos definidos</p>	 <p>Selección de datos Relevantes para los objetivos</p>	 <p>Visualización de los datos y Manipulación</p>	 <p>Regresión-Árboles-Redes Neuronales</p>	 <p>Patrones -Modelos-Conocimiento</p>	
	10%	30%	20%	30%	10%	

Desarrollo del modelo Machine Learning para predecir el Riesgo crediticio de un cliente en la empresa FUTECH PERU S.A.C, 2022

En la investigación realizada empieza con un estudio de la problemática desde un enfoque holístico partiendo desde una revisión literaria de la tecnología de software de Machine Learning, Metodología KDD y el riesgo crediticio de un cliente. Se seleccionaron las bases de datos históricas que relacionan características personales, socioeconómicas y laborales de los clientes, los datos fueron seleccionados, procesados y transformado con el propósito de obtener un conjunto de datos depurado y listo para la aplicación de algoritmos de Minería de Datos. Se descubrieron reglas y perfiles personales, socioeconómicos y laborales de los clientes, se utilizaron tres modelos de predicción como: Decision Tree, Random Forest, Support Vector Machine, Naïves Bayes, K Nearest Neighbor y como herramienta se utilizó la Suite de Anaconda para implementación para la minería de datos, los patrones resultantes fueron interpretados, evaluados, y finalmente usados para soportar la toma de decisiones en el proyecto actual, permitiendo mejorar los valores de precisión, sensibilidad y Especificidad.

Se planteó el desarrollo de un Sistema web con el propósito de ayudar al personal del área de evaluación crediticia para predecir el riesgo crediticio de los clientes en la empresa FUTECH PERU S.A.C. Este sistema utilizó la tecnología de software Machine Learning y la metodología KDD, lo cual permitirá llegar a tener una mejor precisión, sensibilidad y Especificidad con respecto a la predicción del riesgo crediticio de un cliente.

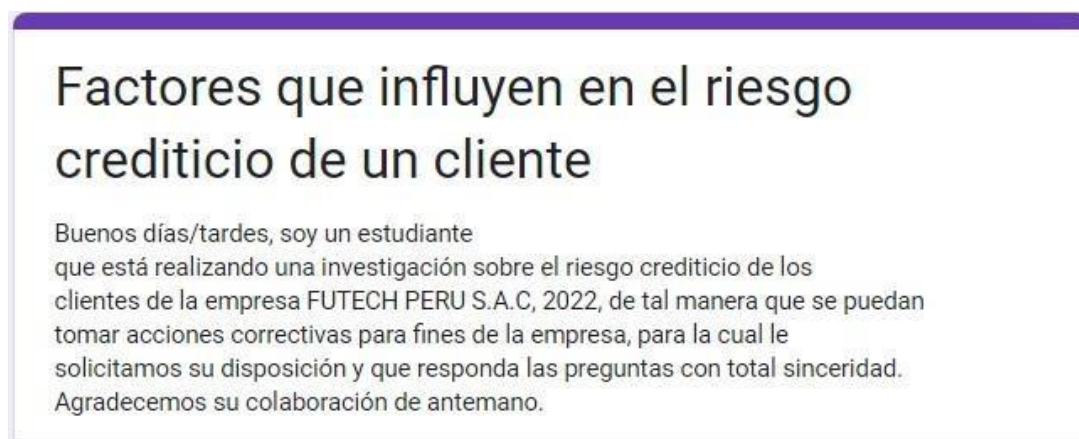
Uso de la Suite de Anaconda y la herramienta de “Jupyter Notebook” con Python para realizar ciencia de datos con el archivo de datos de la empresa Futech(bdfutech.csv).

Etapa 1: Etapa de selección de datos

El fin de esta investigación es predecir el riesgo crediticio de un cliente, de tal manera que se logre identificar aquellos clientes con probabilidades de alto riesgo crediticio, buscando que la empresa tome las mejores decisiones para sobreguarda el patrimonio de la empresa.

Para la obtención de los datos de cliente, se realizó un cuestionario que consta de 27 preguntas los cuales estuvieron relacionadas con determinantes personales (12 preguntas), laborales (7 preguntas), financiera (3 preguntas), socio-culturales (3 preguntas) y variable objetiva (1 pregunta), en el ANEXO N° 13 se puede observar en detalle las preguntas. Estas preguntas fueron transcritas a google formulario y enviadas a los clientes de la empresa en mención, obteniendo un total de 500 registros.

Fig. 11 Cuestionario utilizando Google Forms



Factores que influyen en el riesgo crediticio de un cliente

Buenos días/tardes, soy un estudiante que está realizando una investigación sobre el riesgo crediticio de los clientes de la empresa FUTECH PERU S.A.C, 2022, de tal manera que se puedan tomar acciones correctivas para fines de la empresa, para la cual le solicitamos su disposición y que responda las preguntas con total sinceridad. Agradecemos su colaboración de antemano.

Fuente: Elaboración propia

Etapa 2: Etapa de pre-procesamiento / limpieza de datos

Luego de ejecutar el cuestionario, se pudo recolectar 500 registros clientes, los cuales fueron exportados a la herramienta SPSS Statistic.

En primer lugar, se realizó la definición de las variables, las cuales están asociadas a las preguntas del cuestionario, se definió que los valores sean de tipo numérico, la anchura de 8, en la columna "Decimales" se utilizó para la variable Ingreso_bruto_mensual 1 decimal. Luego en la columna "Valores", se añadió las alternativas por cada pregunta y su equivalencia en números, para la columna "Perdidos" se utilizó el valor 99. También en la columna "Medida" se seleccionó entre ordinal, escalar y nominal según corresponde y finalmente en la columna "Rol" se seleccionó el tipo de dato, es decir si es de entrada, destino, etc. En nuestro caso se eligió para todas las variables, tipo Entrada.

Figura 12: Definición de variables - SPSS Statistics

Nombre	Tipo	Anchura	Deci...	Etiqueta	Valores	Perdidos	Col...	Alineación	Medida	Rol
3 Distrito	Númérico	8	0	¿A qué distrito perteneces?	{1, Ate}...	99	12	Izquierda	Nominal	Entrada
4 EstadoCivil	Númérico	8	0	¿Cuál es tu estado civil?	{1, Soltero}...	Ninguno	12	Izquierda	Nominal	Entrada
5 Vivienda	Númérico	8	0	¿Tipo de vivienda?	{1, Alquilada}...	Ninguno	12	Izquierda	Nominal	Entrada
6 NivelInstrucción	Númérico	8	0	¿Cuál es su nivel instrucción?	{1, Primarios}...	Ninguno	12	Izquierda	Nominal	Entrada
7 Cargo	Númérico	8	0	¿Cuál es su cargo actual?	{1, Emplead}...	Ninguno	12	Izquierda	Nominal	Entrada
8 TiempoLaborando	Númérico	8	0	¿Cuánto tiempo lleva laborando?	Ninguno	Ninguno	12	Izquierda	Escala	Entrada
9 ModalidadContrato	Númérico	8	0	¿Cuál es su modalidad de contrato laboral?	{1, Tiempo ...	Ninguno	12	Derecha	Nominal	Entrada
10 SectorSocioeconómico	Númérico	8	0	¿Tipo de sector socioeconómico donde labora?	{1, Industrial}...	Ninguno	12	Derecha	Nominal	Entrada
11 SituaciónLaboral	Númérico	8	0	¿Cuál es su situación laboral?	{1, Dependie}...	Ninguno	12	Derecha	Nominal	Entrada
12 IngresoBrutoMensual	Númérico	8	1	¿Cuál es su ingreso bruto mensual?	Ninguno	Ninguno	12	Izquierda	Escala	Entrada
13 OtrosIngresos	Númérico	8	1	¿Cuál es su monto de otros ingresos?	Ninguno	Ninguno	12	Izquierda	Escala	Entrada
14 SituaciónLaboral_conyugue	Númérico	8	0	¿Cuál es su situación laboral conyugue?	Ninguno	Ninguno	12	Izquierda	Nominal	Entrada
15 Cargoactual_conyugue	Númérico	8	0	¿Cuál es su cargo actual conyugue?	Ninguno	Ninguno	12	Derecha	Nominal	Entrada
16 TiempoLaborando_conyugue	Númérico	8	0	¿Cuánto tiempo lleva laborando conyugue?	Ninguno	Ninguno	12	Izquierda	Escala	Entrada
17 Modalidad_contratolaboral_conyugue	Númérico	8	0	¿Cuál es su modalidad de contrato laboral conyugue?	{1, Tiempo C...	Ninguno	12	Izquierda	Nominal	Entrada
18 IngresoBruto_mensual_conyugue	Númérico	8	1	¿Cuál es su ingreso bruto mensual conyugue?	Ninguno	Ninguno	12	Derecha	Escala	Entrada
19 OtrosIngresos_conyugue	Númérico	8	1	¿Cuál es su monto de otros ingresos conyugue?	Ninguno	Ninguno	12	Derecha	Escala	Entrada
20 Productos_otras_institucionesfinancieras	Númérico	8	0	¿Tiene productos con otras instituciones financieras?	{0, No}...	Ninguno	12	Derecha	Nominal	Entrada
21 Bienes_casa_vehiculo	Númérico	8	0	¿Cuenta con Vehículo, Casa u otros bienes?	{0, No}...	Ninguno	12	Derecha	Nominal	Entrada
22 Depósito_plazofijo	Númérico	8	0	¿Cuenta con depósito a plazo fijo?	{0, No}...	Ninguno	12	Derecha	Nominal	Entrada
23 Cantidad_Integrantes_grupofamiliar	Númérico	8	0	¿Cuántos integrantes hay en su grupo familiar?	Ninguno	Ninguno	12	Derecha	Escala	Entrada
24 Cantidad_integrantes_grupofamiliar_trabajan	Númérico	8	0	¿Cuántos integrantes de su grupo familiar trabajan?	Ninguno	Ninguno	12	Derecha	Escala	Entrada
25 Cantidad_Integrantes_grupofamiliar_estudian	Númérico	8	0	¿Cuántos integrantes de su grupo familiar estudian?	Ninguno	Ninguno	12	Derecha	Escala	Entrada
26 Riesgocrediticio	Númérico	8	0	¿Cuál es el tipo de riesgo crediticio?	{1, Bajo}...	Ninguno	12	Derecha	Nominal	Objetivo

Fuente: Elaboración propia

A continuación, se da detalle sobre el valor asignado por cada variable:

Genero	
Femenino	2
Masculino	1
Edad	
18 a 27	1
28 a 38	2
39 a 49	3
50 a 60	4
Distrito	
Ate	1
Barranco	2
Chorrillos	3
Otros	4
EstadoCivil	
Soltero	1
Casado	2
Divorciado	3
Viudo	4
Vivienda	
Alquilada	1
Familiar	2
Propia	3
NivelInstruccion	
Primaria	1
Secundaria	2
Técnica	3
Universitaria	4
Cargo	
Empleado	1
Ejecutivo	2

Funcionario	3
Gerente General	4
TiempoLaborando	
1 a 3	1
4 a 5	2
6 a más	3
ModalidadContrato	
Tiempo Parcial	1
Tiempo Completo	2
SectorSocioeconomico	
Industria	1
Agricultura	2
Ganadería	3
Transporte	4
Pesca	5
SituacionLaboral	
Estudiante	1
Casa	2
Jubilado	3
Independiente	4
Dependiente	5
IngresoBrutoMensual	
1025 a 2500	1
2501 a 4000	2
4001 a 8000	3
8001 a 10000	4
10001 a más	5
OtrosIngresos	
1025 a 2500	1
2501 a 4000	2
4001 a 8000	3
8001 a 10000	4
10001 a más	5
SituacionLaboral_conyugue	
Estudiante	1
Casa	2
Jubilado	3
Independiente	4
Dependiente	5
Cargo_Conyugue	
Empleado	1
Ejecutivo	2
Funcionario	3
Gerente General	4
TiempoLaborando_Conyugue	
1 a 3	1
4 a 5	2
6 a más	3
Modalidad_contratolaboral_conyugue	
Tiempo Parcial	1
Tiempo Completo	2
IngresoBrutoMensual_Conyugue	
1025 a 2500	1
2501 a 4000	2
4001 a 8000	3

8001 a 10000	4
10001 a más	5
OtrosIngresos_Conyugue	
1025 a 2500	1
2501 a 4000	2
4001 a 8000	3
8001 a 10000	4
10001 a más	5
Prestamos_otras_institucionesfinancieras	
Si	1
No	2
Bienes_casa_vehiculo	
Si	2
No	1
Depósito_plazofijo	
Si	2
No	1
Cantidad_Integrantes_grupofamiliar	
1 a 2	1
3 a 5	2
6 a más	3
Cantidad_integrantes_grupofamiliar_trabajan	
1 a 2	1
3 a 5	2
6 a más	3
Cantidad_integrantes_grupofamiliar_estudian	
1 a 2	3
3 a 5	2
6 a más	1

En segundo lugar, de los datos exportados de google formulario a Excel, equivalentes a un total de 500 registros clientes. Se hizo un análisis de los datos, corroborando que no existiera datos nulos, también se homogenizaron las respuestas con relación a las siguientes preguntas:

- ¿Cuál es tu edad?, la respuesta debe ser de tipo numérico
Se corrigió el registro N°49 y 54

- ¿A qué distrito perteneces?

Se homogenizaron las respuestas para todos los registros

- ¿Cuál es su cargo?

Se homogenizaron las respuestas para todos los registros se corrigió el registro N°49 y 54

- ¿Cuál es tú modalidad de contrato?

Se homogenizaron las respuestas para todos los registros se corrigió el registro N°7 y 11

Figura 13: Datos importados de Google Forms a Excel

A	B	C	D	E	F	G	H	I	J	K	
Femenino	51	Chorrillos	Soltero	Alquilada	Primaria	Ejecutivo		2	Tiempo Parcial	Pesca	Dependiente
Masculino	22	Carcado de lirn	Soltero	Propia	Universitario	Gerente		1	Tiempo indefinido	Pesca	Independiente
Masculino	39	Ate	Soltero	Familiar	Universitario	Empleado		1	Tiempo Completo	Agricola	Independiente
Masculino	36	Carcado de lirn	Soltero	Propia	Técnico	Empleado		3	Tiempo Completo	Transporte	Dependiente
Femenino	42	Barranco	Viudo	Alquilada	Secundaria	Gerente		3	Tiempo Parcial	Pesca	Jubilado
Femenino	54	Carcado de lirn	Viudo	Alquilada	Secundaria	Empleado		1	Tiempo Parcial	Agricola	Dependiente
Femenino	20	Carcado de lirn	Soltero	Alquilada	Universitario	Gerente		2	Tiempo Completo	Transporte	Casa
Femenino	56	Ate	Casado	Alquilada	Primaria	Empleado		3	Tiempo temporal	Pesca	Dependiente
Femenino	18	Ate	Casado	Propia	Secundaria	Gerente		3	Tiempo Completo	Pesca	Estudiante
Femenino	23	Carcado de lirn	Viudo	Alquilada	Técnico	Funcionario		1	Tiempo Completo	Industria	Dependiente
Femenino	34	Carcado de lirn	Soltero	Alquilada	Universitario	Colaborador		2	Tiempo Completo	Industria	Jubilado
Femenino	50	Chorrillos	Soltero	Propia	Primaria	Empleado		1	Tiempo Completo	Industria	Casa
Femenino	19	Carcado de lirn	Casado	Alquilada	Universitario	Empleado		2	Tiempo Parcial	Pesca	Dependiente
Femenino	47	Carcado de lirn	Viudo	Alquilada	Técnico	Funcionario		3	Tiempo Completo	Pesca	Dependiente
Femenino	36	Ate	Viudo	Propia	Técnico	Trabajador		1	Tiempo Parcial	Pesca	Dependiente
Masculino	35	Carcado de lirn	Viudo	Propia	Universitario	Empleado		2	Tiempo Completo	Industria	Estudiante
Femenino	41	Barranco	Soltero	Familiar	Primaria	Funcionario		2	Tiempo Parcial	Transporte	Casa
Femenino	27	Carcado de lirn	Viudo	Propia	Primaria	Funcionario		2	Tiempo Completo	Industria	Casa
Masculino	19	Carcado de lirn	Soltero	Alquilada	Técnico	Empleado		1	Tiempo Completo	Agricola	Estudiante
Femenino	36	Ate	Viudo	Alquilada	Secundaria	Funcionario		2	Tiempo Parcial	Pesca	Independiente
Masculino	55	Ate	Viudo	Familiar	Técnico	Funcionario		1	Tiempo Completo	Transporte	Casa
Masculino	59	Carcado de lirn	Viudo	Familiar	Técnico	Ejecutivo		1	Tiempo Parcial	Agricola	Dependiente
Femenino	50	Ate	Viudo	Alquilada	Secundaria	Funcionario		2	Tiempo Completo	Pesca	Dependiente
Masculino	56	Ate	Viudo	Familiar	Universitario	Empleado		2	Tiempo Parcial	Pesca	Dependiente
Masculino	49	Chorrillos	Viudo	Familiar	Universitario	Gerente		1	Tiempo Completo	Transporte	Dependiente
Femenino	49	Ate	Casado	Propia	Primaria	Empleado	1 años	1	Tiempo Parcial	Transporte	Estudiante
Masculino	30	Carcado de lirn	Viudo	Propia	Universitario	Gerente		2	Tiempo Parcial	Pesca	Dependiente
Femenino	25	Ate	Soltero	Familiar	Primaria	Empleado		1	Tiempo Parcial	Pesca	Estudiante
Masculino	28	Carcado de lirn	Viudo	Familiar	Universitario	Empleado		1	Tiempo Parcial	Transporte	Casa
Masculino	54	Chorrillos	Casado	Alquilada	Universitario	Funcionario		2	Tiempo Parcial	Agricola	Casa
Femenino	26	Barranco	Viudo	Familiar	Universitario	Funcionario		1	Tiempo Completo	Transporte	Dependiente
Masculino	50	Carcado de lirn	Casado	Familiar	Secundaria	Funcionario		1	Tiempo Parcial	Agricola	Estudiante
Femenino	38	Ate	Viudo	Familiar	Universitario	Gerente		1	Tiempo Completo	Industria	Independiente
Masculino	46	Ate	Divorciado	Alquilada	Universitario	Gerente		2	Tiempo Completo	Industria	Jubilado
Femenino	56	Barranco	Viudo	Alquilada	Universitario	Funcionario		1	Tiempo Parcial	Transporte	Estudiante
Masculino	21	Chorrillos	Viudo	Propia	Universitario	Gerente	2 años	1	Tiempo Parcial	Agricola	Dependiente
Masculino	53	Chorrillos	Viudo	Alquilada	Primaria	Gerente		1	Tiempo Completo	Pesca	Dependiente
Masculino	32	Carcado de lirn	Soltero	Propia	Secundaria	Empleado		2	Tiempo Completo	Transporte	Jubilado
Femenino	42	Ate	Soltero	Familiar	Técnico	Funcionario		3	Tiempo Parcial	Pesca	Estudiante
Masculino	55	Carcado de lirn	Soltero	Familiar	Universitario	Gerente		3	Tiempo Completo	Pesca	Independiente
Masculino	45	Carcado de lirn	Divorciado	Propia	Primaria	Gerente		2	Tiempo Completo	Pesca	Estudiante
Masculino	56	Ate	Casado	Propia	Universitario	Funcionario		2	Tiempo Completo	Pesca	Independiente
Masculino	47	Carcado de lirn	Viudo	Propia	Primaria	Gerente		2	Tiempo Completo	Industria	Dependiente
Masculino	50 años	Carcado de lirn	Soltero	Familiar	Técnico	Funcionario		2	Tiempo Completo	Pesca	Estudiante
Femenino	55	Carcado de lirn	Soltero	Alquilada	Universitario	Empleado		2	Tiempo Completo	Industria	Dependiente
Femenino	43	Barranco	Viudo	Propia	Universitario	Gerente		2	Tiempo Parcial	Pesca	Dependiente

Fuente: Elaboración propia

Luego de realizar la tarea mencionada anteriormente se importó los datos aSPSS
Statistic

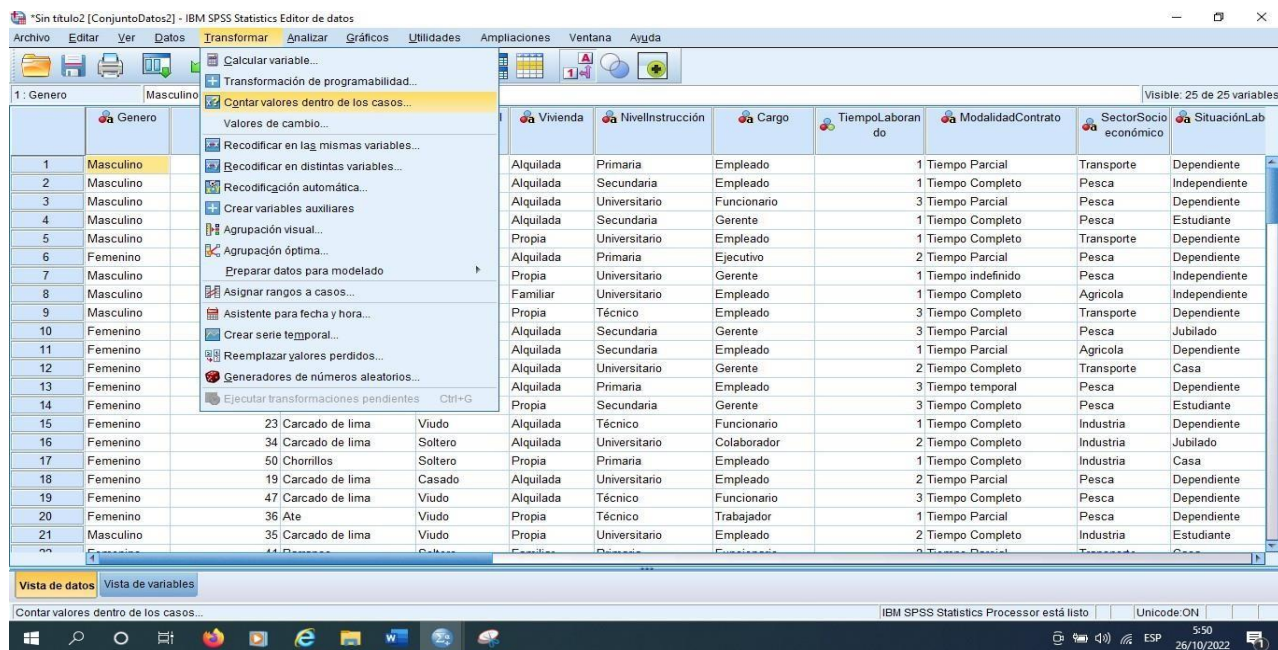
Figura 14: Datos importados de Excel a SPSS Statistics

	Genere	Edad	Distrito	EstadoCivil	Vivienda	NivelInstrucción	Cargo	TiempoLaboran do	ModalidadContrato	SectorSocio económico	SituaciónLab
1	Masculino	48	Barranco	Viudo	Alquilada	Primaria	Empleado	1	Tiempo Parcial	Transporte	Dependiente
2	Masculino	44	Carcado de lima	Viudo	Alquilada	Secundaria	Empleado	1	Tiempo Completo	Pesca	Independiente
3	Masculino	56	Carcado de lima	Soltero	Alquilada	Universitario	Funcionario	3	Tiempo Parcial	Pesca	Dependiente
4	Masculino	39	Ate	Viudo	Alquilada	Secundaria	Gerente	1	Tiempo Completo	Pesca	Estudiante
5	Masculino	39	Carcado de lima	Viudo	Propia	Universitario	Empleado	1	Tiempo Completo	Transporte	Dependiente
6	Femenino	51	Chorrillos	Soltero	Alquilada	Primaria	Ejecutivo	2	Tiempo Parcial	Pesca	Dependiente
7	Masculino	22	Carcado de lima	Soltero	Propia	Universitario	Gerente	1	Tiempo indefinido	Pesca	Independiente
8	Masculino	39	Ate	Soltero	Familiar	Universitario	Empleado	1	Tiempo Completo	Agricola	Independiente
9	Masculino	36	Carcado de lima	Soltero	Propia	Técnico	Empleado	3	Tiempo Completo	Transporte	Dependiente
10	Femenino	42	Barranco	Viudo	Alquilada	Secundaria	Gerente	3	Tiempo Parcial	Pesca	Jubilado
11	Femenino	54	Carcado de lima	Viudo	Alquilada	Secundaria	Empleado	1	Tiempo Parcial	Agricola	Dependiente
12	Femenino	20	Carcado de lima	Soltero	Alquilada	Universitario	Gerente	2	Tiempo Completo	Transporte	Casa
13	Femenino	56	Ate	Casado	Alquilada	Primaria	Empleado	3	Tiempo temporal	Pesca	Dependiente
14	Femenino	18	Ate	Casado	Propia	Secundaria	Gerente	3	Tiempo Completo	Pesca	Estudiante
15	Femenino	23	Carcado de lima	Viudo	Alquilada	Técnico	Funcionario	1	Tiempo Completo	Industria	Dependiente
16	Femenino	34	Carcado de lima	Soltero	Alquilada	Universitario	Colaborador	2	Tiempo Completo	Industria	Jubilado
17	Femenino	50	Chorrillos	Soltero	Propia	Primaria	Empleado	1	Tiempo Completo	Industria	Casa
18	Femenino	19	Carcado de lima	Casado	Alquilada	Universitario	Empleado	2	Tiempo Parcial	Pesca	Dependiente
19	Femenino	47	Carcado de lima	Viudo	Alquilada	Técnico	Funcionario	3	Tiempo Completo	Pesca	Dependiente
20	Femenino	36	Ate	Viudo	Propia	Técnico	Trabajador	1	Tiempo Parcial	Pesca	Dependiente
21	Masculino	35	Carcado de lima	Viudo	Propia	Universitario	Empleado	2	Tiempo Completo	Industria	Estudiante
22	Femenino	44	Barranco	Soltero	Familiar	Técnico	Funcionario	2	Tiempo Parcial	Transporte	Casa

Fuente: Elaboración propia

Así mismo, a través de la herramienta se realizó una cuantificación de ítems no contestados, para lo cual se realizó lo siguiente:

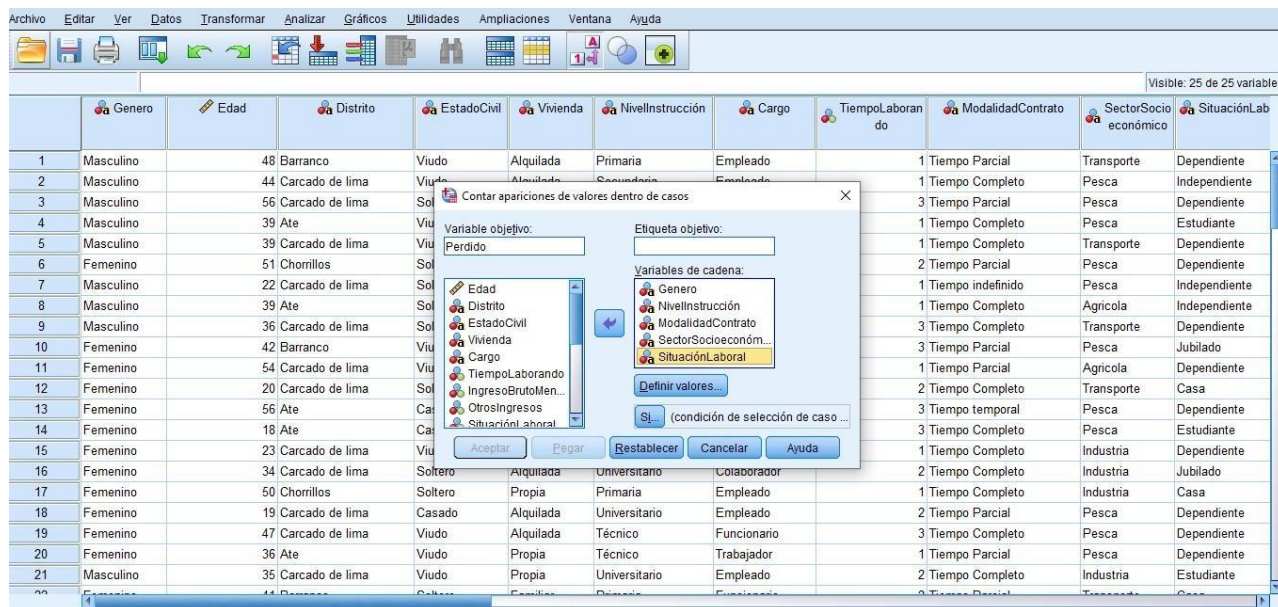
Figura 15: Cuantificación de ítems no contestados parte 1 - SPSS Statistics



Fuente: Elaboración propia

Figura 16 Cuantificación de ítems no contestados parte 2 - SPSS Statistics

Se agrega una variable que obtendrá la cantidad de datos perdidos por cada registro



Fuente: Elaboración propia

Figura 17: No se evidencio que hubo ítems nulos en el total de registros (500) parte 2 - SPSS Statistics

	Genero	Edad	Distrito	EstadoCivil	Vivienda	NivelInstrucción	Cargo	TiempoLaborado	ModalidadContrato	SectorSocioeconómico	SituaciónLaboral
1	Masculino	48	Barranco	Viudo	Alquilada	Primaria	Empleado	1	Tiempo Parcial	Transporte	Dependiente
2	Masculino	44	Cercado de lima	Viudo	Alquilada	Secundaria	Empleado	1	Tiempo Completo	Pesca	Independiente
3	Masculino	56	Cercado de lima	Soltero	Alquilada	Universitario	Funcionario	3	Tiempo Parcial	Pesca	Dependiente
4	Masculino	39	Ate	Viudo	Alquilada	Secundaria	Gerente	1	Tiempo Completo	Pesca	Estudiante
5	Masculino	39	Cercado de lima	Viudo	Propia	Universitario	Empleado	1	Tiempo Completo	Transporte	Dependiente
6	Femenino	51	Chorrillos	Soltero	Alquilada	Primaria	Ejecutivo	2	Tiempo Parcial	Pesca	Dependiente
7	Masculino	22	Cercado de lima	Soltero	Propia	Universitario	Gerente	1	Tiempo indefinido	Pesca	Independiente
8	Masculino	39	Ate	Soltero	Familiar	Universitario	Empleado	1	Tiempo Completo	Agricola	Independiente
9	Masculino	36	Cercado de lima	Soltero	Propia	Técnico	Empleado	3	Tiempo Completo	Transporte	Dependiente
10	Femenino	42	Barranco	Viudo	Alquilada	Secundaria	Gerente	3	Tiempo Parcial	Pesca	Jubilado
11	Femenino	54	Cercado de lima	Viudo	Alquilada	Secundaria	Empleado	1	Tiempo Parcial	Agricola	Dependiente
12	Femenino	20	Cercado de lima	Soltero	Alquilada	Universitario	Gerente	2	Tiempo Completo	Transporte	Casa
13	Femenino	56	Ate	Casado	Alquilada	Primaria	Empleado	3	Tiempo temporal	Pesca	Dependiente
14	Femenino	18	Ate	Casado	Propia	Secundaria	Gerente	3	Tiempo Completo	Pesca	Estudiante
15	Femenino	23	Cercado de lima	Viudo	Alquilada	Técnico	Funcionario	1	Tiempo Completo	Industria	Dependiente
16	Femenino	34	Cercado de lima	Soltero	Alquilada	Universitario	Colaborador	2	Tiempo Completo	Industria	Jubilado
17	Femenino	50	Chorrillos	Soltero	Propia	Primaria	Empleado	1	Tiempo Completo	Industria	Casa
18	Femenino	19	Cercado de lima	Casado	Alquilada	Universitario	Empleado	2	Tiempo Parcial	Pesca	Dependiente
19	Femenino	47	Cercado de lima	Viudo	Alquilada	Técnico	Funcionario	3	Tiempo Completo	Pesca	Dependiente
20	Femenino	36	Ate	Viudo	Propia	Técnico	Trabajador	1	Tiempo Parcial	Pesca	Dependiente
21	Masculino	35	Cercado de lima	Viudo	Propia	Universitario	Empleado	2	Tiempo Completo	Industria	Estudiante

Fuente: Elaboración propia

Tabla 31: Nombre de variables para el entrenamiento del modelo predictivo.

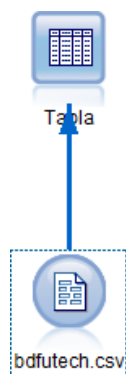
N°	Variable	descripción de la variable	Uso del modelo
1	Edad	Edad del cliente	entrada
2	Vivienda	Tipo de vivienda propia,alquilada o familiar	Entrada
3	NivelInstruccion	Nivel de instrucción académica	Entrada
4	Cargo	Cargo de trabajo	Entrada
5	Situacionlaboral	Situación laboral dependiente o independiente	Entrada
6	Modalidadcontrato	Modalidad de contrato parcial, completo	Entrada
7	Ingresobrutomensual	Ingresos brutos del empelado	Entrada
8	Otroingresos	Otros ingresos que percibe el cliente	Entrada
9	Deposito_plazofijo	Indica si cuenta con un ahorro	Entrada
10	Bienes_casa_vehiculo	Garantía del cliente	Entrada
11	Productos_otras_entidadesfiancieras	Tiene prestamos en otras entidades	entrada
12	resultado	Resultado del estado del riesgo crediticio	salida

Etapa 3: Etapa de transformación y reducción

En esta etapa, se utilizó la herramienta SPSS Modeler para crear el modelo predictivo Machine Learning, así mismo se deshabilitaron algunas variables no influyentes.

Inicialmente se cargó la base de datos creada a partir de la herramienta SPSS Statistic, para lo cual se utilizó un “nodo origen” de tipo “Archivo de Statistic”, para validar la importación de datos se utilizó un “nodo resultado” de tipo Tabla

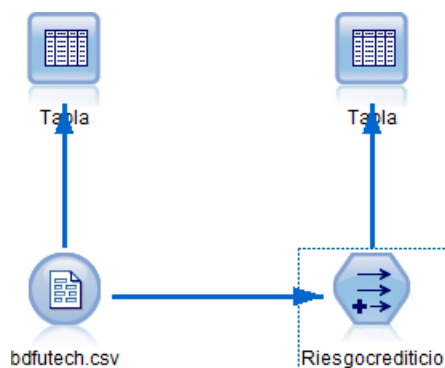
Figura 18: Nodo origen de datos



Fuente: Elaboración propia

Luego se utilizó un “nodo de operaciones con campos”, específicamente el nodo derivar, en el cual se insertó lógica para clasificar a los clientes de bajo y alto riesgo crediticio, esto se realizó luego de identificar a la variable objetivo, para nuestro estudio es el promedio ponderado del estudiante.

Figura 19: Transformación de datos con nodo derivar - SPSS ModelerFuente



Fuente: Elaboración propia

Finalmente se conectó el nodo de origen de datos hacia el nodo derivar de tal manera que se realiza la transformación del promedio ponderado. A continuación, una imagen de muestra

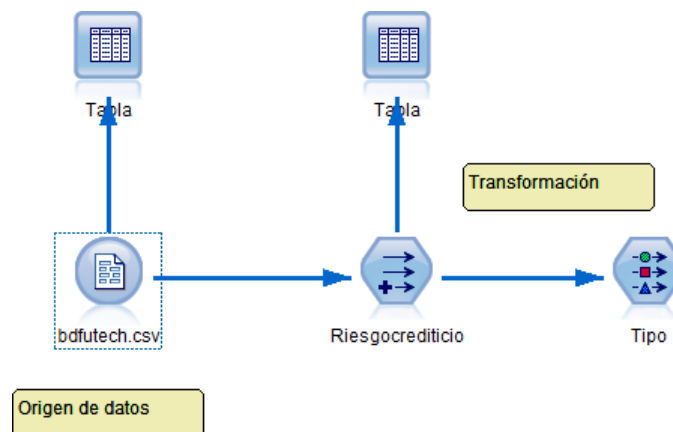
Figura 20: Ejemplo de la transformación de datos - SPSS Modeler

	Genero	Edad	Distrito	EstadoCivil	Vivienda	NivelInstrucción	Cargo	TiempoLaborando	ModalidadContrato	SectorSocioeconómico	SituaciónLaboral	IngresoBruto
1	2	3	2	1	3	4	1	2	2	1	3	
2	2	2	4	3	3	2	3	1	2	5	3	
3	2	4	4	1	3	4	2	1	1	3	4	
4	2	2	2	2	2	1	4	1	2	1	4	
5	2	3	4	1	2	4	4	2	1	3	2	
6	1	3	1	1	3	1	4	1	1	3	5	
7	1	4	4	2	3	4	4	2	1	2	1	
8	1	3	1	4	3	4	1	1	1	2	3	
9	2	2	1	1	1	2	3	1	2	4	5	
10	2	2	4	2	3	3	1	3	1	1	5	
11	2	2	4	3	1	4	3	2	1	5	4	
12	2	2	4	3	1	3	2	1	2	3	1	
13	2	2	4	4	3	4	2	2	1	2	3	
14	1	1	3	2	3	1	4	1	1	3	3	
15	1	1	3	1	3	4	2	1	2	1	3	
16	1	3	2	3	1	2	1	2	1	2	5	
17	2	3	2	4	3	2	3	2	1	4	4	
18	2	3	1	4	3	1	1	2	1	4	1	
19	2	4	2	1	1	2	4	2	2	1	5	
20	2	3	4	1	1	2	4	2	1	2	4	
21	1	1	1	2	3	4	3	1	2	2	2	
22	2	4	1	3	3	2	1	2	2	3	4	
23	2	2	2	4	1	3	1	2	2	1	2	

Fuente: Elaboración propia

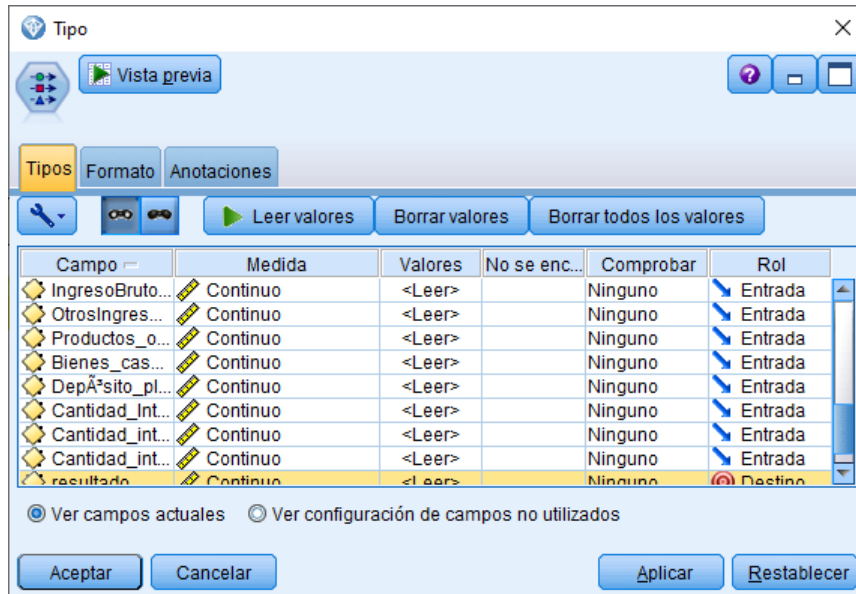
Luego se utilizó el “nodo Tipo”, con la finalidad de elegir nuestras variables de tipo entrada y destino, así mismo se cargaron los valores numéricos para cada variable.

Figura 21: Aplicación del nodo tipo - SPSS Modeler



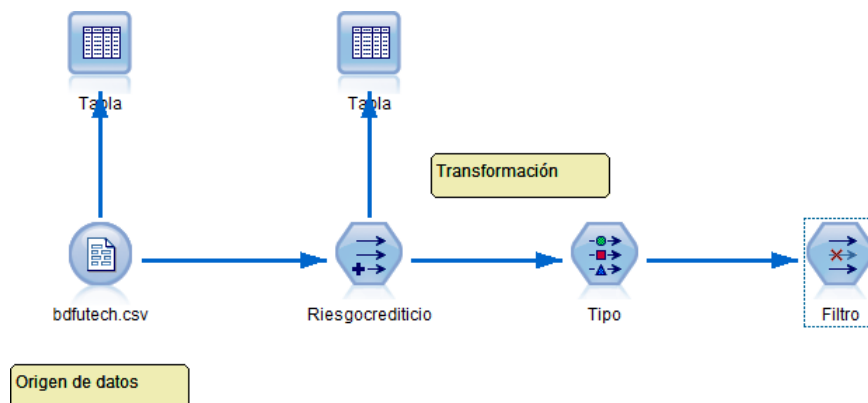
Fuente: Elaboración propia

Figura 22: Ejemplo del nodo tipo - SPSS Modeler



Fuente: Elaboración propia

Figura 23: Aplicación del nodo filtro - SPSS Modeler



Fuente: Elaboración propia

Etapa 4: Minería de datos

En esta etapa, se eligió distintos algoritmos de aprendizaje automático con la finalidad de comparar resultados y obtener el mejor modelo predictivo. En esta investigación se usó Decision Tree, Random Forest, Support Vector Machine, Naibes Bayes, K Vecino más Cercano. Estos algoritmos son proporcionados por la herramienta SPSS Modeler.

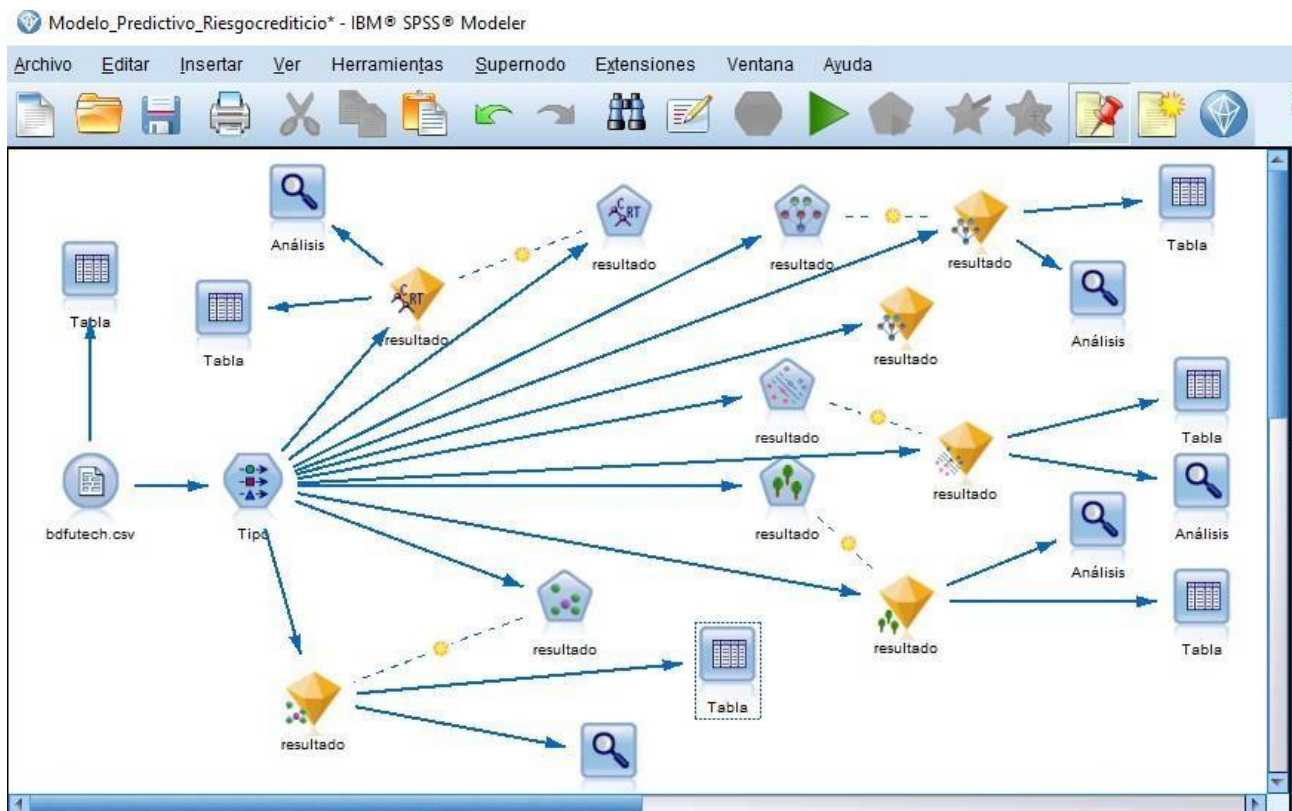
Figura 24: Algoritmos de aprendizaje automático - SPSS Modeler



Fuente: Elaboración propia

A continuación, se muestra los modelos creados por cada algoritmo de aprendizaje:

Figura 25: Proyecto de rendimiento predicción riesgo crediticio - SPSS Modeler

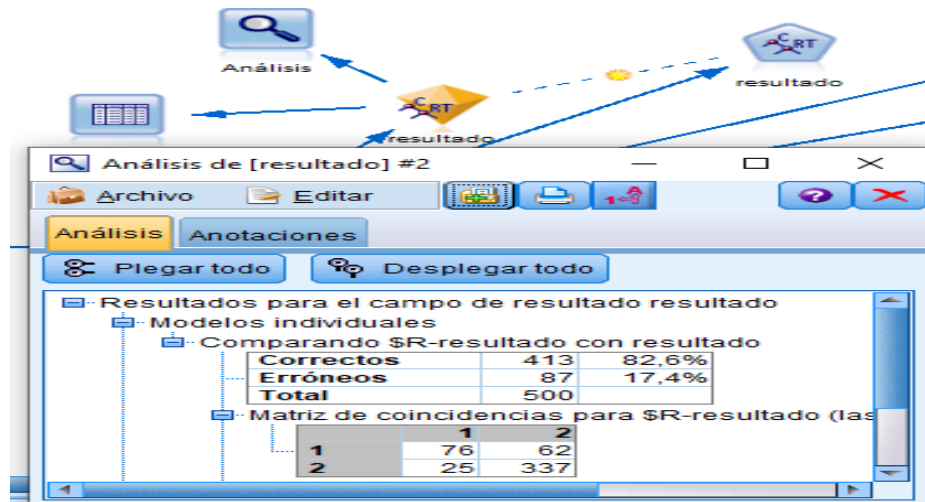


Fuente: Elaboración propia

Etapa 5: Interpretación

Decision Tree, la herramienta nos proporciona información sobre que variables son de mayor relevancia con relación a la variable resultado. Además, nos muestra que el modelo alcanzó una precisión de 82.6% y una tasa de error de 17.4%

Figura 26: Precisión de modelo utilizando Decision Tree - SPSS Modeler



Fuente: Elaboración propia

Random Forest, la herramienta nos proporciona información sobre que variables son de mayor relevancia con relación a la variable resultado. Además, nos muestra que el modelo alcanzó una precisión de 96.2% y una tasa de error de 3.8%

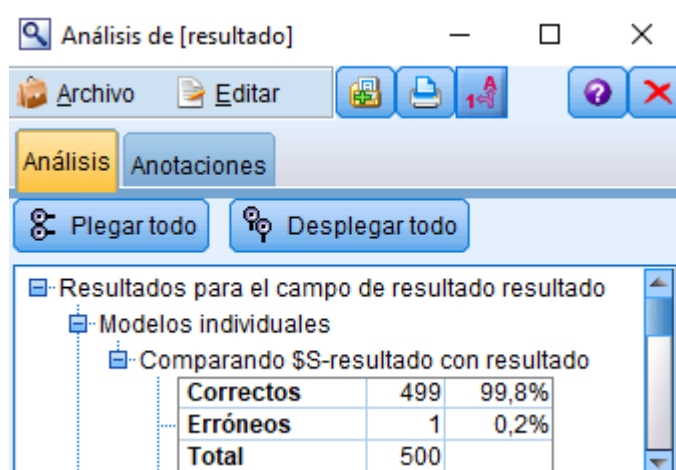
Figura 27: Precisión de modelo utilizando Random Forest - SPSS Modeler



Fuente: Elaboración propia

Support Vector Machine, la herramienta nos proporciona información sobre que variables son de mayor relevancia con relación a la variable resultado. Además, nos muestra que el modelo alcanzó una precisión de 98.8% y una tasa de error de 0.2%.

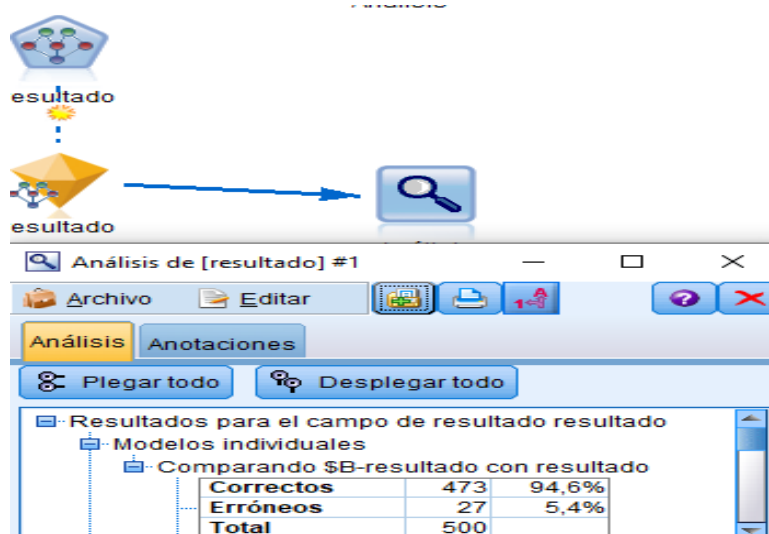
Figura 28: Precisión de modelo utilizando Support Vector Machine



SPSS Modeler Fuente: Elaboración propia

Naibes Bayes, la herramienta nos proporciona información sobre que variables son de mayor relevancia con relación a la variable resultado. Además, nos muestra que el modelo alcanzó una precisión de 94.6% y una tasa de error de 5.4%.

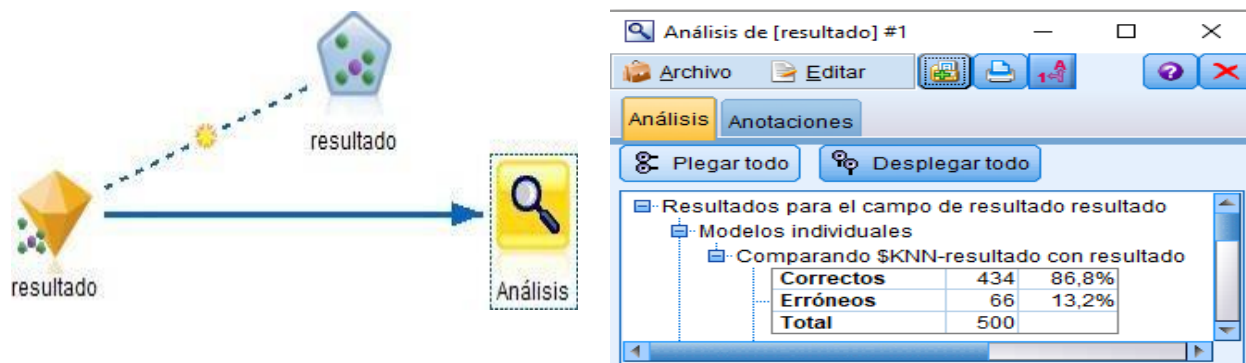
Figura 29: Precisión de modelo utilizando Naibes Bayes - SPSS Modeler



Fuente: Elaboración propia

K Vecino más Cercano, la herramienta nos proporciona información sobre que variables son de mayor relevancia con relación a la variable resultado. Además, nos muestra que el modelo alcanzó una precisión de 86,8% y una tasa de error de 13,2%.

Figura 30: Precisión de modelo utilizando K Vecino más Cercano - SPSS Modeler



Fuente: Elaboración propia

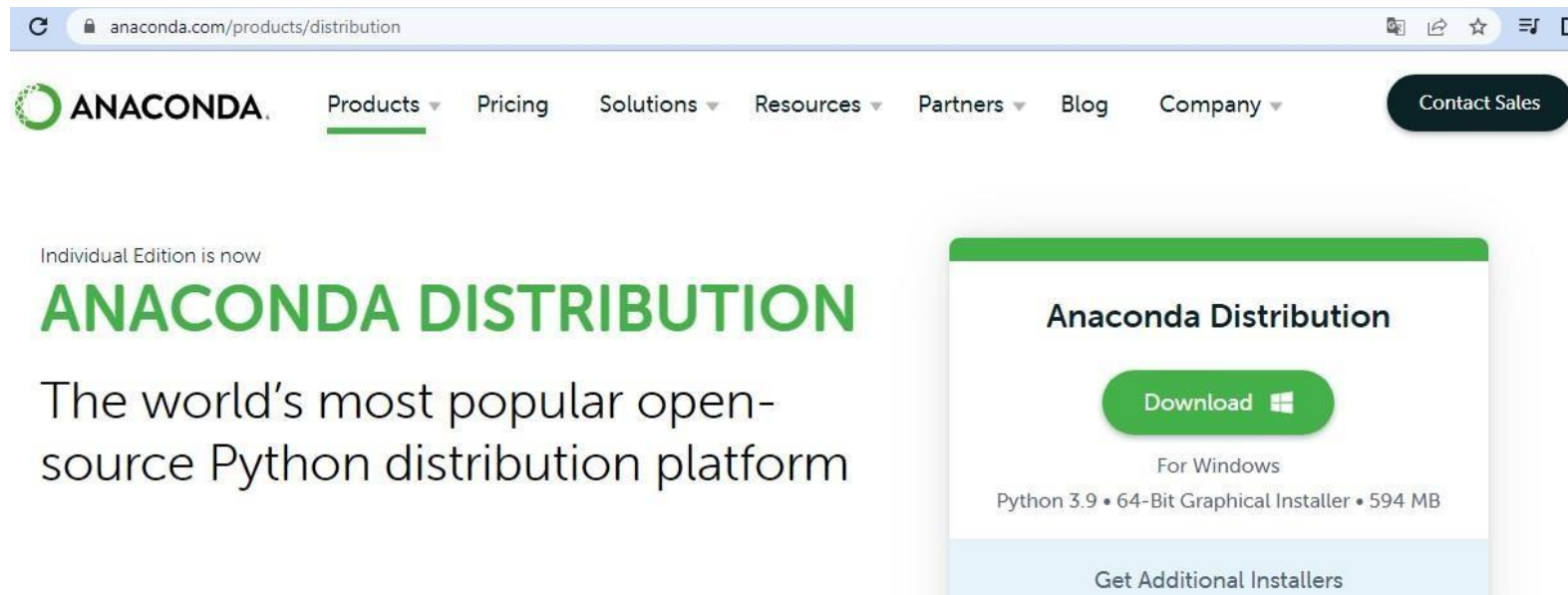
Herramientas utilizadas:

a) Instalación y configuración de la Suite de Anaconda Navigator versión 2.2.0 en una máquina con Windows 10 son:

Ingresar al enlace. <https://www.anaconda.com/products/distribution>

Dar clic en la opción Download, Utilizar la configuración predeterminada que brinda el programa.

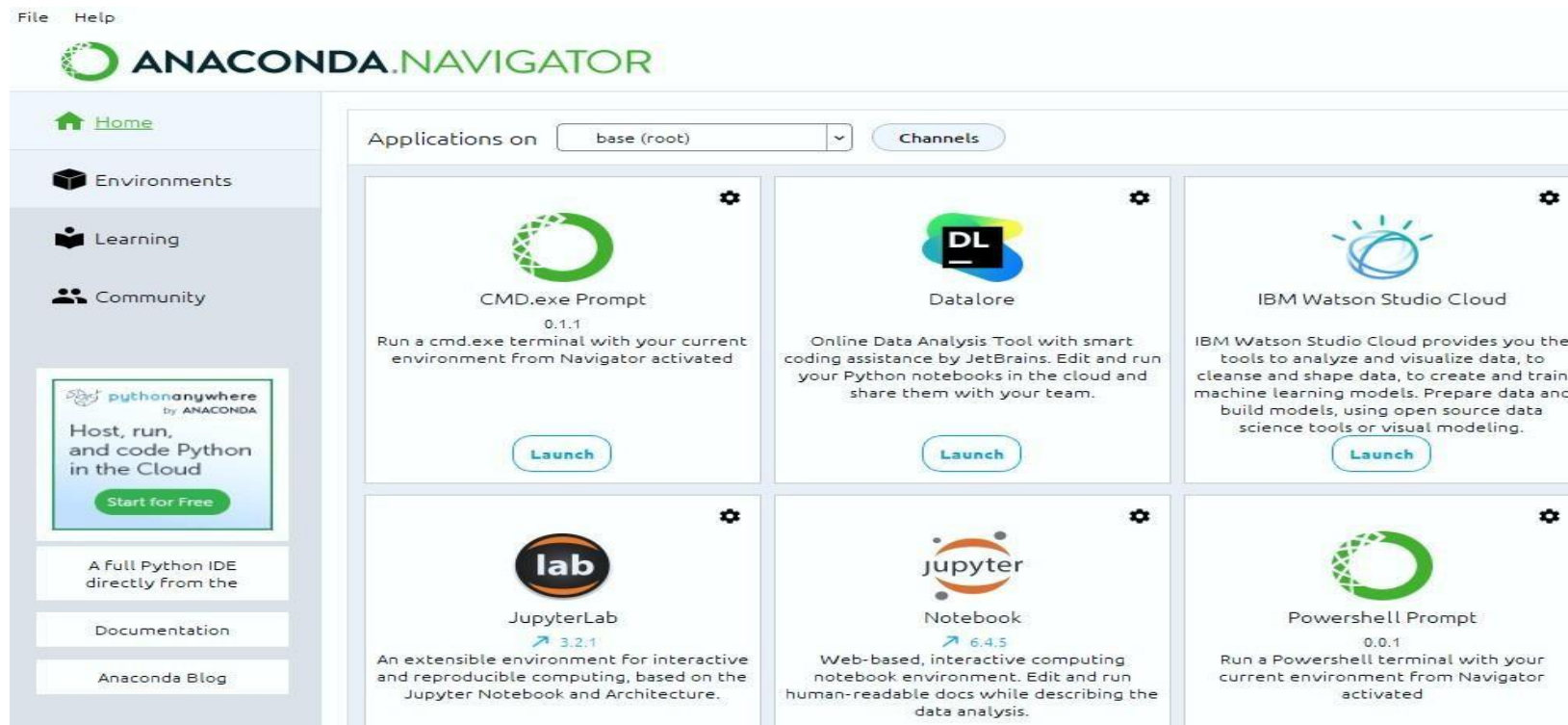
Figura 31. Entorno de la Suite de Anaconda



Fuente: Anaconda Distribution(2022)

b) Ingresar a la Suite de la herramienta Anaconda y elegir la herramienta “Jupyter Notebook”.

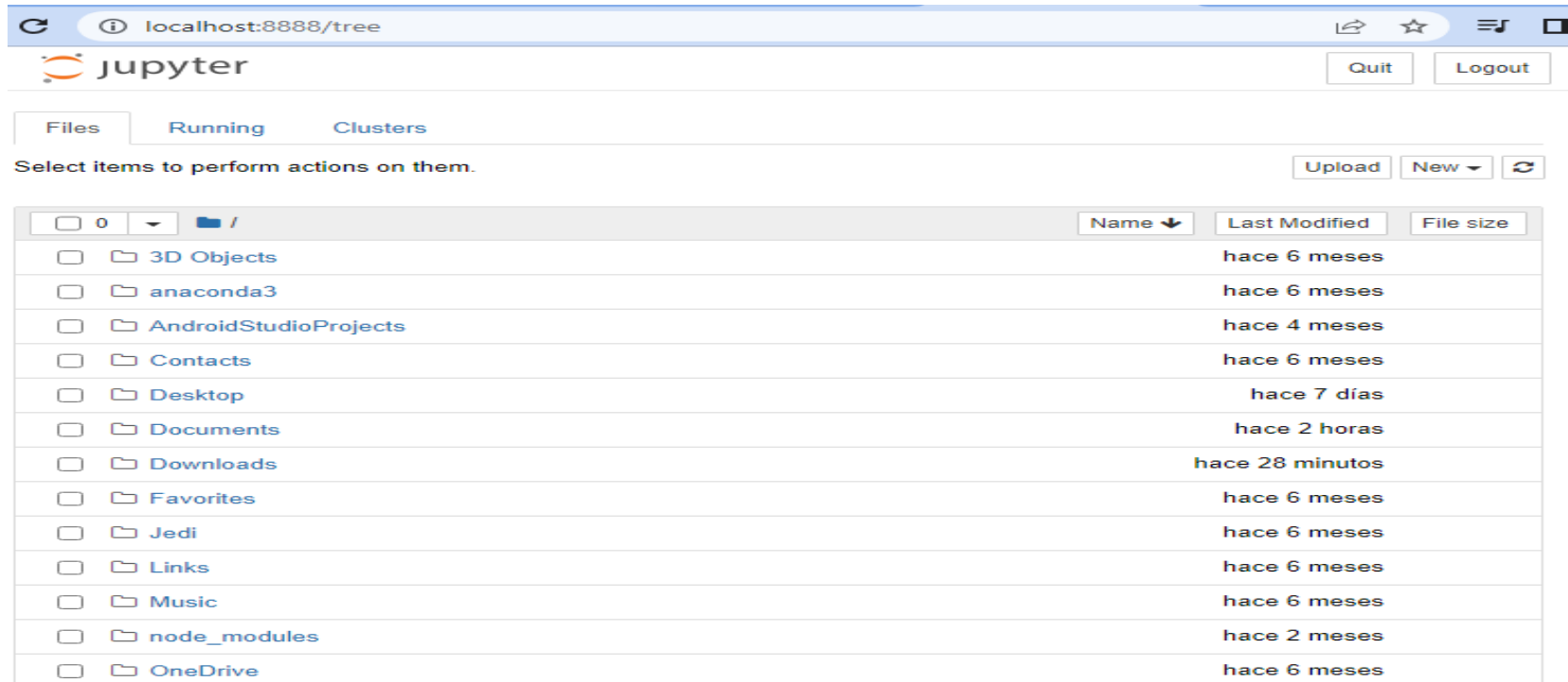
Figura 32. Entorno de Jupyter Notebook



Fuente: Anaconda Distribution(2022)

c) Crear un nuevo archivo en “Jupyter Notebook” para el desarrollo de la ciencia de datos con Python.

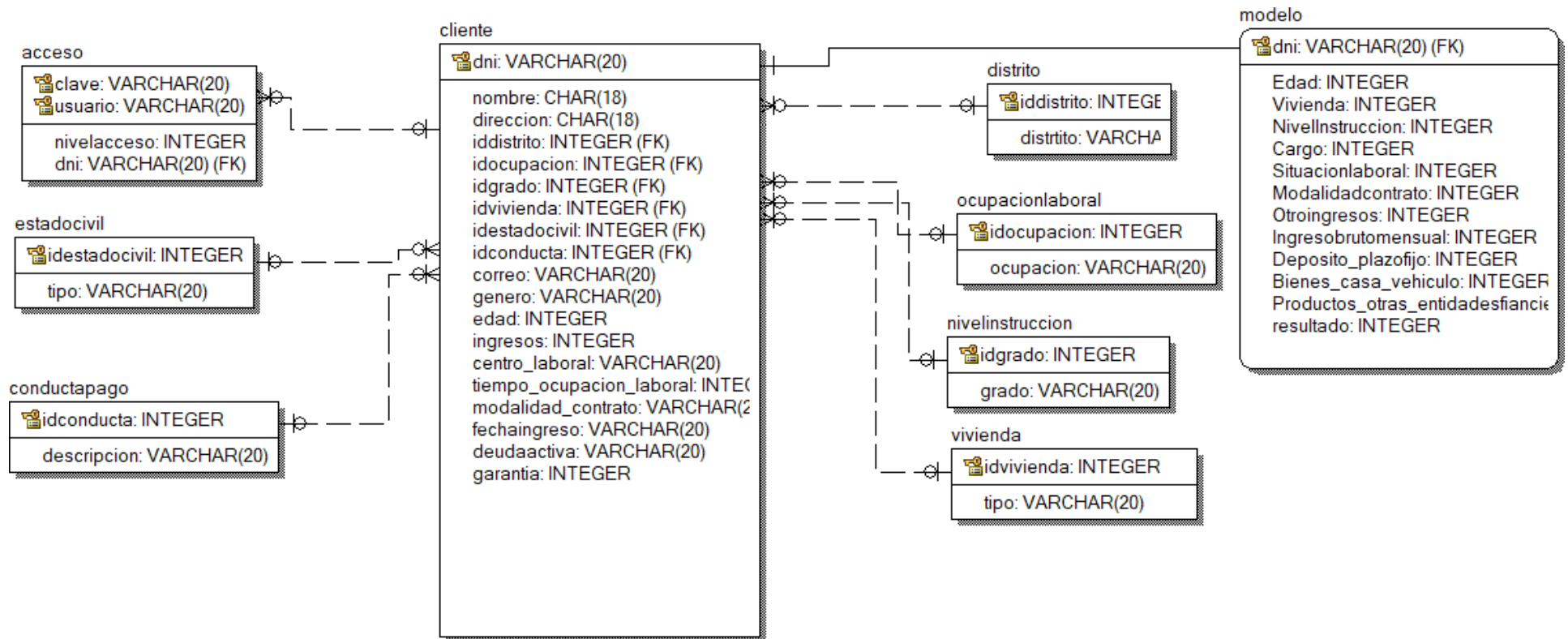
Figura 33. Entorno de Servidor local de Jupyter Notebook



Fuente: Anaconda Distribution(2022)

Anexo 17. Diseño del modelo físico de la base de datos.

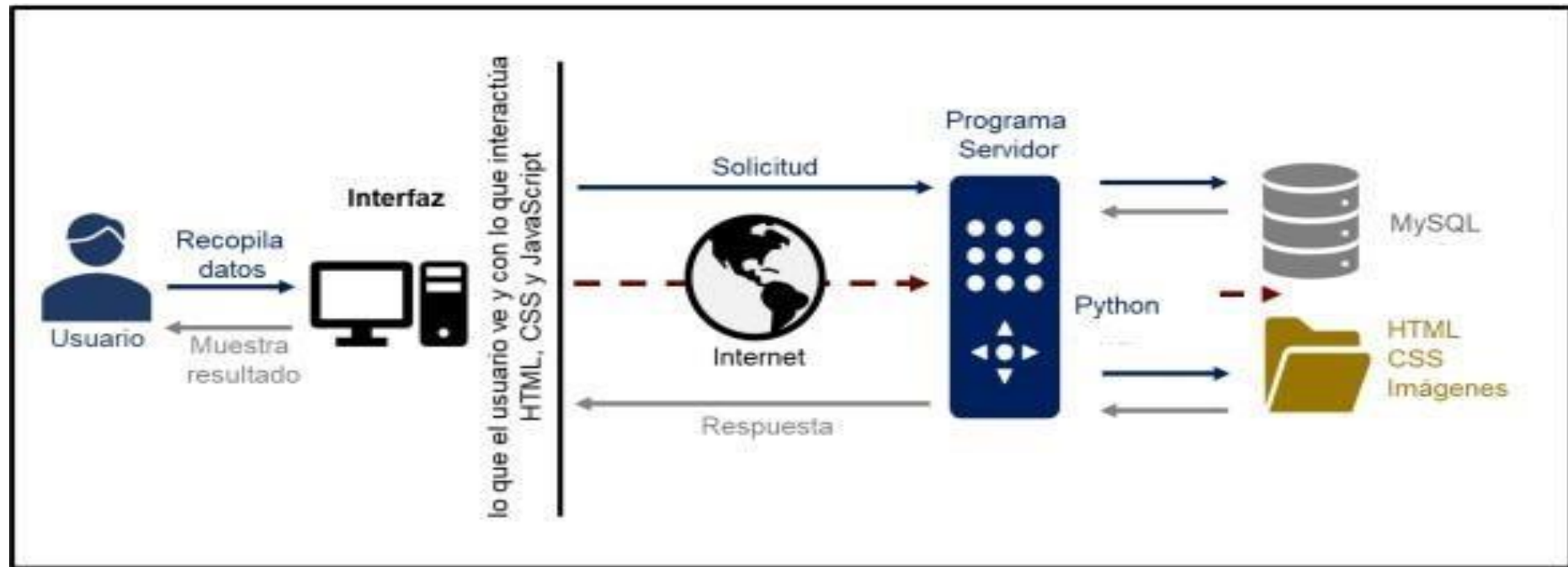
Figura 34. Diseño del modelo Físico de la base de datos



Fuente: Elaboración propia.

Anexo 18. Arquitectura del sistema o aplicaciones web.

Figura 35. Arquitectura del sistema o aplicaciones web



Fuente: Charalla Cutipa (2022)

Anexo 19. Interfaces del Sistema Web.

Figura 36. Verificación de credenciales para el acceso al sistema.



The image shows a login form with a light gray background. At the top center, the word "Login" is written in a large, bold, black font. Below this, there are two input fields. The first field is preceded by a yellow key icon and the text "Ingresar usuario". The second field is preceded by a yellow padlock icon and the text "Ingresar password". Both fields are empty white rectangles with thin black borders. Below the second field is a dark blue button with rounded corners and the text "Iniciar sesión" in white, bold font.

Fuente: Elaboración propia.

Figura 37. Interfaz de principal del uso de la Solución web.



Fuente: Elaboración propia.

Figura 38. Interfaz de la aplicación para la predicción del riesgo crediticio del cliente

Datos personales del cliente				
DNI: 40245173	Nombres: Juan	Fecha Registro: 05/11/2022 📅		
Dirección: Av. Las Palmeras # 253	Distrito: Barranco ▼	Correo: juan@hotmail.com	Teléfono: 123-4567	Celular: 987654321
Edad: 50	Genero: Masculino ▼	Estado Civil: Casado ▼	Tipo Vivienda: Propia ▼	Nivel Instrucción: Universitaria ▼
Datos laborales del cliente				
Cargo laboral: Empleado ▼	Centro Laboral: Senati	Tiempo ocupación Laboral: 6..más ▼	Modalidad de Contrato: Tiempo Completo ▼	
Sector laboral: Industrial ▼	Situación laboral: Dependiente ▼	Ingresos bruto mensual: 5000	Otro ingresos: 3000	
Datos laborales del conyugue				
Cargo laboral: Empleado ▼	Tiempo ocupación Laboral: 6..más ▼	Modalidad de Contrato: Tiempo Completo ▼		
Sector laboral: Industrial ▼	Situación laboral: Dependiente ▼	Ingresos bruto mensual: 3000	Otro ingresos: 2000	
Datos financieros del cliente				
Prestamos en entidades financieras: No ▼	Garantía: Bienes, Casa, Vehículo: Si ▼	Deposito a Plazo fijo: Si ▼	Cantidad Integrantes familiar trabajan: 3-5 ▼	Cantidad Integrantes familiar estudian: 1-2 ▼

Enviar

Fuente: Elaboración propia.

Figura 39. Resultados de la ejecución de la aplicación web para la predicción del riesgo crediticio del cliente.

← → ↻ ⓘ 127.0.0.1:5000/evaluacion

Cliente de Bajo Riesgo Crediticio

Código de la Aplicación Web:

Archivo Login.html

```
<html>
  <head>
    <meta charset="UTF-8">
    <meta name="viewport" content="width=device-width, initial-scale=1">
    <title>Riesgo crediticio</title>
    <link rel="stylesheet" href="{{
url_for('static',filename='css/vaidroll.css') }}">
  </head>
  <body>
    <div class="cajafuera" align="center">

    <div class="formulariocaja">
    <form method="post" action="/acceso">
    <div class="formtitulo">Login</div>
    <div align="left" class="textoscajas">&#128273; Ingresar usuario</div>
    <input type="text" name="txtusuario" class="cajaentradatexto">
    <div align="left" class="textoscajas">
    &#128274; Ingresar password
    </div>
    <input type="password" name="txtclave"
    class="cajaentradatexto">
    <input type="submit" value="Iniciar sesión" class="botonenviar">
    </form>
    </div>
    <div class="autor">
    © 2022 Formulario Login. Todos los derechos reservados | FUTECH PERU
    <div>
    </div>
    </body>
  </html>
```

Archivo index.html

```
<!DOCTYPE html>
<html>
<head>
<meta charset="utf-8">
<title>Evaluación Crediticia</title>
<link rel="stylesheet" href="{{url_for('static', filename= 'css/estilos.css')}}">
</head>
<body>
  <div class="contenedor">
    <header>
      <div class="logo">
        <img src = "{{url_for ('static', filename = './img/logo4.png')}}"
width = "100" height = "100" alt = ""/>
```

```

        <a href="#" style="font-size:30px">FUTECH PERU</a>
    </div>
    <nav>
        <a href="#">Inicio</a>
        <a href="{{url_for('formulario')}}">Evaluacion</a>
        <a href="{{url_for('formulario2')}}">Contactenos</a>
    </nav>
</header>
<section class="main">
    <article>
        <h2 class="titulo">Gestión Comercial</h2>
        <p>
            Nos dirigimos a cualquier socio comercial que cuente con RUC,
            incluyendo las personas naturales con negocio.
        </p>
    </article>
    <article>
        <img src = "{{url_for ('static', filename = './img/banner3.png')}}"
width = "600" height = "300" alt = ""/>
    </article>
</section>
<aside>
    <div class="widget">
        <div class="imagen">
            
        </div>
    </div>
    <div class="widget">
        <div class="imagen">
            
        </div>
    </div>
</aside>
<footer>
    2022 - FUTECH PERU &copy; Todos los Derechos Reservados
</footer>
</div>
</body>
</html>

```

Archivo evaluacion.html

```
<!doctype html>
<html>
<head>
<meta charset="utf-8">
<title>Formulario y Validación</title>
<link rel="stylesheet" href="{{ url_for('static',filename='css/estilos4.css')
}}">
</head>
<body>
  <form action="/evaluacion" method="post">
    <table class="registro" width="950px" border="1px">
      <tr style="background-color:#2c3e50;color:aliceblue;">
        <td id="reg1" colspan="5">Datos personales del cliente</td>
      </tr>
      <tr>
        <td>DNI:<br>
        <input type="text" name="xdni" maxlength="8"
        pattern="[0-9]+" required/></td>
        <td>Nombres:<br>
        <input type="text" name="xnombres" pattern="[a-z,A-Z]+" required/>
      </td>
        <td colspan="3">Fecha Registro:<br>
        <input type="date" name="xfecha"/></td>
      </tr>
      <tr>
        <td>Dirección:<br>
        <input type="text" name="xdireccion" required/> </td>
        <td>Distrito:<br>
        <select name="xdis">
          <option selected>Seleccionar</option>
          <option value="1">Ancón</option>
          <option value="2">Ate</option>
          <option value="3">Barranco</option>
          <option value="4">Breña</option>
          <option value="5">Carabayllo</option>
          <option value="6">Chorrillos</option>
          <option value="7">Cieneguilla</option>
          <option value="8">Comas</option>
          <option value="9">El Agustino</option>
          <option value="10">Independencia
          </option>
        </select>
        <br>
      </td>
        <input type="tel" name="xtel" maxlength="8" pattern="[0-9]{3}[-]{1}[0-
9]{4}"/></td>
        <td>Celular:<br>
        <input type="cel" name="xcel" maxlength="9" pattern="[0-9]"/></td>
      <br>
    </table>
  </form>
</body>
</html>
```

```

        <td colspan="3">Edad<br>
        <input type="number" name="xedad"/></td>
</tr>
</td>
<td>Genero:<br>
    <select name="xgenero">
        <option selected>Seleccionar</option>
        <option value="1">Masculino</option>
        <option value="2">Femenino</option>
    </select>
</td>
<br>
<td>Estado Civil:<br>
    <select name="xestadocivil">
        <option selected>Seleccionar</option>
        <option value="1">Soltero</option>
        <option value="2">Casado</option>
        <option value="3">Divorciado</option>
        <option value="4">Viudo</option>
    </select>
</td>
<td>Tipo Vivienda:<br>
    <select name="xvivienda">
        <option selected>Seleccionar</option>
        <option value="1">Alquilada</option>
        <option value="2">Familia</option>
        <option value="3">Propia</option>
    </select>
    <br>
</td>
<td colspan="2">Nivel Instrucción:<br>
    <select name="xnivelinstruccion">
        <option selected>Seleccionar</option>
        <option value="1">Primaria</option>
        <option value="2">Secundaria</option>
        <option value="3">Técnica</option>
        <option value="4">Universitaria</option>
    </select>
</td>
</tr>
<tr>
</td>
<tr style="background-color:#2c3e50;color:aliceblue;">
    <td id="reg2" colspan="5">Datos laborales del cliente</td>
</tr>
<tr>
    <td>Cargo laboral<br>
    <select name="xoculab">
        <option selected>Seleccionar</option>
        <option value="1">Empleado</option>
        <option value="2">Ejecutivo</option>

```

```

                <option value="3">Funcionario</option>
                <option value="4">Gerente General</option>
            </select>
        </td>
        <td>Centro Laboral<br>
            <input type="text" name="xcentrolaboral" pattern="[a-z,A-Z]+" required/>
        </td>
        <td>Tiempo ocupación Laboral<br>
            <select name="xtol">
                <option selected>Seleccionar</option>
                <option value="1">1-3</option>
                <option value="2">4-5</option>
                <option value="3">6..más</option>
            </select>
        </td>
        <td colspan="2">Modalidad de Contrato<br>
            <select name="xmoculab">
                <option selected>Seleccionar</option>
                <option value="1">Tiempo Parcial</option>
                <option value="2">Tiempo Completo</option>
            </select>
        </td>
    </tr>
    <tr>
        <td>Sector laboral<br>
            <select name="xtseclab">
                <option selected>Seleccionar</option>
                <option value="1">Industrial</option>
                <option value="2">Agricultura</option>
                <option value="3">Ganadería</option>
                <option value="4">Transporte</option>
                <option value="5">Pesca</option>
            </select>
        </td>
        <td>Situación laboral<br>
            <select name="xstitlab">
                <option selected>Seleccionar</option>
                <option value="1">Estudiante</option>
                <option value="2">Casa</option>
                <option value="3">Jubilado</option>
                <option value="4">Independiente</option>
                <option value="5">Dependiente</option>
            </select>
        </td>
        <td>Ingresos bruto mensual<br>
            <input type="number" name="xibm" pattern="[0-9]+" required/>
        </td>
        <td colspan="2">Otro ingresos<br>
            <input type="number" name="xotrosingresos" pattern="[0-9]+" required/>
        </td>
    </tr>

```

```

<tr style="background-color:#2c3e50;color:aliceblue;">
  <td id="reg2" colspan="5">Datos laborales del conyugue</td>
  <tr>
    <td>Cargo laboral<br>
    <select name="xoculabc">
      <option selected>Seleccionar</option>
      <option value="1">Empleado</option>
      <option value="2">Ejecutivo</option>
      <option value="3">Funcionario</option>
      <option value="4">Gerente General</option>
    </select>
  </td>
    <td>Tiempo ocupación Laboral<br>
    <select name="xtolc">
      <option selected>Seleccionar</option>
      <option value="1">1-3</option>
      <option value="2">4-5</option>
      <option value="3">6..más</option>
    </select>
  </td>
    <td colspan="2">Modalidad de Contrato<br>
    <select name="xmoculabc">
      <option selected>Seleccionar</option>
      <option value="1">Tiempo Parcial</option>
      <option value="2">Tiempo Completo</option>
    </select>
  </td>
  </tr>
  <tr>
    <td>Sector laboral<br>
    <select name="xtseclabc">
      <option selected>Seleccionar</option>
      <option value="1">Industrial</option>
      <option value="2">Agricultura</option>
      <option value="3">Ganadería</option>
      <option value="4">Transporte</option>
      <option value="5">Pesca</option>
    </select>
  </td>
    <td>Situación laboral<br>
    <select name="xstitlabc">
      <option selected>Seleccionar</option>
      <option value="1">Estudiante</option>
      <option value="2">Casa</option>
      <option value="3">Jubilado</option>
      <option value="4">Independiente</option>
      <option value="5">Dependiente</option>
    </select>
  </td>
    <td>Ingresos bruto mensual<br>
    <input type="number" name="xibmc" pattern="[0-9]+" required/>
  </td>
  </tr>

```



```

</td>
<td colspan="2">Otro ingresos<br>
  <input type="number" name="xotrosingresosc" pattern="[0-9]+" required/>
</td>
</tr>
<tr style="background-color:#2c3e50;color:aliceblue;">
<td id="reg3" colspan="5">Datos financieros del cliente</td>
</tr>
<tr>
  <td>Prestamos en entidades financieras:<br>
    <select name="xprofin">
      <option selected>Seleccionar</option>
      <option value="1">Si</option>
      <option value="2">No</option>
    </select>
  </td>
  <td>Garantía: Bienes, Casa, Vehículo:
  <br>
    <select name="xgarantia">
      <option selected>Seleccionar</option>
      <option value="2">Si</option>
      <option value="1">No</option>
    </select>
  </td>
  <td>Deposito a Plazo fijo:
  <br>
    <select name="xplazofijo">
      <option selected>Seleccionar</option>
      <option value="2">Si</option>
      <option value="1">No</option>
    </select>
  </td>
  <td>Cantidad Integrantes familiar trabajan<br>
    <select name="xcift">
      <option selected>Seleccionar</option>
      <option value="1">1-2</option>
      <option value="2">3-5</option>
      <option value="3">6..más</option>
    </select>
  </td>
  <td>Cantidad Integrantes familiar estudian<br>
    <select name="xcife">
      <option selected>Seleccionar</option>
      <option value="3">1-2</option>
      <option value="2">3-5</option>
      <option value="1">6..más</option>
    </select>
  </td>
</tr>
</table>
<input type="submit" name="bt1" class="botonenviar" value="Enviar"/>

```

```
<br>
<h1 id="resultado"></h1>
</form>
</body>
</html>
```

Archivo index.py

```
import random
from urllib import request
from flask_mysql import MySQL
from flask import Flask, render_template, request, url_for, redirect
import pandas as pd
from sklearn.linear_model import LogisticRegression
from sklearn.ensemble import RandomForestClassifier
from sklearn.model_selection import train_test_split
from sklearn import ensemble
from sklearn import metrics
from sklearn import linear_model
from tkinter import messagebox as MessageBox
import mysql.connector

app=Flask(__name__)
mysql2 = MySQL()
# MySQL configurations
app.config['MYSQL_USER'] = 'root'
app.config['MYSQL_PASSWORD'] = ''
app.config['MYSQL_DB'] = 'bdfutech'
app.config['MYSQL_HOST'] = 'localhost'
mysql2=MySQL(app)

@app.route('/')
def home():
    return render_template('login.html')
    # return render_template('index.html')

@app.route('/acceso', methods=['GET', 'POST'])
def acceso():
    usuario=request.form['txtusuario']
    clave=request.form['txtclave']
    cur=mysql2.connection.cursor()
    cur.execute('select * from acceso where usuario="'+usuario+'" and
clave="'+clave+'')
    cur.fetchall()
    rows_affected=cur.rowcount
    if rows_affected>0:
        return render_template('index.html')
    else:
        return render_template('login.html')
```

```

@app.route('/formulario')
def formulario():
    return render_template('evaluacion.html')

@app.route('/evaluacion',methods=['POST'])
def evaluacion():
    #utilizar el modelo de predicción
    miConexion = mysql.connector.connect( host='localhost', user= 'root',
password='', db='bdfutech')
    #captura de datos del formulario
    ptgenero=request.form['xgenero']
    ptedad=int(request.form['xedad'])
    pdistrito=request.form['xdis']
    ptestcivil=request.form['xestadocivil']
    ptvivienda=request.form['xvivienda']
    ptnivelinst=request.form['xnivelinstruccion']
    ptoculab=request.form['xoculab']
    pttol=request.form['xtol']
    ptmoculab=request.form['xmoculab']
    ptseclab=request.form['xtseclab']
    ptsitlab=request.form['xstitlab']
    ptibm=request.form['xibm']
    ptotrosingm=request.form['xotrosingresos']
    #conyuque
    cptoculab=request.form['xoculabc']
    cpttol=request.form['xtolc']
    cptomoculab=request.form['xmoculabc']
    cptseclab=request.form['xtseclabc']
    cptsitlab=request.form['xstitlab']
    cptibm=request.form['xibmc']
    cptotrosingm=request.form['xotrosingresosc']

    #financiera
    ptprofin=request.form['xprofin']
    ptgarantia=request.form['xgarantia']
    ptplazofijo=request.form['xplazofijo']

    #socio-cultural
    ptcift=request.form['xcift']
    ptcife=request.form['xcife']

    ptedad=4
    ptibm=3
    ptotrosingm=3
    cptibm=3
    cptotrosingm=3

```

#Entrenamiento del modelo de Support Vector Machine #utilizar el modelo de predicción

```
datos= pd.read_sql_query("SELECT
Edad,EstadoCivil,Vivienda,NivelInstruccion,Cargo,TiempoLaborando,ModalidadContrato,
SituacionLaboral,IngresoBrutoMensual,OtrosIngresos,Producto_financieras,Bienes_casa
_vehiculo,Deposito_plazofijo,resultado FROM riesgocrediticio",miConexion)
dataframe=pd.DataFrame(datos)
x=(dataframe[["Edad","EstadoCivil","Vivienda","NivelInstruccion","Cargo","Tie
mpoLaborando","ModalidadContrato","SituacionLaboral","IngresoBrutoMensual","OtrosIn
gresos","Producto_financieras","Bienes_casa_vehiculo","Deposito_plazofijo"]])
y=(dataframe["resultado"])
x_train,x_test,y_train,y_test = train_test_split(x, y,
test_size=0.3,random_state=0)
model= ensemble.RandomForestClassifier()
model.fit(x_train,y_train)

datanew={'Edad':[ptedad],
        'EstadoCivil':[ptestcivil],
        'Vivienda':[ptvivienda],
        'NivelInstruccion':[ptnivelinst],
        'Cargo':[ptmocolab],
        'TiempoLaborando':[pttol],
        'ModalidadContrato':[ptmocolab],
        'SituacionLaboral':[ptsitlab],
        'IngresoBrutoMensual':[ptibm],
        'OtrosIngresos':[ptotrosingm],
        'Producto_financieras':[ptprofin],
        'Bienes_casa_vehiculo':[ptgarantia],
        'Deposito_plazofijo':[ptplazofijo]
}
clientenew=pd.DataFrame(datanew,columns=['Edad','EstadoCivil','Vivienda','Niv
elInstruccion','Cargo','TiempoLaborando','ModalidadContrato','SituacionLaboral','In
gresoBrutoMensual','OtrosIngresos','Producto_financieras','Bienes_casa_vehiculo','D
eposito_plazofijo'])
prediccion=model.predict(clientenew)
xprediccion=int(prediccion)
cursor=miConexion.cursor()

if xprediccion>1:
    mensaje='Cliente de Alto Riesgo Crediticio'
else:
    mensaje='Cliente de Bajo Riesgo Crediticio'

return "<h1>"+mensaje+"</h1>"
@app.route('/formulario2')
def formulario2():
    return render_template('consulta.html')

@app.route('/consulta',methods=['POST'])
def consulta():
```

```

pdni=request.form['txtdni']
cur=mysql2.connection.cursor()
cur.execute('select * from evaluacion where dni='+pdni+')
data=cur.fetchall()
rows_affected=cur.rowcount
if rows_affected>0:
    mensaje="El cliente ya se encuentra registrado en la base de datos"
    d("<h1>"+mensaje+"</h1>")
else:
    d=redirect(url_for('formulario'))
return d

@app.route('/listado/')
def listado():
    return render_template('listado.html')

if_name == '_main_':
    app.run(debug=True)

```

Archivo estilo.css

```

@charset "utf-8";
/* CSS Document */
body{
    background-color:#e9e9e9;
}
.contenedor{
    width:90%;
    height:90%;
    max-width:1000px;
    background:#ccc;
    margin:auto;
    margin-top:60px;
    /*el contenedor es "flexible" utilizando display:flex o inline-flex;*/
    display:flex;
    /*dirección y ajuste tamaño de los elementos hijos*/
    flex-flow:row wrap;
}
header{
    width:100%;
    padding:20px;
    background-color:#2c3e50;
    display:flex;
    /*espacio entre los elementos que están dentro del header*/
    justify-content:space-between;
    align-items:center;
    flex-flow:row wrap;
}

```

```

}
header .logo img{
  width:100px;
  vertical-align: top;
}

header .logo a{
  color:#fff;
  text-decoration:none;
  line-height:50px;
}
header nav{
  width:60%;
  display:flex;
  flex-flow:row wrap;
  align-items:center;
}
header nav a{
  color:#fff;
  background-color:#039617;
  padding:10px;
  text-decoration:none;
  text-align:center;
  /*los elementos toman el ancho de la etiqueta nav*/
  flex-grow:1;
}
header nav a:hover{
  background-color:#e74c3c;
}
.main{
  background-color:#fff;
  padding:20px;
  /*dimensión de la caja*/
  flex:1 1 60%;
}
aside{
  background-color:#039617;
  padding:20px;
  flex:1 1 30%;
  display:flex;
  flex-flow:column wrap;
  justify-content:flex-start;
}
aside .widget{
  height:150px;
  margin:10px;
  margin-top:25px;
  background-color:#fff;
}
.imagen{
  margin-top:30px;
}

```

```

}
footer{
  width:100%;
  padding:20px;
  background-color:#e9e9e9;
  font-weight:bold;
  display:flex;
  justify-content:center;
}

```

Archivo estilo2.css

```

@charset "utf-8";
/* CSS Document */
body{font-family:Verdana,Geneva,sans-serif}
legend{font-size:20px}
#form1{
  width:40%;
  margin:auto;
  background:#039617;
  color:#fff;
  font-size:16px;
  padding:20px;
  border:solid 10px LightSlateGray;
  border-radius:10px;
  -webkit-border-radius:10px;
}
#form1 fieldset{
  margin-left:5%;
}
#form1 [required]{
  border:solid 1px #FF0000;
}
#form1 input{
  margin-left:30px;
}

```

Archivo estilo3.css

```

@charset "utf-8";
/* CSS Document */
img{
  width:80px;
  height:80px;
}
.tabla{
  width:70%;
  border:1px solid;
  margin-top:50px;
  margin-left:180px;
}

```

```

}
.tabla th{
    background:#f10609;
    color:#ffffff;
    font-size:20px;
}
.tabla,th,td{
    border:1px solid;
    border-collapse: collapse;
}
th,td{
    padding: 5px;
    text-align: center;
}
.tabla tr:hover{
    background:#f86163;
}
.tabla td:hover{
    background:#f86163;
    color:#ffffff;
}

```

Archivo estilo4.css

```

.registro
{
    margin-top:10px;
    margin-left:200px;
}

.botonenviar
{
    width: 15%;
    cursor: pointer;
    margin-top: 20px;
    border: 0;
    outline: none;
    border-radius: 10px;
    border:1px solid black;
    font-size:16px;
    color:white;
    background-color: #2c3e50;
    margin-left:200px;
    font-weight: bold;
}
#reg1{
    font-size:25px;
    font-weight: bold;
}
#reg2{

```



```
font-size:25px;
font-weight: bold;
}
#reg3{
font-size:25px;
font-weight: bold;
}
```

Archivo vaidroll.css

```
*
{
    box-sizing: border-box;
    font-family: sans-serif;
    color:black;
}
body
{
    margin: 0;
    padding: 0;
    background: #f2f2f2;
}
.cajafuera
{
    width: 100vw;
    height: 100vh;
    display: grid;
    overflow: hidden;
}
.formulariocaja
{
    background-color: #f3f3f3;
    width: 400px;
    height: auto;
    position: relative;
    margin: auto;
    padding: 1em;
    border-radius: 5px;
    color:white;
    border:0.1em solid black;
}
input
{
    display: block;
    text-align: left;
    box-sizing: border-box;
}
```

```
.cajaentradatexto{
  width: 80%;
  padding: 10px;
  font-size:1em;
  border-radius:5px;
  border:1px solid black;
  color:black;
  font-weight: bold;
}

.formtitulo
{
  font-size:2em;
  font-weight: bold;
  padding-bottom:0.8em;
  color:black;
}

a
{
  text-decoration: none;
  cursor:pointer;
  color:#1A3A83;
  font-weight: bold;
}

.af
{
  margin-right:10%;
  margin-top:5%;
}

.botonenviar
{
  width: 80%;
  padding: 10px 30px;
  cursor: pointer;
  display: block;
  margin-top: 10px;
  border: 0;
  outline: none;
  border-radius: 10px;
  border:1px solid black;
  font-size:16px;
  color:white;
  background-color: #2c3e50;
  text-align:center;
  margin:5%;
  font-weight: bold;
}
```

```
img
{
  width: 150px;
}
.imgv
{
  padding:20px;
}
.imgv img
{
  cursor: pointer;
  max-width: 20%;
  height:auto;
  margin-right:10px;
  margin-left:10px;
}

.textoscajas
{
  margin-left:8%;
  font-weight: bold;
  margin-top:2%;
  margin-bottom:2%;
  color:black
}

.autor
{
  color:white;
}
```



UNIVERSIDAD CÉSAR VALLEJO

**FACULTAD DE INGENIERÍA Y ARQUITECTURA
ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS**

Declaratoria de Autenticidad del Asesor

Yo, QUIÑONES NIETO YAMIL ALEXANDER, docente de la FACULTAD DE INGENIERÍA Y ARQUITECTURA de la escuela profesional de INGENIERÍA DE SISTEMAS de la UNIVERSIDAD CÉSAR VALLEJO SAC - LIMA NORTE, asesor de Tesis titulada: "Machine Learning para predecir el riesgo crediticio de un cliente en la empresa FUTECH PERU S.A.C, 2022

", cuyo autor es YARANGA VITE ITALO PAUL, constato que la investigación tiene un índice de similitud de 17.00%, verificable en el reporte de originalidad del programa Turnitin, el cual ha sido realizado sin filtros, ni exclusiones.

He revisado dicho reporte y concluyo que cada una de las coincidencias detectadas no constituyen plagio. A mi leal saber y entender la Tesis cumple con todas las normas para el uso de citas y referencias establecidas por la Universidad César Vallejo.

En tal sentido, asumo la responsabilidad que corresponda ante cualquier falsedad, ocultamiento u omisión tanto de los documentos como de información aportada, por lo cual me someto a lo dispuesto en las normas académicas vigentes de la Universidad César Vallejo.

LIMA, 19 de Noviembre del 2022

Apellidos y Nombres del Asesor:	Firma
QUIÑONES NIETO YAMIL ALEXANDER DNI: 42863390 ORCID: 0000-0003-4474-0556	Firmado electrónicamente por: YQUINONES el 05- 12-2022 15:53:55

Código documento Trilce: TRI - 0446296