



UNIVERSIDAD CÉSAR VALLEJO

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

**Machine Learning para Mejorar el Proceso de Prestación Bancaria
en una Entidad Financiera**

**TESIS PARA OBTENER EL TÍTULO PROFESIONAL DE:
Ingeniero de Sistemas**

AUTOR:

Cano Chuqui, Jorge (orcid.org/0000-0003-4809-6008)

ASESORES:

Dr. Chumpe Agosto, Juan Brues Lee (orcid.org/0000-0001-7466-9872)

Ms. Erick Giovanni Flores Chacón (orcid.org/0000-0002-4028-8059)

LÍNEA DE INVESTIGACIÓN:

Sistema de Información y Comunicaciones

LÍNEA DE RESONSABILIDAD SOCIAL UNIVERSITARIA:

Desarrollo económico, empleo y emprendimiento

LIMA – PERÚ

2022

Dedicatoria

Esta tesis la dedico a la II.EE. N° 0069 Machu Picchu 0069, los profesores Niser Minaya Quiroz, Alberto Rojas Sauñe y Maria Ysabel Alvarez Huari que conforman el mejor recuerdo que me llevé de la secundaria antes de empezar una carrera universitaria, a mi madre Maria Luz Chuqui Vargas y mi padre Jorge Alfredo Cano Vasquez, y para todas aquellas personas que creen en el valor de la perseverancia.

Agradecimientos

Agradezco al ingeniero Dr. Joel Visurraga Agüero porque sugirió que el machine learning podría ser para el cliente que tiene que llenar documentos para un préstamo bancario ya no lo haga de esa manera sino con Machine Learning y me dio la curiosidad de investigador sobre aquello. También agradezco la disponibilidad del ingeniero Chumpe Agosto, Juan Brues Lee, Erick Giovanni Flores Chacón y Rafael Eduardo Sebastian Benites Gomez por sus enseñanzas.

Índice de contenidos

	Página
Dedicatoria	ii
Agradecimientos	iii
Índice de contenidos	iv
Índice de tablas	v
Índice de figuras	vii
Índice de anexos	xi
Índice de abreviaturas	xii
Resumen	xiii
Abstract	xiv
I. INTRODUCCIÓN	1
II. MARCO TEÓRICO	7
III. METODOLOGÍA	34
3.1. Tipo y diseño de investigación	35
3.1.1. Tipo de investigación	35
3.1.2. Diseño de investigación	36
3.2. Variables y operacionalización	37
3.3. Población, muestra y muestreo	38
3.4. Técnicas e instrumentos de recolección de datos	39
3.5. Procedimientos	40
3.6. Método de análisis de datos	43
3.7. Aspectos éticos	43
IV. RESULTADOS	44
V. DISCUSIÓN	62
CONCLUSIONES	69
RECOMENDACIONES	71
REFERENCIAS	74
ANEXOS	81

Índice de tablas

	Página
Tabla 1 Indicadores de la investigación	19
Tabla 2 Comparación del AS – IS y TO – BE.....	20
Tabla 3 Statistical Modeling vs. Machine Learning.....	24
Tabla 4 Descripciones del modelo de proceso CRISP-DM	26
Tabla 5 Fortalezas y limitaciones entre las metodologías	29
Tabla 6 Definición conceptual por indicadores de la variable dependiente	38
Tabla 7 Unidad de análisis por indicador	39
Tabla 8 Técnicas e instrumentos de recolección de datos.....	40
Tabla 9 Resultados obtenidos de la preprueba y postprueba	45
Tabla 10 Datos de la preprueba del tiempo de espera en el proceso de prestación bancaria	49
Tabla 11 Datos de la postprueba del tiempo de espera en el proceso de prestación bancaria.....	49
Tabla 12 T de Student para la Hipótesis 1	51
Tabla 13 Datos de la preprueba del costo de personal en el proceso de prestación bancaria	52
Tabla 14 Datos de la postprueba del costo de personal en el proceso de prestación bancaria.....	52
Tabla 15 T de Student para la Hipótesis 2	54
Tabla 16 Datos de la preprueba de la satisfacción del cliente en el proceso de prestación bancaria.....	55
Tabla 17 Datos de la postprueba de la satisfacción del cliente en el proceso de prestación bancaria.....	55
Tabla 18 Datos de la preprueba de la eficiencia de evaluación en el proceso de prestación bancaria.....	59
Tabla 19 Datos de la postprueba de la eficiencia de evaluación en el proceso de prestación bancaria.....	59
Tabla 20 T de Student para de la Hipótesis 4	61

Tabla 21 Matriz de consistencia	82
Tabla 22 Matriz de operacionalización	83
Tabla 23 I1: Tiempo de espera.....	107
Tabla 24 I2: Costo de personal	109
Tabla 25 Tabla de frecuencias preprueba de la satisfacción del cliente para la entidad financiera	111
Tabla 26 Tabla de frecuencias postprueba de la satisfacción del cliente para la entidad financiera	112
Tabla 27 I4: Eficiencia de evaluación	113
Tabla 28 Aplicación de técnicas estadísticas	120
Tabla 29 Análisis de los nodos del árbol de decisión	126
Tabla 30 Tooltips del sistema.....	149
Tabla 31 Diccionario de datos	160

Índice de figuras

	Página
Figura 1 Diagrama Ishikawa.....	4
Figura 2 Flujo de trabajo y arquitectura de la información crediticia.....	13
Figura 3 Tipos de riesgo y clasificación de los modelos de riesgo de crédito	14
Figura 4 Contenido principal de un préstamo.....	15
Figura 5 Modelo de ciclo de vida de CRISP-DM	25
Figura 6 Fases del proceso SEMMA.....	27
Figura 7 Modelo de ciclo de vida de KDD	28
Figura 8 Generalización del Machine Learning supervisado	30
Figura 9 Descenso de gradiente	31
Figura 10 Ejemplo de formato CSV.....	31
Figura 11 Patrón de un árbol de decisión.....	33
Figura 12 Metodología asociada a un problema organizacional	35
Figura 13 Modelo de diseño de investigación	36
Figura 14 Prueba de normalidad preprueba y postprueba del tiempo de espera .	47
Figura 15 Prueba de normalidad preprueba y postprueba del costo de personal	47
Figura 16 Prueba de normalidad preprueba y postprueba de satisfacción del cliente.....	48
Figura 17 Prueba de normalidad preprueba y postprueba de la eficiencia de evaluación	48
Figura 18 Gráfica de distribución postprueba del tiempo de espera	50
Figura 19 Cálculo para contrastación del tiempo de espera	50
Figura 20 Cálculo de criterios de aceptación de la hipótesis del tiempo de espera	51
Figura 21 Gráfica de distribución postprueba del costo de personal.....	53
Figura 22 Cálculo para contrastación del costo de personal.....	53
Figura 23 Cálculo de criterios de aceptación de la hipótesis del costo de personal	54
Figura 24 Gráfica de distribución postprueba de la satisfacción del cliente	56
Figura 25 Cálculo para contrastación de la satisfacción del cliente	57

Figura 26 Cálculo de criterios de aceptación de la hipótesis de la satisfacción del cliente	57
Figura 27 Mann-Whitney preprueba y postprueba de satisfacción del cliente para la entidad financiera	58
Figura 28 Gráfica de distribución postprueba de la eficiencia de evaluación	60
Figura 29 Cálculo para contrastación de la eficiencia de evaluación	60
Figura 30 Cálculo de criterios de aceptación de la hipótesis de la eficiencia de evaluación	61
Figura 31 Gráfico circular preprueba y postprueba de la satisfacción del cliente para la entidad financiera	111
Figura 32 Edad de individuos %	122
Figura 33 Ganancia deseada de individuos %	123
Figura 34 Cantidad de manutención de individuos %	123
Figura 35 Monto de individuos %	124
Figura 36 Edad a jubilarse de individuos %	124
Figura 37 Árbol de decisión	125
Figura 38 Código para obtener datos de prueba	126
Figura 39 Código para construir la matriz de confusión	127
Figura 40 Resultado de matriz de confusión	127
Figura 41 Código para obtener precisión de matriz de confusión	128
Figura 42 Resultados de precisión de matriz de confusión	128
Figura 43 Accesos para visualizar entorno de colab	129
Figura 44 Entorno de Colab	129
Figura 45 Fases de la metodología CRISP-DM	130
Figura 46 Estructura del proyecto	131
Figura 47 App.py_1	131
Figura 48 App.py_2	132
Figura 49 index.html_1	132
Figura 50 index.html_2	133
Figura 51 index.html_3	133
Figura 52 index.html_4	134
Figura 53 index.html_5	134
Figura 54 index.html_6	135

Figura 55 index.html_7.....	135
Figura 56 index.html_8.....	136
Figura 57 index.html_9.....	136
Figura 58 index.html_10.....	137
Figura 59 index.html_11.....	137
Figura 60 index.html_12.....	138
Figura 61 index.html_13.....	138
Figura 62 index.html_14.....	139
Figura 63 index.html_15.....	139
Figura 64 index.html_16.....	140
Figura 65 index.html_17.....	140
Figura 66 index.html_18.....	141
Figura 67 index.html_19.....	141
Figura 68 request.py.....	142
Figura 69 Estilos.css_1.....	142
Figura 70 Estilos.css_2.....	143
Figura 71 Estilos.css_3.....	143
Figura 72 Estilos.css_4.....	144
Figura 73 Estilos.css_5.....	144
Figura 74 Estilos.css_6.....	145
Figura 75 Estilos.css_7.....	145
Figura 76 Consola de PythonAnywhere.....	146
Figura 77 PythonAnywhere_1.....	146
Figura 78 PythonAnywhere_2.....	147
Figura 79 PythonAnywhere_3.....	147
Figura 80 PythonAnywhere_4.....	148
Figura 81 Interfaz del aplicativo.....	148
Figura 82 Botón principal para predecir.....	150
Figura 83 Interfaz del aplicativo con datos.....	150
Figura 84 Interfaz del aplicativo con resultado.....	151
Figura 85 Validación de campos.....	151
Figura 86 Interfaz del aplicativo para satisfacción.....	152
Figura 87 Modal de satisfacción.....	152

Figura 88 uID del documento generado	153
Figura 89 uID del documento generado en Firestore	153
Figura 90 Indicadores en el sistema.....	154
Figura 91 Diagrama de casos de uso.....	155
Figura 92 Diagrama de componentes	155
Figura 93 Diagrama de secuencia.....	156
Figura 94 Diagrama de colaboración	157
Figura 95 Diagrama de despliegue	157
Figura 96 Algoritmo innovador	158
Figura 97 Arquitectura tecnológica.....	159
Figura 98 Diseño de base de datos.....	160
Figura 99 Proceso de prestación bancaria de Banco Alfin AS-IS.....	161
Figura 100 Proceso de prestación bancaria de Banco Alfin TO-BE	161
Figura 101 Evidencia de capacitación N°1	162
Figura 102 Evidencia de capacitación N°2.....	162
Figura 103 Evidencia de capacitación N°3.....	162

Índice de anexos

	Página
Anexo 1: Matriz de consistencia.....	82
Anexo 2: Matriz de operacionalización.....	83
Anexo 3: Certificado de validez de los indicadores para medir la variable prestación bancaria	84
Anexo 4: Tabla de evaluación de expertos: Lenguaje de programación	90
Anexo 5: Tabla de evaluación de expertos: Metodología de desarrollo	93
Anexo 6: Tabla de evaluación de expertos: Gestor de base de datos	96
Anexo 7: Ficha de observación del indicador TDE: Tiempo de espera	99
Anexo 8: Ficha de observación del indicador CDP: Costo de personal	101
Anexo 9: Ficha de observación del indicador CDP: Satisfacción del cliente	103
Anexo 10: Ficha de observación del indicador EFE: Eficiencia de evaluación....	105
Anexo 11: Análisis a profundidad del tiempo de espera	107
Anexo 12: Análisis a profundidad del costo de personal	109
Anexo 13: Análisis a profundidad de la satisfacción del cliente	111
Anexo 14: Análisis a profundidad de la eficiencia de evaluación	113
Anexo 15: Constancia de autorización de investigación de tesis	115
Anexo 16: Constancia de autorización de uso de datos en proyecto de tesis.....	116
Anexo 17: Reportes en Excel.....	117
Anexo 18: Metodología operativa.....	118
Anexo 19: Ingeniería de software.....	155
Anexo 20: Proceso AS-IS y TO-BE	161
Anexo 21: Capacitación antes de implementar retroalimentación.....	162
Anexo 22: Carta de aceptación	163
Anexo 23: Artículo de investigación	164
Anexo 24: Pantallazo software Turnitin	165
Anexo 25: Actas del producto	166

Índice de abreviaturas

Sigla	Significado	Pág.
ML	Machine Learning	5
TDSP	Team Data Science Process	8
KDD	Knowledge Discovery in Databases	9
SEMMA	Sample, Explore, Modify, Model y Assess	12
CRISP-DM	Cross Industry Standard Process for Data Mining	12
DM	Data Mining	29
CTI	Ciencia, Tecnología e Innovación	43
EDA	Event-driven architecture	102

Resumen

El objetivo de la presente investigación fue mejorar el proceso de prestación bancaria en una entidad financiera implementando Machine Learning.

La investigación corresponde al enfoque cuantitativo de tipo aplicada y de diseño preexperimental, la muestra fue de 30 procesos, el instrumento fue una ficha de observación que recopiló datos como tiempo de espera, costo de personal, satisfacción de cliente y eficiencia de evaluación.

Para determinar la normalidad de los datos se aplicó la prueba de Anderson Darling determinándose que son paramétricos el tiempo de espera, costo de personal y eficiencia de evaluación, y no paramétricos la satisfacción del cliente, y con esto se pueden aplicar las pruebas T de Student y U de Mann Whitney para la validación de las hipótesis.

Habiéndose obtenido el p valor de 0.000 en el tiempo de espera, costo de personal y eficiencia de evaluación mediante la prueba T de Student, así como haber alcanzado el p valor de 0.001 en la satisfacción del cliente a través de la prueba U de Mann Whitney, se evidencia la mejora del 100% de los indicadores propuestos con lo que se logró el objetivo de la investigación. Se recomienda otras investigaciones con diseño longitudinal.

Palabras claves: aprendizaje automático¹, calificación crediticia², evaluación de riesgos³, algoritmos⁴, inteligencia artificial⁵.

Abstract

The objective of this research was to improve the banking process in a financial institution by implementing Machine Learning.

The research corresponds to the quantitative approach of applied type and pre-experimental design, the sample was 30 processes, the instrument was an observation sheet that collected data such as waiting time, personnel cost, customer satisfaction and evaluation efficiency.

To determine the normality of the data, the Anderson Darling test was applied, determining that waiting time, personnel cost and evaluation efficiency were parametric, and customer satisfaction was nonparametric, and with this the Student's t-test and Mann Whitney U-test could be applied to validate the hypotheses.

Having obtained a p-value of 0.000 for waiting time, personnel cost and evaluation efficiency through the Student's t-test, as well as having reached a p-value of 0.001 for customer satisfaction through the Mann Whitney U-test, 100% improvement of the proposed indicators is evident, thus achieving the objective of the research. Further research with longitudinal design is recommended.

Keywords: machine learning¹, credit rating², risk assessment³, algorithms⁴, artificial intelligence⁵.

I. INTRODUCCIÓN

Los temas relacionados con la realidad problemática en la que se evidenció problemas para direccionar los resultados para el cliente en cuanto a los préstamos de grandes montos o de bajos montos se desarrollaron en el primer capítulo. También se justificó teórica, metodológica y tecnológicamente la investigación. Como resultado, se propuso como reto determinar el impacto de la puesta en marcha de un sistema para aumentar la evaluación de crédito personal y, por último, se presentan cuestiones particulares sobre las condiciones que deben cumplirse para que la información se considere de alta calidad.

En la sociedad globalizada existe una falta de capacidad para dar cuenta de la complejidad del tiempo en las operaciones comerciales reales. Para acceder a los préstamos, los solicitantes deben mantenerse registrados a una dimensión socioeconómica básica regular para devolver el dinero.

Los solicitantes se preocupan de que sean o no buenos pagadores y asegurarse de que puedan calificar como esperan según sus actividades como cliente. (Chen, Guo y Zhao 2021) señalan que la calificación es la información que proviene de las bancas o entidades financieras reguladas por la superintendencia de bancos que se establecen en función del grado de atraso en el pago de las deudas que fluctúan de normal hasta pérdida en el sistema financiero.

Además, se suele denominar un puntaje de riesgos que va del 1 al 999, que mide específicamente la posibilidad de que una persona pueda caer en morosidad en un periodo que es normalmente de 12 meses. También un componente importante de la calificación es la deuda de la tarjeta de crédito del cliente, que es probable que se clasifique como vencida incluso cuando la tarjeta no ha incumplido los préstamos (Chen, Guo y Zhao 2021).

Por otro lado a nivel nacional, probablemente los morosos no tengan una única deuda y tengan que cumplir con la totalidad de sus obligaciones por el pago, (Mori 2021) señala que a medida que va avanzando el tiempo no solamente la deuda se va distanciando en el tiempo, llegado los 2 años tiene que salir o desaparecer de esa deuda y además debe cumplir rigurosamente las obligaciones que poco a poco le permita mejorar su situación y ya la deuda deja de ser relevante y sale por vencimiento de plazo.

Asimismo, (Mori 2021) señala que el concepto de deudas es inatendible para la persona común, donde hay un resumen de todas las deudas en Infocorp que se

pagaron con retraso, que muestran aquellas canceladas y las todavía no. La ley establece un plazo máximo de 5 años para mostrar información de Infocorp con excepción de personas que cayeron en morosidad que pueden resolver su situación, pero se pueden poner al día, ya que principalmente fomenta la responsabilidad del deudor con un nuevo plazo de 2 años (Mori 2021).

Por otra parte, las personas que no tienen morosidad no tienen deudas, mientras que los morosos acumulan deudas no resueltas, las personas que caen en morosidad ya resuelven la situación y por medio de este simple hecho los bancos no aceptan brindar préstamos a aquellas personas que son morosas que a una persona en una mejor situación o que no es morosa (Mori 2021).

De esta manera, quedan registros, pero con situaciones distintas y el supuesto de que un cliente tiene una deuda y no lo ha pagado, determina la alta probabilidad de calificar a un cliente como moroso (Mori 2021).

En ese sentido, (Mori 2021) plantea que la calificación entonces detecta la morosidad histórica como un control de deuda que la tiene, la tuvo o no un cliente, por lo tanto, los bancos se hacen responsables constantemente de pérdidas inesperadas que excedan las pérdidas planificadas, tal es el caso, de prestar dinero y que este no sea devuelto.

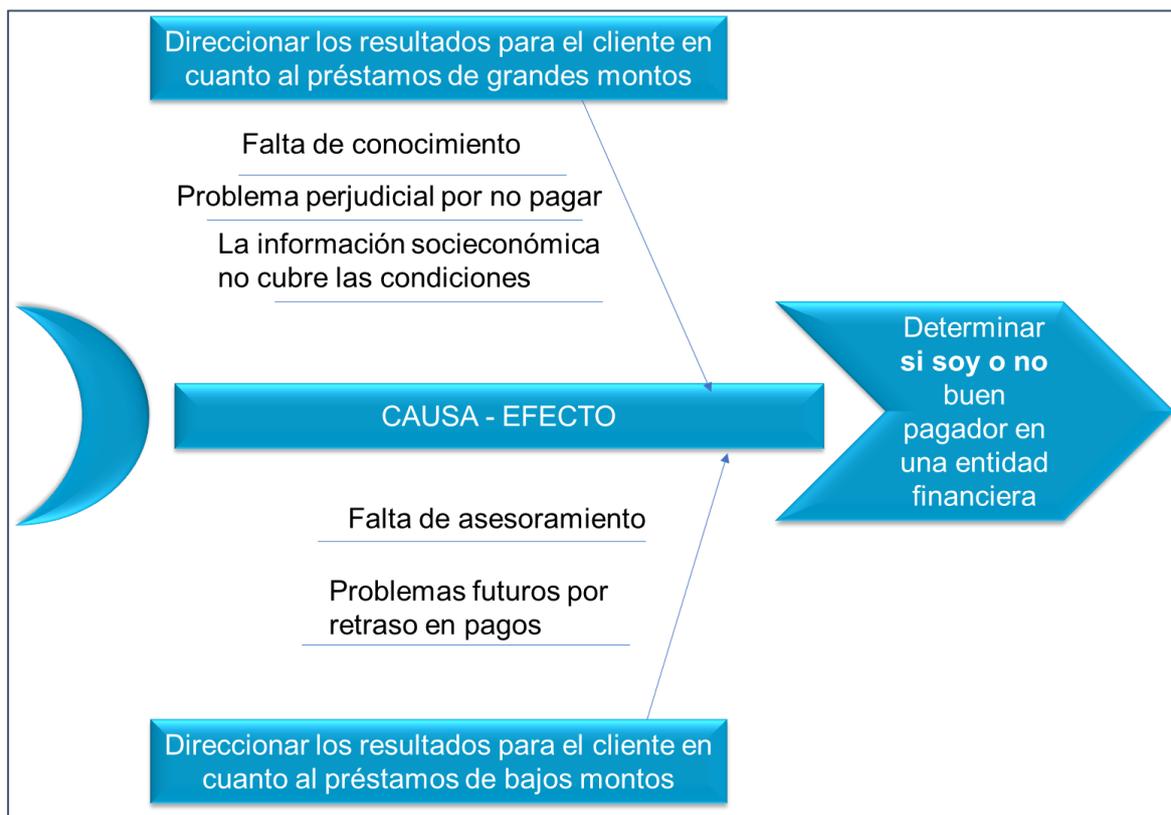
Por otra parte a nivel institucional, hay que remediar esta carencia reduciendo los costes laborales, lo que permite a los bancos sostener pérdidas imprevistas que superan las presupuestadas. Por otro lado, la organización tiene curiosidad por saber si los clientes están muy de acuerdo o muy en desacuerdo con su satisfacción, para ello se debe estudiar la satisfacción del cliente.

Adicionalmente, el tiempo de espera precisa de la importancia de que un cliente presente sus documentos fue alta y crucial, por este motivo se fija generalmente en el tiempo (minutos), ante esto se estudió el tiempo de espera. Asimismo, sobre el desempeño anterior de los préstamos se encontró el factor de mejora significativa de esta información que permita ganar tiempo en determinar el posible cliente moroso o el buen pagador, por ello, se estudió la eficiencia de evaluación en una entidad financiera.

En relación con la problemática expuesta se determinó el problema del investigador antes de abordar la pregunta, identificando las causas fundamentales a través del diagrama de causa – efecto de Ishikawa en la siguiente figura:

Figura 1

Diagrama Ishikawa



Fuente: Elaboración propia

En consecuencia se propuso como problema general ¿En qué medida el uso de Machine Learning mejorará el Proceso de prestación bancaria en una entidad financiera?, también se formuló los siguientes problemas específicos: i) ¿En qué medida la aplicación de Machine Learning disminuirá el tiempo de espera en una entidad financiera?, ii) ¿En qué medida el uso de Machine Learning reducirá el costo de personal en una entidad financiera?, iii) ¿En qué medida el uso de Machine Learning aumentará la satisfacción del cliente en una entidad financiera?, iv) ¿En qué medida el uso de Machine Learning elevará la eficiencia de evaluación en una entidad financiera?

En cuanto a la justificación económica, la investigación fue lo suficientemente fuerte para llevarla a cabo y sustentó los beneficios que derivaron de ella, dentro de lo cual, producto de la realización de la investigación, las aseguradoras utilizan regularmente la IA para medir mejor las características de los clientes y mejorar sus sistemas de calificación. Dentro de las instituciones de crédito, las aplicaciones son comunes y van desde el análisis de riesgo crediticio hasta la detección de fraudes

y el cumplimiento normativo, incluido el análisis económico y financiero (Doerr, Gambacorta y Serena Garralda 2021, p. 8). Para la entidad financiera que deseó elegirlo o informarse acerca de Machine learning que se ejecutó y mostró al cliente si es o no buen pagador, se depositó el ML en un sistema web que fue gratuito para las personas en general, dentro de esto contuvo información adicional que complementó al sistema. Tras acceder a esta información el banco logra el ahorro de costos porque menos personas operaron el proceso para que funcionara.

Asimismo, a nivel operativo, la investigación fue relevante porque la calificación crediticia es el proceso de evaluar la solvencia de un individuo y refleja el riesgo crediticio y también se decide si aprobar o rechazar la solicitud de un individuo (Ampountolas et al. 2021, p. 1). A través de la investigación se resolvió problemas prácticos en la solicitud de documentos para aplicar a un préstamo bancario, permitiendo que la predicción determine si el cliente es o no buen pagador en una entidad financiera y de esta manera fue más rápida, y por ende recibieron las personas una atención en un tiempo menor que el habitual, asimismo, permitió a la persona el llenado sus datos socioeconómicos en el sistema y de manera óptima se hizo la predicción respectiva.

Con respecto a la justificación tecnológica, la investigación ayudó a resolver algunos problemas de capacidad de las empresas para implementar Machine Learning a sus procesos, por ello, en esta investigación se dió a conocer en mayor medida el comportamiento de los sistemas objetivos de calificación crediticia que tienden a ser superiores a los sistemas subjetivos para predecir un cliente con mayor precisión. Además, un sistema de crédito basado en criterios objetivos permite un análisis automatizado estandarizado y contribuye al crecimiento del banco (de Castro Vieira et al. 2019, p. 2). El desarrollo de la presente investigación sirvió para mejorar el proceso de prestación bancaria en una entidad financiera, mediante el uso de un sistema con Machine Learning. Los resultados que se obtuvieron del sistema con Machine Learning, tras la implementación en el proceso de prestación bancaria, se pudo presentar una solución tecnológica que fue posible de ser adaptado para el banco Alfin.

Con relación a la justificación social, a pesar de las preocupaciones anteriores y de cierta cobertura mediática negativa de que se trate de una corrupción o aprovechamiento para obtener datos del cliente, fue importante

realizar el estudio para enfatizar que la automatización del proceso de prestación bancaria brinda una importante oportunidad de mejora para el Perú. La clasificación algorítmica elimina el proceso de toma de decisiones por manos humanas (Varley y Belle 2021, p. 3). Mediante la implementación del sistema con Machine Learning se pudo generar un beneficio mediante la mejora del proceso de prestación bancaria en una entidad financiera, para lograr el desarrollo óptimo de las actividades del cliente. Esto a su vez mejoró la calidad del servicio brindado, con la utilización de parámetros de predicción socioeconómicos para determinar si el cliente es o no buen pagador al cual pudieron acceder los ciudadanos del país.

Con respecto a los objetivos de investigación se planteó el objetivo general: Mejorar el proceso de prestación bancaria en una entidad financiera implementando Machine Learning. También se propuso los siguientes objetivos específicos: i) Disminuir el tiempo de espera del proceso de prestación bancaria en una entidad financiera implementando Machine Learning; ii) Reducir el costo de personal del proceso de prestación bancaria en una entidad financiera implementando Machine Learning; iii) Aumentar la satisfacción del cliente del proceso de prestación bancaria en una entidad financiera implementando Machine Learning e iv) Elevar la eficiencia de evaluación del proceso de prestación bancaria en una entidad financiera implementando Machine Learning.

Finalmente, se planteó la hipótesis de investigación: Si se usa Machine Learning, entonces mejora el proceso de prestación bancaria en una entidad financiera. Asimismo, se planteó las siguientes hipótesis específicas: i) Si se usa Machine Learning, entonces disminuye el tiempo de espera del proceso de prestación bancaria en una entidad financiera; ii) Si se usa Machine Learning, entonces reduce el costo de personal del proceso de prestación bancaria en una entidad financiera; iii) Si se usa Machine Learning, entonces aumenta la satisfacción del cliente del proceso de prestación bancaria en una entidad financiera e iv) Si se usa Machine Learning, entonces eleva la eficiencia de evaluación del proceso de prestación bancaria en una entidad financiera.

II. MARCO TEÓRICO

En este capítulo se presentaron los estudios relacionados con el trabajo de investigación. Se localizaron varios estudios a nivel nacional e internacional entre artículos, tesis y diferentes libros, donde se priorizó el problema, la problemática y los resultados dentro de las revisiones realizadas y se explicaron las teorías relacionadas con el tema. También se detallan conceptos sobre el aprendizaje automático, metodologías, herramientas e instrumentos que se utilizaron para la investigación. Se realizó una búsqueda exhaustiva en varios sitios, repositorios, documentos, bases de datos y otras fuentes para encontrar todos los datos utilizados en este estudio.

En referencia a los antecedentes nacionales previos revisados sobre Machine Learning y el proceso de prestación bancaria, se tiene a (Chávez 2021), en su investigación “Modelos de machine learning para identificar factores asociados a la adquisición de un crédito efectivo en una entidad financiera”, elaborado en la Universidad Nacional Mayor de San Marcos; (Chávez 2021) elaboró una investigación para identificar los factores que influyen en la adquisición de crédito efectivo por parte del cliente, determinar cuáles de ellos son susceptibles de aceptar el producto y gestionarlos de forma proactiva y contextualizada. (Chávez 2021) obtuvo resultados en donde se considera la metodología TDSP, clientes con más de 180 días de mora, la muestra es de 2,765 clientes, muestreo probabilístico, El modelo xgboost logró una sensibilidad de 69,9° frente a los 67,3°. El modelo lightgbm logró muy buenos indicadores en los modelos de aprendizaje automático xgboost y lightgbm. Esto sugiere que en el mundo de las personas que usan créditos aceptados y válidos, se hace uso del algoritmo xgboost para calcular con máxima precisión. Este trabajo previo sirvió para discutir el indicador Eficiencia de Evaluación.

Asimismo, (Pérez 2021), en su investigación “Modelo de Scoring para la segmentación de clientes morosos usando minería de datos en una empresa de cobranzas del Perú”, elaborado en la Universidad Nacional Mayor de San Marcos; (Pérez 2021) analizó el problema de no conocer el segmento de clientes con plazos vencidos por mora de más de 180 días, lo que generaba mayores costos operativos. Utilizando minería de datos y análisis predictivo, se ha desarrollado un

modelo de puntuación que segmenta a los clientes morosos que tienen 180 días o más de mora, reduciendo costos a través de la reasignación de carteras de cobranza. (Pérez 2021) obtuvo resultados en donde se considera los clientes con más de 180 día de mora, la muestra es de 2,765 clientes, la metodología aplicada fue Knowledge Discovery in Databases (KDD), el porcentaje de precisión (64,66%) es una estimación que ayuda a las instituciones microfinancieras a tomar decisiones, ya que se realiza de manera subjetiva y sin ninguna medición que permita conocer de manera más sencilla y clara el posible comportamiento del prestatario al momento de solicitar un préstamo de consumo, concluyendo así que para reducir los costos mensuales de aproximado el 19.5%, representan S/. 92.160 que el Departamento Interno generará, en promedio, una eficiencia mensual de 6,46% y adquirirá 192 clientes más que el Departamento Externo. Este trabajo previo sirvió para discutir el indicador Costo de Personal.

De igual manera, según (Alarcón, Eduardo; Mora 2022), en su investigación “Modelo para la evaluación del riesgo crediticio para los clientes de las microfinancieras del Perú”, elaborada en la Universidad Peruana de Ciencias Aplicadas; (Alarcón, Eduardo; Mora 2022) implementaron un modelo de pronóstico de riesgo crediticio no renovable para instituciones de microfinanzas en Perú. (Alarcón, Eduardo; Mora 2022) analizaron la alta tasa de morosidad de las instituciones de microfinanzas para los bancos en Perú (7,97% sin el estímulo y 4,1% con el estímulo), concluyendo que al momento de evaluar a los clientes, los métodos tradicionales de evaluación serán reemplazados por métodos tecnológicos más efectivos, que no solo reducen el índice de infracciones sino también en la implementación de sus procesos. Este trabajo previo sirvió para discutir el indicador Satisfacción del Cliente.

También, según (De la Cruz 2020), en su investigación “Predicción de adquisición de un préstamo personal bancario a través del canal de televentas utilizando el algoritmo random forest”, elaborado en la Universidad Nacional Agraria la Molina; (De la Cruz 2020) desarrollaron la predicción de la compra de un préstamo bancario personal a través de un canal de telemarketing utilizando el algoritmo Random Forest, su proceso está vinculado a una distribución quincenal

que el banco comparte con el campo de actividad de la campaña de préstamos, sin embargo, este objetivo no se cumplió de manera efectiva, porque este segmento no funcionó correctamente; los resultados obtenidos por (De la Cruz 2020) al tomar en cuenta el algoritmo Random Forest, muestra de 991.619 clientes que se registraron con tarjeta de crédito y tenían un préstamo preaprobado en una entidad financiera, valores aceptables para la sensibilidad y la especificidad de los datos experimentales y sintéticos; el poder predictivo del modelo implementado asegura un buen desempeño futuro, se concluye que el estudio ha desarrollado un modelo de clasificación para predecir si el cliente comprará o rechazará un oficial de crédito bancario vía canal de venta telefónica utilizando el algoritmo Random Forest, esto conlleva a lograr incrementar las ventas de campaña desde que se empezó a gestionar la base de datos con este modelo. Este trabajo previo sirvió para discutir el indicador Satisfacción del Cliente.

Asimismo, según (Aceituno 2019), en su investigación “Modelo predictivo de análisis de riesgo crediticio usando machine learning en una entidad del sector micro financiero”, elaborado en la Universidad Nacional del Altiplano; (Aceituno 2019) realizó una investigación de Machine Learning para mejorar el nivel de asertividad en el otorgamiento de microcréditos en una organización del rubro micro financiero, realizando también el análisis de la problemática que presenta la entidad financiera en cuanto al otorgamiento de microcréditos, a pesar de la presencia de personal especializado conocido como asesores comerciales, donde se puede observar que esta unidad tiene clientes atrasados en el pago de las cuotas. Mediante los estudios realizados por el autor se consideró la investigación no experimental de tipo transversal, la muestra de 15,015 registros de clientes de la entidad, manifestando, asimismo por (Aceituno 2019), obtener un nivel de asertividad que es de 93,72%, en los modelos de Regresión Logística (86,07%), Random Forest (66,35%), Support Vector Machines (84,44%), Decision Trees (88,80%) y kNearest Neighbor (65,98%), concluyendo que el modelo más convincente de microcrédito en instituciones de microfinanzas es la Red Neuronal Artificial. Este trabajo previo sirvió para discutir el indicador Eficiencia de Evaluación.

En el plano internacional, los antecedentes revisados sobre la variable Machine Learning y el proceso de prestación bancaria, se tiene a (Chen, Katchova y Zhou 2021), en su investigación “Predicción de la morosidad de los préstamos agrícolas utilizando métodos de aprendizaje automático”, elaborado en la Universidad Estatal de Ohio; (Chen, Katchova y Zhou 2021) elaboraron una investigación sobre Machine learning para predecir qué bancos tendrán un alto riesgo en términos de tasas de incumplimiento asociadas con los préstamos agrícolas. (Chen, Katchova y Zhou 2021) obtuvieron resultados en donde se considera la regresión logística, bancos con 3% como tasa media-alta para préstamos comerciales y a nivel estatal, la muestra de 131 431 observaciones de años bancarios, clasificando correctamente a un banco, ya sea que tenga o no un alto riesgo de préstamo agrícola en el 61% de los casos, las condiciones financieras y operativas a nivel bancario y los indicadores macroeconómicos, como la tasa de interés, que son más importantes que las características agrícolas agregadas disponibles públicamente al predecir el riesgo de morosidad de un banco. Este trabajo previo sirvió para discutir el indicador Tiempo de Espera.

También, según (Ma et al. 2018), en su investigación “Estudio sobre una predicción del incumplimiento de préstamos de la red P2P basado en los algoritmos de aprendizaje automático LightGBM y XGboost de acuerdo con diferentes limpiezas de datos de alta dimensión”, elaborado en la University of Finance and Economics; (Ma et al. 2018) mencionan que es difícil predecir el riesgo de incumplimiento de préstamos con precisión de forma manual; se estudió la morosidad basada en criterios cualitativos y aspectos cuantitativos, se mejoró el sistema de índices para representar la confiabilidad de los prestatarios o proyectos de préstamo en Lending Club, también aportando a analizar los algoritmos LightGBM y XGboost. (Ma et al. 2018) obtuvieron resultados en donde se considera 8.505.916 observaciones de solicitud de préstamos rechazadas que no han sido examinadas; el ingreso anual representa la fuerza económica de una persona, que maximiza la capacidad del prestatario para pagar la deuda a tiempo cada mes, concluyendo que el algoritmo LightGBM es mejor que el algoritmo XGboost con una tasa de error del 19,9% y una precisión del 80,1%. Este trabajo previo sirvió para discutir el indicador Eficiencia de Evaluación.

Asimismo, según (Tariq et al. 2019), en su investigación “Modelo de predicción de impago de préstamos utilizando Sample, Explore, Modificar, modelar y evaluar (SEMMA)”, desarrollado en la Universidad de Tecnología e Innovación de Asia Pacífico; (Tariq et al. 2019), realizaron una investigación la alta proporción de préstamos incobrables es muy crítica en el sector financiero, realizando una investigación exhaustiva y desarrollar un modelo para predecir los incumplimientos de los préstamos, considerando las metodologías denominadas KDD, CRISP-DM y SEMMA. (Tariq et al. 2019) mediante los estudios realizados se seleccionó SEMMA por sus características significativas en cuanto a la predicción de la morosidad en el sector financiero. Este trabajo previo sirvió para discutir el indicador Costo de Personal.

También, según (Ereiz 2019), en su investigación “Predicción de préstamos predeterminados usando (OptiML)”, elaborado en la Universidad Dzemal Bjedic de Moslar; (Ereiz 2019) analizó el proceso de evaluación de riesgo crediticio estándar basado en el cálculo lineal de una pequeña cantidad de indicadores. Como objetivo que tuvo fue desarrollar machine learning para el cliente y puede usarse para administrar no solo el riesgo crediticio, sino también otros riesgos. (Ereiz 2019) también definió una muestra de 3612 préstamos, el modelo de puntuación obtiene correctamente que solo el 1,8 % de los préstamos en mora estarán en mora, lo que permitió predecir un préstamo que será tan bueno como malo (falso positivo) no es tan costoso como predecir un préstamo que será tan malo como bueno (falso negativo). Este trabajo previo sirvió para discutir el indicador Tiempo de Espera.

De igual manera, según (Liu 2022), en su investigación “Evaluación del procesamiento humano de la información en las decisiones de préstamo: un enfoque de aprendizaje automático”, desarrollado en la Universidad de Chicago; (Liu 2022) desarrolló una investigación para predecir con un error del 5% a los prestatarios que regresan por un segundo préstamo, los oficiales de crédito no tienen el incentivo para prestar a prestatarios de baja calidad para formar relaciones para que luego puedan sostener a estos prestatarios, se empleó como método los árboles de decisión potenciados por gradiente (GBM); maximizó el reembolso de cada préstamo. (Liu 2022) obtuvo resultados positivos como generar una tasa de

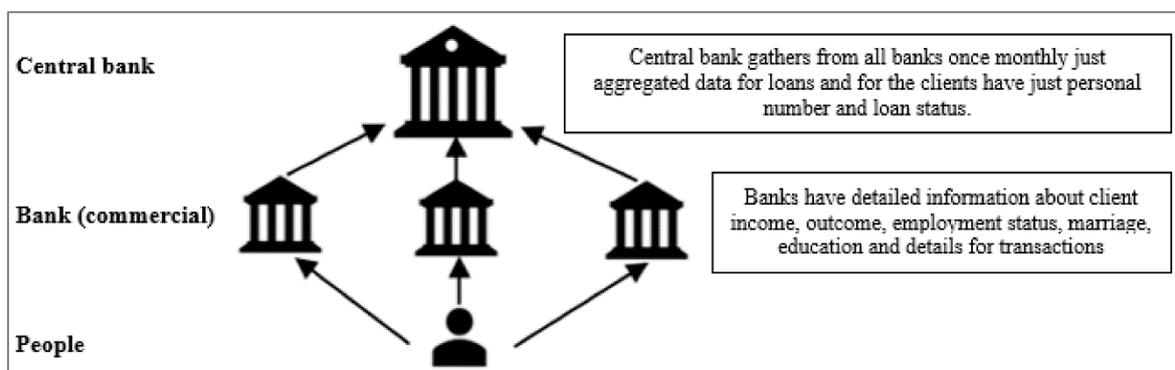
ganancia del 21,5% y el aumento del 38% de la ganancia observada, concluyendo que, usando el modelo de aprendizaje automático como punto de referencia, encuentra que la atención limitada y la reacción exagerada a la información destacada explica en gran medida la debilidad del préstamo de crédito. Este trabajo previo sirvió para discutir el indicador Tiempo de Espera.

Con respecto al marco teórico, la variable proceso de prestación bancaria se fundamentó de los siguientes conceptos:

Según (Shi et al. 2019), tradicionalmente, el cliente debe presentar su prueba de identidad, declaración de ingresos, dirección, estado de deuda anterior al banco A. Luego, el banco evalúa al cliente, si el cliente cumple con sus requisitos, puede otorgarle el préstamo del automóvil al cliente. Sin embargo, el método tradicional puede costar mucho tiempo y esfuerzo al cliente y al banco. El cliente tiene que preparar los documentos relacionados y por su parte el Banco tiene que verificar estos documentos. Además, la información enviada por el cliente solo puede solicitar un préstamo en el banco A, no puede usar sus documentos proporcionados al banco A en otro banco. Si el cliente desea adquirir un producto financiero de otro banco o entidad financiera tiene que presentar los documentos y esperar nuevamente el proceso de verificación. (p. 68). El flujo de un proceso de prestación bancaria se muestra en la Figura 2.

Figura 2

Flujo de trabajo y arquitectura de la información crediticia

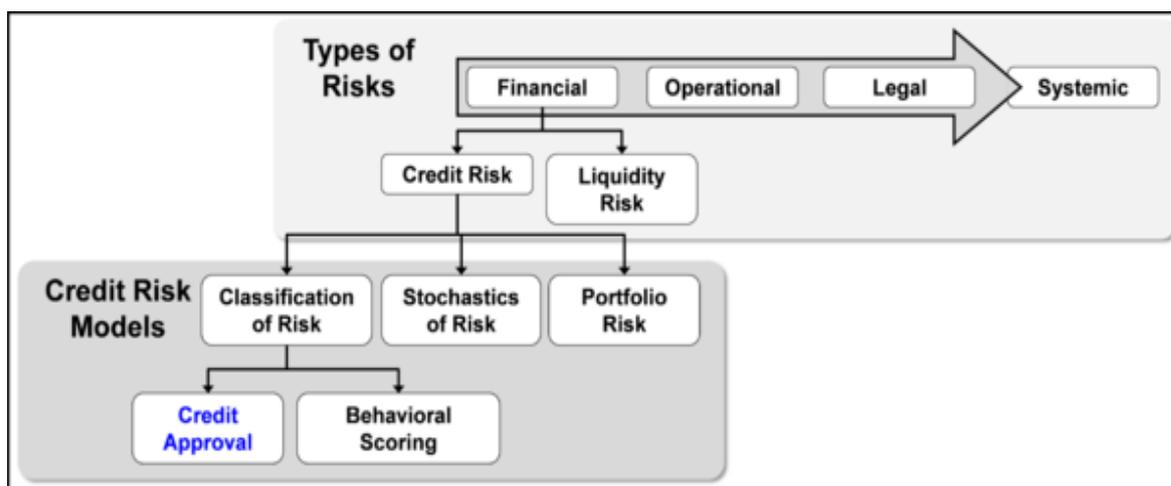


Fuente: (Doko, Kalajdziski y Mishkovski 2021)

Con relación al proceso de prestación bancaria, (Meikang 2018) manifiesta que la solicitud de préstamo es iniciada por el prestatario. La solicitud de préstamo especifica la cantidad y la fecha de pago del préstamo. Las solicitudes de préstamo se envían a través de contratos de préstamo en toda la red y cualquier cuenta de préstamo puede recibir solicitudes de préstamo. Las cuentas interesadas en solicitar un préstamo pueden realizar una evaluación de riesgos basada en la prueba de la garantía del préstamo del prestatario por parte de un fideicomisario público del gobierno y proporcionar el monto del préstamo y la tasa de interés adecuados. Cuando el prestamista acepta la solicitud de préstamo, el contrato inteligente deduce temporalmente el monto del préstamo correspondiente de la cuenta del prestamista. (p. 27). Los tipos de riesgos de un proceso de prestación bancaria se muestran en la Figura 3.

Figura 3

Tipos de riesgo y clasificación de los modelos de riesgo de crédito



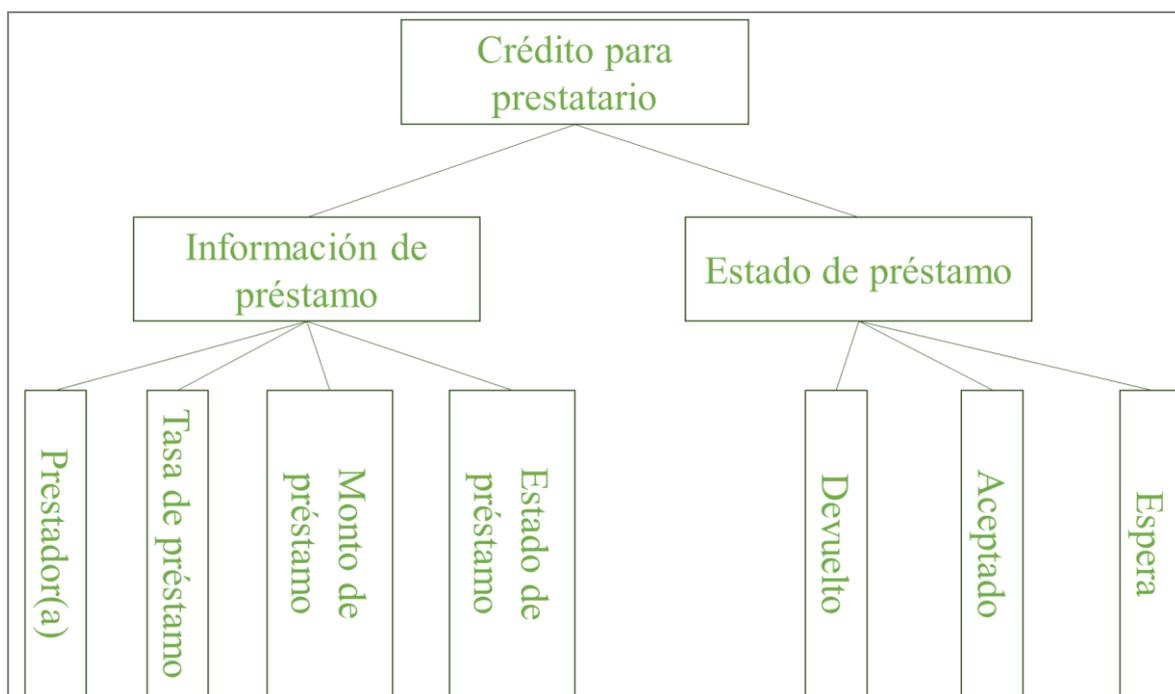
Fuente: (Fenerich et al. 2020)

Por otro lado, (Sen y Rajagopal 2020) señalan que los bancos están comercializando agresivamente los préstamos hipotecarios debido al movimiento a la baja de las tasas de interés, los precios inmobiliarios estables, las concesiones fiscales del gobierno, una vida mejor y la mentalidad positiva de las personas hacia los préstamos. Las iniciativas gubernamentales que se espera que impulsen el crecimiento del sector inmobiliario son la Ley de Bienes Raíces, la Ley de

Adquisición de Tierras y Reasentamiento y los Fideicomisos de Inversión en Bienes Raíces. El avance tecnológico y el acortamiento de los ciclos económicos han potenciado la importancia de los activos intangibles, especialmente el conocimiento, como factor crítico en la toma de decisiones para la industria bancaria. (p. 272). El contenido de un préstamo se muestra en la Figura 4.

Figura 4

Contenido principal de un préstamo



Fuente: (Yang et al. 2018)

Primera dimensión, lo económico: para (ЖЕPHOCEK 2020) un modelo matemático que proporciona a los usuarios números que caracterizan su nivel de riesgo crediticio y les permite comparar con las probabilidades de incumplimiento esperadas de las pequeñas empresas permite extrapolar los siguientes márgenes de tipos de puntuación:

1. Fase 1 - Puntuación de la aplicación:
 - Analizar los datos personales de un prestatario potencial y
 - evaluar los riesgos sin devolución y
 - elegir más formas de cooperación.

2. Fase 2 - Puntaje de conducta:

- Analizar información sobre el comportamiento de los prestatarios y
- evaluar los riesgos sin devolución y
- tomar medidas preventivas tempranas.

3. Fase 3 - Puntos de recogida:

- Elegir los métodos de cobro de deudas más efectivos y
- optimizar el costo de la liquidación de deudas incobrables y
- evaluar el proceso de crecimiento y
- unificar el proceso de toma de decisiones para otorgar crédito en todo el banco y
- disponer de una base de datos electrónica de las solicitudes rechazadas y
- datos históricos accesibles sobre transacciones de crédito y
- que automatice el manejo de solicitudes electrónicas.

A partir de las fases correspondientes a los tipos de puntuación se induce que la dimensión a la que pertenecen es a económico porque todas estas actividades se pueden realizar por un personal del proceso de prestación bancaria, pero con Machine Learning no se necesitaría de personal y por lo tanto los costos se reducirían; pero para que esto sea un hecho se debe medir el indicador que es costo de personal.

Segunda dimensión, la satisfacción: según (Олехнович 2022) podemos distinguir las siguientes razones y posibles formas de atracción cuando los clientes compran servicios financieros:

- Fase 1 - Intereses: el deseo de enriquecerse, de aumentar los activos (se puede expresar, por ejemplo, el deseo de depositar el dinero disponible en un banco con una tasa de interés alta) y
- Fase 2 - Reducción de riesgos: la necesidad de sentirse seguro y confiable, estabilidad garantizada (por ejemplo, comprar diferentes planes de seguro y jubilación) y

- Fase 3 - Concienciación: exploración de medidas relacionadas con la formación de la propia nación, mejora de la fama, imagen (servicios en instituciones financieras líderes) y
- Fase 4 - Conveniencia: el deseo de simplificar sus actividades diarias (el deseo de utilizar un sistema de pago y transferencia bancaria) y
- Fase 5 - Liberad: la necesidad de independencia en todas las áreas de actividad (uso de tarjetas bancarias para el pago).

Dado estas fases por las que el cliente deposita su confianza en una entidad bancaria lo que más destaca son las tasas de intereses porque cuando un cliente se acerca a pedir un préstamo bancario casi siempre pregunta por la tasa de interés que tiene el banco y cuando esta tasa es alta en comparación con otros bancos, suelen buscar mejores alternativas porque les resulta más económico, pero también hay otra fase como es la reducción de riesgos la cual permite saber si el cliente llegara a una etapa en que acumulará moras o se retrasará en sus pagos, muchas veces el cliente ignora la tasa de interés y esto con el tiempo lo perjudica cuando su economía se desestabiliza ya sea por el trabajo o problemas en el domicilio. Estas fases nos permiten inducir que pertenecen a una dimensión de satisfacción ya que finalmente la entidad bancaria presta sus servicios y el cliente deposita su confianza en que podrá pagar, a esto se suma el indicador de satisfacción del cliente, porque si no pudiera pagar o si lo clasificaran como moroso, este cliente manifestaría su insatisfacción con el personal y hasta podría invertir mucho tiempo buscando que los bancos le acepten el préstamo, pero con Machine Learning esto se soluciona puesto que el cliente ya no tendrá que ir por todos los bancos a preguntar y entregar documentos engorrosos que se sabe que tardan en llegar a manos de los analistas y personal que acredita a un cliente de buen pagador, sino que con Machine Learning implementado, el cliente solo tendrá que ingresar sus datos socioeconómicos en el sistema y se le mostrará si es o no buen pagador para un banco y de esta forma puede conocer cuál es su estado económicamente y a partir de los resultados pueda tomar mejores decisiones con efectividad.

Tercera dimensión, el préstamo: según (ЖЕПНОСЕК 2020) La principal fuente de datos son los datos personales del cliente en el momento de la solicitud del préstamo, estos incluyen lo siguiente:

1. Fase 1 - Historial de pagos pasados (registro de crédito del cliente): severidad, relevancia, ausencia de deuda vencida y
2. Fase 2 - Saldo pendiente (registro de crédito del cliente): intereses sobre el uso y
3. Fase 3 - Antigüedad del historial crediticio: antigüedad, cuentas de crédito y
4. Fase 4 - Deseo de obtener nuevo crédito (registro de crédito del cliente): número de solicitudes de crédito recientes y
5. Fase 5 - Tipo de préstamo utilizado (perfil crediticio del cliente): número de préstamos rotativos, subcuentas y
6. Fase 6 - Datos demográficos (perfil de la empresa): cuánto tiempo ha existido la empresa y
7. Fase 7 - Datos financieros (perfil de la empresa): índice de endeudamiento.

Las fases descritas dan por seguro que corresponden a una dimensión de préstamo, en grandes rasgos representan el arjé de las evaluaciones de crédito personal que realiza el personal cuando toma una cantidad de solicitudes de préstamos, las evalúa y determinada cuáles son correctas y cuáles no tienen suficiente información para continuar la evaluación, y esto corresponde con el indicador de eficiencia de evaluación, porque rendir este tipo de evaluaciones requiere que todo el personal planifique realmente cuántas solicitudes de préstamo bancario podrán atender en un día y cuántos fueron los esperados, y lo necesitan hacer constantemente porque el proceso llevado a cabo manualmente está sujeto a errores humanos o variables externas que pueden dificultar realizar la revisión de los clientes. Cuando se conoce si el cliente es o no buen pagador, aparentemente todos los datos proporcionados por el cliente son correctos y ello permitió evaluar la solicitud de préstamo respectiva tomando en cuenta los parámetros socioeconómicos del cliente que hacen referencia a muchos papeles que entrega el cliente llegando a ser un proceso que les toma entre 1 a 2 días para presentar toda la información que satisface cada parámetro que lo contemple la entidad

financiera, pero esto se dificulta cuando hay demoras y en un día la eficiencia puede llegar a ser menor al 50 o 25%, lo que genera problemas en el banco porque también puede ocurrir que se necesita a ese personal para otras funciones o como un rol multifuncional y que puede aparecer de forma espontánea, pero esto con Machine Learning ya no es necesario, porque el cliente evita enviar muchos documentos para cumplir los requisitos de evaluación, sino que puede directamente ingresar sus parámetros socioeconómicos en pantalla y mientras el Machine Learning se ejecute en el sistema, el cliente puede realizar su solicitud mientras los datos se procesan en tiempo real para que el cliente pueda observar en el sistema en solo minutos si es o no buen pagador.

Así mismo, se consideró de utilidad el conocer los indicadores y los datos de pre-prueba promedio de la presente investigación:

Tabla 1

Indicadores de la investigación

Indicador	Datos de pre-prueba (Promedio)	Comentario
Tiempo de espera	6 minutos (Redondear (Fracción decimal (Tiempo real (hh:mm:ss)), 4)	
Costo de personal	3 soles (Redondear (Costo de personal por minuto x Tiempo de espera), 4))	
Satisfacción del cliente	Indeciso (Resultado más frecuente (escala de Likert))	No contrastado indicador cualitativo
Eficiencia de evaluación	60% (Tiempo obtenido/Tiempo objetivo) x 100)	

Fuente: Elaboración propia

En función de lo planteado anteriormente, se puede ver de una forma la situación actual (AS-IS) y la situación propuesta (TO-BE) en la siguiente tabla:

Tabla 2*Comparación del AS – IS y TO – BE*

Situación Actual (AS – IS)	Situación Propuesta (TO – BE)
Tiempos altos de espera en la prestación bancaria	Tiempos bajos de espera en la prestación bancaria
Costos altos de personal en la prestación bancaria	Costos bajos de personal en la prestación bancaria
Satisfacción baja en la prestación bancaria	Satisfacción alta en la prestación bancaria
Eficiencia baja en la prestación bancaria	Eficiencia alta en la prestación bancaria

Fuente: Elaboración propia

Para la investigación se han considerado cuatro (4) indicadores como son: tiempo de espera, costo de personal, satisfacción del cliente y eficiencia de evaluación, a continuación, se detallarán conceptos y fórmulas respectivamente:

Según (Fuentes y Zúñiga 2010), el tiempo de espera en todas las solicitudes a partir de 300 soles son enviadas al departamento de riesgos antes de ser aprobadas por el administrador. Esto tiene por objeto reducir los errores en los documentos que puede generar el personal sin experiencia. El período de aprobación de las solicitudes de préstamo se llega a ampliar de 2 días a 5 días, lo que lo hace menos competitiva a la entidad financiera (p. 94). El tiempo de espera de cada cliente se determina individualmente, luego se redondea al minuto más cercano y se introduce en el formato hh:mm:ss. Para ello se utiliza la siguiente fórmula:

$$\text{Tiempo de espera} = \text{Redondear} (\text{Fracción decimal} (\text{Tiempo real} (\text{hh:mm:ss})), 4)$$

Siguiendo el concepto mencionado anteriormente, podemos entonces especificar que el tiempo total mostrado es el tiempo de espera para que los clientes reciban una atención en función a la evaluación que realiza el personal en un tiempo determinado, siendo el resultado el tiempo de espera por cliente en minutos.

Según (Fuentes y Zúñiga 2010), el costo de personal llega a implicar varias generalizaciones dado que la implementación de un programa de mitigación de riesgos operacionales requiere inversión en actividades tales como la elaboración de perfiles de consultores de negocios, capacitación continua de consultores de negocios y programas de reclutamiento de nuevos consultores (p. 97). Se afirma que prestar atención a la calificación crediticia individual de cada cliente para cada solicitud de préstamo supone un costo de personal por minuto. Para representarlo se utiliza la siguiente fórmula:

$$\text{Costo de personal} = \text{Redondear} (\text{Costo de personal por minuto} \times \text{Tiempo de espera}), 4)$$

De acuerdo con el concepto anterior, el costo de personal en soles (moneda del Perú) se calcula como el costo de personal por minuto que es constante por cada vez que este personal procesa los datos proporcionados por el cliente multiplicado por el tiempo que debe esperar el cliente para que el personal le mencione o le muestre sus resultados de la calificación.

Según (Fuentes y Zúñiga 2010), la satisfacción de los clientes en cada una de las etapas del proceso de crédito de las microfinanzas garantiza el cumplimiento de los objetivos de cada una de ellas, a excepción de las áreas encargadas de la realización de las tareas propias y de la gestión de las aprobaciones de los préstamos, que se encuentran en el área encargada de la administración pero tienen un nivel de gravedad insuficiente y controlan su eficiencia del 20% al 39% (p. 42). Utilizando la pregunta; ¿Qué tanto le gusto el proceso? Empleando la escala

de Likert con los siguientes valores: Me gustó mucho (5), Me gustó (4), Neutral (3), Me disgustó (2), Me disgustó mucho (1)

Satisfacción del Cliente = Resultado más frecuente (escala de Likert)

Siguiendo el concepto mencionado anteriormente, podemos determinar que la satisfacción promedio del cliente después de interactuar con el proceso se mide como el resultado más frecuente por cada cliente basado en la escala de Likert al culminar el proceso.

Según (Fuentes y Zúñiga 2010), la eficiencia del control se mide en términos de eficiencia de la evaluación, y cada nivel de eficiencia recibe una puntuación y una escala. Cuando alcanza una puntuación de utilidad superior al 80%, recibe una calificación de excelente (5 puntos), y cuando cae por debajo del 20%, recibe una calificación de deficiente (1 punto). Cuando ya se tienen los datos del cliente y el empleado está procesando una respuesta para este cliente, y con ello se verifica visualmente el tiempo previsto entre el tiempo real para lograr una eficiencia de evaluación. Esto se representa con la siguiente fórmula:

$$\text{Eficiencia de Evaluación} = \left(\frac{\text{Tiempo obtenido}}{\text{Tiempo objetivo}} \right) \times 100$$

Siguiendo el concepto anterior, podemos especificar que los minutos estimados que se muestran se basa en un promedio de espera y que es constante, lo cual es utilizado para establecer una expectativa de tiempo que no debe ser excedida, de ser superada, se generaría retrasos en el proceso, en tanto, si los minutos reales que provienen de una realidad con el tiempo de espera son menores a los minutos reales, habría una eficiencia positiva para el proceso en cuestión, pero si los minutos reales son más altos que los minutos estimados, la eficiencia de

evaluación bajaría de su porcentaje, es así, que los mencionados parámetros del producto de la división son multiplicados 100 para obtener un valor porcentual.

Con respecto al marco teórico, la variable Machine Learning se fundamentó de los siguientes conceptos:

Según (Cardin 2021), Machine Learning es un programa de inteligencia artificial (IA) que proporciona al marco la capacidad de aprender y trabajar con aptitud sin ser cambiado explícita y automáticamente. Muchos utilizan Machine Learning para abordar el progreso de la vida. Implica técnicas que son a la vez competentes pero modernas. Las técnicas de aprendizaje comienzan con los siguientes datos o conocimientos: modelo empírico rápido o entrenamiento para tomar mejores decisiones más adelante, proporcionar y buscar planos de información. A esto se espera que las computadoras puedan adaptarse y comprender sus actividades sin intervención humana.

Para (Taddy 2019), Machine Learning es el campo de pensamiento sobre cómo crear automáticamente predicciones sólidas a partir de datos complejos. Esto está estrechamente relacionado con las estadísticas modernas y, de hecho, muchas de las mejores ideas para el aprendizaje automático provienen de estadísticos (lazos, árboles, bosques, etc.). La gama completa de ML se ajusta utilizando experimentos fuera de la muestra para probar la efectividad del modelo entrenado en el conjunto de datos para predecir nuevos datos. También hay esfuerzos recientes para aumentar la transparencia del aprendizaje automático, pero los practicantes sabios evitan dar implicaciones estructurales a los parámetros del modelo ajustado.

(Sosnovshchenko, Oleksandr Baiev 2018) entienden al Machine Learning como un subcampo de la inteligencia artificial (IA) que ha logrado grandes avances en la última década y sigue siendo un tema de investigación candente. Esta es un área de conocimiento sobre la construcción de algoritmos que pueden aprender de los datos y mejorar las tareas que realizan. El aprendizaje automático permite que las computadoras deriven algoritmos para tareas y extraigan patrones ocultos de los datos. El aprendizaje automático se conoce con diferentes nombres en

diferentes comunidades de investigación, incluido el análisis predictivo, la extracción de datos, el aprendizaje estadístico y el reconocimiento de patrones. Ergo, se puede argumentar que existen diferencias sutiles entre estos términos, pero básicamente todos se superponen siempre que sea posible usarlos (algoritmos) indistintamente (p. 10). Las capacidades del aprendizaje automático se muestran en la Tabla 3.

Tabla 3

Statistical Modeling vs. Machine Learning

Propiedad	Inferencia estadística	Machine learning supervisado
Objetivo	Modelos causales con poder explicativo	Rendimiento de la predicción, a menudo con un poder explicativo limitado
Datos	Los datos son generados por un modelo.	El proceso de generación de datos es desconocido.
Marco de trabajo	Probabilístico	Algorítmica y Probabilística
Expresividad	Típicamente lineal	no lineal
Modelo de selección	Basado en criterios de información	Optimización numérica
Escalabilidad	Limitado a datos de menor dimensión	Escala a datos de entrada de alta dimensión
Robustes	Propenso al ajuste excesivo	Diseñado para un rendimiento fuera de la muestra
Diagnóstico	Extensivo	Limitado

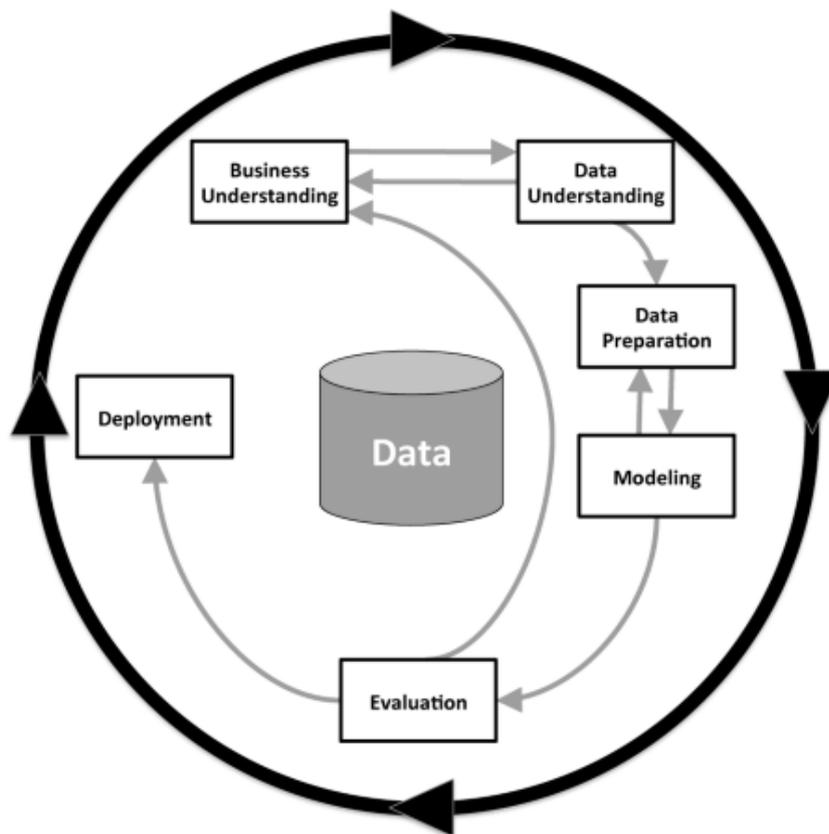
Fuente: (Dixon, Halperin y Bilokon 2020)

Para la investigación se han considerado tres (3) tipos de metodologías como son: CRISP-DM, SEMMA, KDD, a continuación, se detallarán los siguientes conceptos:

Según (Pérez López 2021) la metodología CRISP-DM se describe como modelo de proceso jerárquico que consiste en una serie de tareas descritas en cuatro niveles de abstracción (de general a específico): fases de tareas generales, tareas especiales e instancias de procesos. En la metodología CRISP-DM (Cross-Industry Standard Process for Data Mining) el orden de las fases no es exacto. El ciclo de vida de un proyecto de minería de datos consta de seis fases, ver Figura 5. El resultado de cada fase determina la siguiente fase o la tarea específica de la fase que debe realizarse. Las flechas muestran las dependencias más importantes y comunes entre fases.

Figura 5

Modelo de ciclo de vida de CRISP-DM



Fuente: (Kelleher, Mac y Arcy 2020)

El círculo exterior en la figura anterior representa la naturaleza periódica de la extracción de datos. La minería de datos hace más que simplemente implementar

una solución, de hecho, la información oculta (lecciones aprendidas), ver Tabla 4, durante el proceso y las soluciones brindadas a menudo pueden generar nuevas preguntas que están más enfocadas en el negocio.

Tabla 4

Descripciones del modelo de proceso CRISP-DM

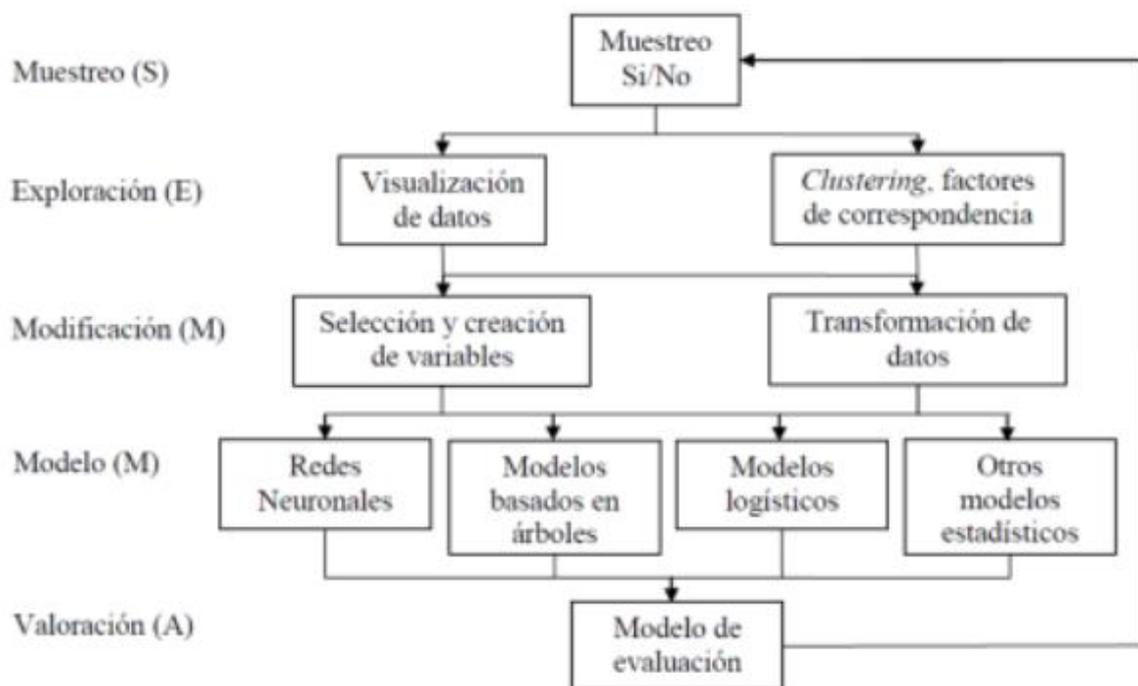
Fase	Breve descripción
Comprensión de la empresa	<p>Para obtener una descripción general de los recursos disponibles y necesarios, debe evaluar la salud de su negocio. Esto implica definir el propósito de la minería de datos, que es uno de los aspectos más importantes de esta fase. En esta fase, debemos explicar los patrones de minería de datos (clasificación, etc.) y los criterios de éxito de la minería de datos (precisión, etc.), y luego se debe crear un plan de proyecto vinculante.</p>
Comprensión de los datos	<p>Recopilar, investigar, explicar y verificar la calidad de los datos de las fuentes de datos es una tarea importante durante esta fase. Especifica que el manual del usuario describe la tarea de describir los datos a través del análisis estadístico. Como resultado, se determinan sus atributos y relaciones.</p>
Preparación de los datos	<p>La preparación de datos debe hacerse definiendo criterios de inclusión y exclusión para esta etapa. La degradación de la calidad de los datos se puede solucionar mediante la limpieza de datos según el modelo utilizado (definido en el primer paso), por lo que es necesario establecer un atributo obtenido de todos los datos en este paso antes de poder hacer algunas formas dependientes del modelo.</p>
Modelado	<p>La fase de modelado de datos selecciona tecnologías de modelado para crear casos de prueba y modelos. Como regla general, puede utilizar cualquier tecnología de minería de datos. Su elección depende de cuestiones operativas y datos. Lo más importante es explicar la decisión de construir el modelo después de establecer ciertos parámetros.</p>
Evaluación	<p>Para evaluar el modelo, es necesario analizarlo desde la perspectiva de los criterios de evaluación y de allí seleccionar el mejor modelo. Por lo tanto, se debe definir una nueva acción para interpretar los resultados, y se debe probar el proceso desde el principio para obtener mejores resultados.</p>
Despliegue	<p>La fase de despliegue generalmente se enumera en el manual del usuario, que puede ser el informe final o el componente de software. El manual del usuario explica que la fase de implementación incluye la planificación, la supervisión y el mantenimiento de la implementación.</p>

Fuente: (Solano et al. 2022)

Con respecto a la metodología SEMMA, para (Vicente Vírseda, Juan Antonio Gonzáles Arias y Parra Rodríguez, Francisco Beltrán Pascual 2019), en el mundo empresarial, cada empresa suele elegir e implementar una de una metodología particular, que será de aplicación universal a todos los proyectos de minería de datos. Además de la metodología CRISP-DM, también existe el conocido SEMMA, las diferentes fases que lo caracterizan, así como sus definiciones específicas, que pueden ser más apropiadas dependiendo del proyecto en particular que se quiera emprender, ver Figura 6.

Figura 6

Fases del proceso SEMMA



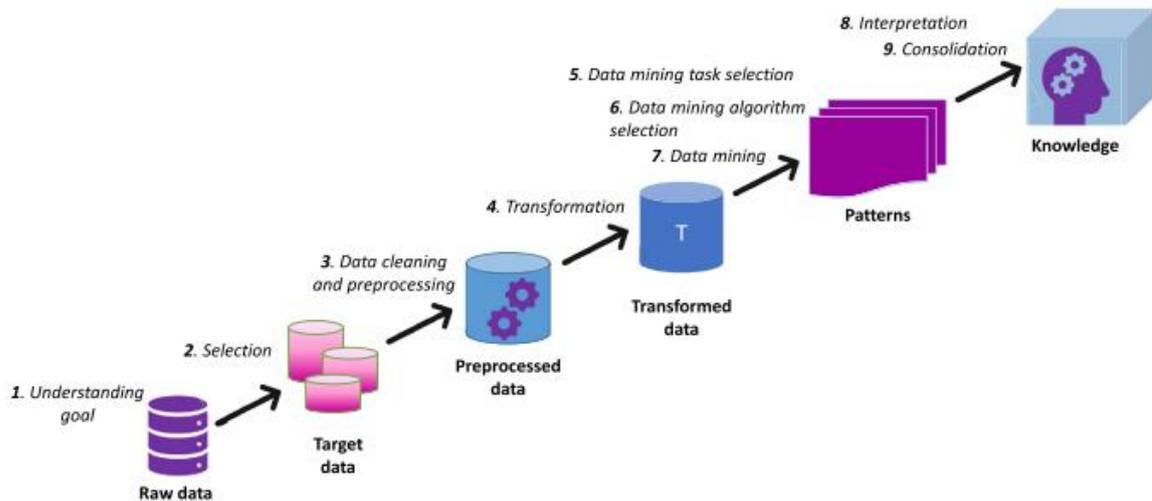
Fuente: (Vicente Vírseda, Juan Antonio Gonzáles Arias y Parra Rodríguez, Francisco Beltrán Pascual 2019)

En relación a la metodología KDD, (Petr y Fathalhrman 2020), señalan que ofrece un proceso de nueve pasos, visto principalmente como una metodología basada en la investigación. Esto implica tanto la evaluación como la interpretación de modelos (posiblemente conocimiento) y la selección del preprocesamiento, muestreo y pronóstico de datos antes de la etapa de extracción de datos. Mientras

que algunas de estas nueve fases se centran en la toma de decisiones o el análisis, otras son transformaciones de datos en la cadena de datos-información-conocimiento. Como se mencionó, KDD es un "proceso no trivial de identificación de modelos de datos válidos, novedosos, potencialmente útiles y, en última instancia, comprensibles" (Fayyad et al., 1996). La descripción del proceso KDD también proporciona una descripción general de los pasos de este proceso, disponible en la Figura 7.

Figura 7

Modelo de ciclo de vida de KDD



Fuente: (Petr y Fathalhrman 2020)

De las metodologías compartidas hay una particularidad en el modelamiento de los datos, de cara a que en el caso de basarse en un proceso que puede ser observable y experimentable, la metodología CRISP-DM detalla cómo aplicarla específicamente en términos de Machine Learning, en realidad a un desarrollo de Deep Learning es recomendable que sea aplicado con KDD; pero en la evaluación va a ser diferente si tratamos con datos socioeconómicos donde hay ciertas preocupaciones como no acceder a ciertos datos por protección o seguridad, y si no se tiene acceso a estos datos va a afectar a las otras etapas, ver Tabla 5.

Tabla 5*Fortalezas y limitaciones entre las metodologías*

Metodología	Fortalezas	Limitaciones
CRISP-DM	<ul style="list-style-type: none"> • Proceso de identificación clara, anticipándose a los posibles problemas. • Fase de finalización para transformar el problema en términos de datos. • Rentabilidad de grandes proyectos. • Admite múltiples técnicas de minería de datos. • Repite el procedimiento. • Posee documentación con casos típicos de uso que ayudan en este proceso. 	<ul style="list-style-type: none"> • Proceso largo y complicado. • Las fases de preparación de datos y modelado de flujo de datos difieren de la minería de datos estática tradicional debido a su naturaleza temporal.
KDD	<ul style="list-style-type: none"> • El proceso es repetitivo. • Admite varias técnicas de DM, como redes neuronales. 	<ul style="list-style-type: none"> • Requiere conocimientos previos en DM.
SEMMA	<ul style="list-style-type: none"> • Buena atención al usuario. • Admite varias técnicas de DM. • Repite el procedimiento. 	<ul style="list-style-type: none"> • Está diseñado para trabajar con el motor SAS® Enterprise Miner™, pero surgen problemas si ocurre un caso atípico. • No tiene en cuenta la comprensión de un problema por parte de una empresa.

Fuente: (Dåderman y Rosander 2018)

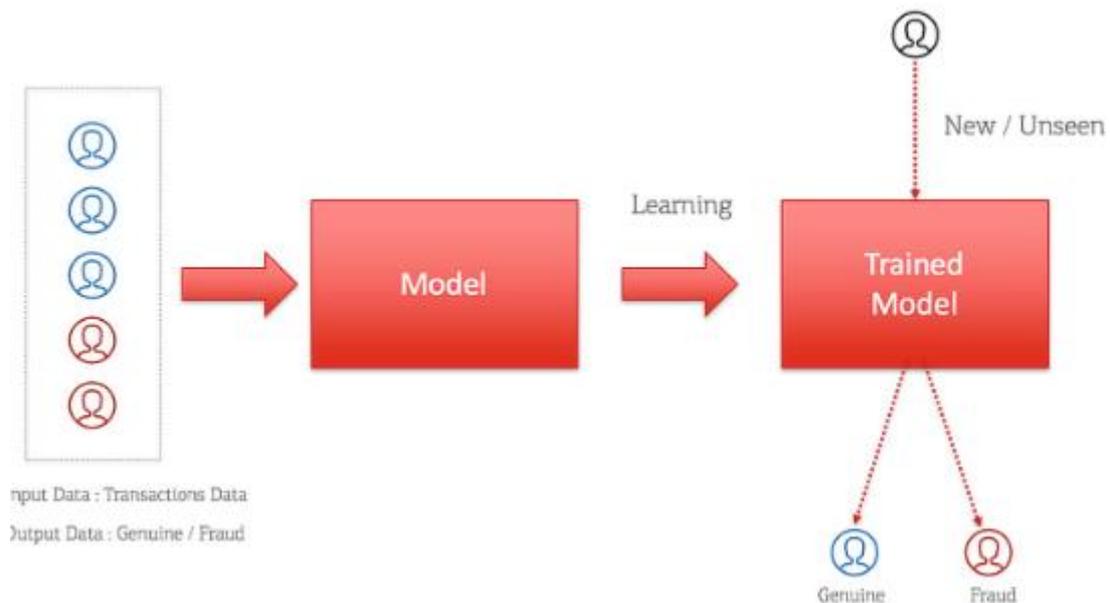
Para la investigación se ha considerado la infraestructura que es para Machine Learning, a continuación, se detallan los siguientes conceptos:

Según (Singh 2020), Machine Learning supervisado es una categoría importante del Machine Learning que genera valor para muchas aplicaciones y negocios. En este tipo de entrenamiento, el modelo se entrena con datos que ya tienen las etiquetas o salidas adecuadas. En resumen, como se muestra en la

Figura 8, trata de representar la relación entre los datos de entrada y salida de una manera que sea bien generalizada para los datos invisibles. El entrenamiento del modelo se realiza comparando la salida real con la salida esperada y optimizando la función para reducir el error general entre la salida real y la esperada.

Figura 8

Generalización del Machine Learning supervisado



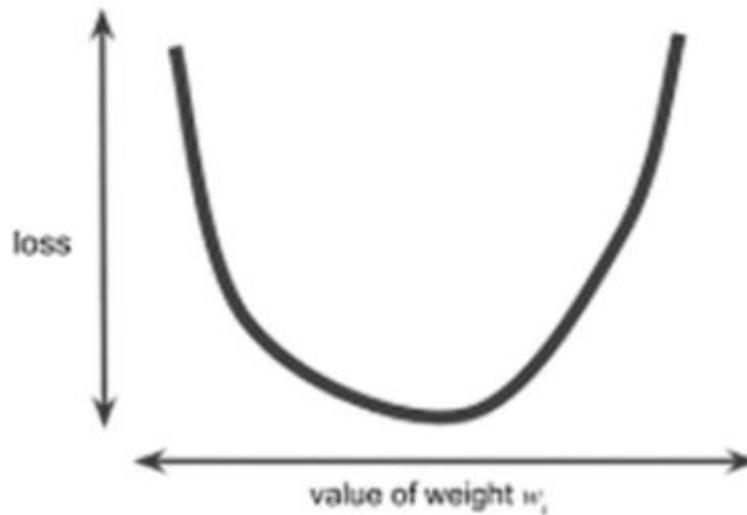
Fuente: (Singh 2020)

Menciona a su vez, (Singh 2020), que después de todo, los modelos de Machine Learning son tan buenos como las pérdidas predecibles. Existen diferentes tipos de funciones de pérdida que pertenecen a una categoría particular de problemas y, en la mayoría de los casos, una tarea común de clasificación o regresión intenta minimizar el error cuadrático medio y registrar pérdidas durante el entrenamiento y la validación cruzada. Si se imagina la pérdida como una curva, como se muestra en la Figura 9, el método de descenso más pronunciado lo ayuda a alcanzar el punto donde el valor de la pérdida alcanza su valor mínimo. A esto se comienza con un punto aleatorio basado en los pesos o parámetros iniciales del modelo y se va moviendo en la dirección en la que el modelo comienza a encogerse. Una cosa a tener en cuenta es que el método de descenso más

pronunciado toma grandes pasos a medida que se aleja del mínimo real, pero cuando alcanza valores cercanos, el tamaño del paso se vuelve muy pequeño para que no se pierda el mínimo.

Figura 9

Descenso de gradiente



Fuente: (Singh 2020)

Según (Kolodiazhnyi 2020), el formato más común para presentar datos estructurados se llama CSV. Este formato no es diferente a un archivo de texto con una matriz bidimensional, donde los valores de una línea están separados por comas y las líneas se colocan en cada nueva línea, ver la Figura 10:

Figura 10

Ejemplo de formato CSV

```
1, 2, 3, 4  
5, 6, 7, 8  
9, 10, 11, 12
```

Fuente: (Kolodiazhnyi 2020)

Por otra parte, (Kneusel 2021) menciona que la ventaja de este formato de archivo es su estructura simple, puede ser procesado por muchas herramientas de software, es legible por humanos y es compatible con muchas plataformas informáticas diferentes. La desventaja es la falta de compatibilidad con datos estructurados complejos y multidimensionales, análisis lento en comparación con los formatos binarios. Asimismo, para (Kneusel 2021), Python, en su explicación más simple, es una declaración secuencial agrupada en bloques por sangría. Estructuras de datos como números, cadenas, tuplas, listas, diccionarios, etc. Una estructura de control que contiene bloques iflife, bucles for, bucles while, sentencias y bloques tryexcept. Una función con funciones anidadas opcionales. Una gran biblioteca de módulos importables. Vamos a explicar cada una de estas funciones.

Igualmente para (Kneusel 2021), todos los modelos tradicionales de Machine Learning se encuentran en la gran biblioteca Scikit-learn, sklearn, que a menudo se llama cuando se carga en Python y no se debe capitalizar como scikit-learn. Esto se debe a que es una forma de que los autores hagan referencia a scikit-learn en su documentación. Esta biblioteca utiliza matrices NumPy que permite implementar una interfaz estandarizada para muchos modelos diferentes de aprendizaje automático y muchas otras características que no tocamos. A medida que se familiarice con el aprendizaje automático y las herramientas que lo respaldan, se recomienda leer la documentación oficial de scikit-learn.

Por otro lugar, en una programación de alto nivel, las decisiones son la mejor manera de analizar datos en columnas específicas (resultados) para determinar la estructura que reduce los datos a una estructura de árbol de decisión. Para crear un árbol, se verifica los valores de datos de los demás para encontrar el valor de la columna, que se puede dividir mejor en dos grupos individuales. Los datos con cientos de filas o decenas de columnas tardan más en encontrar los mejores resultados, pero para los árboles de decisión no es problema, puesto que también pueden manejar datos más grandes (Clinton 2019).

Frente a lo expuesto, los árboles de decisión se utilizan en la minería de datos para descubrir patrones de información en los datos. Una vez creados, los árboles de decisión se pueden usar para predecir el resultado de nuevos datos usando patrones observados en los datos a partir de los cuales se creó el árbol. De esta forma, un árbol de decisión puede considerarse como una estructura de datos para almacenar experiencia (Clinton 2019).

Figura 11

Patrón de un árbol de decisión

	1	2	3	4	5	6	
1	2	3	4	5	6	7	
2	3	4	5	6	7	8	
3	4	5	6	7	8	9	
4	5	6	7	8	9	10	
5	6	7	8	9	10	11	
6	7	8	9	10	11	12	

Fuente: (Clinton 2019)

En consecuencia, los árboles de decisión son muy buenos para encontrar patrones que conduzcan a resultados específicos en los datos originales. Finalmente, esta fortaleza es también su debilidad, ya que los datos sesgados por muchas muestras de un tipo particular pueden dar como resultado un árbol desequilibrado que favorece a la mayoría y dispersa a la minoría (Clinton 2019).

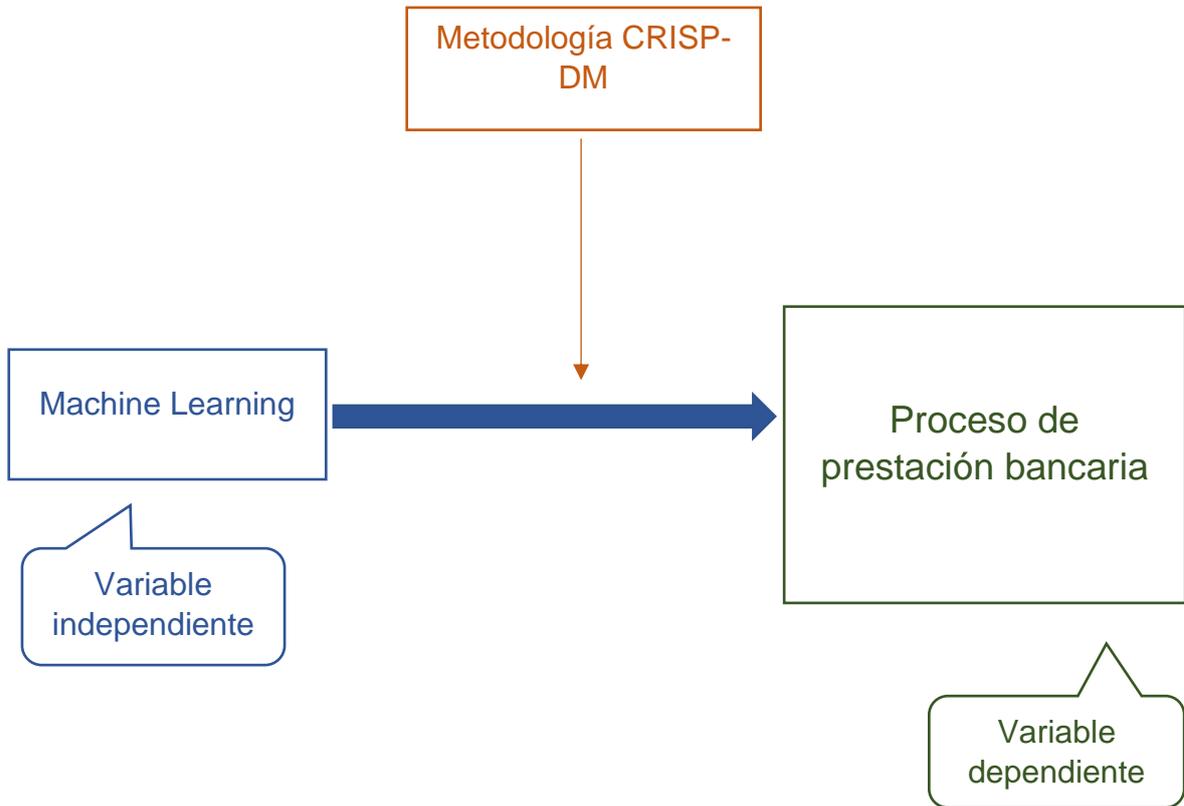
III. METODOLOGÍA

3.1. Tipo y diseño de investigación

3.1.1. Tipo de investigación

Figura 12

Metodología asociada a un problema organizacional



Indicadores	Pre prueba	Post prueba
Tiempo de Espera	6 minutos	2 minutos
Costo de Personal	3 soles	1 sol
Satisfacción del Cliente	Indeciso	Totalmente de acuerdo
Eficiencia de Evaluación	60%	20%

Fuente: Elaboración propia

Esta investigación se enfocó a un nivel aplicado, pues sólo se implementó la aplicación de Machine Learning para la solución del problema principal, en su caso el Proceso de Prestación Bancaria; se creó una guía y documento para aplicar la resolución de problemas del proceso de prestación de servicios bancarios en un banco (González Castellano, Yll Lavín y Curiel Lorenzo 2003).

La investigación presentó un enfoque cuantitativo dado que los hechos se recogieron y analizaron mediante conceptos y variables cuantificables, donde la recogida de información fundamental y la elección del modelo más adecuado fueron lo que permitió interpretar la realidad de forma más objetiva (Neill y Cortez Suárez 2018). Con base en el autor en esta investigación se recogieron los diseños para el análisis con certeza de las hipótesis formuladas.

3.1.2. Diseño de investigación

Se utilizó un diseño de preprueba/posprueba con grupo experimental (pre experimental). El procedimiento para la manipulación de la variable se presentó así:

Figura 13

Modelo de diseño de investigación



Fuente: Elaboración propia

Se obtuvieron los datos de pre Prueba (O₁) y luego de implementar el estímulo (X) se obtuvieron los datos post Prueba (O₂). En su caso, si los datos de O₂ son mejores a los datos de O₁, las hipótesis planteadas serán aceptadas.

En referencia a cómo se validaron las hipótesis se tomó en cuenta el silogismo hipotético donde si las hipótesis alternas se aceptan de las específicas (Beuchot 2018), la general se acepta a partir de las pruebas realizadas a los datos que se redondearon a 4 decimales para resultados más significativos (Casares 2021, p. 18).

3.2. Variables y operacionalización

Variable independiente: Machine Learning

Definición conceptual: Según (Taddy 2019), Machine Learning es el campo de pensamiento sobre cómo crear automáticamente predicciones sólidas a partir de datos complejos. Esto está estrechamente relacionado con las estadísticas modernas y, de hecho, muchas de las mejores ideas para el aprendizaje automático provienen de estadísticos (regresión, árboles, clasificación, etc.).

Esta variable no se operacionalizó porque no fue objeto de análisis. De ahí que, para esta variable se elaboró un sistema web como sistema de conocimiento efectivo con Machine learning utilizado para soporte del modelo predictivo, automatizado para los procesos de software en el banco al punto de hacer una evaluación del cliente más eficaz mediante un proceso de machine learning. Dicho sistema fue aplicado al grupo experimental con el objetivo de mejorar el proceso de prestación bancaria de esta entidad financiera.

Variable dependiente: Proceso de prestación bancaria

Definición conceptual: Según (Meikang 2018), la solicitud de préstamo es iniciada por el prestatario. La solicitud de préstamo especifica la cantidad y la fecha de pago del préstamo. Las solicitudes de préstamo se envían a través de contratos de préstamo en toda la red y cualquier cuenta de préstamo puede realizar solicitudes de préstamo.

Definición operacional: La variable proceso de prestación bancaria está constituida por cuatro dimensiones y cuatro indicadores: tiempo (tiempo de espera), económico (costo de personal), satisfacción (satisfacción del cliente) y préstamo (eficiencia de evaluación), asimismo se presenta 30 mediciones por indicador, que serán evaluados por una prueba, cuya puntuación por indicador depende del índice de valores al ser realizado en una ficha de observación, ver anexo 1. Para lo suyo se llevaron a cabo las siguientes etapas 1) Entendimiento de los datos; 2) Selección de los datos; 3).

Una vez identificadas las variables, la conceptualización para Machine Learning no aplica, mientras que, para el Proceso de Prestación Bancaria, cada uno de sus indicadores fue definido en la tabla 6:

Tabla 6

Definición conceptual por indicadores de la variable dependiente

Indicador	Definición conceptual	Escala de medición
Tiempo de espera	El tiempo de espera no se trata del tiempo en que el personal está registrando o esperando al cliente en la cola, sino, se refiere a que el cliente ya proporcionó los datos al personal y este personal está evaluando para entregarle una respuesta al cliente.	Razón
Costo de personal	Es el costo que se tiene del analista y el administrador por sus servicios prestamos correspondientes a los minutos cedidos a un cliente determinado que se encuentra en tiempo de espera.	Razón
Satisfacción del cliente	Es la que permite conocer la experiencia del cliente en el proceso de prestación bancaria.	Ordinal
Eficiencia de evaluación	Es el porcentaje que estima lo bueno que fue el tiempo de espera en base a una expectativa contra una realidad.	Razón

Fuente: Elaboración propia

3.3. Población, muestra y muestreo

3.3.1. Población

La población en este estudio la conformaron todos los procesos (transacciones) de prestación bancaria en una entidad financiera formalizada con flujo de caja a nivel nacional. Debido a que no se pueden contar los procesos antes mencionados (Pande 2004, p. 135-136), se tuvo que N (Población) fue indeterminada.

- **Criterios de inclusión:** Las personas que se tomaron en cuenta son los que les interesa conocer su diagnóstico para aplicar a un crédito, que tenga acceso a internet y tenga un dispositivo móvil o de escritorio.

- **Criterios de exclusión:** Las personas que no se tomaron en cuenta son menores de 18, no tienen internet y no tienen un dispositivo móvil o de mesa.

3.3.2. Muestra

La muestra utilizada en este estudio consistió en 30 procesos (transacciones) de Prestación Bancaria en una entidad financiera formalizada con flujo de caja a nivel nacional (Pande 2004, p. 135-136)

3.3.3. Muestreo

La investigación estuvo enfocada a una muestra no probabilística porque se basó en el criterio de realizar el reconocimiento de los datos y el conocimiento que tuvo el investigador sobre la población (Hernández Collado y Baptista Lucio 2014).

3.3.4. Unidad de análisis

La unidad de análisis resulta de la descomposición de la variable objeto de estudio (proceso de prestación bancaria) y el conjunto de índices de esta unidad de análisis en una serie de conceptos observables (Dihigo 2021, p. 83) como en la tabla 7:

Tabla 7

Unidad de análisis por indicador

INDICADORES	ÍNDICES	UNIDAD DE OBSERVACIÓN
Tiempo de espera.	[0... 13]	Tiempo
Costo de personal.	[0...7]	Tiempo y Costo de personal por minuto
Satisfacción del cliente.	[Totalmente de acuerdo; De acuerdo; Indeciso; En desacuerdo; Totalmente en desacuerdo]	Satisfacción
Eficiencia de evaluación.	[0...100]	Expectativa y realidad de los tiempos

Fuente: Elaboración propia

3.4. Técnicas e instrumentos de recolección de datos

La técnica para esta investigación es aquella de campo que permite estudiar al fenómeno desde su propio ambiente, asimismo, requiere gran capacidad para

interpretar toda la información recopilada en el campo de investigación (Hernández Collado y Baptista Lucio 2014). La técnica e instrumento se menciona a continuación:

Tabla 8

Técnicas e instrumentos de recolección de datos

Técnica	Instrumento
Observación indirecta (campo):	
<ul style="list-style-type: none"> • Consulta a base de datos 	<ul style="list-style-type: none"> • Ficha de observación

Fuente: Elaboración propia

La observación directa como su propio nombre lo dice, permite que la muestra sea observada por sus características y comportamientos. En este caso, se utilizará una base de datos para hacer una selección aleatoria de los 30 clientes que conforma la técnica y después de obtener los 30 clientes se evaluará cada uno de los 30 procesos (transacciones) en una ficha de observación (Pande 2004).

3.5. Procedimientos

En relación con el modo de recolección de información de los datos Pre Prueba, se pidió los datos al Supervisor de Help Desk, Jesús Martín Ronceros Salhuana del banco Alfin del Perú S.A, que facilitó los datos filtrados del 01 de junio del 2022, ver Anexo 13. Para la pre Prueba se realizaron 5 etapas que a continuación se señalan:

Etapa 1 (Entendimiento de los datos). Los datos facilitados indicaban un número de clientes por sucursal para diferentes procesos, dentro de lo cual, en Ventanilla, por ejemplo, para acceder a los datos correspondientes al proceso de prestación bancaria, con la confirmación del Supervisor de Help Desk, se debían filtrar los datos por sucursal y filial para que se mostrará la información de los clientes que hicieron trámites para préstamos, ya que, otros códigos de filial se utilizaban para otros procesos, como retirar todo el dinero o pagar con cuenta, en función de la sucursal seleccionada.

Etapa 2 (Selección de los datos). Con los clientes visibles, se procedió a contar por cada sucursal los resultados, de modo que por sucursal se obtuvo N cantidad de clientes y todas estas cantidades fueron tabuladas en un Excel (ver anexo 13) para extraer exactamente 30 filas, es decir, 30 diferentes cantidades.

Etapa 3 (Normalización de los datos). Con dicha cantidad se trabajó por indicador con variables de apoyo como promedio de atención neto, tiempo total de horas laborales, e interviniendo en los cálculos con cantidad de clientes que se obtuvo en la etapa anterior, luego el tiempo total de atención y el total de todos los tiempos de espera para obtener un resultado de tiempo de espera por sucursal, pero dado que se necesitaban los datos por cliente, la figura pasó de tener una cantidad de clientes a no tenerla y en su reemplazo se agregó un criterio para este indicador (tiempo de espera) que mostrar el tiempo en su expresión a formato [Horas, minutos y segundos], de modo que se complementaba con el tiempo de espera obtenido que estaba en decimales en una primera instancia.

Luego, tras evaluar este tiempo, se pudo trabajar de forma armónica el costo de personal, que está en función del sueldo que gana el Analista y el Administrador que operan en el proceso de prestación bancaria de forma manual de unos 133 soles por día, a reducirse a un sueldo que ganan dicho personal por minuto, de modo que el tiempo obtenido en formato de minutos, se multiplicaba con este costo por minuto de personal para que su producto sea el resultado del costo de personal. Por otro lado, el indicador eficiencia de evaluación se trabajó en función de los datos obtenidos en la primera instancia de la presente etapa, para que se le reste a los tiempos estimados (expectativa) los tiempos reales (realidad) obtenidos del tiempo de espera.

Finalmente, una ficha de observación se aplicó para conocer que tanto le gustó proceso a los clientes, pero a salidas del banco, no concretamente dentro, ya que a estos encuestados se les explicó que se quería conocer su expectativa o experiencia del proceso de prestación bancaria no manifestándole a la persona que tuvo en adelante atendiéndola ya que el banco Alfin no realiza preguntas de satisfacción al cliente y por ello se realizó a puertas del banco de manera que las respuestas de estos clientes dependan de una interacción con el banco en momentos recientes.

En relación con el modo de recolección de información se elaboró la programación para el estímulo (Post Prueba), la cual consta de 5 etapas que a continuación se presentan:

Etapa 1 (Capacitación). Se empieza solicitando el apoyo del grupo de personas indicando que tomará 10 minutos por capacitación (incluye el uso del sistema y la encuesta de una pregunta a la persona) de 30 personas de diferentes familias a ser presentada a nivel gerencial al banco Alfin. Se espera que tengamos 3 minutos para capacitarlo sobre lo referente al sistema. Se espera que tome menos de 5 minutos que la persona termine de usar el sistema hasta obtener sus resultados que es el objetivo del sistema demo. Se espera que tarde 1 minuto en tomarle la encuesta a la persona. Un minuto se reserva para la despedida y una foto para la evidencia. Si son 10 minutos por capacitación puede tomar 5 minutos continuar con otra, dando lugar a que esto tomaría 7.5 horas aproximadamente.

Etapa 2 (Monitoreo). Si son 30 personas y las tratamos individualmente, perfectamente podemos tomar las medidas que necesitamos para el llenado de la ficha de observación de la investigación, con lo cual, se tomaría concretamente el tiempo desde que inicia la persona a llenar sus datos hasta que se le muestren sus resultados. Este tiempo se registrará en la ficha de observación, con cual se hará uso de los 5 minutos máximo para este fin, dando a lugar que esto tomaría alrededor de 2.5 horas. Sumado al minuto que se tomaría para la encuesta que es también requerido para el llenado en la ficha de observación, dando a lugar a otros 30 minutos, totalizando 3 horas que tomaría completar esta fase.

Etapa 3 (Acompañamiento). Al ser realizado por zoom el investigador comparte su pantalla por 3 minutos en lo referente a capacitar a la persona y los 5 minutos siguientes esta persona comparte pantalla hasta que se muestren sus resultados del sistema demo, esta fase tomaría 4 horas.

Etapa 4 (Retroalimentación). En el penúltimo minuto (minuto 9) se tomaría la encuesta para medir su satisfacción por el proceso, esta fase tomaría 30 minutos

Etapa 5 (Normalización). Con las medidas del tiempo y la satisfacción, nos resta obtener las medidas de otros indicadores, para estos dos es necesario el tiempo que ya obtuvimos, con lo cual podemos normalizar para dar por completado las fichas de observación posttest (véase Anexo 21).

3.6. Método de análisis de datos

Estadística descriptiva: Se utilizaron estadísticas descriptivas para caracterizar las cantidades y los porcentajes de los resultados producidos en todo el proceso bancario, así como para cada uno de sus indicadores, y para mostrar los datos del grupo experimental. Para lo suyo se generaron las tablas de los indicadores antes y después, asimismo, se introdujeron los datos de la satisfacción en gráfico circular.

Estadística inferencial: Se utilizó la prueba de Anderson-Darling para establecer la normalidad del conjunto de datos, ya que está adecuadamente calibrada para las pruebas que ejecutaron en el programa Minitab. Dado que los resultados indicaron que una de las distribuciones de los indicadores no era normalizada, se siguió con la utilización de pruebas no paramétricas para muestras independientes como la prueba U de Mann-Whitney, lo que permitió comprobar la hipótesis de investigación, mientras que la prueba t de Student se utilizó para evidenciar las distribuciones de los indicadores que mostraban un comportamiento normal, en consecuencia se comprobaron las hipótesis cuantitativas. La herramienta estadística Minitab 2.0 fue empleada para procesar los datos.

3.7. Aspectos éticos

Es un objetivo que se cumplan los aspectos éticos antes de publicarse el trabajo de investigación para lo cual se toma en cuenta la referencia en la RCUN 760-2007 en 2021 por la Universidad César Vallejo en el artículo N° 15 cuya propiedad intelectual fue parte característica de la investigación. Se promovió de igual manera, la realización del examen de certificación de Conducta Responsable en Investigación, la cual corresponde el autor de la presente investigación en el portal CTI Vitae en conformidad con el plan curricular del investigador de la carrera profesional. Dentro de 5 o 10 años, la carta de presentación será esta investigación que procede de elevados fines que tiene la profesión con el Colegio de Ingenieros del Perú.

IV. RESULTADOS

A continuación, se mostraron las medidas de los indicadores para la preprueba y postprueba:

Tabla 9

Resultados obtenidos de la preprueba y postprueba

N°	I 1: Tiempo de espera		I 2: Costo de personal		I 3: Satisfacción del cliente		I 4: Eficiencia de evaluación	
	Preprueba	Postprueba	Preprueba	Postprueba	Preprueba	Postprueba	Preprueba	Postprueba
1	8.6207	2.1833	2.7586	0.6987	5	4	86.2070	21.8330
2	1.7391	1.4667	0.5565	0.4693	4	3	17.3910	14.6670
3	2.2727	1.2833	0.7273	0.4107	4	5	22.7270	12.8330
4	4.2105	2.9000	1.3474	0.9280	1	5	42.1050	29.0000
5	9.2857	2.2667	2.9714	0.7253	3	4	92.8570	22.6670
6	2.5581	2.5667	0.8186	0.8213	4	4	25.5810	25.6670
7	11.6000	2.0167	3.7120	0.6453	1	3	116.0000	20.1670
8	7.4194	2.7167	2.3742	0.8693	3	4	74.1940	27.1670
9	2.2727	2.9500	0.7273	0.9440	5	5	22.7270	29.5000
10	13.4783	1.9000	4.3131	0.6080	1	4	134.7830	19.0000
11	2.2727	1.5833	0.7273	0.5067	4	5	22.7270	15.8330
12	6.3636	1.0333	2.0364	0.3307	2	5	63.6360	10.3330
13	5.0000	1.2500	1.6000	0.4000	4	5	50.0000	12.5000

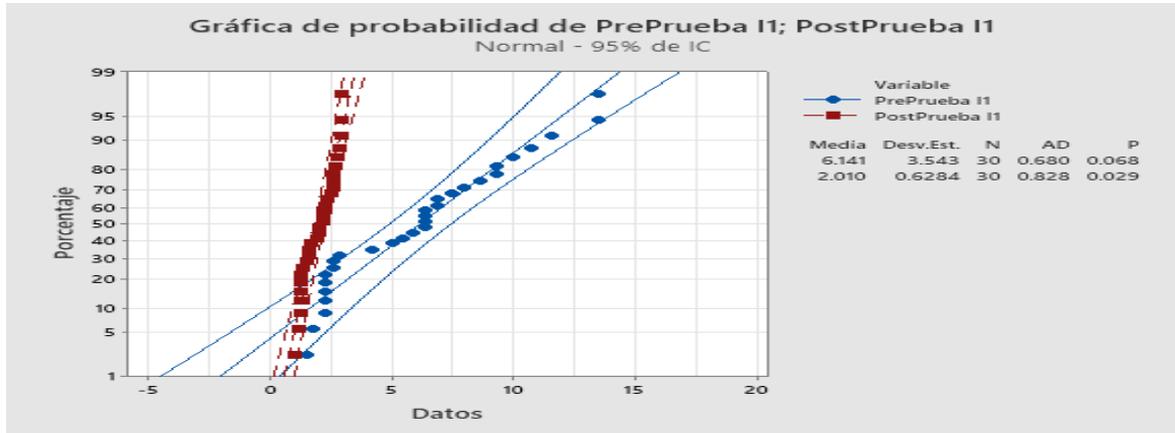
14	6.3636	1.5500	2.0364	0.4960	2	3	63.6360	15.5000
15	6.3636	2.8500	2.0364	0.9120	5	3	63.6360	28.5000
16	2.5581	2.9000	0.8186	0.9280	1	5	25.5810	29.0000
17	6.8750	2.2667	2.2000	0.7253	3	5	68.7500	22.6670
18	10.0000	1.2167	3.2000	0.3893	5	5	100.0000	12.1670
19	9.2857	2.5833	2.9714	0.8267	3	3	92.8570	25.8330
20	5.8824	1.2333	1.8824	0.3947	1	4	58.8240	12.3330
21	6.8750	2.4833	2.2000	0.7947	5	5	68.7500	24.8330
22	1.4894	1.2333	0.4766	0.3947	3	5	14.8940	12.3330
23	2.2727	1.3333	0.7273	0.4267	4	3	22.7270	13.3330
24	5.4286	2.5833	1.7372	0.8267	4	5	54.2860	25.8330
25	13.4783	2.1833	4.3131	0.6987	3	4	134.7830	21.8330
26	10.7692	2.3333	3.4461	0.7467	3	3	107.6920	23.3330
27	8.0000	1.6333	2.5600	0.5227	3	5	80.0000	16.3330
28	6.3636	2.6500	2.0364	0.8480	3	5	63.6360	26.5000
29	2.8571	1.9667	0.9143	0.6293	5	5	28.5710	19.6670
30	2.2727	1.1833	0.7273	0.3787	3	3	22.7270	11.8330

Fuente: Elaboración propia

I 1: Tiempo de espera

Figura 14

Prueba de normalidad preprueba y postprueba del tiempo de espera



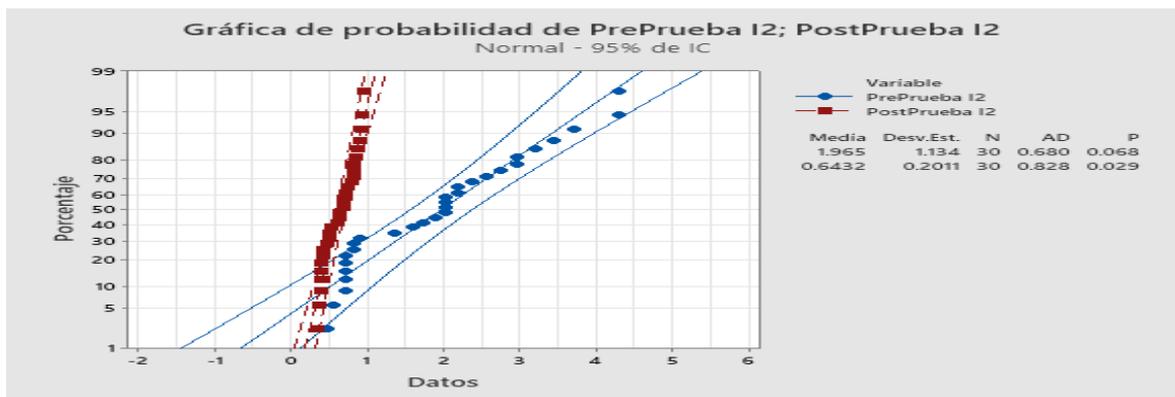
Fuente: Elaboración propia

Se vio que para el indicador, en la preprueba y postprueba respectivamente fue p (0.068 y 0.029) $>$ α (0.005). Por lo tanto, los valores del indicador tuvieron un comportamiento normal (paramétricos).

I 2: Costo de personal

Figura 15

Prueba de normalidad preprueba y postprueba del costo de personal



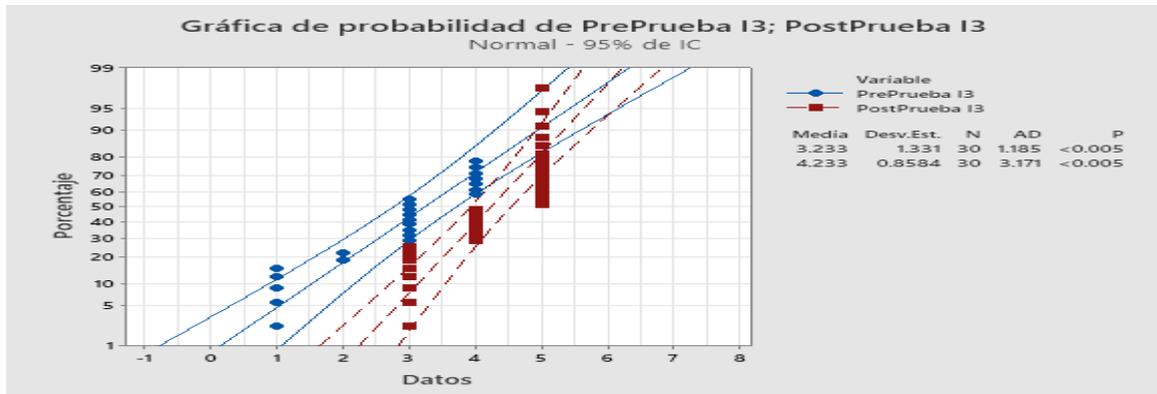
Fuente: Elaboración propia

Se vio que para el indicador, en la preprueba y postprueba respectivamente fue p (0.068 y 0.029) $>$ α (0.005). Por lo tanto, los valores del indicador tuvieron un comportamiento normal (paramétricos).

I 3: Satisfacción del cliente

Figura 16

Prueba de normalidad preprueba y postprueba de satisfacción del cliente



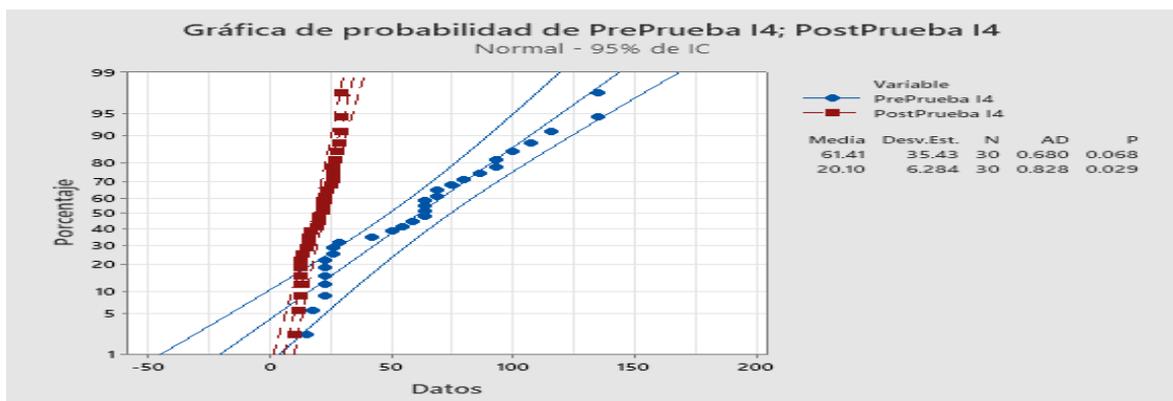
Fuente: Elaboración propia

Se vio que para el indicador, en la preprueba y postprueba respectivamente $> \alpha$ (0.005). Por lo tanto, los valores del indicador no tuvieron un comportamiento normal (no paramétricos).

I 4: Eficiencia de evaluación

Figura 17

Prueba de normalidad preprueba y postprueba de la eficiencia de evaluación



Fuente: Elaboración propia

Se vio que para el indicador, en la preprueba y postprueba respectivamente fue p (0.027 y 0.029) $> \alpha$ (0.005). Por lo tanto, los valores del indicador tuvieron un comportamiento normal (no paramétricos).

A. Contrastación para el I 1: Tiempo de espera.

Se realizó una medición sin el uso de Machine Learning (preprueba) y otra con el uso de un Machine Learning (postprueba):

Tabla 10

Datos de la preprueba del tiempo de espera en el proceso de prestación bancaria

PrePrueba	8.6207	1.7391	2.2727	4.2105	9.2857
	2.5581	11.6000	7.4194	2.2727	13.4783
	2.2727	6.3636	5.0000	6.3636	6.3636
	2.5581	6.8750	10.0000	9.2857	5.8824
	6.8750	1.4894	2.2727	5.4286	13.4783
	10.7692	8.0000	6.3636	2.8571	2.2727

Fuente: Elaboración propia

Tabla 11

Datos de la postprueba del tiempo de espera en el proceso de prestación bancaria

PostPrueba	2.1833	1.4667	1.2833	2.9000	2.2667
	2.5667	2.0167	2.7167	2.7167	2.7167
	1.5833	1.0333	1.2500	1.5500	2.8500
	2.9000	2.2667	1.2167	2.5833	1.2333
	2.4833	1.2333	1.2333	1.2333	2.1833
	2.1833	1.6333	2.6500	1.9667	1.1833

Fuente: Elaboración propia

a) Planteamiento de la hipótesis Nula y Alterna

H0: Si se usa Machine Learning, entonces incrementa el tiempo espera en una entidad financiera.

Ha: Si se usa Machine Learning, entonces disminuye el tiempo espera en una entidad financiera.

μ_1 : Media poblacional del tiempo de espera en la preprueba.

μ_2 : Media poblacional del tiempo de espera en la postprueba.

$H_0: \mu_1 \leq \mu_2$

$H_a: \mu_1 > \mu_2$

b) Criterios de decisión

Figura 18

Gráfica de distribución postprueba del tiempo de espera

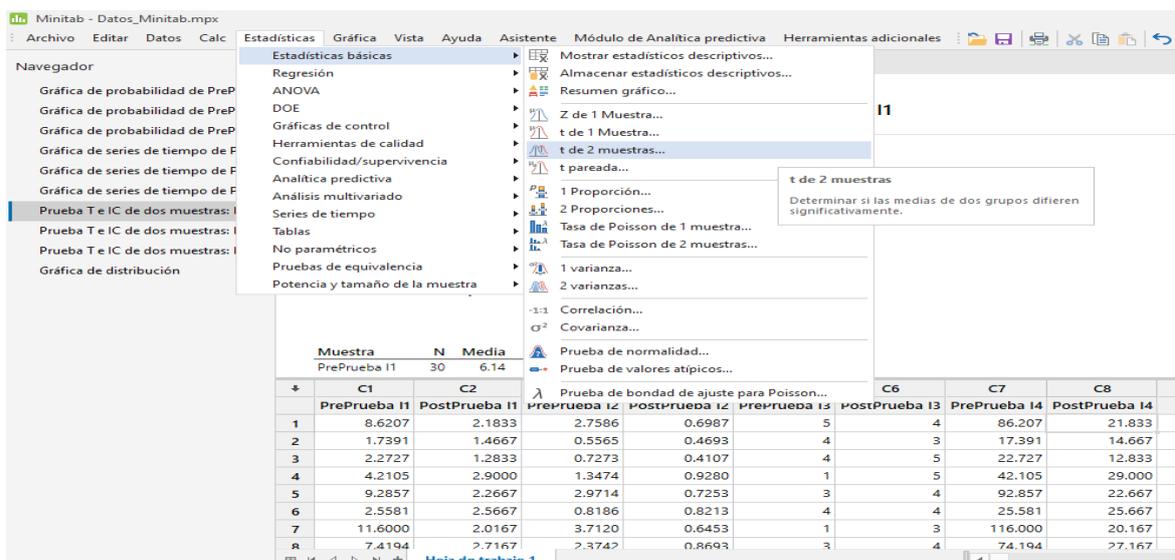


Fuente: Elaboración propia

c) Cálculo: Prueba t para medias de las dos muestras

Figura 19

Cálculo para contrastación del tiempo de espera

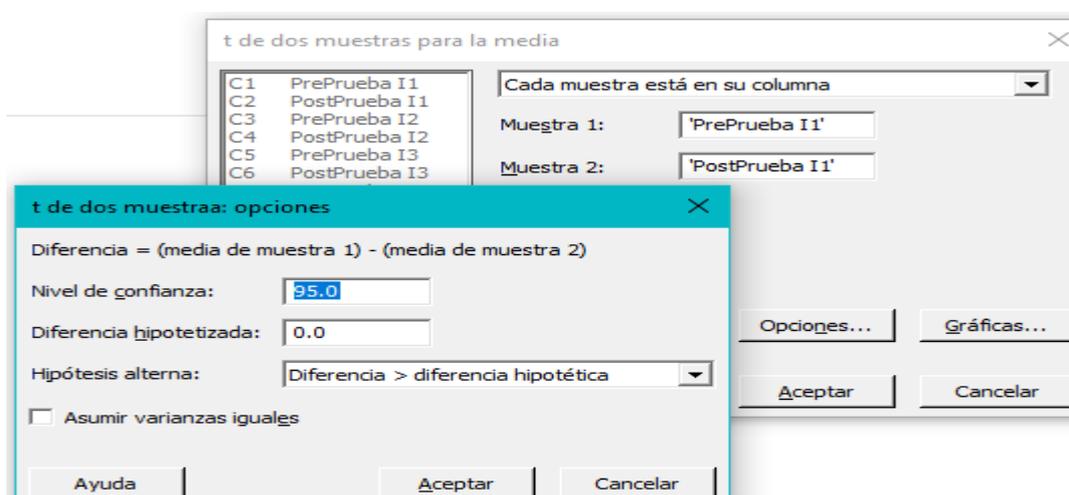


Fuente: Elaboración propia

Se eligió los datos de la muestra de la preprueba del I 1. En segundo lugar, los datos de la muestra de la postprueba del I 1.

Figura 20

Cálculo de criterios de aceptación de la hipótesis del tiempo de espera



Fuente: Elaboración propia

Tabla 12

T de Student para la Hipótesis 1

	Preprueba	Postprueba
Media (\bar{x})	6.14	2.010
Desviación Estándar (S)	3.54	0.628
Observaciones (n)	30	30
Diferencia hipotética de las medias	0.000	
t calculado: t_c	6.29	
p-valor (una cola)	0.000	
Valor crítico de $t_{\alpha/2}$ (una cola): t_t	3.043	

Fuente: Elaboración propia

d) Decisión estadística

Puesto que el valor-p ($0.000 < \alpha (0.05)$), los resultados proporcionaron suficiente evidencia para rechazar la hipótesis nula (H_0), y la hipótesis alterna (H_a) fue cierta. La prueba resultó ser significativa.

Se enriqueció el análisis de la prueba de hipótesis, realizando el siguiente otro análisis, lo cual se puede profundizar en el anexo 11.

B. Contratación para el I 2: Costo de personal

Se realizó una medición sin el uso de Machine Learning (preprueba) y otra con el uso de un Machine Learning (postprueba):

Tabla 13

Datos de la preprueba del costo de personal en el proceso de prestación bancaria

PrePrueba	2.7586	0.5565	0.7273	1.3474	2.9714
	0.8186	3.7120	2.3742	0.7273	4.3131
	0.7273	2.0364	1.6000	2.0364	2.0364
	0.8186	2.2000	3.2000	2.9714	1.8824
	2.2000	0.4766	0.7273	1.7372	4.3131
	3.4461	2.5600	2.0364	0.9143	0.7273

Fuente: Elaboración propia

Tabla 14

Datos de la postprueba del costo de personal en el proceso de prestación bancaria

PostPrueba	0.6987	0.4693	0.4107	0.9280	0.7253
	0.8213	0.6453	0.8693	0.9440	0.6080
	0.5067	0.3307	0.4000	0.4960	0.9120
	0.9280	0.7253	0.3893	0.8267	0.3947
	0.7947	0.3947	0.4267	0.8267	0.6987
	0.7467	0.5227	0.8480	0.6293	0.3787

Fuente: Elaboración propia

a) Planteamiento de la hipótesis Nula y Alternativa

Ho: Si se usa Machine Learning, entonces aumenta el costo de personal en una entidad financiera.

Ha: Si se usa Machine Learning, entonces reduce el costo de personal en una entidad financiera.

μ_1 : Media poblacional del costo de personal en la preprueba.

μ_2 : Media poblacional del costo de personal en la postprueba.

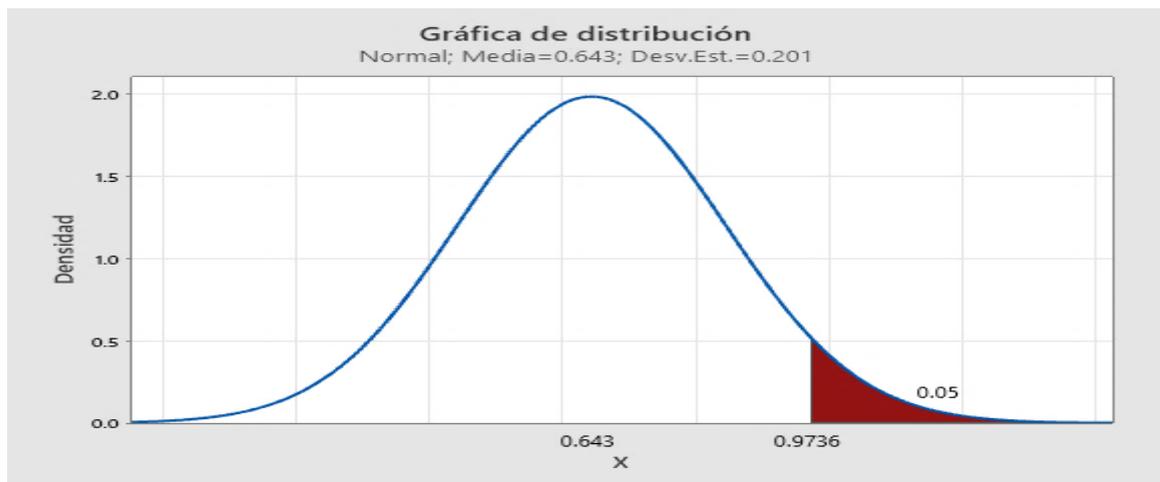
$H_0: \mu_1 \leq \mu_2$

$H_a: \mu_1 > \mu_2$

b) Criterios de decisión

Figura 21

Gráfica de distribución postprueba del costo de personal

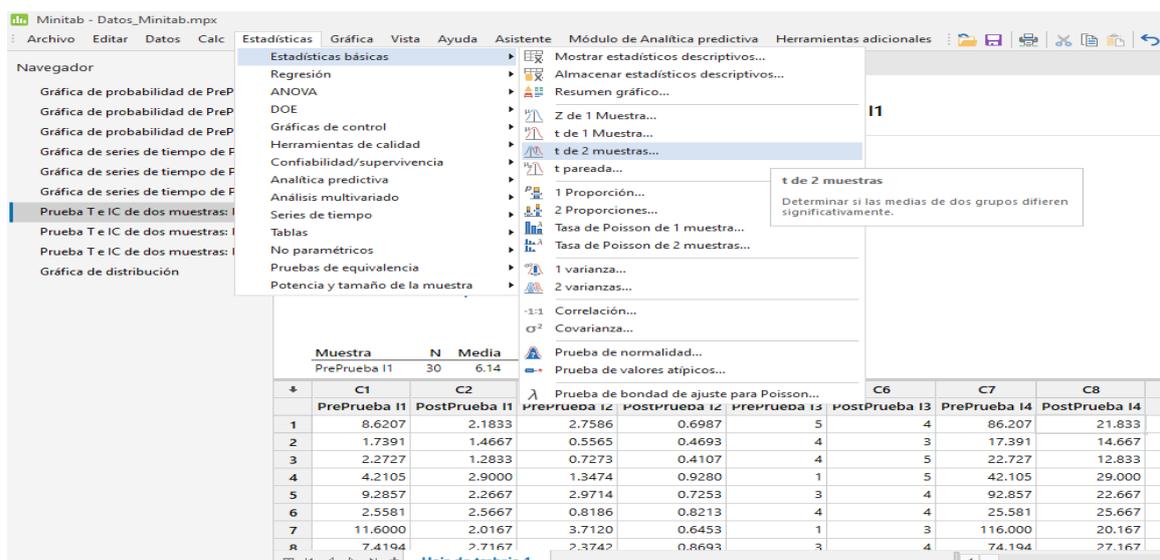


Fuente: Elaboración propia

c) Cálculo: Prueba t para medias de las dos muestras

Figura 22

Cálculo para contrastación del costo de personal

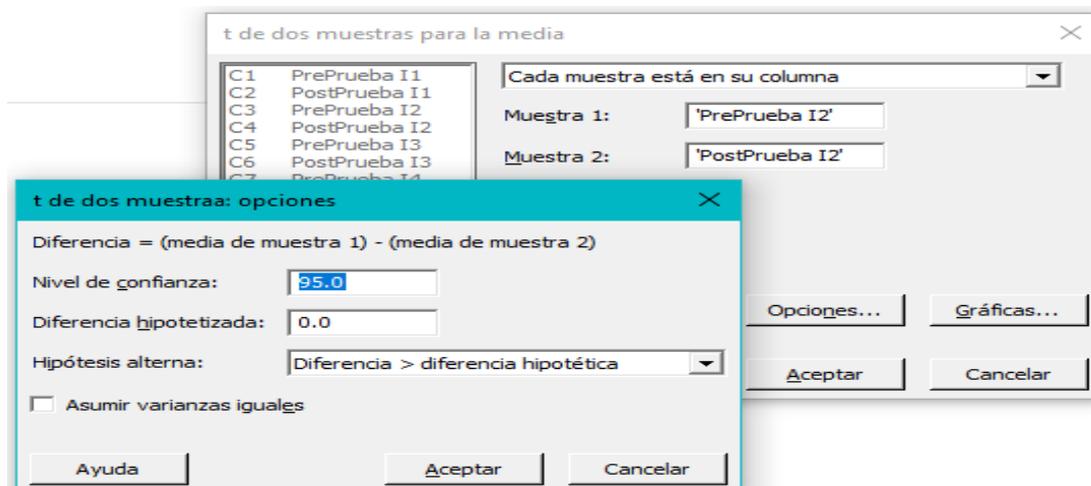


Fuente: Elaboración propia

Se eligió los datos de la muestra de la preprueba del I 2. En segundo lugar, los datos de la muestra de la postprueba del I 2.

Figura 23

Cálculo de criterios de aceptación de la hipótesis del costo de personal



Fuente: Elaboración propia

Tabla 15

T de Student para la Hipótesis 2

	Preprueba	Postprueba
Media (\bar{x})	1.97	0.643
Desviación Estándar (S)	1.13	0.201
Observaciones (n)	30	30
Diferencia hipotética de las medias	0.000	
t calculado: t_c	6.29	
p-valor (una cola)	0.000	
Valor crítico de $t_{\alpha/2}$ (una cola): t_t	0.9736	

Fuente: Elaboración propia

e) Decisión estadística

Puesto que el valor-p ($0.000 < \alpha (0.05)$), los resultados proporcionaron suficiente evidencia para rechazar la hipótesis nula (H_0), y la hipótesis alterna (H_a) fue cierta. La prueba resultó ser significativa.

Se enriqueció el análisis de la prueba de hipótesis, realizando el siguiente otro análisis, lo cual se puede profundizar en el anexo 12.

C. Contratación para el I 3: Satisfacción del cliente

Se realizó una medición sin el uso de Machine Learning (preprueba) y otra con el uso de un Machine Learning (postprueba):

Leyenda:

1 = Totalmente en desacuerdo

2 = En desacuerdo

3 = Indeciso

4 = De acuerdo

5 = Totalmente de acuerdo

Tabla 16

Datos de la preprueba de la satisfacción del cliente en el proceso de prestación bancaria

PrePrueba	5	4	4	1	3
	3	1	3	5	1
	4	2	4	2	5
	1	3	5	3	1
	5	3	4	4	3
	3	3	3	5	3

Fuente: Elaboración propia

Tabla 17

Datos de la postprueba de la satisfacción del cliente en el proceso de prestación bancaria

PostPrueba	4	3	5	5	4
	4	3	4	5	4
	5	5	5	3	3
	5	5	5	3	4
	5	5	3	5	4
	3	5	5	5	3

Fuente: Elaboración propia

a) Planteamiento de la hipótesis Nula y Alternativa

H_0 : Si se usa Machine Learning, entonces reduce la satisfacción del cliente en una entidad financiera.

H_a : Si se usa Machine Learning, entonces aumenta la satisfacción del cliente en una entidad financiera.

μ_1 : Media poblacional de la satisfacción del cliente en la preprueba.

μ_2 : Media poblacional de la satisfacción del cliente en la postprueba.

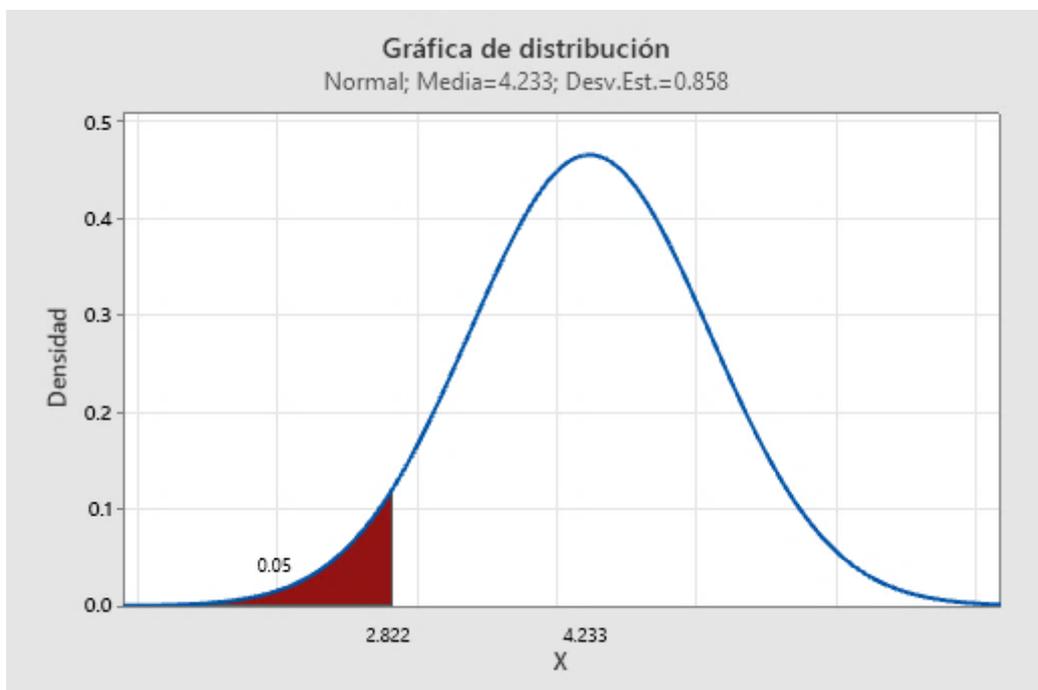
$H_0: \mu_1 \geq \mu_2$

$H_a: \mu_1 < \mu_2$

b) Criterios de decisión

Figura 24

Gráfica de distribución postprueba de la satisfacción del cliente

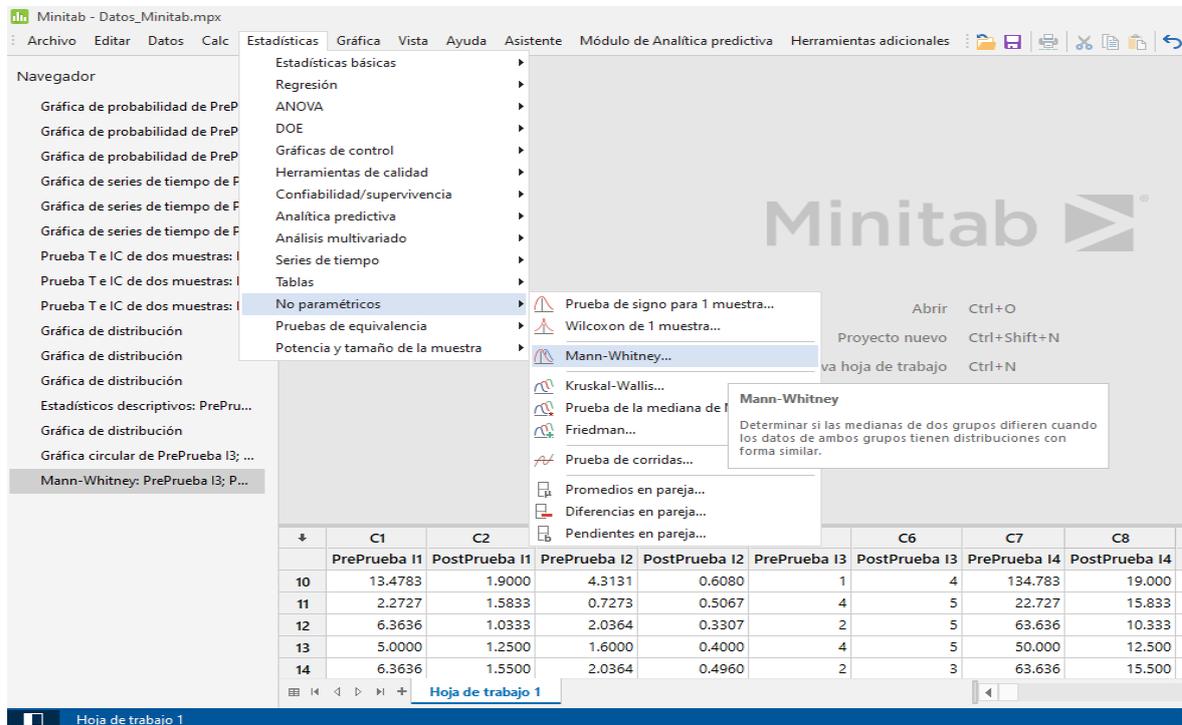


Fuente: Elaboración propia

c) **Cálculo:** Prueba U de Mann-Whitney para las medias de dos muestras

Figura 25

Cálculo para contrastación de la satisfacción del cliente

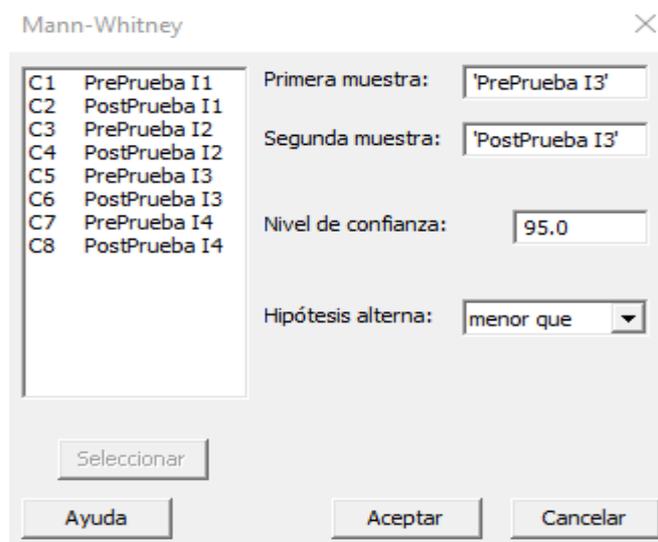


Fuente: Elaboración propia

Se eligió los datos de la muestra de la preprueba del I 3. En segundo lugar, los datos de la muestra de la postprueba del I 3.

Figura 26

Cálculo de criterios de aceptación de la hipótesis de la satisfacción del cliente



Fuente: Elaboración propia

Figura 27

Mann-Whitney preprueba y postprueba de satisfacción del cliente para la entidad financiera

HOJA DE TRABAJO 1

Mann-Whitney: PrePrueba I3; PostPrueba I3

Método

η_1 : mediana de PrePrueba I3

η_2 : mediana de PostPrueba I3

Diferencia: $\eta_1 - \eta_2$

Estadísticas descriptivas

Muestra	N	Mediana
PrePrueba I3	30	3.0
PostPrueba I3	30	4.5

Estimación de la diferencia

Límite superior para la Confianza		
Diferencia	diferencia	lograda
-1	-0.0000000	95.04%

Prueba

Hipótesis nula $H_0: \eta_1 - \eta_2 = 0$

Hipótesis alterna $H_1: \eta_1 - \eta_2 < 0$

Método	Valor W	Valor p
No ajustado para empates	720.50	0.002
Ajustado para empates	720.50	0.001

Fuente: Elaboración propia

En estos resultados, la hipótesis nula indicó que la diferencia entre las medianas preprueba y postprueba de la satisfacción del cliente es 0. Puesto que el valor p fue 0.001, que fue menor que el nivel de significancia de 0.05, los resultados proporcionaron suficiente evidencia para rechazar la hipótesis nula (H_0), y la hipótesis alterna (H_a) fue cierta. La prueba resultó ser significativa.

Se enriqueció el análisis de la prueba de hipótesis, realizando el siguiente otro análisis, lo cual se puede profundizar en el anexo 13.

D. Contratación para el I 4: Eficiencia de evaluación

Se realizó una medición sin el uso de Machine Learning (preprueba) y otra con el uso de un Machine Learning (postprueba):

Tabla 18

Datos de la preprueba de la eficiencia de evaluación en el proceso de prestación bancaria

PrePrueba	86.2070	17.3910	22.7270	42.1050	92.8570
	25.5810	116.0000	74.1940	22.7270	134.7830
	22.7270	63.6360	50.0000	63.6360	63.6360
	25.5810	68.7500	100.0000	92.8570	58.8240
	68.7500	14.8940	22.7270	54.2860	134.7830
	107.6920	80.0000	63.6360	28.5710	22.7270

Fuente: Elaboración propia

Tabla 19

Datos de la postprueba de la eficiencia de evaluación en el proceso de prestación bancaria

PostPrueba	21.8330	14.6670	12.8330	29.0000	22.6670
	25.6670	20.1670	27.1670	29.5000	19.0000
	15.8330	10.3330	12.5000	15.5000	28.5000
	29.0000	22.6670	12.1670	25.8330	12.3330
	24.8330	12.3330	13.3330	25.8330	21.8330
	23.3330	16.3330	26.5000	19.6670	11.8330

Fuente: Elaboración propia

a) Planteamiento de la hipótesis Nula y Alterna

H0: Si se usa Machine Learning, entonces disminuye la eficiencia de evaluación en una entidad financiera.

Ha: Si se usa Machine Learning, entonces eleva la eficiencia de evaluación en una entidad financiera.

μ_1 : Media poblacional de la eficiencia de evaluación en la preprueba.

μ_2 : Media poblacional de la eficiencia de evaluación en la postprueba.

$H_0: \mu_1 \leq \mu_2$

$H_a: \mu_1 > \mu_2$

b) Criterios de decisión

Figura 28

Gráfica de distribución postprueba de la eficiencia de evaluación

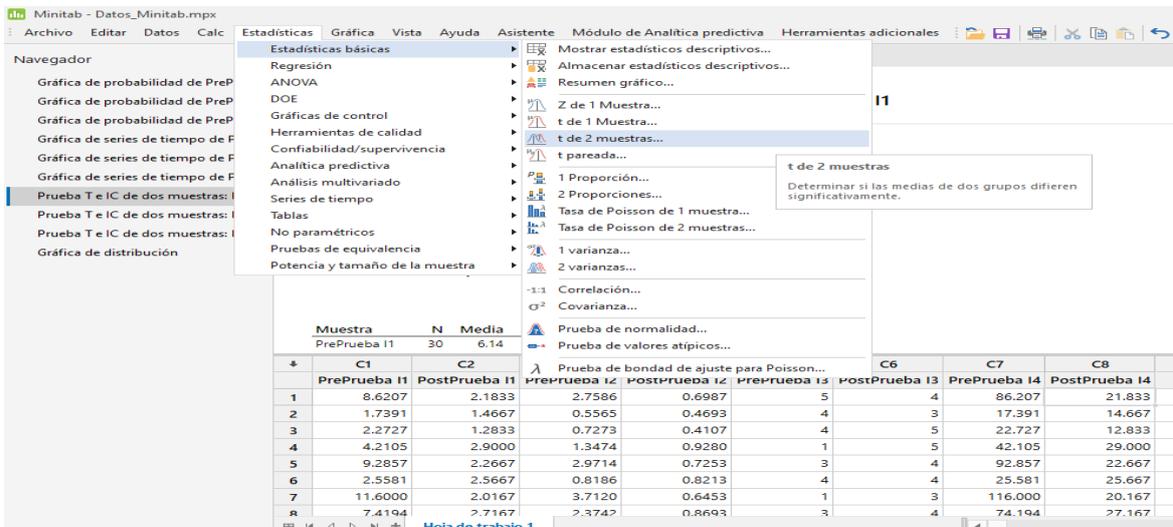


Fuente: Elaboración propia

d) Cálculo: Prueba t para medias de las dos muestras

Figura 29

Cálculo para contrastación de la eficiencia de evaluación

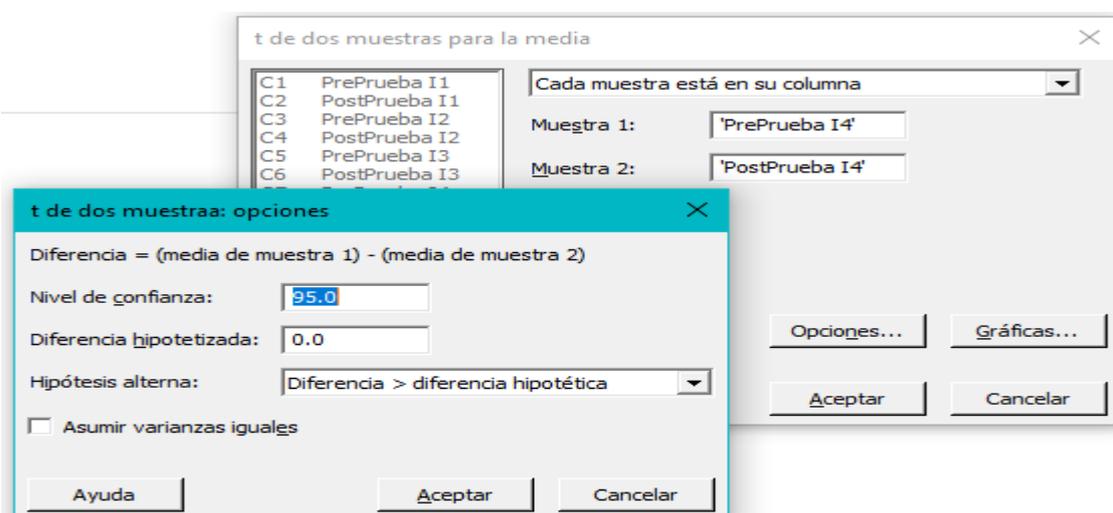


Fuente: Elaboración propia

Se eligió los datos de la muestra de la preprueba del I 4. En segundo lugar, los datos de la muestra de la postprueba del I 4.

Figura 30

Cálculo de criterios de aceptación de la hipótesis de la eficiencia de evaluación



Fuente: Elaboración propia

Tabla 20

T de Student para de la Hipótesis 4

	Preprueba	Postprueba
Media (\bar{x})	61.4	20.10
Desviación Estándar (S)	35.4	6.28
Observaciones (n)	30	30
Diferencia hipotética de las medias	0.000	
t calculado: t_c	6.29	
p-valor (una cola)	0.000	
Valor crítico de $t_{\alpha/2}$ (una cola): t_t	30.43	

Fuente: Elaboración propia

f) Decisión estadística

Puesto que el valor-p ($0.000 < \alpha (0.05)$), los resultados proporcionaron suficiente evidencia para rechazar la hipótesis nula (H_0), y la hipótesis alterna (H_a) fue cierta. La prueba resultó ser significativa.

Se enriqueció el análisis de la prueba de hipótesis, realizando el siguiente otro análisis, lo cual se puede profundizar en el anexo 14.

V. DISCUSIÓN

Los resultados de la actual tesis demostraron el logro de la mejora del proceso de prestación bancaria si se usa Machine Learning basado en el silogismo hipotético (Beuchot 2018).

Con referencia a la hipótesis específica 1, al realizar el análisis inferencial con la prueba T de student de dos muestras, el valor-p ($0.000 < \alpha (0.05)$), los resultados proporcionaron suficiente evidencia para rechazar la hipótesis nula (H_0), y la hipótesis alterna (H_a) fue cierta. Es más, según (Chen, Katchova y Zhou, 2021) predijo en el 61% de los casos correctamente para 131 431 observaciones analizadas por su Machine Learning durante 23 años, y para que se aceptara esta hipótesis específica se tuvo que considerar la menor pero necesaria cantidad de parámetros a completar por el usuario desde un sistema, para que el tiempo de espera sea el menor. Concretamente 11 características se usaron en el Machine Learning de la presente tesis; mientras que los mencionados autores utilizaron 27 características para su modelo de Machine Learning. En ese sentido, la investigación es superior, porque se infiere que 1.5 de transacciones en promedio se realizaron por día en el banco analizado por (Chen, Katchova y Zhou, 2021), mientras que con el Machine Learning implementado para una entidad financiera, en promedio a un usuario le tomó 2.010 segundos (ver anexo 11) ejecutar el proceso, esto evidencia que se puede llegar a tener una mayor cantidad de transacciones en menos tiempo que la investigación realizada por (Chen, Katchova y Zhou, 2021).

Resultados similares también los obtuvo (Ereiz, 2019) que obtiene una precisión muy alta en lo que respecta a señalar que 3612 préstamos se mantendrán con la misma mora sin importar que permute en el tiempo. De igual forma, la investigación es superior, porque el autor ingresa 29 atributos a su modelo de Machine Learning, planteándola de un modo distinto porque no aplica un estímulo a los clientes, sino estimula su sistema de OptiML para ayudarlo a encontrar cuál es el mejor algoritmo de clasificación. En cambio, en esta investigación, la predicción obtenida (70%) compite con la precisión del autor (75.2%), sin embargo, el autor no aplica el estímulo a las personas, por el contrario, en esta investigación

si se aplica el estímulo a los individuos obteniéndose un promedio de 2.010 segundos para completar una transacción (ver anexo 11).

También resulta de gran importancia recabar los resultados de (Liu, 2022) donde obtiene un 21.5% de tasa de ganancia que implica que ganaron más clientes y un aumento de 38% de ganancia observada que explica la capacidad de la atención desde el acercamiento de los clientes al banco hasta llegar a compartirle una decisión basado en el Machine Learning del autor. En ese sentido los resultados del autor merecen ser citados en futuras investigaciones porque detectó que existió una gran debilidad en cuanto al préstamo de crédito, el mismo factor detectado en la tesis actual, que para estos temas de Machine Learning en procesos de prestación bancaria se requiere pensar en la debilidad y una de ellas, detectada por (Liu, 2022) fue la atención limitada, la cual se relaciona muy estrechamente con el tiempo de espera, llegándose a sugerir que la investigación de (Liu, 2022) es una buena referencia para apoyar las investigaciones sobre préstamos de crédito con enfoque de aprendizaje automático.

Con referencia a la hipótesis específica 2, al realizar el análisis inferencial con la prueba T de student de dos muestras, el valor-p ($0.000 < \alpha (0.05)$), los resultados proporcionaron suficiente evidencia para rechazar la hipótesis nula (H_0), y la hipótesis alterna (H_a) fue cierta.

Mas aún, para (Pérez, 2021) quien obtuvo resultados en donde 2765 tuvieron 180 días de mora, lo cual representa una pérdida para el banco en costos por los procedimientos que debe hacer para fomentar la responsabilidad del cliente mediante nuevos plazos de pago, significativos para la investigación realizada porque el porcentaje de precisión obtenido por el autor (66.66%) fue similar al de la tesis actual (70%), sin embargo, el autor no llega a estimular a las personas con su Machine Learning, en cambio, en esta investigación si se estimula a los individuos con el Machine Learning, porque el tipo de metodología que utiliza el autor es KDD, que permite generar conocimiento pero no tiene el objetivo de solucionar problemas en un producto de software que puedan usar las personas, pero es de reconocer que su investigación da lugar a que futuras investigaciones, apliquen la metodología CRISP-DM para desplegar un modelo con una precisión mayor a la obtenida por el

autor (66.66%) para tratar un problema detectado por el mismo, que es el costo de personal, ya que, en la presente investigación se planteó dicho indicador porque el personal de la entidad financiera además de ejecutar procesos manuales para evaluar el préstamo a un cliente, también debe dar seguimiento a los clientes para que se mantengan responsables con sus pagos respectivos y esto le cuesta dinero al banco por el tema de los salarios del personal requerido, en ese sentido, da pie la investigación por (Pérez, 2021) para que además de predecir si el cliente es o no buen pagador basado en su capacidad de pago como se vio en la presente tesis, también se realice la predicción mostrando cuántas moras debe tener el cliente para que el personal ya no realice trabajos extras que implican un constante seguimiento de este cliente, ya que en la investigación realizada se obtuvo como resultados 0.643 soles en promedio que la entidad financiera se ahorra por cliente que consulta su estado de pagador, y si esto se adapta al propuesto por (Pérez, 2021) quien no desplegó pero obtuvo una considerable precisión de su Machine Learning, la entidad financiera que lo integre, en definitiva, se posicionará en mayor medida.

Estas afirmaciones, también se evidencian en el trabajo por (Tariq et al., 2019) que también seleccionaron una metodología para interpretar qué algoritmo es mejor, más no, el despliegue de su algoritmo para uso por parte de las personas. Es más, por (Tariq et al., 2019) se destaca que la predicción de la morosidad en el sector financiero cuenta con características significativas para su emprendimiento, en tal sentido, se recalca que las entidades financieras integren el Machine Learning ya que les ahorra 0.643 soles por cliente a partir de la investigación realizada, y que lo relacionen con la predicción de la morosidad que también cuenta con un futuro prometedor para que las entidades financieras se preparen a más eventos menos predecibles, puesto que también (Tariq et al., 2019) reconocen que la estabilidad económica en cualquier país es sustancial, y el hecho de que los costos de personal bajen como se mostró en los resultados de la tesis actual, implica que surjan otros puestos de trabajo dentro del banco que se deban de estudiar, asimismo se insta a trabajar no con metodologías que utilizaron (Tariq et al., 2019), sino con metodologías que permiten hacer el despliegue de la solución como CRISP-DM.

Con referencia a la hipótesis específica 3, el valor p fue 0.001, que fue menor que el nivel de significancia de 0.05, la decisión fue rechazar la hipótesis nula y concluir que la satisfacción del cliente en el momento preprueba fue menor que la satisfacción en el momento postprueba. Es más, según (Alarcón, Eduardo y Mora, 2022) la tasa de morosidad llega a reducirse con Machine Learning y es porque esta solución le ayuda al cliente a saber en cuántas moras debe pagar para que no se retrase en sus pagos, de ahí que, para que el cliente pague en el tiempo, los autores tuvieron que hacer una evaluación de los métodos tecnológicos para generar el modelo de forma efectiva. Por otra parte, la presente investigación utiliza los métodos tecnológicos basados en Python y Javascript, diferentes a los métodos de (Alarcón, Eduardo y Mora, 2022) que tienen el modelo, pero no lo despliegan, lo cual da a entender que no utilizaron una metodología para el despliegue como CRISP-DM dando cuenta a que solo analizaron características pero dichas características no las compartieron a los clientes para que ellos interactúen con el sistema y este le realice la predicción. Es más, los resultados de (Alarcón, Eduardo y Mora, 2022) no evidencian la mejora de la satisfacción del cliente mediante sus métodos tecnológicos, en cambio, en la tesis actual la satisfacción es una información que se recolecta en el mismo sistema. Por otra parte, los investigadores plantean que sus resultados reducen las infracciones de implementarlo en sus procesos, y esto es significativo para los resultados de la investigación, porque a una menor cantidad de fallas que comete el Machine Learning, los clientes se irán más satisfechos, pero esto no es objeto de discusión en el trabajo por (Alarcón, Eduardo y Mora, 2022), porque no llegó al nivel de medir la satisfacción del cliente cuando su morosidad empezó a subir, que es un resultado negativo para ellos, porque en la tesis actual, aunque se muestre que el cliente no es buen pagador, el sistema le pide que seleccione su satisfacción del proceso.

En esa misma línea (De la Cruz, 2020) afirma que el poder predictivo de su modelo de Machine Learning asegura un buen desempeño para predecir si el cliente comprará o rechazará un oficial de crédito bancario vía canal de venta telefónica, esto es semejante a los resultados que se encontraron en la investigación, pero difieren en el sentido que el canal por el cual se realiza es en un sistema web, sin embargo, logró generar el incremento de las ventas, lo cual es

importante para esta investigación, porque de cara al uso de la empresa; la efectividad de la campaña de las ventas se da fuertemente con sus variables del Machine Learning que le permitió a (De la Cruz, 2020) concluir que sus resultados determinan que la gestión de las llamadas como canal de ventas y el Machine Learning logran una relación estrecha con la satisfacción del cliente porque no hubieran más ventas si los clientes no sentirían que el proceso va mejor.

Con referencia a la hipótesis específica 4, al realizar el análisis inferencial con la prueba T de student de dos muestras, el valor-p ($0.000 < \alpha (0.05)$), los resultados proporcionaron suficiente evidencia para rechazar la hipótesis nula (H_0), y la hipótesis alterna (H_a) fue cierta.

Es más, según (Chávez, 2021) obtuvo resultados similares donde su mejor algoritmo fue XGBoost, y afirma que es uno de los mejores para calcular con máxima precisión los créditos aceptados y válidos. Sin embargo, su aporte se limita a que analiza características socioeconómicas para entrenar un modelo que no llega a implementarse en una solución de software, por el contrario, deja la tarea a que investigaciones, consideren su modelo para que se implemente en un software, lo cual es válido pero difícil de replicar. Esto debido a que (Chávez, 2021) no tienen público su dataset para entrenamiento, más aún, sobre la eficiencia de evaluación, el autor identificó la efectividad de la aceptación de los créditos pero a partir de una muestra en el pasado, diferente a la forma de trabajo de la presente tesis, pues, para que la eficiencia promedio sea 20.100 (ver anexo 14), el recurso que se tuvo que analizar fue el tiempo, y por el autor (Chávez, 2021), no hay observado este recurso. En ese sentido, la investigación es superior, porque su modelo de Machine Learning tiene una precisión de 70% utilizando un algoritmo de Árboles de decisión, mientras que (Chávez, 2021), utilizó un algoritmo XGBoost con una precisión de 69.5% y en la presente tesis se implementa una solución de software para dar conocimiento de la eficiencia de evaluación a una entidad financiera respecto al tiempo real, en tanto, el autor no despliega su modelo a una solución real, más aún, utilizando TDSP como metodología por parte de (Chávez, 2021), que es una metodología que llega a presentar un software pero el autor no lo hace.

Esto se confirma con los resultados de (Aceituno, 2019), quien luego de aplicar su Machine Learning con el modelo de árbol de decisión, mejoró el nivel de asertividad en el otorgamiento de créditos en una entidad financiera. Estos resultados llegan a la recomendación de que su modelo de árbol de decisión con una precisión de 88% sea aprobado en el sector financiero. En otras palabras, el modelo de árbol de decisión es bueno para el tema de prestación bancaria, porque permite la clasificación y visualización de cómo se llega al resultado final. Al igual que la creación del modelo que se realizó, se sabe cuál es el procedimiento para que el usuario llegue al resultado mediante un entorno de análisis de datos, pero no se sabe cómo llegará a esa solución en un entorno de desarrollo de software. En tal sentido, los resultados por (Aceituno, 2019) son similares a la presente investigación porque, en esta investigación también se tuvo la seguridad de que el Machine Learning si es árbol de decisión, cuando se aplica a un sistema, es fácilmente reconocible saber cuáles son los factores por los cuales llega el usuario a su resultado, lo cual permite evaluar con eficiencia estos resultados. No obstante, la presente tesis es superior porque llega a implementar el modelo a una solución de software, mientras que el autor no lo hizo, llegando solo a una propuesta de cuál es el mejor algoritmo, pero concluye que el árbol de decisión es bueno.

Al respecto, (Cela y Cuenca 2019), obtuvieron resultados más parecidos a la presente tesis, destacando un método tecnológico que es adecuado para el personal, esto quiere decir que se muestran resultados desde el punto de vista estadístico para que el personal tome una decisión en base a las probabilidades estimadas que concluye si es seguro o no prestarle dinero al cliente, resultados que no puede interpretar el cliente. Por otro lado, los autores no llegan a mostrar cuál fue el tiempo que tomó registrar sus 20 variables de entrada para que se muestren los resultados en una base de datos. En cambio, en la presente tesis, 11 variables se esperaron introducir por el cliente, lo cual les tomó un tiempo y ya que el tiempo fue menor al objetivo, la eficiencia de evaluación se elevó y todo ello se registra en una base de datos. No obstante (Cela y Cuenca 2019) evidencian que el proceso de aprobación se optimizó en un 70%, lo que significa que para este proceso se utilizó menos tiempo del objetivo. Más aún, que en la presente tesis se obtuvo que la eficiencia promedio fue en el 100% de los casos mejor (ver anexo 14).

CONCLUSIONES

Primero: Si se usa Machine Learning, entonces mejora el proceso de prestación bancaria en una entidad financiera habiéndose obtenido la aceptación de las hipótesis alternativas de la investigación, lo que evidencia la mejora del 100% de los indicadores propuestos y la mejora de la situación actual (AS-IS) con la situación propuesta (TO-BE) en la prestación bancaria como beneficio a la entidad financiera.

Segundo: Si se usa Machine Learning, entonces disminuye el tiempo de espera del proceso de prestación bancaria en una entidad financiera habiéndose obtenido una media de 2.010 en la postprueba, inferior a la media de 6.141 en la preprueba, logrando una disminución del 67.43% del tiempo de espera, lo que evidencia los tiempos bajos de espera en la prestación bancaria como beneficio al cliente.

Tercero: Si se usa Machine Learning, entonces reduce el costo de personal del proceso de prestación bancaria en una entidad financiera habiéndose obtenido una media de 0.643 en la postprueba, inferior a la media de 1.965 en la preprueba, logrando una reducción del 67.34% del costo de personal, lo que evidencia los costos bajos de personal en la prestación bancaria como beneficio a la entidad financiera.

Cuarto: Si se usa Machine Learning entonces aumenta la satisfacción del cliente del proceso de prestación bancaria en una entidad financiera habiéndose obtenido un valor de $p = 0.001 < 0.05$, lo que evidencia la satisfacción alta en la prestación bancaria como beneficio a la generación del conocimiento.

Quinto: Si se usa Machine Learning, entonces eleva la eficiencia de evaluación del proceso de prestación bancaria en una entidad financiera habiéndose obtenido una media de 20.100 en la postprueba, inferior a la media de 61.410 en la preprueba, logrando la elevación del 67.26% de la eficiencia de evaluación, lo que evidencia la eficiencia alta en la prestación bancaria como beneficio a la entidad financiera por un menor uso de tiempo en la evaluación.

RECOMENDACIONES

Primera: A la alta gerencia del banco Alfin, estos resultados facilitan el escenario para proponer políticas de seguridad tales como las reglas para almacenar información específica del cliente como su nombre, dni, teléfono y correo electrónico para que los bancos del Perú tengan la confianza de que dicha información es posible de ser almacenada sin riesgo a vulnerabilidad en la información.

Segunda: A la alta gerencia del banco Alfin, estos resultados facilitan el escenario para adquirir planes anuales de almacenamiento en Firebase para soportar un mayor número de consultas por minuto por los clientes, posteriormente realizar pruebas unitarias y de concurrencia en la fase de testing por los temas de latencia con la finalidad de implementar el proyecto a nivel nacional para el público. De hacerlo correctamente, se tendría una base de datos invulnerable porque la pertenece a Google con la capacidad de almacenamiento ampliada a comparación del Firebase gratuito que permite 1 millón de transacciones.

Tercera: A la alta gerencia del banco Alfin, estos resultados facilitan el escenario para no preocuparse por la concurrencia y presentarlo a poco público para ir atendiendo los temas de concurrencia que preocupa en los proyectos con Machine Learning para dejar feliz al cliente cuando el sistema llegue a más público.

Cuarta: A las futuras investigaciones, estos resultados facilitan el escenario para utilizar CRISP-DM para poder desplegar el proyecto, ya que otras metodologías como KDD o SEMMA, no contribuyen a que el público pueda solucionar sus problemas, tales como conocer si es o no buen pagador y en el despliegue se debe utilizar PythonAnywhere.

Quinta: A las futuras investigaciones, estos resultados facilitan el escenario para utilizar CRISP-DM para el tratamiento de los datos y se vuelva un conocimiento como contribución a la ciencia de los datos, independiente de la ingeniería de software para el aplicativo; diagrama de casos de uso, diagrama de entidad relación, diagrama de componentes, diagrama de despliegue pero que se va a relacionar en la base de datos porque CRISP-DM gestiona el conocimiento.

Sexto: Se recomienda que futuras investigaciones implementen el Machine Learning con 2 o 3 años de antelación antes de recolectar los datos postprueba como parte de un diseño longitudinal, tomando en cuenta que una carrera universitaria dura 5 años en promedio para Ingeniería de Sistemas, el investigador debe tener listo su sistema y funcionando a partir del segundo año e ir recabando información en la base de datos, para ello, también se sugiere que el investigador adquiera un plan premium de Firebase anual ya que actualmente con el Firebase gratuito se pueden llegar a realizar 1 millón de transacciones y después la base de datos no permitirá ninguna otra operación de escritura o sobreescritura.

Séptimo: Se recomienda que futuras investigaciones implementen el Machine Learning con los datos de acuerdo con sus posibilidades, pero que contengan características socioeconómicas ya que utilizó para la tesis datos de la ONP con clientes que ya registraron sus datos socioeconómicos en un sistema guardado a una base de datos MySQL, y porque es del estado peruano hubo la seguridad de emprender el análisis de los datos con esas características. En tanto, se sugiere que si se va a tomar un dataset de internet, debe ser de Kaggle. Para el año en que se realizó esta investigación e hizo la consulta en Kaggle, se encontró un dataset llamado “Loan Data for Dummy Bank” el cual se evaluó pero no tenía sentido utilizarse porque solo se quedaría en temas de minería de datos habiendo hecho las pruebas en el entorno de Colab. En ese sentido, si se busca un dataset, debe priorizarse ante todo que sean datos que pueda llenar el cliente, no un experto en economía social, ya que la mayoría de dataset para temas de prestación bancaria tienen esa complejidad de interpretación, y aunque dejen la descripción de los campos, cabe la posibilidad de que en el entendimiento de los datos se tenga problema, porque no se tiene un referente con quien conversar sobre las columnas en la base de datos. Esto es así porque si se cuenta con una base de datos que se proporciona por un gerente por previa solicitud, se puede luego programar una reunión con ese gerente para obtener una explicación acerca del negocio y los datos de la base de datos, que no se haría con Kaggle.

REFERENCIAS

- ACEITUNO, M., 2019. Modelo predictivo de análisis de riesgo crediticio usando Machine Learning en una entidad del sector microfinanciero. [en línea], Disponible en: <http://repositorio.unap.edu.pe/handle/UNAP/14481>.
- ALARCÓN, EDUARDO; MORA, B., 2022. Modelo para la evaluación del riesgo crediticio para los clientes de las microfinancieras del Perú. [en línea], Disponible en: <https://repositorioacademico.upc.edu.pe/handle/10757/650407>.
- AMPOUNTOLAS, A., NDE, T.N., DATE, P. y CONSTANTINESCU, C., 2021. A machine learning approach for micro-credit scoring. En: cited By 5, *Risks*, vol. 9, no. 3. DOI 10.3390/risks9030050.
- BEUCHOT, M., 2018. *Introducción a la lógica* [en línea]. S.l.: UNAM, Dirección General de Publicaciones y Fomento Editorial. ISBN 9786070249778. Disponible en: <https://books.google.com.pe/books?id=gwqnDwAAQBAJ>.
- CARDIN, J., 2021. *Diving Into Big Data: The Ultimate Guides For Beginners To Use Python And C++ For Machine Learning*. S.l.: s.n. ISBN 9798641685359.
- CASARES, J.L.E., 2021. *Apuntes de probabilidad y estadística* [en línea]. S.l.: José Luis Espinoza Casares. ISBN 9786077523468. Disponible en: <https://books.google.com.pe/books?id=qBwsEAAAQBAJ>.
- CELA, G. y CUENCA, J.P., 2019. Propuesta de modelo de machine learning para la evaluación de riesgo de crédito utilizando algoritmos de predicción para la Cooperativa de Ahorro y Crédito La Merced Ltda. Cuenca. [en línea], Disponible en: <https://www.researchgate.net/publication/337480778>.
- CHÁVEZ, A., 2021. *Modelos de machine learning para identificar factores asociados a la adquisición de un crédito efectivo en una entidad financiera* [en línea]. S.l.: s.n. Disponible en: <https://hdl.handle.net/20.500.12672/17342>.
- CHEN, J., KATCHOVA, A. y ZHOU, C., 2021. Agricultural loan delinquency prediction using machine learning methods. *International Food and Agribusiness Management Review*, DOI <https://doi.org/10.22434/IFAMR2020.0019>.
- CHEN, S., GUO, Z. y ZHAO, X., 2021. Predicting mortgage early delinquency with machine learning methods. *European Journal of Operational Research*, vol.

290, no. 1, pp. 358-372. ISSN 0377-2217. DOI <https://doi.org/10.1016/j.ejor.2020.07.058>.

CLINTON, S., 2019. *Tree-based Machine Learning Algorithms: Decision Trees, Random Forests, and Boosting*. Clinton Sh. S.I.: Clinton Sheppard. ISBN 978-1-975-86097-4.

DÅDERMAN, A. y ROSANDER, S., 2018. *Evaluating Frameworks for Implementing Machine Learning in Signal Processing: A Comparative Study of CRISP-DM, SEMMA and KDD* [en línea]. School of Electrical Engineering and Computer Science (EECS), KTH: s.n. Disponible en: <http://kth.diva-portal.org/smash/get/diva2:1250897/FULLTEXT01.pdf>.

DE CASTRO VIEIRA, J.R., BARBOZA, F., SOBREIRO, V.A. y KIMURA, H., 2019. Machine learning models for credit analysis improvements: Predicting low-income families' default. En: cited By 9, *Applied Soft Computing Journal*, vol. 83. DOI 10.1016/j.asoc.2019.105640.

DE LA CRUZ, F., 2020. Predicción de adquisición de un préstamo personal bancario a través del canal de televentas utilizando el algoritmo Random Forest. [en línea], Disponible en: <https://hdl.handle.net/20.500.12996/4832>.

DIHIGO, J.G., 2021. *Metodología de la investigación para administradores* [en línea]. S.I.: Ediciones de la U. ISBN 9789587625288. Disponible en: <https://books.google.com.pe/books?id=JiwaEAAAQBAJ>.

DIXON, M., HALPERIN, I. y BILOKON, P., 2020. *Machine Learning in Finance From Theory to Practice*. S.I.: Springer Cham. ISBN 978-3-030-41068-1.

DOERR, S., GAMBACORTA, L. y SERENA GARRALDA, J.M., 2021. Big data and machine learning in central banking. *BIS Working Papers* [en línea], Disponible en: <https://econpapers.repec.org/RePEc:bis:biswps:930>.

DOKO, F., KALAJDZISKI, S. y MISHKOVSKI, I., 2021. Credit Risk Model Based on Central Bank Credit Registry Data. *Journal of Risk and Financial Management* [en línea], vol. 14, no. 3. ISSN 1911-8074. DOI 10.3390/jrfm14030138. Disponible en: <https://www.mdpi.com/1911-8074/14/3/138>.

EREIZ, Z., 2019. Predicting Default Loans Using Machine Learning (OptiML). 2019

- 27th Telecommunications Forum (TELFOR). S.l.: s.n., pp. 1-4. DOI 10.1109/TELFOR48224.2019.8971110.
- FENERICH, A., ARNS STEINER, M.T., STEINER NETO, P.J., TOCHETTO, E., TSUTSUMI, D., ASSEF, F.M. y DOS SANTOS, B.S., 2020. Use of machine learning techniques in bank credit risk analysis. *REVISTA INTERNACIONAL DE METODOS NUMERICOS PARA CALCULO Y DISEÑO EN INGENIERIA*, vol. 36, no. 3. ISSN 0213-1315. DOI 10.23967/j.rimni.2020.08.003.
- FUENTES, C. y ZÚÑIGA, I., 2010. *Gestión integral de riesgos operativos en el proceso de créditos microfinancieros* [en línea]. S.l.: s.n. ISBN 978-9972-622-85-4. Disponible en: <https://hdl.handle.net/20.500.12640/143>.
- GONZÁLES CASTELLANO, R., YLL LAVÍN, M. y CURIEL LORENZO, L., 2003. *Metodología de la investigación científica para las ciencias técnicas*. S.l.: s.n.
- HERNÁNDEZ COLLADO, C. y BAPTISTA LUCIO, P., 2014. *Metodología de la investigación*. S.l.: McGRAW-HILL. ISBN 978-1-4562-2396-0.
- KELLEHER, J., MAC, B. y ARCY, A., 2020. *Fundamentals of Machine Learning for Predictive Data Analytics, second edition: Algorithms, Worked Examples, and Case Studies*. The MIT Pr. S.l.: s.n. ISBN 978-0262044691.
- KNEUSEL, R., 2021. *Practical Deep Learning: A Python-Based Introduction*. S.l.: No Starch Press. ISBN 978-1718500747.
- KOLODIAZHNYI, K., 2020. *Hands-On Machine Learning with C++: Build, train, and deploy end-to-end machine learning and deep learning pipelines*. Packt Publ. S.l.: s.n. ISBN 978-1-78995-533-0.
- LIU, M., 2022. Assessing Human Information Processing in Lending Decisions: A Machine Learning Approach. *Accounting Research*, DOI <https://doi.org/10.1111/1475-679X.12427>.
- MA, X., SHA, J., WANG, D., YU, Y., YANG, Q. y NIU, X., 2018. Study on a prediction of P2P network loan default based on the machine learning LightGBM and XGboost algorithms according to different high dimensional data cleaning. En: cited By 122, *Electronic Commerce Research and Applications*, vol. 31, pp. 24-39. DOI 10.1016/j.elerap.2018.08.002.

- MEIKANG, Q., 2018. *Smart Blockchain*. S.l.: s.n. ISBN 978-3-030-05764-0.
- MORI, J., 2021. Calificación en Infocorp. *Diario Gestión* [en línea]. Disponible en: https://www.facebook.com/watch/live/?ref=watch_permalink&v=1115007752304759.
- NEILL, D.A. y CORTEZ SUÁREZ, L., 2018. *Procesos y Fundamentos de la Investigación Científica* [en línea]. S.l.: s.n. ISBN 978-9942-24-093-4. Disponible en: [http://repositorio.utmachala.edu.ec/bitstream/48000/14232/1/Cap.4-Investigación cuantitativa y cualitativa.pdf](http://repositorio.utmachala.edu.ec/bitstream/48000/14232/1/Cap.4-Investigación%20cuantitativa%20y%20cualitativa.pdf).
- PANDE, P., 2004. *Las claves prácticas de SIX SIGMA*. S.l.: s.n. ISBN 9788448140373.
- PÉREZ LÓPEZ, C., 2021. *DATA MINING. The CRISP-DM METHODOLOGY. The CLEM language and IBM SPSS MODELER*. Lulu. S.l.: s.n. ISBN 9781008981652.
- PÉREZ, R., 2021. Modelo de Scoring para la segmentación de clientes morosos usando minería de datos en una empresa de cobranzas del Perú. [en línea], Disponible en: <https://hdl.handle.net/20.500.12672/16845>.
- PETR, S. y FATHALHRMAN, A., 2020. *Knowledge Discovery in Big Data from Astronomy and Earth Observation: Astrogeoinformatics*. S.l.: s.n. ISBN 978-0128191545.
- SEN, S. y RAJAGOPAL, K., 2020. Design of Knowledge Base Model for Home Loan: Case Study of a Bank in India. En: S. SIKDAR, R.C. DAS y R. BHATTACHARYYA (eds.), *Role of IT- ITES in Economic Development of Asia: Issues of Growth, Sustainability and Governance* [en línea]. Singapore: Springer Singapore, pp. 271-282. ISBN 978-981-15-4206-0. Disponible en: https://doi.org/10.1007/978-981-15-4206-0_21.
- SHI, B., ZHAO, X., WU, B. y DONG, Y., 2019. Credit rating and microfinance lending decisions based on loss given default (LGD). En: cited By 16, *Finance Research Letters*, vol. 30, pp. 124-129. DOI 10.1016/j.frl.2019.03.033.
- SINGH, P., 2020. *Deploy Machine Learning Models to Production: With Flask*,

Streamlit, Docker, and Kubernetes on Google Cloud Platform. Apress. S.l.: s.n. ISBN 978-1-48426-6456-8.

SOLANO, J.A., LANCHEROS CUESTA, D.J., UMAÑA IBÁÑEZ, S.F. y CORONADO-HERNÁNDEZ, J.R., 2022. Predictive models assessment based on CRISP-DM methodology for students performance in Colombia - Saber 11 Test. En: 12th International Conference on Emerging Ubiquitous Systems and Pervasive Networks / 11th International Conference on Current and Future Trends of Information and Communication Technologies in Healthcare, *Procedia Computer Science* [en línea], vol. 198, pp. 512-517. ISSN 1877-0509. DOI <https://doi.org/10.1016/j.procs.2021.12.278>. Disponible en: <https://www.sciencedirect.com/science/article/pii/S1877050921025175>.

SOSNOVSHCHENKO, OLEKSANDR BAIEV, O., 2018. *Machine Learning with Swift: Artificial Intelligence for iOS*. S.l.: s.n. ISBN 1787123529, 9781787123526.

TADDY, M., 2019. *Business Data Science: Combining Machine Learning and Economics to Optimize, Automate, and Accelerate Business Decisions*. S.l.: McGraw Hill. ISBN 978-1260452778.

TARIQ, H., SOHAIL, A., ASLAM, U. y BATCHA, N., 2019. Loan Default Prediction Model Using Sample, Explore, Modify, Model, and Assess (SEMMA). *Computational and Theoretical Nanoscience* [en línea], Disponible en: 1546-1955/2019/16/3489/015.

VARLEY, M. y BELLE, V., 2021. Fairness in machine learning with tractable models. *Knowledge-Based Systems* [en línea], vol. 215, pp. 106715. ISSN 0950-7051. DOI <https://doi.org/10.1016/j.knosys.2020.106715>. Disponible en: <https://www.sciencedirect.com/science/article/pii/S0950705120308443>.

VICENTE VÍRSEDA, JUAN ANTONIO GONZÁLES ARIAS, J. y PARRA RODRÍGUEZ, FRANCISCO BELTRÁN PASCUAL, M., 2019. *Métodos de Data Science aplicados a la Economía y a la Dirección y Administración de Empresas*. S.l.: s.n. ISBN 9788436276015.

YANG, Q., XU, F., ZHANG, Y., LIU, F., HU, W. y LIAO, Q., 2018. Design and Implementation of a Loan System Based on Smart Contract. En: M. QIU (ed.),

Smart Blockchain. Cham: Springer International Publishing, pp. 22-31. ISBN 978-3-030-05764-0.

ЖЕРНОСЕК, Н.С., 2020. ПРИМЕНЕНИЕ СКОРИНГОВОЙ МОДЕЛИ ПРИ КРЕДИТОВАНИИ ПРЕДПРИЯТИЙ МАЛОГО БИЗНЕСА КОММЕРЧЕСКИМИ БАНКАМИ. S.l.: s.n. ISBN 978-985-531-747-1.

ОЛЕХНОВИЧ, Л.В., 2022. *Банковский менеджмент и маркетинг* [en línea]. S.l.: s.n. Disponible en: <http://e.biblio.bru.by/handle/1212121212/18472>.

ANEXOS

Anexo 1: Matriz de consistencia

Tabla 21

Matriz de consistencia

PROBLEMAS	OBJETIVOS	HIPÓTESIS	VARIABLES E INDICADORES															
<p>Problema General ¿En qué medida la aplicación de Machine Learning, mejorará el Proceso de prestación bancaria en una entidad financiera?</p> <p>Problemas Específicos ¿En qué medida la aplicación de Machine Learning, disminuirá el tiempo de espera en una entidad financiera? ¿En qué medida la aplicación de Machine Learning, reducirá el costo de personal en una entidad financiera? ¿En qué medida la aplicación de Machine Learning, aumentará la satisfacción del cliente en una entidad financiera? ¿En qué medida la aplicación de Machine Learning, elevará la eficiencia de evaluación en una entidad financiera?</p>	<p>Objetivo General Mejorar el Proceso de prestación bancaria en una entidad financiera implementando Machine Learning</p> <p>Objetivos Específicos Disminuir el tiempo de espera en una entidad financiera implementando Machine Learning Reducir el costo de personal en una entidad financiera implementando Machine Learning Aumentar la satisfacción del cliente en una entidad financiera implementando Machine Learning Elevar la eficiencia de evaluación en una entidad financiera implementando Machine Learning.</p>	<p>Hipótesis General Si se usa Machine Learning, entonces mejorará el Proceso de prestación bancaria en una entidad financiera.</p> <p>Hipótesis Específicas: Si se usa Machine Learning, entonces disminuye el tiempo de espera en una entidad financiera. Si se usa Machine Learning, entonces reduce el costo de personal en una entidad financiera. Si se usa Machine Learning, entonces aumenta la satisfacción del cliente en una entidad financiera. Si se usa Machine Learning, entonces eleva la eficiencia de evaluación en una entidad financiera.</p>	<p>Variable independiente: Machine Learning</p> <p>Variable Dependiente: Proceso de prestación bancaria</p> <table border="1"> <thead> <tr> <th>INDICADORES</th> <th>ÍNDICES</th> <th>UNIDAD DE ANÁLISIS</th> </tr> </thead> <tbody> <tr> <td>Tiempo de espera.</td> <td>[0... 13]</td> <td>Tiempo</td> </tr> <tr> <td>Costo de personal.</td> <td>[0...7]</td> <td>Tiempo y Costo de personal por minuto</td> </tr> <tr> <td>Satisfacción del cliente.</td> <td>[Totalmente de acuerdo; De acuerdo; Indeciso; En desacuerdo; Totalmente en desacuerdo]</td> <td>Satisfacción</td> </tr> <tr> <td>Eficiencia de evaluación.</td> <td>[0...100]</td> <td>Expectativa y realidad de los tiempos</td> </tr> </tbody> </table>	INDICADORES	ÍNDICES	UNIDAD DE ANÁLISIS	Tiempo de espera.	[0... 13]	Tiempo	Costo de personal.	[0...7]	Tiempo y Costo de personal por minuto	Satisfacción del cliente.	[Totalmente de acuerdo; De acuerdo; Indeciso; En desacuerdo; Totalmente en desacuerdo]	Satisfacción	Eficiencia de evaluación.	[0...100]	Expectativa y realidad de los tiempos
INDICADORES	ÍNDICES	UNIDAD DE ANÁLISIS																
Tiempo de espera.	[0... 13]	Tiempo																
Costo de personal.	[0...7]	Tiempo y Costo de personal por minuto																
Satisfacción del cliente.	[Totalmente de acuerdo; De acuerdo; Indeciso; En desacuerdo; Totalmente en desacuerdo]	Satisfacción																
Eficiencia de evaluación.	[0...100]	Expectativa y realidad de los tiempos																
MÉTODO	POBLACIÓN Y MUESTRA	TÉCNICAS E INSTRUMENTOS	MÉTODO DE ANÁLISIS DE DATOS															
<p>Tipo de investigación Aplicada</p> <p>Nivel de investigación Descriptiva Explicativa</p> <p>Métodos de investigación Campo Pre experimental</p>	<p>Población Para esta investigación N es Indeterminado.</p> <p>Tipo de muestreo: Muestra No probabilística</p> <p>Tamaño de la Muestra: 30 procesos de Prestación Bancaria en una entidad financiera formalizada con flujo de caja a nivel nacional</p>	<p>Variable dependiente Prestación bancaria</p> <p>Técnica Observación indirecta</p> <p>Instrumento Autor: Jorge Cano Chuqui Instrumento: Ficha de observación Cantidad de indicadores: 4 Ámbito de la aplicación: Procesos (transacciones) Forma de administración: Individual</p>	<p>Descriptiva Se utilizaron estadísticas descriptivas para caracterizar las cantidades y los porcentajes de los resultados producidos en todo el proceso bancario, así como para cada uno de sus indicadores, y para mostrar los datos del grupo experimental.</p> <p>Inferencial Se utilizó la prueba de Anderson-Darling para establecer la normalidad del conjunto de datos, ya que está adecuadamente calibrada para las pruebas que ejecutaría el programa Minitab. Dado que los resultados indicaron que una de las distribuciones de los indicadores era anormal, se siguió la utilización de pruebas no paramétricas para muestras independientes como la prueba U de Mann-Whitney, lo que permitió comprobar la hipótesis de investigación cualitativa, mientras que la prueba t de Student se utilizó para evidencias las distribuciones de los indicadores que mostraban un comportamiento normal, en consecuencia se comprobaron las hipótesis cuantitativas. La herramienta estadística Minitab 2.0 fue empleada para procesar los datos.</p>															

Fuente: Elaboración propia

Anexo 2: Matriz de operacionalización

Tabla 22

Matriz de operacionalización

VARIABLES	DEFINICIÓN CONCEPTUAL	DEFINICIÓN OPERACIONAL	DIMENSIÓN	INDICADOR	ESCALA DE MEDICIÓN
Machine Learning	Según (Taddy 2019), Machine Learning es el campo de pensamiento sobre cómo crear automáticamente predicciones sólidas a partir de datos complejos. Esto está estrechamente relacionado con las estadísticas modernas y, de hecho, muchas de las mejores ideas para el aprendizaje automático provienen de estadísticos (regresión, clasificación, bosques, etc.).				
Proceso de Prestación bancaria	Según (Meikang 2018), la solicitud de préstamo es iniciada por el prestatario. La solicitud de préstamo especifica la cantidad y la fecha de pago del préstamo. Las solicitudes de préstamo se envían a través de contratos de préstamo en toda la red y cualquier cuenta de préstamo puede recibir solicitudes de préstamo.	La variable proceso de prestación bancaria está constituida por cuatro dimensiones y cuatro indicadores: tiempo (tiempo de espera), económico (costo de personal), satisfacción (satisfacción del cliente) y préstamo (eficiencia de evaluación), asimismo se presenta 30 mediciones por indicador, que serán evaluados por una prueba, cuya puntuación por indicador depende del índice de valores al ser realizado en una ficha de observación, ver anexo 1. Para lo suyo se llevaron a cabo las siguientes etapas 1) Entendimiento de los datos; 2) Selección de los datos; 3).	TIEMPO	$TDE = \text{Rendonder}(\text{Fracción decimal (TR)}, 4)$ TDE = Tiempo de espera TR = Tiempo real (hh:mm:ss)	Razón
			ECONÓMICO	$CDP = \text{Rendonder}((CPM * TDE), 4)$ CDP = Costo de personal CPM = Costo de personal por minuto TDE = Tiempo de espera	Razón
			SATISFACCIÓN	$SDC = RMF$ SDC = Satisfacción del cliente RMF = Resultado más frecuente (escala de Likert)	Ordinal
			PRÉSTAMO	$EFE = \frac{TOj}{TOt} * 100$ EFE = Eficiencia de evaluación TO = Tiempo objetivo TO = Tiempo obtenido	Razón

Fuente: Elaboración propia

Anexo 3: Certificado de validez de los indicadores para medir la variable prestación bancaria

CERTIFICADO DE VALIDEZ DE CONTENIDO DEL INSTRUMENTO QUE MIDE:

N°	DIMENSIONES	PERTINENCIA ¹		RELEVANCIA ²		CLARIDAD ³		OBSERVACIONES
		SI	NO	SI	NO	SI	NO	
	DIMESIÓN 1: TIEMPO	SI	NO	SI	NO	SI	NO	
1	<p>Tiempo de Espera (Fuentes y Zúñiga 2010)</p> <p style="text-align: center;">$TDE = \text{Rendonder}(\text{Fracción decimal}(\text{TR}), 0)$</p> <p>TDE = Tiempo de espera TR = Tiempo real (hh:mm:ss)</p>	X		X		X		
	DIMENSIÓN 2: ECONÓMICO	SI	NO	SI	NO	SI	NO	
2	<p>Costo de Personal (Fuentes y Zúñiga 2010)</p> <p style="text-align: center;">$CDP = \text{Rendonder}(CPM * TDE)$</p> <p>CDP = Costo de personal CPM = Costo de personal por minuto TDE = Tiempo de espera</p>	X		X		X		
	DIMENSIÓN 3: SATISFACCIÓN							
3	<p>Satisfacción del Cliente (Fuentes y Zúñiga 2010)</p>	X		X		X		

	$SDC = RMF$							
	SDC = Satisfacción del cliente RMF = Resultado más frecuente (escala de Likert)							
	DIMENSIÓN 4: PRÉSTAMO							
4	Eficiencia de Evaluación (Fuentes y Zúñiga 2010) $EFE = \frac{TOt}{TOj} * 100$ EFE = Eficiencia de evaluación TOj = Tiempo objetivo TOt = Tiempo obtenido	X		X		X		

Observación (precisar si hay suficiencia):

Opinión de aplicabilidad: aplicable [X] aplicable después de corregir [] no aplicable []

Apellidos y Nombres del experto validador: Mg. Crispín Sánchez Iván **DNI:** 09926119

28 de Junio de 2022

Especialidad del validador: Ingeniero de sistemas



¹**Pertinencia:** El ítem corresponde al concepto teórico formulado.

²**Relevancia:** El ítem es apropiado para representar el componente o dimensión específica del constructo

³**Claridad:** Se entiende sin dificultad alguna el enunciado del ítem, es conciso, exacto y directo.

Nota: Suficiencia, se dice de suficiencia cuando los ítems planteados son suficientes para medir la dimensión.

Firma del Experto Informante.

CERTIFICADO DE VALIDEZ DE CONTENIDO DEL INSTRUMENTO QUE MIDE:

N°	DIMENSIONES	PERTINENCIA ¹		RELEVANCIA ²		CLARIDAD ³		OBSERVACIONES
		SI	NO	SI	NO	SI	NO	
	DIMESIÓN 1: TIEMPO							
1	<p>Tiempo de Espera (Fuentes y Zúñiga 2010)</p> <p style="text-align: center;">$TDE = \text{Rendonder}(\text{Fracción decimal}(\text{TR}), 0)$</p> <p>TDE = Tiempo de espera TR = Tiempo real (hh:mm:ss)</p>	X		X		X		
	DIMENSIÓN 2: ECONÓMICO							
2	<p>Costo de Personal (Fuentes y Zúñiga 2010)</p> <p style="text-align: center;">$CDP = \text{Rendonder}(CPM * TDE)$</p> <p>CDP = Costo de personal SPM = Costo de personal por minuto TDE = Tiempo de espera</p>	X		X		X		
	DIMENSIÓN 3: SATISFACCIÓN							
3	<p>Satisfacción del Cliente (Fuentes y Zúñiga 2010)</p>	X		X		X		

	$SDC = RMF$							
	SDC = Satisfacción del cliente RMF = Resultado más frecuente (escala de likert)							
	DIMENSIÓN 4: PRÉSTAMO							
4	Eficiencia de Evaluación (Fuentes y Zúñiga 2010) $EFE = \frac{TOt}{TOj} * 100$ EFE = Eficiencia de evaluación TOj = Tiempo objetivo TOt = Tiempo obtenido	X		X		X		

Observación (precisar si hay suficiencia):

Opinión de aplicabilidad: aplicable [X] aplicable después de corregir [] no aplicable []

Apellidos y Nombres del experto validador: Mg. Ogozi Auqui, José Antonio **DNI:** 42870080

23 de Mayo de 2022

Especialidad del validador: Maestro en Gestión de Tecnologías de Información

¹**Pertinencia:** El ítem corresponde al concepto teórico formulado.

²**Relevancia:** El ítem es apropiado para representar el componente o dimensión específica del constructo

³**Claridad:** Se entiende sin dificultad alguna el enunciado del ítem, es conciso, exacto y directo.

Nota: Suficiencia, se dice de suficiencia cuando los ítems planteados son suficientes para medir la dimensión.



Firma del Experto Informante.

CERTIFICADO DE VALIDEZ DE CONTENIDO DEL INSTRUMENTO QUE MIDE:

N°	DIMENSIONES	PERTINENCIA ¹		RELEVANCIA ²		CLARIDAD ³		OBSERVACIONES
		SI	NO	SI	NO	SI	NO	
	DIMESIÓN 1: TIEMPO							
1	<p>Tiempo de Espera (Fuentes y Zúñiga 2010)</p> <p style="text-align: center;">$TDE = \text{Rendondear}(\text{Fracción decimal}(\text{TR}), 0)$</p> <p>TDE = Tiempo de espera TR = Tiempo real (hh:mm:ss)</p>	X		X		X		
	DIMENSIÓN 2: ECONÓMICO							
2	<p>Costo de Personal (Fuentes y Zúñiga 2010)</p> <p style="text-align: center;">$CDP = \text{Rendondear}(SPM * TDE)$</p> <p>CDP = Costo de personal SPM = Suelo por minuto TDE = Tiempo de espera</p>	X		X		X		
	DIMENSIÓN 3: SATISFACCIÓN							
3	<p>Satisfacción del Cliente (Fuentes y Zúñiga 2010)</p>	X		X		X		

	$SDC = RMF$							
	SDC = Satisfacción del cliente RMF = Resultado más frecuente (escala de likert)							
	DIMENSIÓN 4: PRÉSTAMO							
4	Eficiencia de Evaluación (Fuentes y Zúñiga 2010) $EFE = \frac{TOt}{TOj} * 100$ EFE = Eficiencia de evaluación TOj = Tiempo objetivo TOt = Tiempo obtenido	X		X		X		

Observación (precisar si hay suficiencia):

Opinión de aplicabilidad: aplicable [X] aplicable después de corregir [] no aplicable []

Apellidos y Nombres del experto validador: Dr. Petrlik Azabache, Iván Carlo **DNI:** 10140461

23 de Mayo de 2022

Especialidad del validador: Doctor en Ingeniería de Sistemas

¹**Pertinencia:** El ítem corresponde al concepto teórico formulado.

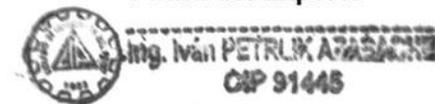
²**Relevancia:** El ítem es apropiado para representar el componente o dimensión específica del constructo

³**Claridad:** Se entiende sin dificultad alguna el enunciado del ítem, es conciso, exacto y directo.

Nota: Suficiencia, se dice de suficiencia cuando los ítems planteados son suficientes para medir la dimensión.



Firma del Experto



Firma del Experto Informante.

Anexo 4: Tabla de evaluación de expertos: Lenguaje de programación

**TABLA DE EVALUACIÓN DE EXPERTOS
LENGUAJE DE PROGRAMACIÓN**

Apellidos y Nombres del Experto:	Crispín Sánchez Iván
Título y/o Grado Académico:	Ingeniero de Sistemas
Doctor [] Magister [x] Ingeniero [] Licenciado [] Otro []	
Fecha:	28/06/2022

Autor:

Cano Chuqui, Jorge

1. MALO 2. REGULAR 3. BUENO

Mediante la tabla de evaluación de expertos. Usted tiene la facultad de evaluar el Lenguaje de programación

ÍTEM	PREGUNTAS	LENGUAJE DE PROGRAMACIÓN		
		C++	PYTHON	JAVA
1	¿Permite trabajar con librerías relacionadas al machine learning?	2	3	2
2	¿Maneja la disponibilidad de librerías?	2	3	3
3	¿Compatibilidad con IDE de desarrollo?	2	3	3
4	¿Permite minimizar el código de programación?	2	3	2
PUNTUACIÓN		8	12	9

FIRMA DEL EXPERTO



SUGERENCIAS

--

TABLA DE EVALUACIÓN DE EXPERTOS
LENGUAJE DE PROGRAMACIÓN

Apellidos y Nombres del Experto:

Ogosi Auqui José Antonio

Título y/o Grado Académico:

Ingeniero de Sistemas

Doctor [] Magister [x] Ingeniero [] Licenciado [] Otro []

Fecha:

23/05/2022

Autor:

Cano Chuqui, Jorge

1. MALO 2. REGULAR 3. BUENO

Mediante la tabla de evaluación de expertos. Usted tiene la facultad de evaluar el Lenguaje de programación

ÍTEM	PREGUNTAS	LENGUAJE DE PROGRAMACIÓN		
		C++	PYTHON	JAVA
1	¿Permite trabajar con librerías relacionadas al machine learning?	1	3	1
2	¿Maneja la disponibilidad de librerías?	2	2	3
3	¿Compatibilidad con IDE de desarrollo?	3	2	3
4	¿Permite minimizar el código de programación?	1	3	1
PUNTUACIÓN		7	10	8

FIRMA DEL EXPERTO



SUGERENCIAS

**TABLA DE EVALUACIÓN DE EXPERTOS
LENGUAJE DE PROGRAMACIÓN**

Apellidos y Nombres del Experto:

Petrlík Azabache Iván Carlo

Título y/o Grado Académico:

Ingeniero de Sistemas

Doctor Magister Ingeniero Licenciado Otro

Fecha:

23/05/2022

Autor:

Cano Chuqui, Jorge

1. MALO 2. REGULAR 3. BUENO

Mediante la tabla de evaluación de expertos. Usted tiene la facultad de evaluar el Lenguaje de programación

ÍTEM	PREGUNTAS	LENGUAJE DE PROGRAMACIÓN		
		C++	PYTHON	JAVA
1	¿Permite trabajar con librerías relacionadas al machine learning?	1	3	1
2	¿Maneja la disponibilidad de librerías?	1	3	2
3	¿Compatibilidad con IDE de desarrollo?	2	3	2
4	¿Permite minimizar el código de programación?	1	3	1
PUNTUACIÓN		5	12	6

FIRMA DEL EXPERTO



SUGERENCIAS

Anexo 5: Tabla de evaluación de expertos: Metodología de desarrollo

TABLA DE EVALUACIÓN DE EXPERTOS
METODOLOGÍA DE DESARROLLO

Apellidos y Nombres del Experto:

Crispín Sánchez Iván

Título y/o Grado Académico:

Ingeniero de Sistemas

Doctor [] Magister [x] Ingeniero [] Licenciado [] Otro []

Fecha:

28/06/2022

Autor:

Cano Chuqui, Jorge

1. MALO 2. REGULAR 3. BUENO

Mediante la tabla de evaluación de expertos. Usted tiene la facultad de evaluar el Lenguaje de programación

ÍTEM	PREGUNTAS	METODOLOGÍA DE DESARROLLO		
		KDD	CRISP-DM	SEMMA
1	¿Proceso de identificación clara anticipándose a los posibles problemas?	2	3	2
2	¿Admite múltiples técnicas de minería de datos?	2	3	3
3	¿Repite el procedimiento?	3	3	3
4	¿Posee documentación con casos típicos de uso que ayudan en este proceso?	2	3	3
PUNTUACIÓN		11	12	11

FIRMA DEL EXPERTO



SUGERENCIAS

TABLA DE EVALUACIÓN DE EXPERTOS
METODOLOGÍA DE DESARROLLO

Apellidos y Nombres del Experto:

Ogosi Auqui José Antonio

Título y/o Grado Académico:

Ingeniero de Sistemas

Doctor [] Magister [x] Ingeniero [] Licenciado [] Otro []

Fecha:

23/05/2022

Autor:

Cano Chuqui, Jorge

1. MALO 2. REGULAR 3. BUENO

Mediante la tabla de evaluación de expertos. Usted tiene la facultad de evaluar el Lenguaje de programación

ÍTEM	PREGUNTAS	METODOLOGÍA DE DESARROLLO		
		KDD	CRISP-DM	SEMMA
1	¿Proceso de identificación clara anticipándose a los posibles problemas?	1	2	2
2	¿Admite múltiples técnicas de minería de datos?	2	2	2
3	¿Repite el procedimiento?	1	3	1
4	¿Posee documentación con casos típicos de uso que ayudan en este proceso?	3	3	3
PUNTUACIÓN		7	10	8

FIRMA DEL EXPERTO



SUGERENCIAS

TABLA DE EVALUACIÓN DE EXPERTOS
METODOLOGÍA DE DESARROLLO

Apellidos y Nombres del Experto:

Petrlík Azabache Iván Carlo

Título y/o Grado Académico:

Ingeniero de Sistemas

Doctor Magister Ingeniero Licenciado Otro

Fecha:

23/05/2022

Autor:

Cano Chuqui, Jorge

1. MALO 2. REGULAR 3. BUENO

Mediante la tabla de evaluación de expertos. Usted tiene la facultad de evaluar el Lenguaje de programación

ÍTEM	PREGUNTAS	METODOLOGÍA DE DESARROLLO		
		KDD	CRISP-DM	SEMMA
1	¿Proceso de identificación clara anticipándose a los posibles problemas?	2	3	2
2	¿Admite múltiples técnicas de minería de datos?	2	3	3
3	¿Repite el procedimiento?	1	3	2
4	¿Posee documentación con casos típicos de uso que ayudan en este proceso?	2	3	2
PUNTUACIÓN		7	12	9

FIRMA DEL EXPERTO



SUGERENCIAS

Anexo 6: Tabla de evaluación de expertos: Gestor de base de datos

TABLA DE EVALUACIÓN DE EXPERTOS

GESTOR DE BASE DE DATOS

Apellidos y Nombres del Experto:

Crispín Sánchez Iván

Título y/o Grado Académico:

Ingeniero de Sistemas

Doctor [] Magister [x] Ingeniero [] Licenciado [] Otro []

Fecha:

28/06/2022

Autor:

Cano Chuqui, Jorge

1. MALO 2. REGULAR 3. BUENO

Mediante la tabla de evaluación de expertos. Usted tiene la facultad de evaluar el Gestor de Base de datos

ÍTEM	PREGUNTAS	Gestor de Base de datos		
		IBMDB2	FIREBASE	ORACLE
1	¿Qué tan bueno es administrando la información?	2	3	2
2	¿Los recursos que consume lo hace?	3	3	3
3	¿Hay buena documentación?	3	3	3
4	¿En el manejo de base de datos relacional o no relacional es?	3	3	2
PUNTUACIÓN		11	12	10

FIRMA DEL EXPERTO



SUGERENCIAS

TABLA DE EVALUACIÓN DE EXPERTOS
GESTOR DE BASE DE DATOS

Apellidos y Nombres del Experto:

Ogosi Auqui José Antonio

Título y/o Grado Académico:

Ingeniero de Sistemas

Doctor [] Magister [x] Ingeniero [] Licenciado [] Otro []

Fecha:

23/05/2022

Autor:

Cano Chuqui, Jorge

1. MALO 2. REGULAR 3. BUENO

Mediante la tabla de evaluación de expertos. Usted tiene la facultad de evaluar el Gestor de Base de datos

ÍTEM	PREGUNTAS	Gestor de Base de datos		
		IBMDB2	FIREBASE	ORACLE
1	¿Qué tan bueno es administrando la información?	1	1	2
2	¿Los recursos que consume lo hace?	1	3	1
3	¿Hay buena documentación?	2	2	2
4	¿En el manejo de base de datos relacional o no relacional es?	2	2	2
PUNTUACIÓN		6	8	7

FIRMA DEL EXPERTO



SUGERENCIAS

**TABLA DE EVALUACIÓN DE EXPERTOS
GESTOR DE BASE DE DATOS**

Apellidos y Nombres del Experto:

Petrlík Azabache Iván Carlo

Título y/o Grado Académico:

Ingeniero de Sistemas

Doctor [X] Magister [] Ingeniero [] Licenciado [] Otro []

Fecha:

23/05/2022

Autor:

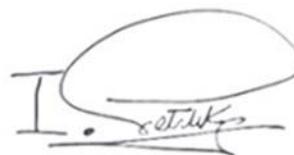
Cano Chuqui, Jorge

1. MALO 2. REGULAR 3. BUENO

Mediante la tabla de evaluación de expertos. Usted tiene la facultad de evaluar el Gestor de Base de datos

ÍTEM	PREGUNTAS	Gestor de Base de datos		
		IBMDB2	FIREBASE	ORACLE
1	¿Qué tan bueno es administrando la información?	2	2	3
2	¿Los recursos que consume lo hace?	2	3	1
3	¿Hay buena documentación?	2	3	2
4	¿En el manejo de base de datos relacional o no relacional es?	3	3	2
PUNTUACIÓN		9	11	8

FIRMA DEL EXPERTO



SUGERENCIAS

Anexo 7: Ficha de observación del indicador TDE: Tiempo de espera

MACHINE LEARNING PARA MEJORAR EL PROCESO DE PRESTACIÓN BANCARIA EN UNA ENTIDAD FINANCIERA	
<u>DIMENSIÓN:</u>	TIEMPO
<u>INDICADOR:</u>	TIEMPO DE ESPERA
<u>INVESTIGADOR:</u>	CANO CHUQUI JORGE
<u>LUGAR DE ESTUDIO:</u>	BANCO ALFIN
<u>FÓRMULA:</u>	$TDE = \text{Rondear} (\text{Fracción decimal } (TR), 4)$ TDE = Tiempo de espera TR = Tiempo real (hh:mm:ss)

Indicador 1 TDE: Tiempo de espera

N	PRE PRUEBA		POST PRUEBA	
	TDEa	TR	TDEd	TR
1	8.6207	00:08:37	2.1833	00:02:11
2	1.7391	00:01:44	1.4667	00:01:28
3	2.2727	00:02:16	1.2833	00:01:17
4	4.2105	00:04:12	2.9000	00:02:54
5	9.2857	00:09:17	2.2667	00:02:16
6	2.5581	00:02:33	2.5667	00:02:34
7	11.6000	00:11:36	2.0167	00:02:01
8	7.4194	00:07:25	2.7167	00:02:43
9	2.2727	00:02:16	2.9500	00:02:57
10	13.4783	00:13:28	1.9000	00:01:54
11	2.2727	00:02:16	1.5833	00:01:35
12	6.3636	00:06:21	1.0333	00:01:02
13	5.0000	00:05:00	1.2500	00:01:15
14	6.3636	00:06:21	1.5500	00:01:33
15	6.3636	00:06:21	2.8500	00:02:51
16	2.5581	00:02:33	2.9000	00:02:54
17	6.8750	00:06:52	2.2667	00:02:16

18	10.0000	00:10:00	1.2167	00:01:13
19	9.2857	00:09:17	2.5833	00:02:35
20	5.8824	00:05:52	1.2333	00:01:14
21	6.8750	00:06:52	2.4833	00:02:29
22	1.4894	00:01:29	1.2333	00:01:14
23	2.2727	00:02:16	1.3333	00:01:20
24	5.4286	00:05:25	2.5833	00:02:35
25	13.4783	00:13:28	2.1833	00:02:11
26	10.7692	00:10:46	2.3333	00:02:20
27	8.0000	00:08:00	1.6333	00:01:38
28	6.3636	00:06:21	2.6500	00:02:39
29	2.8571	00:02:51	1.9667	00:01:58
30	2.2727	00:02:16	1.1833	00:01:11

Anexo 8: Ficha de observación del indicador CDP: Costo de personal

MACHINE LEARNING PARA MEJORAR EL PROCESO DE PRESTACIÓN BANCARIA EN UNA ENTIDAD FINANCIERA	
<u>DIMENSIÓN:</u>	ECONÓMICO
<u>INDICADOR:</u>	COSTO DE PERSONAL
<u>INVESTIGADOR:</u>	CANO CHUQUI JORGE
<u>LUGAR DE ESTUDIO:</u>	BANCO ALFIN
<u>FÓRMULA:</u>	$CDP = \text{Rendondear} ((CPM * TDE), 4)$ <p style="text-align: center;"> CDP = COSTO DE PERSONAL CPM = COSTO DE PERSONAL POR MINUTO TDE = TIEMPO DE ESPERA </p>

Indicador 2 CDP: Costo de personal

N	PRE PRUEBA			POST PRUEBA		
	CPDa	CPM	TDE	CDPd	CPM	TDE
1	2.7586	0.32	8.6207	0.6987	0.32	2.1833
2	0.5565	0.32	1.7391	0.4693	0.32	1.4667
3	0.7273	0.32	2.2727	0.4107	0.32	1.2833
4	1.3474	0.32	4.2105	0.9280	0.32	2.9000
5	2.9714	0.32	9.2857	0.7253	0.32	2.2667
6	0.8186	0.32	2.5581	0.8213	0.32	2.5667
7	3.7120	0.32	11.6000	0.6453	0.32	2.0167
8	2.3742	0.32	7.4194	0.8693	0.32	2.7167
9	0.7273	0.32	2.2727	0.9440	0.32	2.9500
10	4.3131	0.32	13.4783	0.6080	0.32	1.9000
11	0.7273	0.32	2.2727	0.5067	0.32	1.5833
12	2.0364	0.32	6.3636	0.3307	0.32	1.0333
13	1.6000	0.32	5.0000	0.4000	0.32	1.2500
14	2.0364	0.32	6.3636	0.4960	0.32	1.5500
15	2.0364	0.32	6.3636	0.9120	0.32	2.8500

16	0.8186	0.32	2.5581	0.9280	0.32	2.9000
17	2.2000	0.32	6.8750	0.7253	0.32	2.2667
18	3.2000	0.32	10.0000	0.3893	0.32	1.2167
19	2.9714	0.32	9.2857	0.8267	0.32	2.5833
20	1.8824	0.32	5.8824	0.3947	0.32	1.2333
21	2.2000	0.32	6.8750	0.7947	0.32	2.4833
22	0.4766	0.32	1.4894	0.3947	0.32	1.2333
23	0.7273	0.32	2.2727	0.4267	0.32	1.3333
24	1.7372	0.32	5.4286	0.8267	0.32	2.5833
25	4.3131	0.32	13.4783	0.6987	0.32	2.1833
26	3.4461	0.32	10.7692	0.7467	0.32	2.3333
27	2.5600	0.32	8.0000	0.5227	0.32	1.6333
28	2.0364	0.32	6.3636	0.8480	0.32	2.6500
29	0.9143	0.32	2.8571	0.6293	0.32	1.9667
30	0.7273	0.32	2.2727	0.3787	0.32	1.1833

Anexo 9: Ficha de observación del indicador CDP: Satisfacción del cliente

MACHINE LEARNING PARA MEJORAR EL PROCESO DE PRESTACIÓN BANCARIA EN UNA ENTIDAD FINANCIERA	
<u>DIMENSIÓN:</u>	SATISFACCIÓN
<u>INDICADOR:</u>	SATISFACCIÓN DEL CLIENTE
<u>INVESTIGADOR:</u>	CANO CHUQUI JORGE
<u>LUGAR DE ESTUDIO:</u>	BANCO ALFIN
<u>FÓRMULA:</u>	$SDC = RMF$ SDC = SATISFACCIÓN DEL CLIENTE RMF = RESULTADO MÁS FRECUENTE (ESCALA DE LIKERT)

Indicador 3 SDC: Satisfacción del personal

N	PRE PRUEBA		POST PRUEBA	
	SDCa	RMF	SDCD	RMF
1	5	Totalmente de acuerdo	4	De acuerdo
2	4	De acuerdo	3	Indeciso
3	4	De acuerdo	5	Totalmente de acuerdo
4	1	Totalmente en desacuerdo	5	Totalmente de acuerdo
5	3	Indeciso	4	De acuerdo
6	4	De acuerdo	4	De acuerdo
7	1	Totalmente en desacuerdo	3	Indeciso
8	3	Indeciso	4	De acuerdo
9	5	Totalmente de acuerdo	5	Totalmente de acuerdo
10	1	Totalmente en desacuerdo	4	De acuerdo
11	4	De acuerdo	5	Totalmente de acuerdo
12	2	En desacuerdo	5	Totalmente de acuerdo
13	4	De acuerdo	5	Totalmente de acuerdo

14	2	En desacuerdo	3	Indeciso
15	5	Totalmente de acuerdo	3	Indeciso
16	1	Totalmente en desacuerdo	5	Totalmente de acuerdo
17	3	Indeciso	5	Totalmente de acuerdo
18	5	Totalmente de acuerdo	5	Totalmente de acuerdo
19	3	Indeciso	3	Indeciso
20	1	Totalmente en desacuerdo	4	De acuerdo
21	5	Totalmente de acuerdo	5	Totalmente de acuerdo
22	3	Indeciso	5	Totalmente de acuerdo
23	4	De acuerdo	3	Indeciso
24	4	De acuerdo	5	Totalmente de acuerdo
25	3	Indeciso	4	De acuerdo
26	3	Indeciso	3	Indeciso
27	3	Indeciso	5	Totalmente de acuerdo
28	3	Indeciso	5	Totalmente de acuerdo
29	5	Totalmente de acuerdo	5	Totalmente de acuerdo
30	3	Indeciso	3	Indeciso

Anexo 10: Ficha de observación del indicador EFE: Eficiencia de evaluación

MACHINE LEARNING PARA MEJORAR EL PROCESO DE PRESTACIÓN BANCARIA EN UNA ENTIDAD FINANCIERA	
<u>DIMENSIÓN:</u>	PRÉSTAMO
<u>INDICADOR:</u>	EFICIENCIA DE EVALUACIÓN
<u>INVESTIGADOR:</u>	CANO CHUQUI JORGE
<u>LUGAR DE ESTUDIO:</u>	BANCO ALFIN
<u>FÓRMULA:</u>	$EFE = \frac{TOt}{TOj} * 100$ <p>EFE = EFICIENCIA DE EVALUACIÓN TOj = TIEMPO OBJETIVO TOt = TIEMPO OBTENIDO</p>

Indicador 4 EFE: Eficiencia de evaluación

N	PRE PRUEBA			POST PRUEBA		
	EFEa	TOj	TOt	EFEd	TOj	TOt
1	86.2070	10	8.6207	21.8330	10	2.1833
2	17.3910	10	1.7391	14.6670	10	1.4667
3	22.7270	10	2.2727	12.8330	10	1.2833
4	42.1050	10	4.2105	29.0000	10	2.9000
5	92.8570	10	9.2857	22.6670	10	2.2667
6	25.5810	10	2.5581	25.6670	10	2.5667
7	116.0000	10	11.6000	20.1670	10	2.0167
8	74.1940	10	7.4194	27.1670	10	2.7167
9	22.7270	10	2.2727	29.5000	10	2.9500
10	134.7830	10	13.4783	19.0000	10	1.9000
11	22.7270	10	2.2727	15.8330	10	1.5833
12	63.6360	10	6.3636	10.3330	10	1.0333
13	50.0000	10	5.0000	12.5000	10	1.2500
14	63.6360	10	6.3636	15.5000	10	1.5500
15	63.6360	10	6.3636	28.5000	10	2.8500

16	25.5810	10	2.5581	29.0000	10	2.9000
17	68.7500	10	6.8750	22.6670	10	2.2667
18	100.0000	10	10.0000	12.1670	10	1.2167
19	92.8570	10	9.2857	25.8330	10	2.5833
20	58.8240	10	5.8824	12.3330	10	1.2333
21	68.7500	10	6.8750	24.8330	10	2.4833
22	14.8940	10	1.4894	12.3330	10	1.2333
23	22.7270	10	2.2727	13.3330	10	1.3333
24	54.2860	10	5.4286	25.8330	10	2.5833
25	134.7830	10	13.4783	21.8330	10	2.1833
26	107.6920	10	10.7692	23.3330	10	2.3333
27	80.0000	10	8.0000	16.3330	10	1.6333
28	63.6360	10	6.3636	26.5000	10	2.6500
29	28.5710	10	2.8571	19.6670	10	1.9667
30	22.7270	10	2.2727	11.8330	10	1.1833

Anexo 11: Análisis a profundidad del tiempo de espera

Tabla 23

I1: Tiempo de espera

Preprueba	Postprueba		
8.621	2.183	2.183	2.183
1.739	1.467	1.467	1.467
2.273	1.283	1.283	1.283
4.211	2.900	2.900	2.900
9.286	2.267	2.267	2.267
2.558	2.567	2.567	2.567
11.600	2.017	2.017	2.017
7.419	2.717	2.717	2.717
2.273	2.950	2.950	2.950
13.478	1.900	1.900	1.900
2.273	1.583	1.583	1.583
6.364	1.033	1.033	1.033
5.000	1.250	1.250	1.250
6.364	1.550	1.550	1.550
6.364	2.850	2.850	2.850
2.558	2.900	2.900	2.900
6.875	2.267	2.267	2.267
10.000	1.217	1.217	1.217
9.286	2.583	2.583	2.583
5.882	1.233	1.233	1.233
6.875	2.483	2.483	2.483
1.489	1.233	1.233	1.233
2.273	1.333	1.333	1.333
5.429	2.583	2.583	2.583
13.478	2.183	2.183	2.183
10.769	2.333	2.333	2.333
8.000	1.633	1.633	1.633
6.364	2.650	2.650	2.650
2.857	1.967	1.967	1.967
2.273	1.183	1.183	1.183
6.141	2.010		
Meta Planteada	6.000		
N° menor al Promedio	14	0	30
% menor al Promedio	47%	100%	100%

Promedio

Meta Planteada

N° menor al

Promedio

% menor al

Promedio

Fuente: Elaboración propia

La primera columna son los datos de la preprueba que no tienen color como la segunda, tercera y cuarta columna, porque son independientes de este análisis. El 2.010 colocado es el promedio de los datos de la postprueba, lo cual no significa que se tomaron 90 muestras, sino, es el promedio de las 30 muestras establecidas en la sección "3.3. Población, muestra y muestreo". Este 2.010 se relaciona con la segunda columna.

El 6.000 colocado es una meta planteada por el investigador, basado en su empirismo. Este 6.000 se relaciona con la tercera columna.

El 6.141 fue puesto porque es un promedio de los datos de la primera columna que corresponden al momento preprueba. Este 6.141 se relaciona con la cuarta columna.

En la segunda columna de la tabla, las celdas son blancas porque los valores en ellas son mayores al promedio de los datos postprueba, si son menores se resaltan de verdes. En consecuencia, el 47.00% de los tiempos de espera en la postprueba fueron más bajos que el promedio del tiempo de espera en la postprueba. En la tercera columna de la tabla, las celdas son blancas porque los valores en ellas son mayores a la meta planteada por el investigador, si son menores se resaltan de azules. Por consiguiente, el 100.00% de los tiempos de espera en la postprueba fueron menores que el tiempo de espera objetivo establecido. En la cuarta columna de la tabla, las celdas son blancas porque los valores en ellas son mayores al promedio de los datos preprueba, si son menores se resaltan de rojos. En tal sentido, el 100% de los tiempos de espera en la postprueba fueron más cortos que el tiempo promedio de espera en la preprueba. Los valores de la segunda, tercera y cuarta columna de la tabla son iguales porque son los únicos resultados que obtuvimos de la postprueba.

Anexo 12: Análisis a profundidad del costo de personal

Tabla 24

I2: Costo de personal

Preprueba	Postprueba		
2.759	0.699	0.699	0.699
0.557	0.469	0.469	0.469
0.727	0.411	0.411	0.411
1.347	0.928	0.928	0.928
2.971	0.725	0.725	0.725
0.819	0.821	0.821	0.821
3.712	0.645	0.645	0.645
2.374	0.869	0.869	0.869
0.727	0.944	0.944	0.944
4.313	0.608	0.608	0.608
0.727	0.507	0.507	0.507
2.036	0.331	0.331	0.331
1.600	0.400	0.400	0.400
2.036	0.496	0.496	0.496
2.036	0.912	0.912	0.912
0.819	0.928	0.928	0.928
2.200	0.725	0.725	0.725
3.200	0.389	0.389	0.389
2.971	0.827	0.827	0.827
1.882	0.395	0.395	0.395
2.200	0.795	0.795	0.795
0.477	0.395	0.395	0.395
0.727	0.427	0.427	0.427
1.737	0.827	0.827	0.827
4.313	0.699	0.699	0.699
3.446	0.747	0.747	0.747
2.560	0.523	0.523	0.523
2.036	0.848	0.848	0.848
0.914	0.629	0.629	0.629
0.727	0.379	0.379	0.379
1.965	0.643		
Meta Planteada	3.000		
N° menor al Promedio	14	30	30
% menor al Promedio	47%	100%	100%

Promedio

Meta Planteada

N° menor al

Promedio

% menor al

Promedio

Fuente: Elaboración propia

La primera columna son los datos de la preprueba que no tienen color como la segunda, tercera y cuarta columna, porque son independientes de este análisis. El 0.643 colocado es el promedio de los datos de la postprueba, lo cual no significa que se tomaron 90 muestras, sino, es el promedio de las 30 muestras establecidas en la sección "3.3. Población, muestra y muestreo". Este 0.643 se relaciona con la segunda columna.

El 3.000 colocado es una meta planteada por el investigador, basado en su empirismo. Este 3.000 se relaciona con la tercera columna.

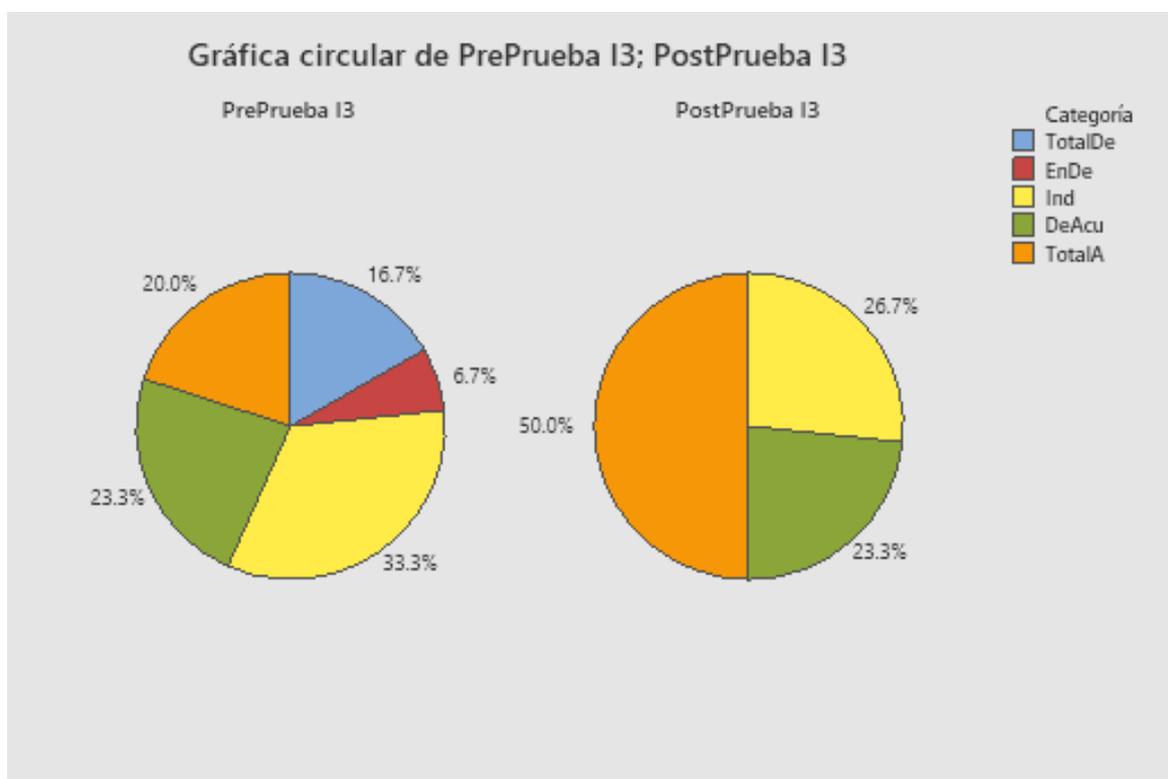
El 1.965 fue puesto porque es un promedio de los datos de la primera columna que corresponden al momento preprueba. Este 1.965 se relaciona con la cuarta columna.

En la segunda columna de postprueba, las celdas son blancas porque los valores en ellas son mayores al promedio de los datos postprueba, si son menores se resaltan de verdes. En consecuencia, el 43.33% de los costos de personal en la postprueba fueron más bajos que el promedio del costo de personal en la postprueba. En la tercera columna de postprueba, las celdas son blancas porque los valores en ellas son mayores a la meta planteada por el investigador, si son menores se resaltan de azules. Por consiguiente, el 100.00% de los costos de personal en la postprueba fueron menores que el costo de personal objetivo establecido. En la cuarta columna de postprueba, las celdas son blancas porque los valores en ellas son mayores al promedio de los datos preprueba, si son menores se resaltan de rojos. En tal sentido, el 100% de los costos de personal en la postprueba fueron más cortos que el costo promedio de personal en la preprueba.

Anexo 13: Análisis a profundidad de la satisfacción del cliente

Figura 31

Gráfico circular preprueba y postprueba de la satisfacción del cliente para la entidad financiera



Fuente: Elaboración propia

Tabla 25

Tabla de frecuencias preprueba de la satisfacción del cliente para la entidad financiera

Estado	Frecuencia
Totalmente en desacuerdo	5
En desacuerdo	2
Indeciso	10
De acuerdo	7
Totalmente de acuerdo	6
Total	30

Fuente: Elaboración propia

Los clientes estuvieron indecisos en el 20,0% de las veces cuando se les pidió en el sistema que calificaran su nivel de satisfacción, que indicaba si era completamente satisfactorio. Los clientes estuvieron de acuerdo en el 23.3% de las veces cuando se les pidió en el sistema que calificaran su nivel de satisfacción, que indicaba si era completamente satisfactorio. Los clientes estuvieron totalmente de acuerdo en el 20.0% de las veces cuando se les pidió en el sistema que calificaran su nivel de satisfacción, que indicaba si era completamente satisfactorio. Los clientes estuvieron totalmente en desacuerdo en el 16.7% de las veces cuando se les pidió en el sistema que calificaran su nivel de satisfacción, que indicaba si era completamente satisfactorio. Los clientes estuvieron en desacuerdo en el 6.7% de las veces cuando se les pidió en el sistema que calificaran su nivel de satisfacción, que indicaba si era completamente satisfactorio.

Tabla 26

Tabla de frecuencias postprueba de la satisfacción del cliente para la entidad financiera

Estado	Frecuencia
Totalmente en desacuerdo	0
En desacuerdo	0
Indeciso	8
De acuerdo	7
Totalmente de acuerdo	15
Total	30

Fuente: Elaboración propia

Los clientes estuvieron totalmente de acuerdo en el 50,0% de las veces cuando se les pidió en el sistema que calificaran su nivel de satisfacción, que indicaba si era completamente satisfactorio. Los clientes estuvieron de acuerdo en el 23.3% de las veces cuando se les pidió en el sistema que calificaran su nivel de satisfacción, que indicaba si era completamente satisfactorio. Los clientes estuvieron indecisos en el 26.7% de las veces cuando se les pidió en el sistema que calificaran su nivel de satisfacción, que indicaba si era completamente satisfactorio.

Anexo 14: Análisis a profundidad de la eficiencia de evaluación

Tabla 27

14: Eficiencia de evaluación

Preprueba	Postprueba		
86.207	21.833	21.833	21.833
17.391	14.667	14.667	14.667
22.727	12.833	12.833	12.833
42.105	29.000	29.000	29.000
92.857	22.667	22.667	22.667
25.581	25.667	25.667	25.667
116.000	20.167	20.167	20.167
74.194	27.167	27.167	27.167
22.727	29.500	29.500	29.500
134.783	19.000	19.000	19.000
22.727	15.833	15.833	15.833
63.636	10.333	10.333	10.333
50.000	12.500	12.500	12.500
63.636	15.500	15.500	15.500
63.636	28.500	28.500	28.500
25.581	29.000	29.000	29.000
68.750	22.667	22.667	22.667
100.000	12.167	12.167	12.167
92.857	25.833	25.833	25.833
58.824	12.333	12.333	12.333
68.750	24.833	24.833	24.833
14.894	12.333	12.333	12.333
22.727	13.333	13.333	13.333
54.286	25.833	25.833	25.833
134.783	21.833	21.833	21.833
107.692	23.333	23.333	23.333
80.000	16.333	16.333	16.333
63.636	26.500	26.500	26.500
28.571	19.667	19.667	19.667
22.727	11.833	11.833	11.833
Promedio	61.410	20.100	
Meta Planteada		60.000	
N° menor al Promedio	19	30	30
% menor al Promedio	63%	100%	100%

Fuente: Elaboración propia

La primera columna son los datos de la preprueba que no tienen color como la segunda, tercera y cuarta columna, porque son independientes de este análisis. El 20.100 colocado es el promedio de los datos de la postprueba, lo cual no significa que se tomaron 90 muestras, sino, es el promedio de las 30 muestras establecidas en la sección "3.3. Población, muestra y muestreo". Este 0.643 se relaciona con la segunda columna.

El 60.000 colocado es una meta planteada por el investigador, basado en su empirismo. Este 60.000 se relaciona con la tercera columna.

El 61.410 fue puesto porque es un promedio de los datos de la primera columna que corresponden al momento preprueba. Este 61.410 se relaciona con la cuarta columna.

En la primera columna de postprueba, las celdas son blancas porque los valores en ellas son mayores al promedio de los datos postprueba, si son menores se resaltan de verdes. En consecuencia, el 63.00% de las eficiencias de evaluación en la postprueba fueron más bajas que el promedio de la eficiencia de evaluación en la postprueba. En la segunda columna de postprueba, las celdas son blancas porque los valores en ellas son mayores a la meta planteada por el investigador, si son menores se resaltan de azules, por consiguiente, el 100.00% de las eficiencias de evaluación en la postprueba fueron menores que eficiencias de evaluación objetivo establecidas. En la tercera columna de postprueba, las celdas son blancas porque los valores en ellas son mayores al promedio de los datos preprueba, si son menores se resaltan de rojos. En tal sentido, el 100% de las eficiencias de evaluación en la postprueba fueron más cortas que la eficiencia promedio de evaluación en la preprueba.

Anexo 15: Constancia de autorización de investigación de tesis



CONSTANCIA DE AUTORIZACIÓN DE INVESTIGACIÓN DE TESIS

Con la presente se autoriza al señor Jorge Cano Chuqui a realizar el trabajo de investigación "Machine learning para mejorar el proceso de prestación bancaria en una entidad financiera" correspondiente en el Banco Alfin, con la finalidad de obtener el grado de Ingeniero de Sistemas,

Se expide la presente a solicitud del interesado, para los fines que se estime conveniente.

Lima, 20 de junio del 2022


Banco Azteca del Perú S.A.
Jesús Martín Ronceros Salhuana
Supervisor de Help Desk

Anexo 16: Constancia de autorización de uso de datos en proyecto de tesis



ONP
Oficina de
Normalización
Previsional

Firmado digitalmente por DIAZ
PICASSO Margarita Maria FAU
20291973851 soft
Motivo: Soy el autor del documento
Fecha: 01.09.2022 09:23:29

“Año del Fortalecimiento de la Soberanía Nacional”

Lima, 23 de Setiembre del 2022

CARTA N° 000168-2022-ONP/GSFP

De: **Margarita María Díaz Picasso**
Gerencia de Supervisión OTI

Para: **Jorge Cano Chuqui**
Ciudadano con DNI: 77929589

Asunto: Autorización para uso de datos en PROYECTO TESIS

Referencia: SOLICITUD N° 00125-2022/EDGT

Tengo a bien dirigirme a usted, con relación al documento de la referencia, mediante el cual el ciudadano Jorge Cano Chuqui, identificado con DNI: 77929589 solicita a esta Gerencia “La Autorización para uso de datos en PROYECTO-TESIS”.

En ese sentido, la Gerencia de Supervisión OTI indica que se ha procedido a la Autorización del correcto uso de datos.

Por consiguiente, se remite el presente documento para los fines pertinentes. No obstante, se sugiere utilizar dicha información para usos estrictamente académicos.

Atentamente.

Firmado digitalmente
MARGARITA MARÍA DIAZ PICASSO
Gerencia de Supervisión
OTI

(MDP/fpt)

Esta es una copia auténtica imprimible de un documento electrónico archivado por la Oficina de Normalización Previsional, aplicando lo dispuesto por el Art. 25 de D.S. 070-2013-PCM y la Tercera Disposición Complementaria Final del D.S. 026- 2016-PCM

Anexo 17: Reportes en Excel

Autoguardado Formato_Buscador_Junio2022 Sistemas--Jess.xlsx • Error al guardar

Archivo Inicio Insertar Disposición de página Fórmulas Datos Revisar Vista Programador Ayuda Comentarios Compartir

A1

A	B	C	D	E	F	G	H	I	J	K	L	M
fecha_base	FcCteNombre	FcCteAPater	FcCteAMaterr	FCDNI	fiPe	fiCar	fiSucurs	fiFol	CU	tien	Sucursal	Region
1/06/2022	SUSAN GABRIELA	ANHUAMAN	ALVAREZ	47552501	6	1	5260	12577	06015260000012577	8252	738252 - DAZ SANTA ANITA	470202 - REGIONAL VENTAS LIMA EST
1/06/2022	NANCY YRMA	VARGAS	MENDIVEL	21505402	6	1	9580	130002	06019580000130002	9580	739580 - BA EKT ICA	470208 - REGIONAL VENTAS ICA
1/06/2022	JULITA FORTUNATA	RIOS	ASCENCIO	15753753	6	1	9470	59939	06019470000059939	9470	739470 - BA EKT HUARAL	470199 - REGIONAL VENTAS NORTE CH
1/06/2022	ZENAIDA	TARAZONA	QUISPE	22415662	6	1	5996	4353	06015996000004353	5996	735996 - BA EKT HUANUCO	479957 - REGIONAL VENTAS CENTRO
1/06/2022	LUZMILA	CESPEDES	DE HERMITA?O	22459301	6	1	5996	33427	06015996000033427	5996	735996 - BA EKT HUANUCO	479957 - REGIONAL VENTAS CENTRO
1/06/2022	RODRIGO CESAR	PINEDA	ZEVALLS	73202494	6	1	2243	44323	06012243000044323	2243	732243 - BA EKT CAÑETE	470208 - REGIONAL VENTAS ICA
1/06/2022	SAMMY JEISON	FLORES	YANQUI	44076273	6	1	2249	77227	06012249000077227	2249	732249 - BA EKT VILLA EL SALVADC	470206 - REGIONAL VENTAS LIMA SUR
1/06/2022	JESICA SILVIA	PLASENCIA	PAICO	46165830	6	1	3826	5612	06013826000005612	8363	738363 - DAZ CAJAMARCA	471021 - REGIONAL VENTAS LAMBAYE
1/06/2022	DIEGO ALBERTH	ASTUQUIPAN	BURBANO	46924544	6	1	5996	22616	06015996000022616	5996	735996 - BA EKT HUANUCO	479957 - REGIONAL VENTAS CENTRO
1/06/2022	JHONY ENRIQUE	QUIROZ	RAMIREZ	42455738	6	1	9849	80177	06019849000080177	9849	739849 - BA EKT LOS OLIVOS	470200 - REGIONAL VENTAS LIMA NOF
1/06/2022	LOURDES DEL CARMEN	MORA	TOLENTINO	42444601	6	1	111	131376	06010111000131376	111	730111 - BA EKT PUENTE PIEDRA	479956 - REGIONAL VENTAS LIMA NOF
1/06/2022	MARIA EUGENIA	TACORA	LLUNGO	43632237	6	1	9629	95396	06019629000095396	9629	739629 - BA EKT ATE VITARTE	470202 - REGIONAL VENTAS LIMA ESTI
1/06/2022	CARLOS RONALD	SARAVIA	TASAYCO	09662138	6	9	8381	14777	06098381000014777	8381	738381 - DAZ ENMANCIPACION	470200 - REGIONAL VENTAS LIMA NOF
1/06/2022	MARIA ISABEL	SALAZAR	PEREZ	48271755	6	1	2246	74452	06012246000074452	2246	732246 - BA EKT LURIN	470206 - REGIONAL VENTAS LIMA SUR
1/06/2022	JUAN DE DIOS	QUISPE	ÑAHUIZ	07895270	6	1	2246	26510	06012246000026510	2246	732246 - BA EKT LURIN	470206 - REGIONAL VENTAS LIMA SUR
1/06/2022	BEATRIZ	RIVEROS	CAMPOS	80180511	6	1	9865	116758	06019865000116758	9865	739865 - BA EKT CANTO GRANDE	470201 - REGIONAL VENTAS SAN JUAN
1/06/2022	MARCELINO	ESPINOZA	CONDORI	23209137	6	1	7870	154403	06017870000154403	7870	737870 - BA EKT VILLA MARIA DEL	470206 - REGIONAL VENTAS LIMA SUR
1/06/2022	GABRIELA	QUISPE	USCAMAYTA	10096250	6	1	2248	26808	06012248000026808	2954	732954 - DAZ JOSE GALVEZ	470206 - REGIONAL VENTAS LIMA SUR
1/06/2022	WENDY SABY	SANDOVAL	RENGIFO	47546625	6	1	9836	48636	06019836000048636	9629	739629 - BA EKT ATE VITARTE	470202 - REGIONAL VENTAS LIMA ESTI
1/06/2022	ROSA YOSILYN	AGUIRRE	RODRIGUEZ	46272994	6	1	4276	81509	06014276000081509	8364	738364 - DAZ TRUJILLO AMERICA	470198 - REGIONAL VENTAS LA LIBERT
1/06/2022	YENNI	GUTIERREZ	CUSIPAUCAR	41483219	6	1	4266	71126	06014266000071126	8366		470198 - REGIONAL VENTAS LA LIBERT
1/06/2022	MARIA JULIA	ROBLES	ROMERO	15724343	6	1	9467	45205	06019467000045205	9467	739467 - BA EKT HUACHO	470199 - REGIONAL VENTAS NORTE CH
1/06/2022	JULIA DORIS	CHUMPITAZ	MORAN	40277033	6	9	8370	10368	06098370000010368	9467	739467 - BA EKT HUACHO	470199 - REGIONAL VENTAS NORTE CH
1/06/2022	EDUARDA	QUISPE	HERRERA	25208811	6	1	4299	53806	06014299000053806	4299	734299 - BA EKT CUZCO	470209 - REGIONAL VENTAS CUSCO
1/06/2022	MARGARITA GUADALUP	SULCA	VERA	18165885	6	1	6498	12809	06016498000012809	4265	734265 - BA EKT TRUJILLO AYACUCI	470198 - REGIONAL VENTAS LA LIBERT
1/06/2022	LOURDES DEL ROSARI	TORREALVA	FUENTES	47528589	6	9	8392	938	06098392000000938	9580	739580 - BA EKT ICA	470208 - REGIONAL VENTAS ICA

Hoja1 Hoja2

Listo Accesibilidad: es necesario investigar Promedio: 44713 Recuento: 10837 Suma: 484510068 100%

Windows taskbar: 17:40 27/10/2022

Anexo 18: Metodología operativa

1) ENTENDIMIENTO DEL NEGOCIO

Descripción

La intención de este reporte es proveer la exploración inicial de los datos provenientes de la Oficina de Normalización Previsional (ONP) para que pueda ser usado para futuros análisis y desarrollo de planes de acción relacionados con entregar un estándar de clasificación del cliente que es hipotético al caso de que el cliente es o no buen pagador.

Enlace dataset:
https://drive.google.com/file/d/1eZZfh4_0js4gGDS22Di2VObiHy6a1O_0/view?usp=share_link

Este reporte contiene el entendimiento de los datos para preparar el dataset, códigos usados, problemas, identificar soluciones y las relaciones. También contiene la visualización de los datos y la exploración de las relaciones entre las variables.

2) PREPARACIÓN DE LOS DATOS

Descripción

El dataset usado fue proveído por la Oficina de Normalización Previsional (ONP) con una variedad de variables como los datos en aspectos de salud, económicos y sociales. Para poder usar los datos para el análisis se necesitó preparar y limpiar los datos.

Proceso

Los procesos de entendimiento de los datos para preparar los datos a ser usados son los siguientes:

1. **Importing Libraries:** Las librerías importadas fueron Pandas, Numpy, Matplotlib, Seaborn, Pandas es crucial para la manipulación de los datos,

habilita la lectura de los datos y la creación de data frames. Numpy es usado en álgebra lineal para hacer operaciones numéricas. Seaborn y Matplotlib son librerías para la visualización de los datos, y así como también la creación de mapas de histogramas y boxplots.

2. **Evaluating structure:** La estructura y los componentes del dato fueron evaluados. El tamaño del dataset fue accesado usando la función `shape ()`. El dataset tiene un total de 999 filas y 13 columnas. Usando la función `head ()` y `tail ()`, respectivamente se mostró una previsualización de los primeros y últimos registros. Todas las columnas fueron escritas en el título del formato y fueron correctamente deletreadas. Las columnas representan esencialmente la variable usada para el análisis. Las variables incluyen: ``idInfoBasica``, ``resultado``, ``edad``, ``sexo``, ``ganancia_deseada``, ``aspira_ganar_mas``, ``cantManutencion``, ``monto``, ``idEnfermedad``, ``tiene_alguna_enfermedad``, ``pertenece_1st_pensiones``, ``edad_jubilarse``, ``idTipoPension``.
3. **Checking for the unique values in columns:** Fue necesario para verificar los valores únicos en las columnas. El valor único representa los valores de las variables. Fue esencial verificar los valores no deseados para mayor precisión de los resultados. Esto fue hecho usando `print ("Columna = ", data['Columna'].unique())`. Cada columna fue verificada para valores únicos.
4. **Data types:** Los correctos tipos de datos son requeridos para hacer el correcto procedimiento de análisis de los datos. Se verificó los tipos de datos usando el proceso `info()` y cambiando los tipos de variables de float a integer usando la función `astype()`.
5. **Data cleaning:** La limpieza de los datos fue el siguiente paso en el procedimiento de análisis de los datos. Los pasos fueron los siguientes a continuación:
 1. Resolver los valores nulos.
 2. Reemplazar los valores de la columna.
 3. Cambiar el tipo de datos.
 4. Cambiar todas las mayúsculas.
 5. Verificar los datos duplicados. En algunos casos, los datos no tienen duplicados.

6. Eliminar los espacios en blanco.
7. Verificar la coherencia de los datos.

Tabla 28

Aplicación de técnicas estadísticas

	Nombre de problema	Localización	Código para identificar	Relaciones y solución
1	Verificar valores únicos en las columnas	Columna ["genero"]	print("idSexo = ", data["idSexo"].unique())	Los valores numéricos en una celda indican un tipo.
2	Cambiar tipo de datos		print(data.dtypes) data['resultado'] = data['resultado'].astype(str)	Cambiar los floats a objetos para el análisis de atos
3	Cambiar todas las mayúsculas		# cambiar todo a mayúsculas data.apply(lambda x: x.astype(str).str.upper())	Para hacer consistencia
4	Valores nulos		# identificación de los valores nulos data.isnull().sum() # eliminar los valores nulos data.dropna(axis=0, inplace = True) data.isnull().sum()	Los valores "null" son identificados. Se elimina el número pequeño que no hace diferencia en el dataset.
5	Valores atípicos en los datos	Columna "monto"	# resolviendo errores en data["monto"] por una mediana # el programa de python obtiene una lista	

			<pre>def Average(lst): return sum(lst) / len(lst) # driver code lst = data["monto"] average = Average(lst) # imprime la lista promedio print("Promedio de la lista =", round(average, 1))</pre>	
6	Reemplazar valores de una columna	Columna "monto"	<pre># reemplazando valores de la columna data["monto"] = data["monto"].replace({'1.0':'670.0'}) print("monto" = ", data["monto"].unique())</pre>	Como anteriormente
7	Espacios en blanco	A lo largo del Dataset	<pre># Eliminar espacios en blanco data.columns = data.columns.str.rstrip() data.columns = data.columns.str.lstrip()</pre>	strip() remueve los caracteres anteriores y posteriores que incluyen espacios en blanco de un string
8	Coherencia de los datos	Columna ["resultado"]	<pre>data[data["resultado"] =="2.0"]</pre>	Verifique si los valores correctos estaban en su lugar, es decir, "2.0" no podría ser un "resultado"

Fuente: Elaboración propia

3) EXPLORACIÓN DE DATOS

Descripción

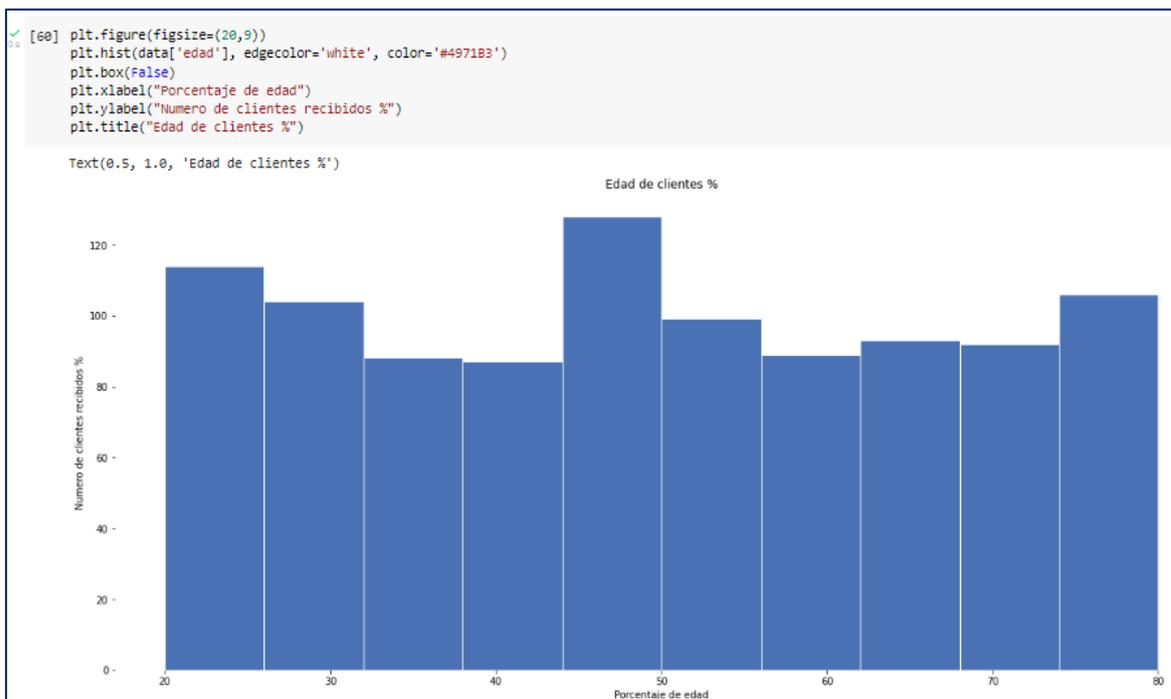
El análisis de exploración de los datos fue el siguiente paso después de la limpieza de los datos. Los conocimientos obtenidos de EDA son útiles

Proceso

1. Análisis univariado. Esto ayuda a obtener una comprensión más profunda de los atributos individuales. Se detectan detalles como la media, la mediana y la moda, y se visualiza mediante un histograma. Un histograma muestra la frecuencia de aparición de una variable categórica.
2. La estadística bivariante analiza la asociación de un par de atributos elegido. Este método se eligió para buscar relaciones entre tres pares de columnas. Las columnas elegidas fueron “Porcentaje de edad” (Figura 12), “Porcentaje de ganancia deseada” (Figura 13), “Porcentaje de cantidad de manutención” (Figura 14), “Porcentaje de cantidad de manutención” (Figura 15), “Monto” (Figura 16) y “Porcentaje de edad a jubilarse” (Figura 17). “Porcentaje de edad” fue escogida para tener en cuenta la edad que se maneja y poder obtener el porcentaje más bajo de edad. “Porcentaje de ganancia deseada” fue seleccionado como la ganancia deseada de los individuos. “Porcentaje de cantidad de manutención” es marcado como la cantidad que pagan los clientes. “Monto” significa que la entidad entrega un dinero al cliente.

Figura 32

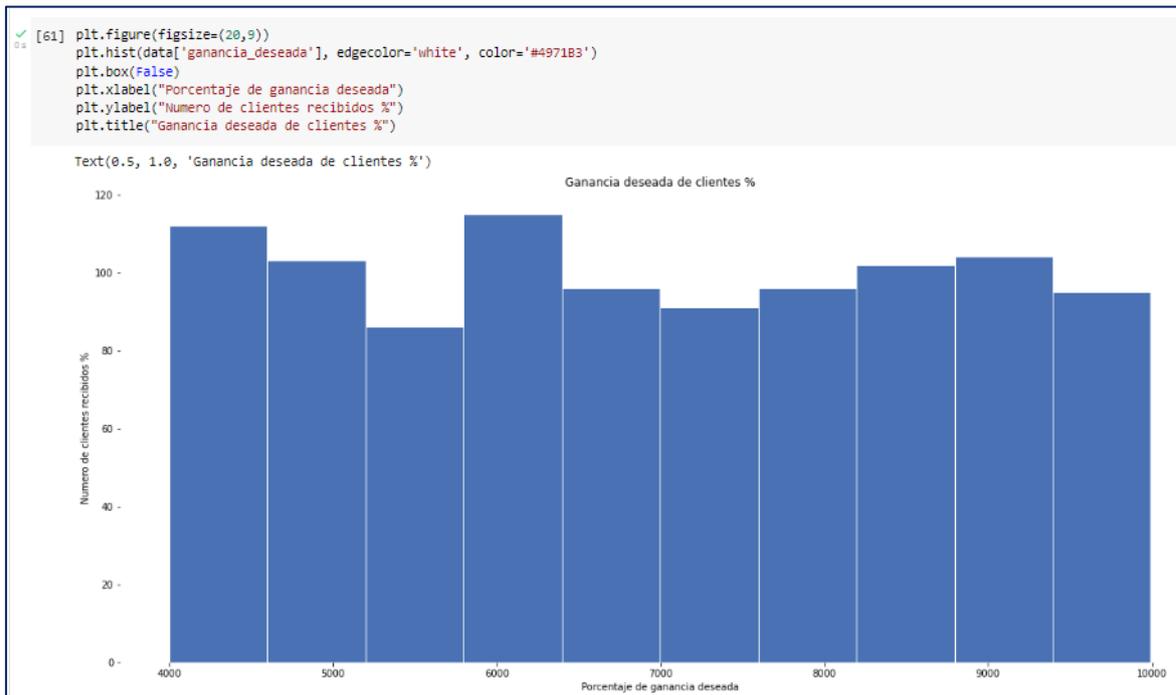
Edad de individuos %



Fuente: Elaboración propia

Figura 33

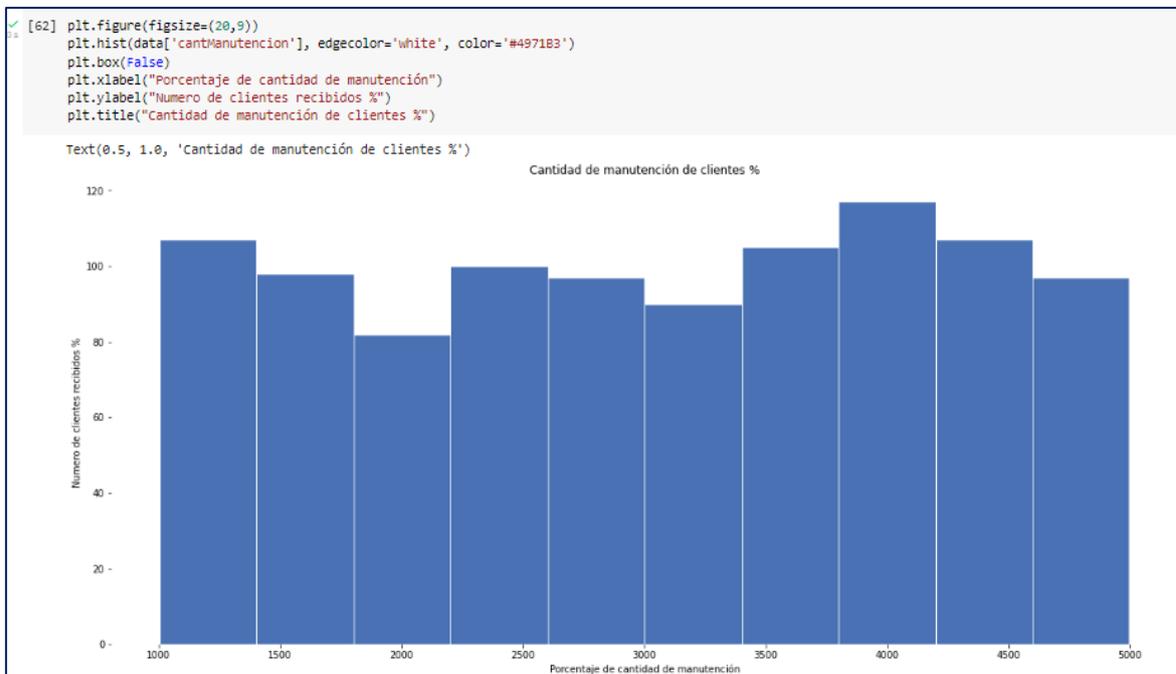
Ganancia deseada de individuos %



Fuente: Elaboración propia

Figura 34

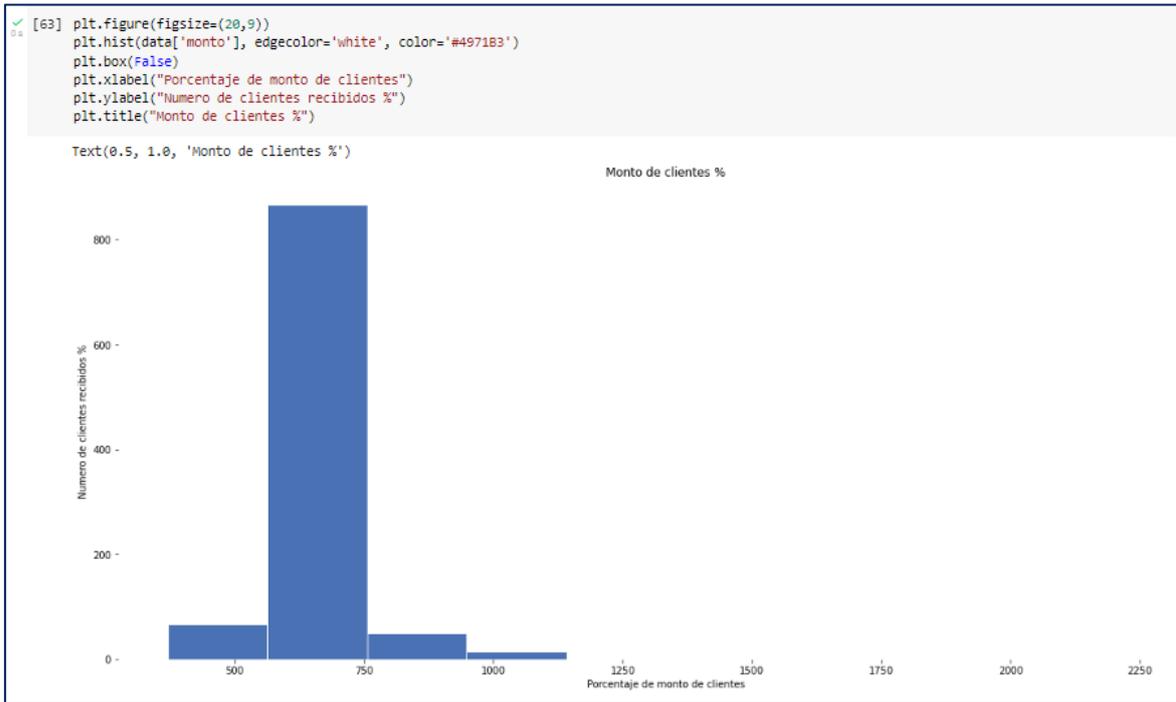
Cantidad de manutención de individuos %



Fuente: Elaboración propia

Figura 35

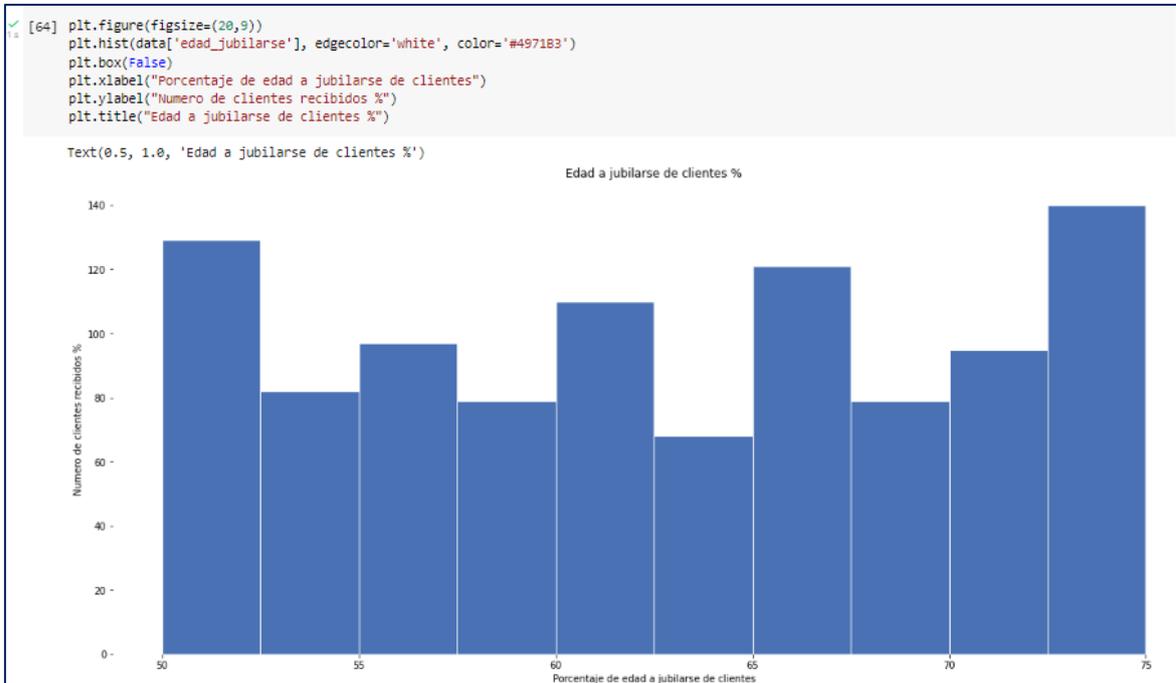
Monto de individuos %



Fuente: Elaboración propia

Figura 36

Edad a jubilarse de individuos %



Fuente: Elaboración propia

4) CREACIÓN DEL MODELO

Descripción

Normalización, el escalado y la discretización. Realizar operaciones que transforman los datos originales (categóricos a nominales), casi siempre de manera reversible, produciendo una nueva versión más conveniente para el análisis CRISP-DM.

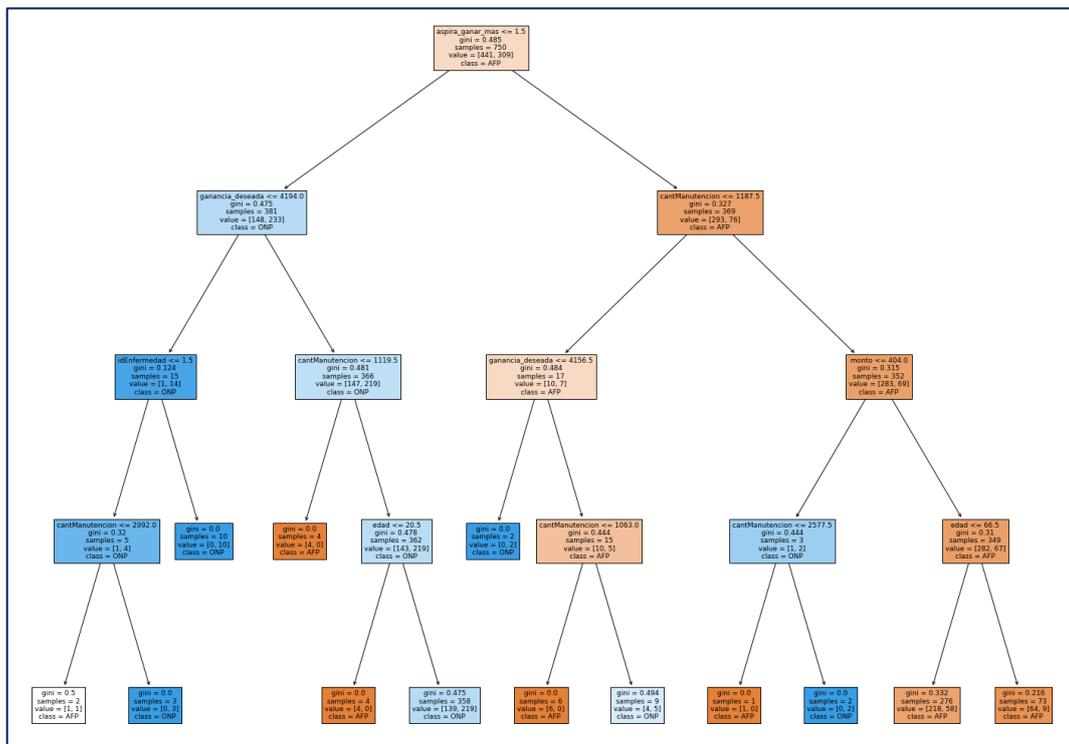
Elegir algoritmo(s) de minería de datos a fin de extraer conocimiento y realizar división de datos en 75% para entrenamiento y 25% pruebas

En función del rendimiento obtenido (matriz de confusión), y del análisis efectuado para determinar qué problemas plantea el modelo, sería preciso realizar ajustes en el algoritmo

Revisar si es problema de volumen de datos, datos desactualizados, no comprendimos el problema y si el modelo es inadecuado.

Figura 37

Árbol de decisión



Fuente: Elaboración propia

Proceso

Ahora entenderemos que significa los datos que están el cuadro

Tabla 29

Análisis de los nodos del árbol de decisión

<code>tiene_alguna_enfermedad <= 1.5</code>	La pregunta
<code>gini = 0.485</code>	La pureza del nodo, menor sea este número tenemos mejor división de clases
<code>samples = 750</code>	Cantidad de individuos que cumplen con estas condicione
<code>value = [441, 309]</code>	Cantidad de individuos que es AFP y ONP
<code>class = AFP</code>	Nos indica la clase con la mayoría de los individuos en el nodo en este caso son AFP

<code>gini = 0.0 samples = 4 value = [4, 0] class = AFP</code>	<code>gini = 0.475 samples = 358 value = [139, 219] class = ONP</code>
--	--

<code>gini = 0.0</code>	Cuando obtenemos $gini = 0.0$ tenemos definido a que clase pertenece esa hoja en este caso cuadro naranja $gini = 0$ hay 4 individuos en AFP.
-------------------------	---

Fuente: Elaboración propia

Calcularemos que tan bueno es para predecir el árbol con los datos brindados para entrenar.

Figura 38

Código para obtener datos de prueba

```
[604] #AHORA VEMOS QUE TAN BUENO ES PARA PREDECIR SEGUN LOS DATOS OBTENIDOS
      y_pred = model.predict(x_test)
      y_pred
```

Fuente: Elaboración propia

Pasamos `y_test` que tiene los valores reales de las clases y `y_pred` que tiene los valores predichos

Figura 39

Código para construir la matriz de confusión

```
[644] #CREAMOS LA MATRIX DE CONFUSION
      matriz_de_confusion = confusion_matrix(y_test,y_pred)#valores reales de la clase, y los valores predichos
      print(matriz_de_confusion)#primera fila clase AFP y segunda la clase ONP
      #pero a la hora de predecir suma vertical AFP y ONP
      # en la diagonal el primero es correctamente predicho en la otra diagonal fue incorrecta
```

Fuente: Elaboración propia

Ahora obtenemos la matriz de confusión

Figura 40

Resultado de matriz de confusión

```
[[110  49]
 [ 28  63]]
```

Fuente: Elaboración propia

Obtenemos $110+49 = 159$ individuos con clase realmente AFP

Obtenemos $28+ 63 = 91$ individuos con clase realmente ONP

Pero a la hora de predecir se obtuvieron $110+28 = 138$ con AFP

Pero a la hora de predecir se obtuvieron $49+63 = 112$ con ONP

Primera diagonal se encuentra fue correctamente predicho \

$110+63 =173$

Segunda diagonal se encuentra la predicción que fue incorrecta. /

$28 +49 =77$

En base a eso calculamos la precisión global y la precisión por cada una de estas clases es decir para AFP y ONP

Figura 41

Código para obtener precisión de matriz de confusión

```
[606] precision_global = np.sum(matriz_de_confusion.diagonal()/np.sum(matriz_de_confusion))  
      print("precisión global: ",precision_global)#me da la precisión global del entrenamiento  
  
[607] precision_AFP = matriz_de_confusion[0,0]/sum(matriz_de_confusion[0,])  
      print("precisión de los AFP",precision_AFP)  
  
[608] precision_ONP = matriz_de_confusion[1,1]/sum(matriz_de_confusion[1,])  
      print("precisión de los ONP",precision_ONP)
```

Fuente: Elaboración propia

Arrojó estos valores:

Figura 42

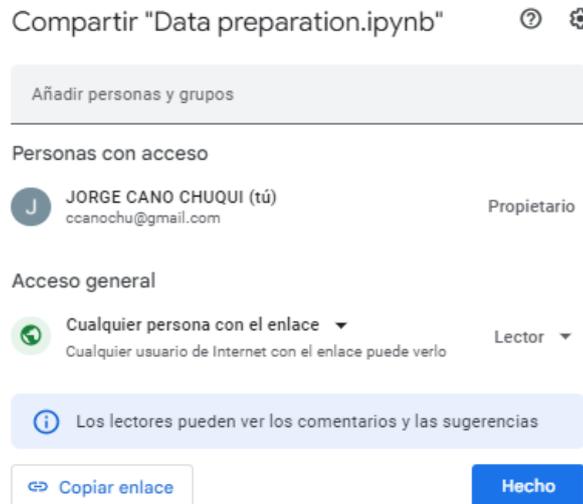
Resultados de precisión de matriz de confusión

```
✓ [645] precision_global = np.sum(matriz_de_confusion.diagonal()/np.sum(matriz_de_confusion))  
0s print("precisión global: ",precision_global)#me da la precisión global del entrenamiento  
  
      precisión global:  0.692  
  
✓ [646] precision_AFP = matriz_de_confusion[0,0]/sum(matriz_de_confusion[0,])  
0s print("precisión de los AFP",precision_AFP)  
  
      precisión de los AFP 0.6918238993710691  
  
✓ [647] precision_ONP = matriz_de_confusion[1,1]/sum(matriz_de_confusion[1,])  
0s print("precisión de los ONP",precision_ONP)  
  
      precisión de los ONP 0.6923076923076923
```

Fuente: Elaboración propia

Figura 43

Accesos para visualizar entorno de colab



Fuente: Elaboración propia

Figura 44

Entorno de Colab

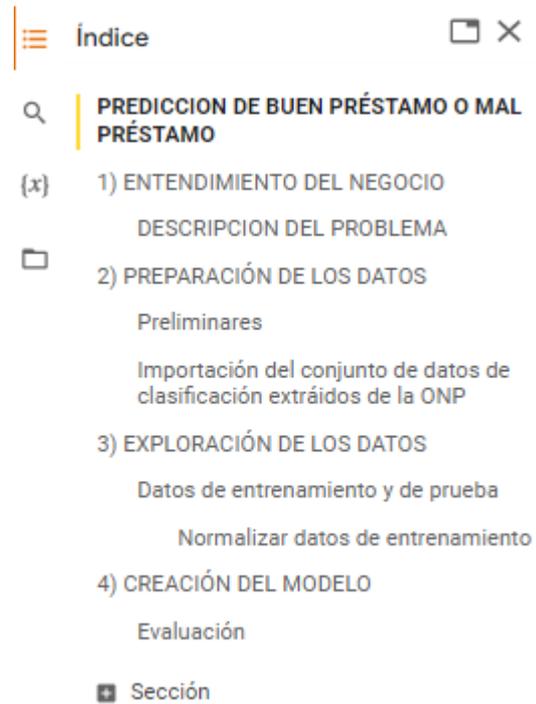


Fuente: Elaboración propia

Enlace: <https://colab.research.google.com/drive/1bgALEayBSrm4ltvDEVi5IJLdw-TjCM0?usp=sharing>

Figura 45

Fases de la metodología CRISP-DM



Índice

Q PREDICCIÓN DE BUEN PRÉSTAMO O MAL PRÉSTAMO

(x) 1) ENTENDIMIENTO DEL NEGOCIO
DESCRIPCIÓN DEL PROBLEMA

2) PREPARACIÓN DE LOS DATOS
Preliminares
Importación del conjunto de datos de clasificación extraídos de la ONP

3) EXPLORACIÓN DE LOS DATOS
Datos de entrenamiento y de prueba
Normalizar datos de entrenamiento

4) CREACIÓN DEL MODELO
Evaluación

Sección

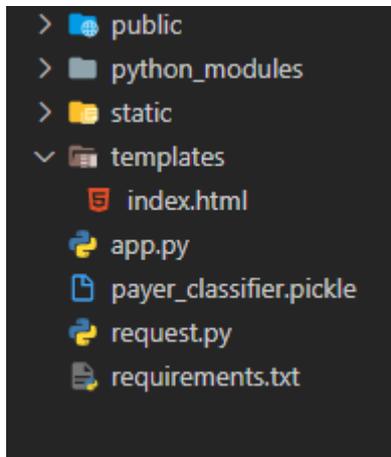
Fuente: Elaboración propia

La siguiente fase de la creación del modelo es el despliegue, a continuación:

5) DESPLIEGUE

Figura 46

Estructura del proyecto



Fuente: Elaboración propia

Crear el archivo app.py para trabajar el modelo con datos json y enviarlo a un api

Figura 47

App.py_1

```
app.py x
1 import numpy as np
2 from flask import Flask, request, jsonify, render_template
3 import pickle
4
5 app = Flask(__name__)
6 model = pickle.load(open('payer_classifier.pickle', 'rb'))
7
8 @app.route('/')
9 def home():
10     return render_template('index.html')
11
12 @app.route('/predict', methods=['POST'])
13 def predict():
14     """
15     For rendering results on HTML GUI
16     """
17     int_features = [int(x) for x in request.form.values()]
18     final_features = [np.array(int_features)]
19     prediction = model.predict(final_features)
20
21     output = round(prediction[0], 2)
22     if output == 1:
23         output = "Se concluye que el cliente es un buen pagador"
24     else:
25         output = "Se concluye que el cliente no es un buen pagador"
26
27     return output;
28
29 # return render_template('index.html', prediction_text='Resultado de la calificación: Condición de {}'.format(output))
30 return render_template('index.html', prediction_text='{}'.format(output))
31
32 @app.route('/predict_api', methods=['POST'])
33 def predict_api():
34     """
```

Fuente: Elaboración propia

Figura 48

App.py_2

```
29 # return render_template('index.html', prediction_text='Resultado de la calificación: Condición de {}'.format(output))
30 return render_template('index.html', prediction_text='{}'.format(output))
31
32 @app.route('/predict_api', methods=['POST'])
33 def predict_api():
34     """
35     For direct API calls through request
36     """
37     data = request.get_json(force=True)
38     prediction = model.predict([np.array(list(data.values()))])
39
40     output = prediction[0]
41
42
43     return jsonify(output)
44
45 if __name__ == "__main__":
46     app.run(debug=True)
```

Fuente: Elaboración propia

Crear el archivo index.html para crear el formulario y enviar los datos por post con fetch

Figura 49

index.html_1

```
1 <!DOCTYPE html>
2 <html lang="es">
3 <head>
4     <meta charset="UTF-8">
5     <meta http-equiv="X-UA-Compatible" content="IE=edge">
6     <meta name="viewport" content="width=device-width, initial-scale=1.0">
7
8     <!-- ===== CSS ===== -->
9     <link rel="stylesheet" href="https://cdn.jsdelivr.net/npm/font-awesome@6.2.1/css/all.min.css" integrity="sha512-M
10 <link rel="stylesheet" href="{{ url_for('static', filename='css/estilos.css') }}" type="text/css">
11 <link rel="icon" type="image/x-icon" href="{{ url_for('static', filename='images/favicon.ico') }}">
12 <link rel="stylesheet" href="https://cdn.jsdelivr.net/npm/limonte-sweetalert2@11.6.9/sweetalert2.css" integrity="
13
14
15     <!-- ===== Icons Scout CSS ===== -->
16 <link rel="stylesheet" href="https://unicons.iconscout.com/release/v4.0.0/css/line.css">
17
18 <!-- Sweet alert Js -->
19 <script src="https://cdn.jsdelivr.net/npm/sweetalert2@11"></script>
20 <script src="https://cdn.jsdelivr.net/npm/limonte-sweetalert2@11.6.9/sweetalert2.js" integrity="sha512-0BFoY+AoBA
21
22 <script rel="text/javascript" src="{{ url_for('static', filename='js/timer.js') }}"></script>
23
24 <!-- Firebase JS -->
25 <script src="https://cdn.jsdelivr.net/npm/font-awesome@6.2.1/js/all.min.js" integrity="sha512-rpL11167T5LJHwp0wa
26
27 <script src="https://www.gstatic.com/firebasejs/4.12.1/firebase.js"></script>
28 <!-- Servicio de Firestore -->
29 <script src="https://www.gstatic.com/firebasejs/4.12.1/firebase-firestore.js"></script>
30
31 <script rel="text/javascript" src="{{ url_for('static', filename='js/firestore.js') }}"></script>
32
33 <title>Diagnóstico de crédito</title>
```

Fuente: Elaboración propia

Figura 50

index.html_2

```
app.py index.html X
34
35 </head>
36 <body onload="Prediccion()">
37   <div class="container">
38     <header>Alfin Nuestro Banco</header>
39     <input type="hidden" id="hidden_codigo_generado"/>
40
41     <!-- <form action="{{ url_for('predict')}}" method="post"> -->
42     <span></span>
43     <div id="form" attr-url-predict="{{ url_for('predict')}}">
44       <input type="hidden" id="resultado_hidden"/>
45       <input type="hidden" id="satisfaccion_hidden"/>
46       <div class="form first">
47         <div class="details personal">
48           <span class="title">Variables de entrada</span>
49           <br>
50           <div class="fields">
51
52             <!-- Primer Campo EDAD -->
53             <div class="input-field">
54               <label class="tooltip">Edad: <span class="tooltiptext">Puedes escoger una edad entre 18 a 100 años,
55               <input type="range" id="edad" class="form-range" name="edad" min="18" max="100" step="1" value="18"
56             </div>
57
58             <!-- Segundo Campo Género -->
59             <div class="input-field">
60               <label class="tooltip">Género: <span class="tooltiptext">Puedes escoger tu género</span> <i class="
61               <select class="form-select select-genero" required name="idSexo" id="genero" onchange="ValorGenero(
62                 <option disabled selected value="">Seleccione el género</option>
63                 <option value="1">Masculino</option>
64                 <option value="2">Femenino </option>
65               </select>
66               <label><output id="label_genero" class="badge badge-primary"></output></label>
67             </div>
```

Fuente: Elaboración propia

Figura 51

index.html_3

```
app.py index.html X
68
69   <!-- Tercer Campo MAYOR SUELDO -->
70   <div class="input-field">
71     <label class="tooltip">¿Aspira a un mayor sueldo? <span class="tooltiptext">Si aspiras a un mayor s
72     <select class="form-select select-mayor-sueldo" required name="aspira_ganar_mas" id="aspira_ganar_m
73       <option disabled selected value="">Seleccione la opción</option>
74       <option value="1">Si</option>
75       <option value="2">No</option>
76     </select>
77     <output id="label_mayor_sueldo" class="badge badge-primary"></output>
78   </div>
79
80   <!-- Cuarto Campo SUELDO EXPECTATIVA -->
81   <div class="input-field">
82     <label class="tooltip">Sueldo expectativa: <span class="tooltiptext">Cuál es tu sueldo expectativa,
83     <input type="range" id="ganancia_deseada" class="form-range" name="ganancia_deseada" min="930" max=
84   </div>
85
86   <!-- Quinto Campo SUELDO ACTUAL -->
87   <div class="input-field">
88     <label class="tooltip">Sueldo actual: <span class="tooltiptext">Cuánto ganas actualmente, en el ban
89     <input type="range" id="cantManutencion" class="form-range" name="cantManutencion" min="930" max="8
90   </div>
91
92   <!-- Sexto Campo MONTO SOLICITADO -->
93   <div class="input-field">
94     <label class="tooltip">Monto solicitado: <span class="tooltiptext">Cuánto es tu monto que solicitas
95     <input type="range" id="monto" class="form-range" name="monto" min="300" max="8000" step="1" value=
96   </div>
97
98   <!-- Séptimo Campo SISTEMA SALUD PREFERIDO -->
99   <div class="input-field">
100     <label class="tooltip">¿Sistema de salud preferido? <span class="tooltiptext">Tenemos otras opcione
101     <select class="form-select select-salud-preferido" required name="idEnfermedad" id="enfermedad">
```

Fuente: Elaboración propia

Figura 52

index.html_4

```
101 <select class="form-select select-salud-preferido" required name="idEnfermedad" id="enfermedad">
102 <option disabled selected value="">Seleccione la opción</option>
103 <option value="1">Privado</option>
104 <option value="2">Público</option>
105 </select>
106 <output id="label_salud_preferido" class="badge badge-primary"></output>
107 </div>
108
109 <!-- Octavo Campo TIENE ALGUNA ENFERMEDAD -->
110 <div class="input-field">
111 <label class="tooltip">¿Tiene alguna enfermedad? <span class="tooltiptext">Si acá tu tienes algunas
112 <select class="form-select select-alguna-enfermedad" required name="tiene_alguna_enfermedad" id="t
113 <option disabled selected value="">Seleccione la opción</option>
114 <option value="1">Si</option>
115 <option value="2">No</option>
116 </select>
117 <output id="label_alguna_enfermedad" class="badge badge-primary"></output>
118 </div>
119
120 <!-- Noveno Campo PERTENECE AL SISTEMA DE PENSIONES -->
121 <div class="input-field">
122 <label class="tooltip">¿Pertenece al sistema de pensiones? <span class="tooltiptext">Si aquí tu est
123 <select class="form-select select-pertenece-pensiones" required name="pertenece_1st_peniones" id="
124 <option disabled selected value="">Seleccione la opción</option>
125 <option value="1">Si</option>
126 <option value="2">No</option>
127 </select>
128 <output id="label_pertenece_pensiones" class="badge badge-primary"></output>
129 </div>
130
131 <!-- Décimo Campo EDAD A JUBILARSE -->
132 <div class="input-field">
133 <label class="tooltip">Edad a jubilarse: <span class="tooltiptext">A qué edad quisieras jubilarte,
134 <input type="range" id="edad_jubilarse" class="form-range" name="edad_jubilarse" min="50" max="75" >
```

Fuente: Elaboración propia

Figura 53

index.html_5

```
135 </div>
136
137 <!-- Onceavo Campo SISTEMA DE PENSIONES QUE LE GUSTA -->
138 <div class="input-field">
139 <label class="tooltip">¿Qué sistema de pensiones le gusta? <span class="tooltiptext">Qué sistema de
140 <select class="form-select select-sistema-pensiones" required name="idTipoPension" id="idTipoPensi
141 <option disabled selected value="">Seleccione la opción</option>
142 <option value="1">AFP</option>
143 <option value="2">ONP</option>
144 </select>
145 <output id="label_sistema_pensiones" class="badge badge-primary"></output>
146 </div>
147 <div class="input-field"></div>
148 </div>
149 </div>
150 <button class="nextBtn" type="button" onclick="guardar()" id="confirmar">
151 <span class="btnText">Predecirme</span>
152 <i class="uil uil-navigator"></i>
153 </button>
154 </div>
155 </div>
156 <!-- </form> -->
157 <h2 class="text-center" id="calificacion_final" style="color: #000; font-weight: bold; font-size: 24px;">{{ predictio
158
159 <div id="label_tiempo_real" class="text-center" style="color: #000;">
160 <p style="font-size: 10px;">
161 Tiempo en el sistema:
162 <span id="minutos">0</span>:<span id="segundos">0</span>
163 </p>
164 </div>
165
166
167 </div>
```

Fuente: Elaboración propia

Figura 54

index.html_6

```
app.py index.html X
169 <script>
170
171 //Configuracion de Firebase Cloud
172 const firebaseConfig = {
173   apiKey: "AizaSyAoMeEGgT4ru35Hasla8ptAjurmXYBhdVM",
174   authDomain: "appmachinelearning-b4022.firebaseio.com",
175   projectId: "appmachinelearning-b4022",
176   storageBucket: "appmachinelearning-b4022.appspot.com",
177   messagingSenderId: "397486173991",
178   appId: "1:397486173991:web:28642d72204fb7ad834302"
179 };
180
181 // Inicializar Firebase en nuestro proyecto
182 firebase.initializeApp(firebaseConfig);
183
184 //Inicializar el servicio de Cloud Firestore de Firebase
185 var db = firebase.firestore();
186
187 function satisfaccion(data){
188   (async () => {
189
190     var that = this;
191     /* inputOptions can be an object or Promise */
192     const inputOptions = new Promise((resolve) => {
193       setTimeout(() => {
194         resolve({
195           'Totalmente en desacuerdo': 'Totalmente en desacuerdo',
196           'En desacuerdo': 'En desacuerdo',
197           'Indeciso': 'Indeciso',
198           'De acuerdo': 'De acuerdo',
199           'Totalmente de acuerdo': 'Totalmente de acuerdo'
200         });
201       }, 1000)
202     });
203   });
204
205   const { value: satisfaccion } = await Swal.fire({
206     title: 'Satisfacción',
207     input: 'radio',
208     inputOptions: inputOptions,
209     confirmButtonText: 'Aceptar',
210     inputValue: value,
211     if (!value) {
212       return 'Necesitas escoger una opción'
213     }
214   }).then((result)=>{
215     console.log(result);
216     if(result.isConfirmed){
217       console.log(result.value);
218       document.getElementById('satisfaccion_hidden').value=result.value;
219       //document.location.reload(true);
220       console.log("Se esta actualizando el id: ", idhiddengenerado);
221       that.actualizar(idhiddengenerado, data, result.value);
222     }
223   });
224
225   })()
226 }
227
228 const generateRandomString = (num) => {
229   const characters = 'ABCDEFGHIJKLMNOPQRSTUVWXYZabcdefghijklmnopqrstuvwxyz0123456789';
230   let result1 = '';
231   const charactersLength = characters.length;
232   for (let i = 0; i < num; i++) {
233     result1 += characters.charAt(Math.floor(Math.random() * charactersLength));
234   }
235 }
```

Fuente: Elaboración propia

Figura 55

index.html_7

```
app.py index.html X
202
203 })
204
205 const { value: satisfaccion } = await Swal.fire({
206   title: 'Satisfacción',
207   input: 'radio',
208   inputOptions: inputOptions,
209   confirmButtonText: 'Aceptar',
210   inputValue: value,
211   if (!value) {
212     return 'Necesitas escoger una opción'
213   }
214 }).then((result)=>{
215   console.log(result);
216   if(result.isConfirmed){
217     console.log(result.value);
218     document.getElementById('satisfaccion_hidden').value=result.value;
219     //document.location.reload(true);
220     console.log("Se esta actualizando el id: ", idhiddengenerado);
221     that.actualizar(idhiddengenerado, data, result.value);
222   }
223 });
224
225 })()
226 }
227
228 const generateRandomString = (num) => {
229   const characters = 'ABCDEFGHIJKLMNOPQRSTUVWXYZabcdefghijklmnopqrstuvwxyz0123456789';
230   let result1 = '';
231   const charactersLength = characters.length;
232   for (let i = 0; i < num; i++) {
233     result1 += characters.charAt(Math.floor(Math.random() * charactersLength));
234   }
235 }
```

Fuente: Elaboración propia

Figura 56

index.html_8

```
app.py index.html X
235     return result1;
236   }
237   let idhiddengeneradoone = generateRandomString(12);
238   const idhiddengenerado = `${idhiddengeneradoone}`
239   document.getElementById('hidden_codigo_generado').value = idhiddengenerado;
240
241   function Prediccion(){
242
243     document.getElementById("confirmar").addEventListener('click', guardar);
244
245     var ruta = window.location.href;
246
247     if(ruta == "http://127.0.0.1:5000/"){
248
249       var alerta = document.getElementsByClassName('swal2-container');
250       document.getElementById('calificacion_final').classList.add('esconder');
251
252       contador_s = 0;
253       contador_m = 0;
254
255       s = document.getElementById('segundos');
256       m = document.getElementById('minutos');
257
258       var cronometro = setInterval( function(){
259
260         if(contador_s == 60)
261         {
262           contador_s = 0;
263           contador_m++;
264           m.innerHTML = contador_m;
265
266           if(contador_m == 0)
267           {
```

Fuente: Elaboración propia

Figura 57

index.html_9

```
app.py index.html X
267     {
268       contador_m = 0;
269     }
270   }
271
272   s.innerHTML = contador_s;
273   contador_s++;
274
275   }, 1000);
276
277   //Invocamos los metodo para extraer el valor de los Select
278   ValorGenero();
279   ValorAspiraMayorSueldo();
280   ValorSaludPreferido();
281   ValorAlgunaEnfermedad();
282   ValorPertenecePensiones();
283   ValorSistemaPensiones();
284
285   }else
286   {
287     Swal.fire({
288
289       title:"Resultado de su evaluación!",
290       // success , error , warning ,info
291
292       text: "{{ prediction_text }}",
293       // Custom message flashed from your flask routes
294
295       icon: "info",
296
297     });
298   }
299
300 }
```

Fuente: Elaboración propia

Figura 58

index.html_10

```
app.py | index.html X
300
301 }
302
303 function ValorGenero(){
304     //Obtener el valor del Select Genero
305     const selectElement = document.querySelector('.select-genero');
306     selectElement.addEventListener('change', (event) => {
307         //valueTexto = event.target.value;|
308         //console.log(valueTexto);
309         const resultado = document.getElementById('label_genero');
310         //resultado.textContent = event.target.value;
311
312         if(event.target.value == 1){
313             resultado.innerHTML = "Masculino";
314         }else{
315             resultado.innerHTML = "Femenino";
316         }
317     });
318 }
319
320 function ValorAspiraMayorSueldo(){
321     //Obtener el valor del Select Genero
322     const selectElement = document.querySelector('.select-mayor-sueldo');
323     selectElement.addEventListener('change', (event) => {
324         //valueTexto = event.target.value;|
325         //console.log(valueTexto);
326         const resultado = document.getElementById('label_mayor_sueldo');
327         //resultado.textContent = event.target.value;
328
329         if(event.target.value == 1){
330             resultado.textContent = "Si";
331         }else{
332             resultado.textContent = "No";
333         }
334     });
335 }
336
337 function ValorSaludPreferido(){
338     //Obtener el valor del Select Genero
339     const selectElement = document.querySelector('.select-salud-preferido');
340     selectElement.addEventListener('change', (event) => {
341         //valueTexto = event.target.value;|
342         //console.log(valueTexto);
343         const resultado = document.getElementById('label_salud_preferido');
344         //resultado.textContent = event.target.value;
345
346         if(event.target.value == 1){
347             resultado.textContent = "Privado";
348         }else{
349             resultado.textContent = "Público";
350         }
351     });
352 }
353
354 function ValorAlgunaEnfermedad(){
355     //Obtener el valor del Select Genero
356     const selectElement = document.querySelector('.select-alguna-enfermedad');
357     selectElement.addEventListener('change', (event) => {
358         //valueTexto = event.target.value;|
359         //console.log(valueTexto);
360         const resultado = document.getElementById('label_alguna_enfermedad');
361         //resultado.textContent = event.target.value;
362
363         if(event.target.value == 1){
364             resultado.textContent = "Si";
365         }else{
366             resultado.textContent = "No";
367         }
368     });
369 }
```

Fuente: Elaboración propia

Figura 59

index.html_11

```
app.py | index.html X
334     });
335 }
336
337 function ValorSaludPreferido(){
338     //Obtener el valor del Select Genero
339     const selectElement = document.querySelector('.select-salud-preferido');
340     selectElement.addEventListener('change', (event) => {
341         //valueTexto = event.target.value;|
342         //console.log(valueTexto);
343         const resultado = document.getElementById('label_salud_preferido');
344         //resultado.textContent = event.target.value;
345
346         if(event.target.value == 1){
347             resultado.textContent = "Privado";
348         }else{
349             resultado.textContent = "Público";
350         }
351     });
352 }
353
354 function ValorAlgunaEnfermedad(){
355     //Obtener el valor del Select Genero
356     const selectElement = document.querySelector('.select-alguna-enfermedad');
357     selectElement.addEventListener('change', (event) => {
358         //valueTexto = event.target.value;|
359         //console.log(valueTexto);
360         const resultado = document.getElementById('label_alguna_enfermedad');
361         //resultado.textContent = event.target.value;
362
363         if(event.target.value == 1){
364             resultado.textContent = "Si";
365         }else{
366             resultado.textContent = "No";
367         }
368     });
369 }
```

Fuente: Elaboración propia

Figura 60

index.html_12

```
app.py index.html X
367     }
368   });
369 }
370
371 function ValorPertenecePensiones(){
372   //Obtener el valor del Select Genero
373   const selectElement = document.querySelector('.select-pertenece-pensiones');
374   selectElement.addEventListener('change', (event) => {
375     //valueTexto = event.target.value;|
376     //console.log(valueTexto);
377     const resultado = document.getElementById('label_pertenece_pensiones');
378     //resultado.textContent = event.target.value;
379
380     if(event.target.value == 1){
381       resultado.textContent = "Si";
382     }else{
383       resultado.textContent = "No";
384     }
385   });
386 }
387
388 function ValorSistemaPensiones(){
389   //Obtener el valor del Select Genero
390   const selectElement = document.querySelector('.select-sistema-pensiones');
391   selectElement.addEventListener('change', (event) => {
392     //valueTexto = event.target.value;|
393     //console.log(valueTexto);
394     const resultado = document.getElementById('label_sistema_pensiones');
395     //resultado.textContent = event.target.value;
396
397     if(event.target.value == 1){
398       resultado.textContent = "AFP";
399     }else{
```

Fuente: Elaboración propia

Figura 61

index.html_13

```
app.py index.html X
400     resultado.textContent = "ONP";
401   }
402   });
403 }
404
405 function alerta_(titulo, texto, icono){
406   Swal.fire({
407     title: titulo,
408     text: texto,
409     icon: icono
410   });
411 }
412
413 function alerta2_(titulo, texto, icono){
414   var that = this;
415   Swal.fire({
416     title: titulo,
417     text: texto,
418     icon: icono,
419     confirmButtonText: 'Aceptar'
420   }).then((result)=>{
421     if(result.isConfirmed){
422       console.log("hola");
423       that.satisfaccion(texto);
424     }
425   });
426 }
427 }
428
429 function actualizar(idRef, data, data2){
430   var dataRef = db.collection("Pruebas").doc(idRef);
431
432   // Set the "capital" field of the city 'DC'
```

Fuente: Elaboración propia

Figura 62

index.html_14

```
app.py index.html X
433     return dataRef.update({
434         resultado: data,
435         satisfaccion: data2
436     })
437     .then(() => {
438         console.log("Document successfully updated!");
439     })
440     .catch((error) => {
441         // The document probably doesn't exist.
442         console.error("Error updating document: ", error);
443     });
444 }
445
446 /* FIREBASE*/
447
448 async function guardar(evt){
449
450
451
452     /* Primer Campo EDAD */
453     var edad = document.getElementById('edad_Y').value;
454     /* Segundo Campo Género */
455     var genero = document.getElementById('label_genero').value;
456     /* Tercer Campo MAYOR SUELDO */
457     var aspira_ganar_mas = document.getElementById('label_mayor_sueldo').value;
458     /* Cuarto Campo SUELDO EXPECTATIVA */
459     var ganancia_deseada = document.getElementById('ganancia_deseada_Y').value;
460     /* Quinto Campo SUELDO ACTUAL */
461     var manutención = document.getElementById('manutencion').value;
462     /* Sexto Campo MONTO SOLICITADO */
463     var monto = document.getElementById('monto_Y').value;
464     /* Séptimo Campo SISTEMA SALUD PREFERIDO */
465     var salud_preferido = document.getElementById('label_salud_preferido').value;
```

Fuente: Elaboración propia

Figura 63

index.html_15

```
app.py index.html X
465     var salud_preferido = document.getElementById('label_salud_preferido').value;
466     /* Octavo Campo TIENE ALGUNA ENFERMEDAD */
467     var enfermedad = document.getElementById('label_alguna_enfermedad').value;
468     /* Noveno Campo PERTENECE AL SISTEMA DE PENSIONES */
469     var pertenece_sistema_pensiones = document.getElementById('label_pertenece_pensiones').value;
470     /* Decimo Campo EDAD A JUBILARSE */
471     var edad_jubilarse = document.getElementById('edjub').value;
472     /* Onceavo Campo SISTEMA DE PENSIONES QUE LE GUSTA */
473     var tipo_pension = document.getElementById('label_sistema_pensiones').value;
474
475     /* Tiempo de Espera del Servicio*/
476     var minutos = document.getElementById('minutos').innerHTML;
477     var segundos = document.getElementById('segundos').innerHTML;
478     let resultado = document.getElementById('resultado_hidden').value;
479     let satisfaccion = document.getElementById('satisfaccion_hidden').value;
480     let hidden_codigo_generado = document.getElementById('hidden_codigo_generado').value;
481
482     if( genero == "" ){
483         alerta("", "Seleccione un genero", "error");
484         return;
485     }
486
487     if( aspira_ganar_mas == "" ){
488         alerta("", "Seleccione si aspira ganar un sueldo mayor", "error");
489         return;
490     }
491
492     if( enfermedad == "" ){
493         alerta("", "Seleccione su sistema de salud preferido", "error");
494         return;
495     }
496
497     if( pertenece_sistema_pensiones == "" ){
498         alerta("", "Seleccione si pertenece a algun sistema de pensiones", "error");
```

Fuente: Elaboración propia

Figura 64

index.html_16

```
app.py index.html X
497 if( pertenece_sistema_pensiones == ) {
498     alerta("", "Seleccione si pertenece a algun sistema de pensiones", "error");
499     return;
500 }
501
502 if( tipo_pension == "" ){
503     alerta("", "Seleccione si pertenece una AFP u ONP", "error");
504     return;
505 }
506
507
508 db.collection("Pruebas").doc(hidden_codigo_generado).set({
509     edad: edad,
510     genero: genero,
511     aspira_ganar_mas: aspira_ganar_mas,
512     ganancia_deseada: ganancia_deseada,
513     manutencion: manutencion,
514     monto: monto,
515     salud_preferido: salud_preferido,
516     enfermedad: enfermedad,
517     pertenece_sistema_pensiones: pertenece_sistema_pensiones,
518     edad_jubilarse: edad_jubilarse,
519     tipo_pension: tipo_pension,
520     tiemporeal: minutos + ":" + segundos,
521     resultado: resultado,
522     satisfaccion: satisfaccion,
523 })
524 .then(function(docRef){
525
526     var edad_ = document.getElementById('edad').value;
527     var genero_ = document.getElementById('genero').value;
528     var aspira_ganar_mas_ = document.getElementById('aspira_ganar_mas').value;
529     var ganancia_deseada_ = document.getElementById('ganancia_deseada').value;
530     var manutencion_ = document.getElementById('cantManutencion').value;
```

Fuente: Elaboración propia

Figura 65

index.html_17

```
app.py index.html X
530 var manutencion_ = document.getElementById('cantManutencion').value;
531 var monto_ = document.getElementById('monto').value;
532 var salud_preferido_ = document.getElementById('enfermedad').value;
533 var enfermedad_ = document.getElementById('tiene_alguna_enfermedad').value;
534 var pertenece_sistema_pensiones_ = document.getElementById('pertenece_1st_peniones').value;
535 var edad_jubilarse_ = document.getElementById('edad_jubilarse').value;
536 var tipo_pension_ = document.getElementById('idTipoPension').value;
537 var url_procesar = document.getElementById("form").getAttribute("attr-url-predict");
538
539 var formData = new FormData();
540 formData.append('edad', edad_);
541 formData.append('idSexo', genero_);
542 formData.append('aspira_ganar_mas', aspira_ganar_mas_);
543 formData.append('ganancia_deseada', ganancia_deseada_);
544 formData.append('cantManutencion', manutencion_);
545 formData.append('monto', monto_);
546 formData.append('idEnfermedad', salud_preferido_);
547 formData.append('tiene_alguna_enfermedad', enfermedad_);
548 formData.append('pertenece_1st_peniones', pertenece_sistema_pensiones_);
549 formData.append('edad_jubilarse', edad_jubilarse_);
550 formData.append('idTipoPension', tipo_pension_);
551
552 var json_data = {
553     'edad': edad,
554     'idSexo': genero,
555     'aspira_ganar_mas': aspira_ganar_mas,
556     'ganancia_deseada': ganancia_deseada,
557     'cantManutencion': manutencion,
558     'monto': monto,
559     'idEnfermedad': salud_preferido,
560     'tiene_alguna_enfermedad': tiene_alguna_enfermedad,
561     'pertenece_1st_peniones': pertenece_sistema_pensiones,
562     'edad_jubilarse': edad_jubilarse,
563     'idTipoPension': tipo_pension_
```

Fuente: Elaboración propia

Figura 66

index.html_18

```
app.py index.html X
563     'idTipoPension':tipo_pension_
564   }
565
566
567
568
569
570   const options = {
571     method: "POST",
572     body: formData
573   }
574
575   var that = this;
576
577   fetch(url_procesar, options)
578     .then(response => {
579       if( response.ok )
580         return response.text()
581       else
582         throw new Error(response.status);
583     })
584     .then(data => {
585
586       //that.satisfaccion();
587       alerta2_("",data, "success");
588       data2 = document.getElementById('satisfaccion_hidden').value;
589       document.getElementById('resultado_hidden').value = data;
590
591       /*
592       setTimeout(function(){
593         document.location.reload(true);
594       }, 8000);
595       */

```

Fuente: Elaboración propia

Figura 67

index.html_19

```
app.py index.html X
584     .then(data => {
585
586       //that.satisfaccion();
587       alerta2_("",data, "success");
588       data2 = document.getElementById('satisfaccion_hidden').value;
589       document.getElementById('resultado_hidden').value = data;
590
591       /*
592       setTimeout(function(){
593         document.location.reload(true);
594       }, 8000);
595       */
596     })
597     .catch(err => {
598       console.log("Error:", err.message);
599     })
600
601
602
603   }).catch(function(error){
604     console.log("Hay un error al momento de crear el documento: ", error);
605   });
606 }
607
608
609
610 </script>
611 </body>
612 </html>

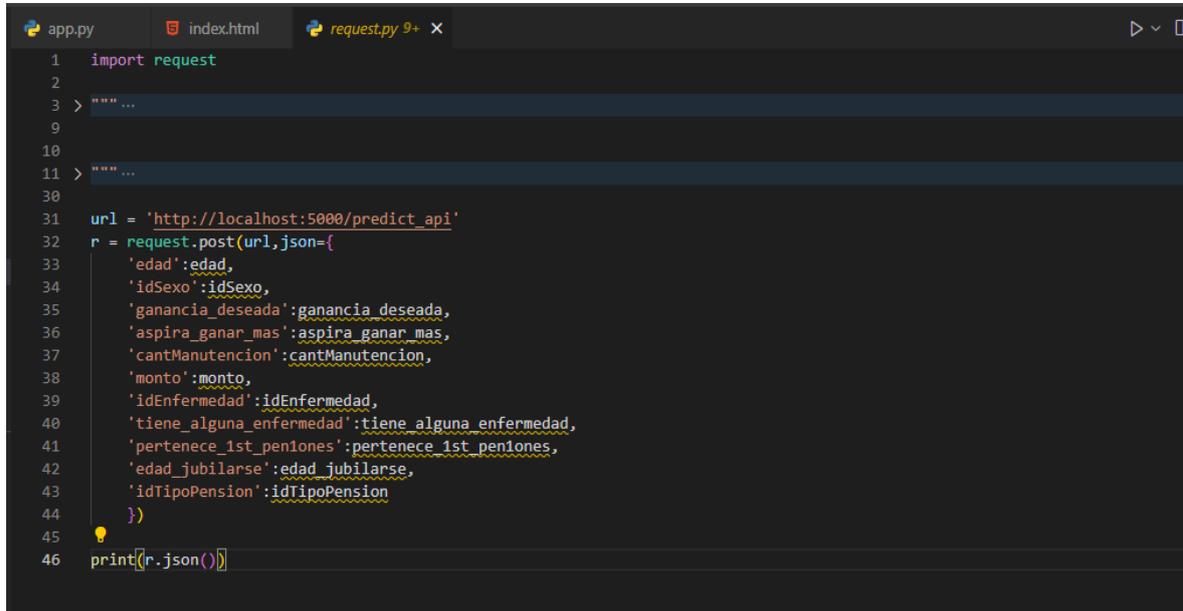
```

Fuente: Elaboración propia

Crear archivo request.py para recibir los datos en formato json

Figura 68

request.py



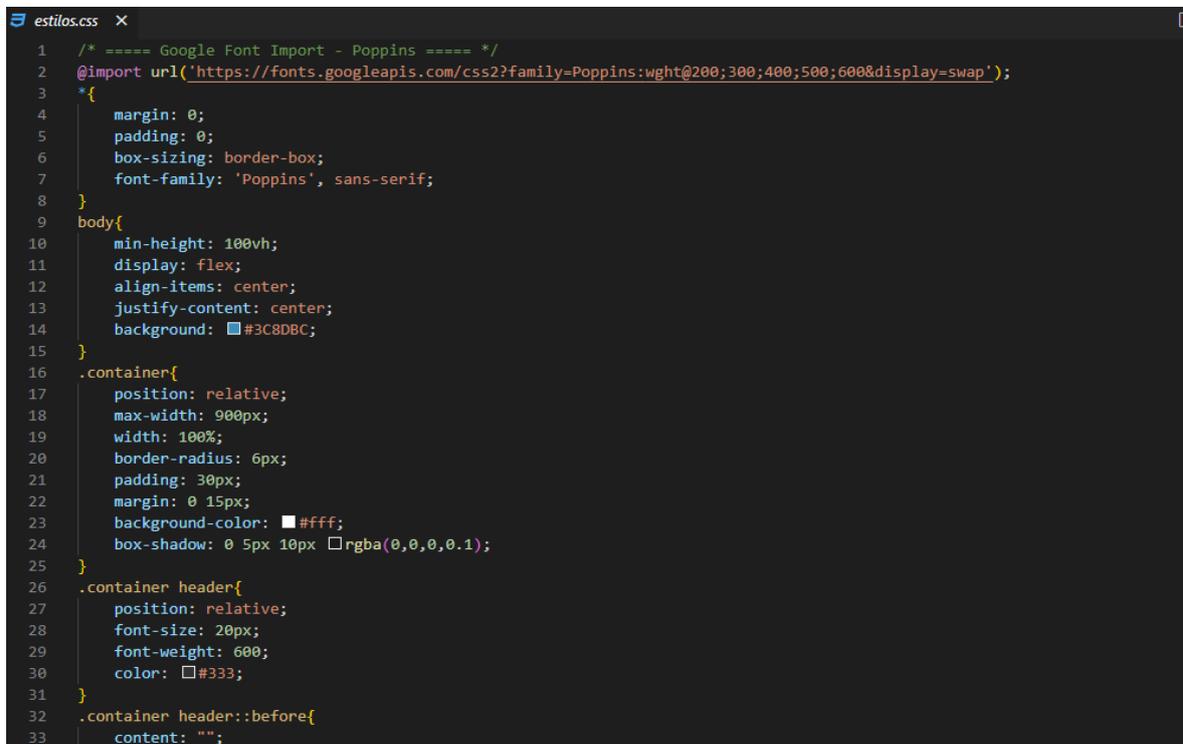
```
1 import request
2
3 > """ ...
9
10
11 > """ ...
30
31 url = 'http://localhost:5000/predict_api'
32 r = request.post(url, json={
33     'edad': edad,
34     'idSexo': idSexo,
35     'ganancia_deseada': ganancia_deseada,
36     'aspira_ganar_mas': aspira_ganar_mas,
37     'cantManutencion': cantManutencion,
38     'monto': monto,
39     'idEnfermedad': idEnfermedad,
40     'tiene_alguna_enfermedad': tiene_alguna_enfermedad,
41     'pertenece_1st_peniones': pertenece_1st_peniones,
42     'edad_jubilarse': edad_jubilarse,
43     'idTipoPension': idTipoPension
44 })
45
46 print(r.json())
```

Fuente: Elaboración propia

Crear archivo estilos.css para dar estilo a la interface

Figura 69

Estilos.css_1



```
1 /* ===== Google Font Import - Poppins ===== */
2 @import url('https://fonts.googleapis.com/css2?family=Poppins:wght@200;300;400;500;600&display=swap');
3 *{
4     margin: 0;
5     padding: 0;
6     box-sizing: border-box;
7     font-family: 'Poppins', sans-serif;
8 }
9 body{
10     min-height: 100vh;
11     display: flex;
12     align-items: center;
13     justify-content: center;
14     background: #3C8DBC;
15 }
16 .container{
17     position: relative;
18     max-width: 900px;
19     width: 100%;
20     border-radius: 6px;
21     padding: 30px;
22     margin: 0 15px;
23     background-color: #fff;
24     box-shadow: 0 5px 10px rgba(0,0,0,0.1);
25 }
26 .container header{
27     position: relative;
28     font-size: 20px;
29     font-weight: 600;
30     color: #333;
31 }
32 .container header::before{
33     content: "";
```

Fuente: Elaboración propia

Figura 70

Estilos.css_2

```
estilos.css x
34     position: absolute;
35     left: 0;
36     bottom: -2px;
37     height: 3px;
38     width: 27px;
39     border-radius: 8px;
40     background-color: #3C8DBC;
41 }
42 .container #form{
43     position: relative;
44     margin-top: 16px;
45     min-height: 490px;
46     background-color: #fff;
47     overflow: hidden;
48 }
49 .container #form .form{
50     position: absolute;
51     background-color: #fff;
52     transition: 0.3s ease;
53 }
54 .container #form .form.second{
55     opacity: 0;
56     pointer-events: none;
57     transform: translateX(100%);
58 }
59 #form.secActive .form.second{
60     opacity: 1;
61     pointer-events: auto;
62     transform: translateX(0);
63 }
64 #form.secActive .form.first{
65     opacity: 0;
66     pointer-events: none;
```

Fuente: Elaboración propia

Figura 71

Estilos.css_3

```
estilos.css x
67     transform: translateX(-100%);
68 }
69 .container #form .title{
70     display: block;
71     margin-bottom: 8px;
72     font-size: 16px;
73     font-weight: 500;
74     margin: 6px 0;
75     color: #333;
76 }
77 .container #form .fields{
78     display: flex;
79     align-items: center;
80     justify-content: space-between;
81     flex-wrap: wrap;
82 }
83 #form .fields .input-field{
84     display: flex;
85     width: calc(100% / 3 - 15px);
86     flex-direction: column;
87     margin: 4px 0;
88 }
89 .input-field label{
90     font-size: 12px;
91     font-weight: 500;
92     color: #2e2e2e;
93 }
94 .input-field input, select{
95     outline: none;
96     font-size: 14px;
97     font-weight: 400;
98     color: #333;
99     border-radius: 5px;
```

Fuente: Elaboración propia

Figura 72

Estilos.css_4

```
estilos.css x
100 border: 1px solid #aaa;
101 padding: 0 15px;
102 height: 42px;
103 margin: 8px 0;
104 }
105 .input-field input :focus,
106 .input-field select:focus{
107     box-shadow: 0 3px 6px #rgba(0,0,0,0.13);
108 }
109 .input-field select,
110 .input-field input[type="date"]{
111     color: #707070;
112 }
113 .input-field input[type="date"]:valid{
114     color: #333;
115 }
116 .container #form button, .backBtn{
117     display: flex;
118     align-items: center;
119     justify-content: center;
120     height: 45px;
121     max-width: 200px;
122     width: 100%;
123     border: none;
124     outline: none;
125     color: #fff;
126     border-radius: 5px;
127     margin: 25px 0;
128     background-color: #3C8DBC;
129     transition: all 0.3s linear;
130     cursor: pointer;
131 }
132 .container #form .btnText{
```

Fuente: Elaboración propia

Figura 73

Estilos.css_5

```
estilos.css x
133 font-size: 14px;
134 font-weight: 400;
135 }
136 #form button:hover{
137     background-color: #3C8DBC;
138 }
139 #form button i,
140 #form .backBtn i{
141     margin: 0 6px;
142 }
143 #form .backBtn i{
144     transform: rotate(180deg);
145 }
146 #form .buttons{
147     display: flex;
148     align-items: center;
149 }
150 #form .buttons button, .backBtn{
151     margin-right: 14px;
152 }
153
154 @media (max-width: 750px) {
155     .container #form{
156         overflow-y: scroll;
157     }
158     .container #form::-webkit-scrollbar{
159         display: none;
160     }
161     #form .fields .input-field{
162         width: calc(100% / 2 - 15px);
163     }
164 }
```

Fuente: Elaboración propia

Figura 74

Estilos.css_6

```
estilos.css x
166 @media (max-width: 550px) {
167     #form .fields .input-field{
168         width: 100%;
169     }
170 }
171
172 /* TIMER */
173 #number {
174     font-size:1px;
175     color: white;
176     display: hidden;
177 }
178 /* TIMER */
179
180
181 /* START TOOLTIP */
182
183 .tooltip {
184     position: relative;
185     display: inline-block;
186     /*border-bottom: 1px dotted black;*/
187 }
188
189 .tooltip .tooltiptext {
190     visibility: hidden;
191     width: 250px;
192     background-color: #555;
193     color: #fff;
194     text-align: center;
195     border-radius: 6px;
196     padding: 5px 0;
197     position: absolute;
198     z-index: 1;
```

Fuente: Elaboración propia

Figura 75

Estilos.css_7

```
estilos.css x
199     bottom: 125%;
200     left: 27%;
201     margin-left: -60px;
202     opacity: 0;
203     transition: opacity 0.3s;
204 }
205
206 .tooltip .tooltiptext::after {
207     content: "";
208     position: absolute;
209     top: 100%;
210     left: 50%;
211     margin-left: -5px;
212     border-width: 5px;
213     border-style: solid;
214     border-color: #555 transparent transparent transparent;
215 }
216
217 .tooltip:hover .tooltiptext {
218     visibility: visible;
219     opacity: 1;
220 }
221 output#label_genero, output#label_mayor_sueldo, output#label_salud_preferido, output#label_alguna_enfermedad,
222     output#label_pertenece_pensiones, output#label_sistema_pensiones {
223     display: none;
224 }
225
226 div#label_tiempo_real {
227     display:none;
228 }
229 .swal2-radio {
230     display: grid !important;
231     text-align: left !important;
232 }
```

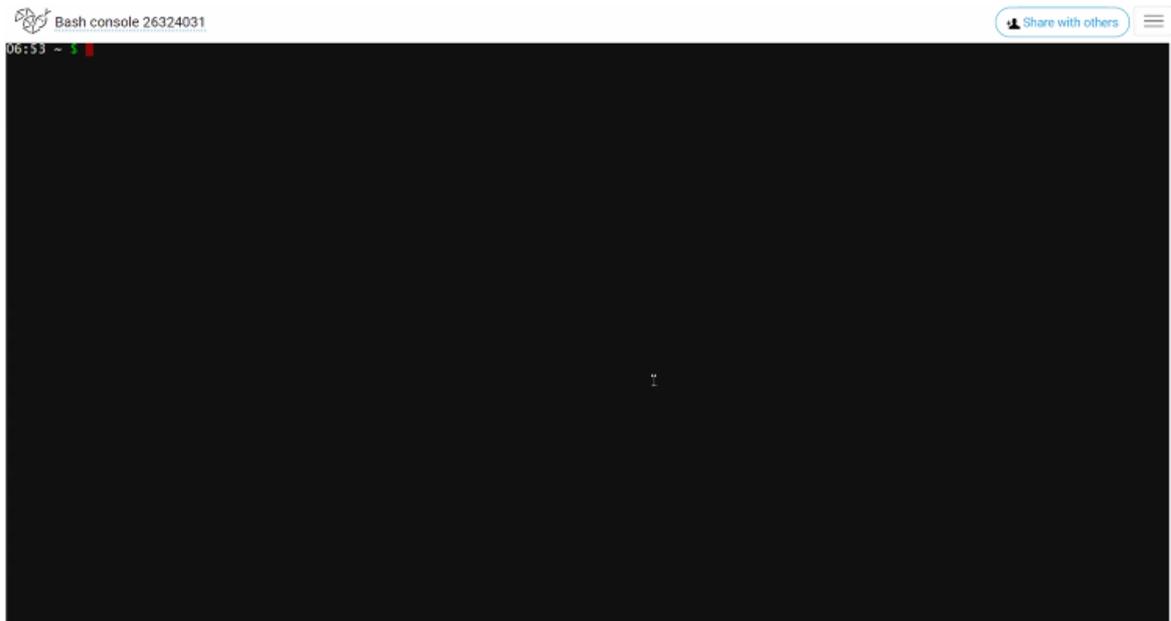
Fuente: Elaboración propia

Despliegue en PythonAnywhere

Se requiere un conocimiento intermedio de las variables de entorno por cada versión desplegada. Al igual, es por la consola, que de no configurarse se podrían generar problemas con el funcionamiento del sistema.

Figura 76

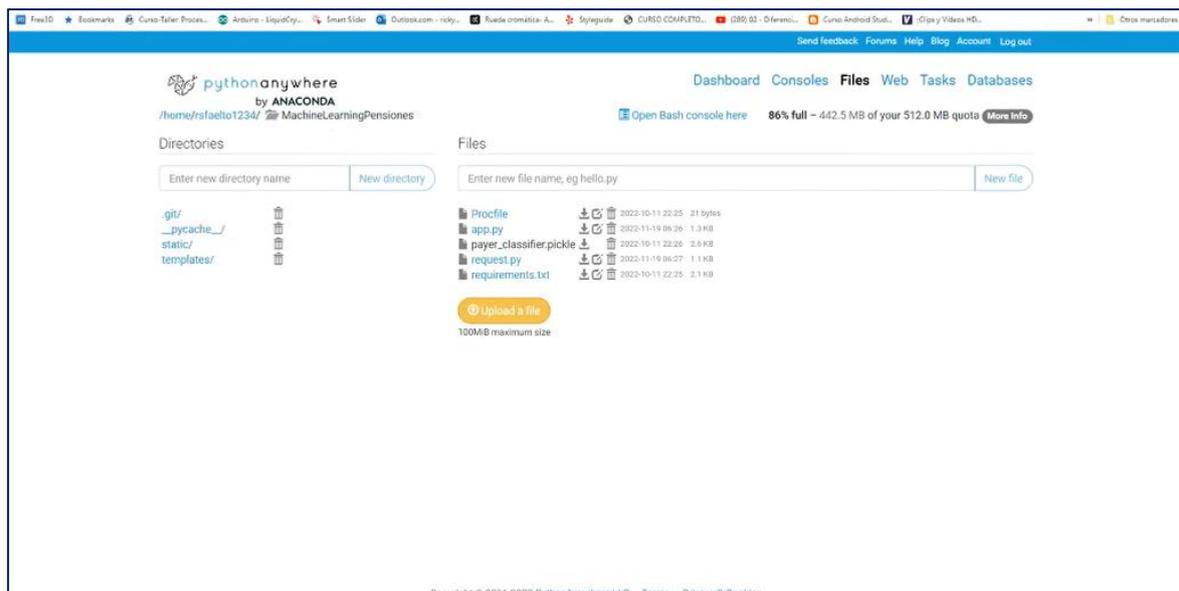
Consola de PythonAnywhere



Fuente: Elaboración propia

Figura 77

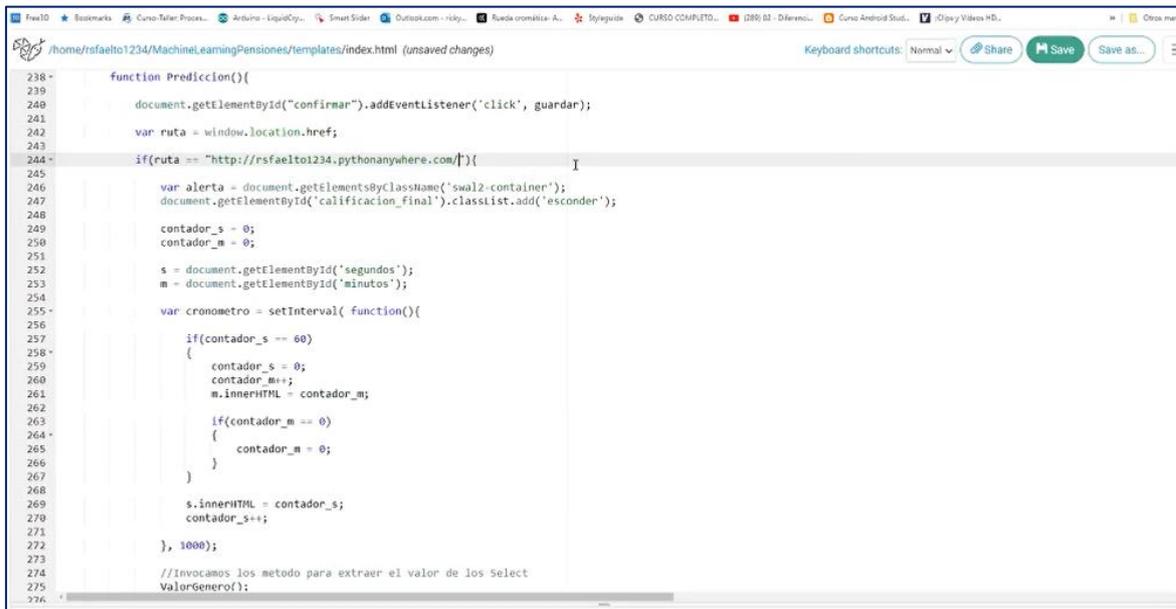
PythonAnywhere_1



Fuente: Elaboración propia

Figura 78

PythonAnywhere_2



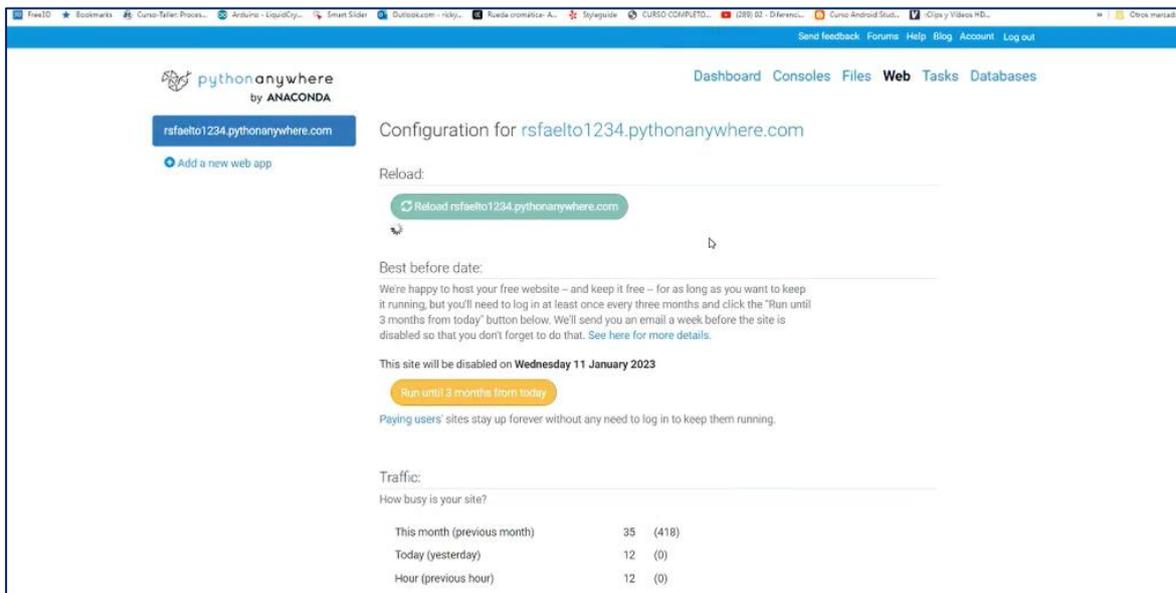
```
238- function Prediccion(){
239-
240-     document.getElementById("confirmar").addEventListener('click', guardar);
241-
242-     var ruta = window.location.href;
243-
244-     if(ruta == "http://rsfaelto1234.pythonanywhere.com/"){
245-
246-         var alerta = document.getElementsByClassName('swal2-container');
247-         document.getElementById("calificacion_final").classList.add('esconder');
248-
249-         contador_s = 0;
250-         contador_m = 0;
251-
252-         s = document.getElementById('segundos');
253-         m = document.getElementById('minutos');
254-
255-         var cronometro = setInterval( function(){
256-
257-             if(contador_s == 60)
258-             {
259-                 contador_s = 0;
260-                 contador_m++;
261-                 m.innerHTML = contador_m;
262-
263-                 if(contador_m == 0)
264-                 {
265-                     contador_m = 0;
266-                 }
267-             }
268-
269-             s.innerHTML = contador_s;
270-             contador_s++;
271-
272-         }, 1000);
273-
274-         //Invocamos los metodo para extraer el valor de los Select
275-         valorGenerof();
276-
```

Fuente: Elaboración propia

En la figura anterior el mismo PythonAnywhere tiene que

Figura 79

PythonAnywhere_3



pythonanywhere by ANACONDA

rsfaelto1234.pythonanywhere.com

Dashboard Consoles Files Web Tasks Databases

Configuration for rsfaelto1234.pythonanywhere.com

Reload:

Reload rsfaelto1234.pythonanywhere.com

Best before date:

We're happy to host your free website – and keep it free – for as long as you want to keep it running, but you'll need to log in at least once every three months and click the "Run until 3 months from today" button below. We'll send you an email a week before the site is disabled so that you don't forget to do that. See [here](#) for more details.

This site will be disabled on **Wednesday 11 January 2023**

Run until 3 months from today

Paying users' sites stay up forever without any need to log in to keep them running.

Traffic:

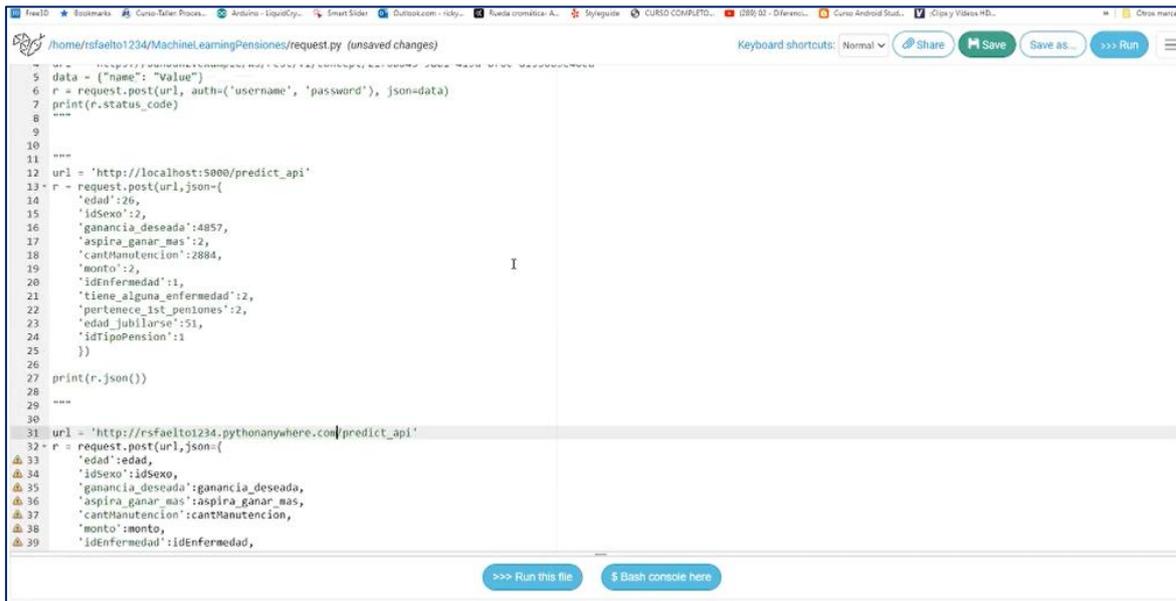
How busy is your site?

Period	Visits	Unique
This month (previous month)	35	(418)
Today (yesterday)	12	(0)
Hour (previous hour)	12	(0)

Fuente: Elaboración propia

Figura 80

PythonAnywhere_4

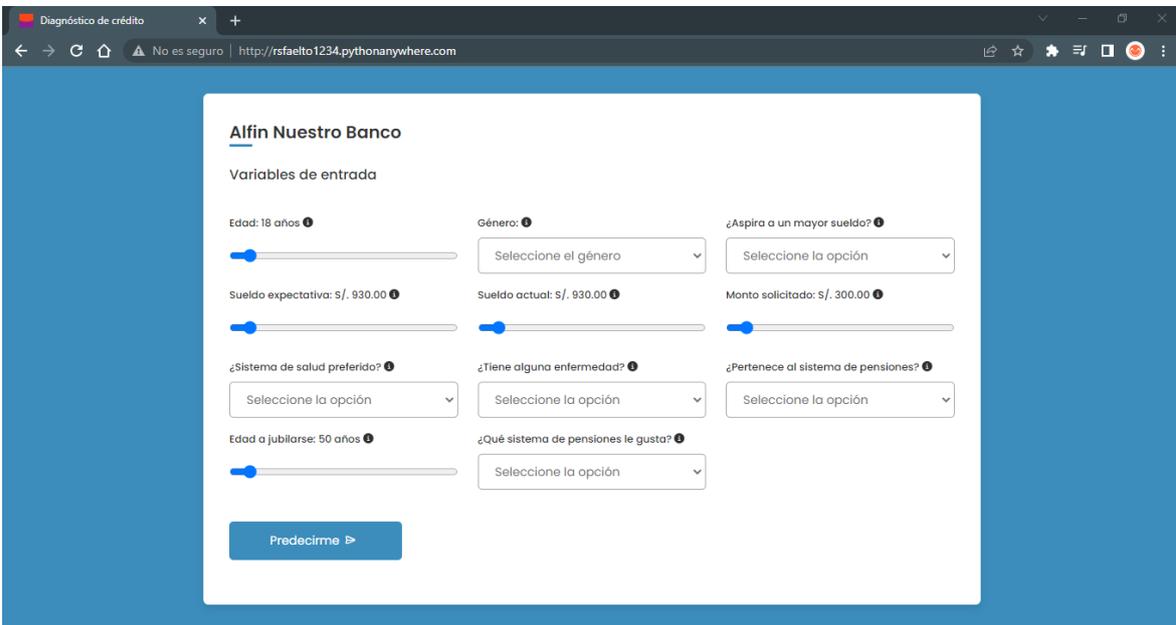


```
5 data = {"name": "Value"}
6 r = request.post(url, auth=('username', 'password'), json=data)
7 print(r.status_code)
8
9
10
11
12 url = 'http://localhost:5000/predict_api'
13 r = request.post(url,json={
14     'edad':26,
15     'idSexo':2,
16     'ganancia_deseada':4857,
17     'aspira_ganar_mas':2,
18     'cantManutencion':2884,
19     'monto':2,
20     'idEnfermedad':1,
21     'tiene_alguna_enfermedad':2,
22     'pertenece_1st_pensiones':2,
23     'edad_jubilarse':51,
24     'idTipoPension':1
25 })
26
27 print(r.json())
28
29
30
31 url = 'http://rfsfaelto1234.pythonanywhere.com/predict_api'
32 r = request.post(url,json={
33     'edad':edad,
34     'idSexo':idSexo,
35     'ganancia_deseada':ganancia_deseada,
36     'aspira_ganar_mas':aspira_ganar_mas,
37     'cantManutencion':cantManutencion,
38     'monto':monto,
39     'idEnfermedad':idEnfermedad,
```

Fuente: Elaboración propia

Figura 81

Interfaz del aplicativo



Diagnóstico de crédito

Alfin Nuestro Banco

Variables de entrada

Edad: 18 años

Género: Seleccione el género

¿Aspira a un mayor sueldo?: Seleccione la opción

Sueldo expectativa: \$/. 930.00

Sueldo actual: \$/. 930.00

Monto solicitada: \$/. 300.00

¿Sistema de salud preferido?: Seleccione la opción

¿Tiene alguna enfermedad?: Seleccione la opción

¿Pertenece al sistema de pensiones?: Seleccione la opción

Edad a jubilarse: 50 años

¿Qué sistema de pensiones le gusta?: Seleccione la opción

Predecirme

Fuente: Elaboración propia

Se utilizaron Tooltips (Información de que es una forma nativa de javascript para poner un mensaje de contexto para los usuarios que requieren una explicación corta de los campos. A continuación se mencionan las etiquetas y los respectivos Tooltips en la siguiente tabla:

Tabla 30*Tooltips del sistema*

Etiquetas	Tootip
Edad	Puedes escoger una edad entre 18 a 100 años, porque 18 es lo mínimo permitido para el préstamo
Género	Puedes escoger tu género
¿Aspira a un mayor sueldo?	Si aspiras a un mayor sueldo a ganar, si quieres más ingresos
Sueldo expectativa	Cuál es tu sueldo expectativa, específicamente cuánto quisieras ganar (aproximadamente)
Sueldo actual	Cuánto ganas actualmente, en el banco está limitado a que ganes por lo menos 930 para que sea el préstamo
Monto solicitado	Cuánto es tu monto que solicitas al banco como préstamo
¿Sistema de salud preferido?	Tenemos otras opciones, si tu prefieres un sistema de salud público o privado
¿Tiene alguna enfermedad?	Si acá tu tienes algunas enfermedad
¿Pertenece al sistema de pensiones?	Si aquí tu estás en un sistema de pensiones, cosa que es probable que no si eres joven
Edad a jubilarse	A qué edad quisieras jubilarte, a qué edad desearías jubilarte
¿Qué sistema de pensiones le gusta?	Qué sistema de pensiones te gusta, el AFP que es el privado o la ONP que es el público

Fuente: Elaboración propia

El resultado que procede a mostrar cuando se presiona en el siguiente botón:

Figura 82

Botón principal para predecir

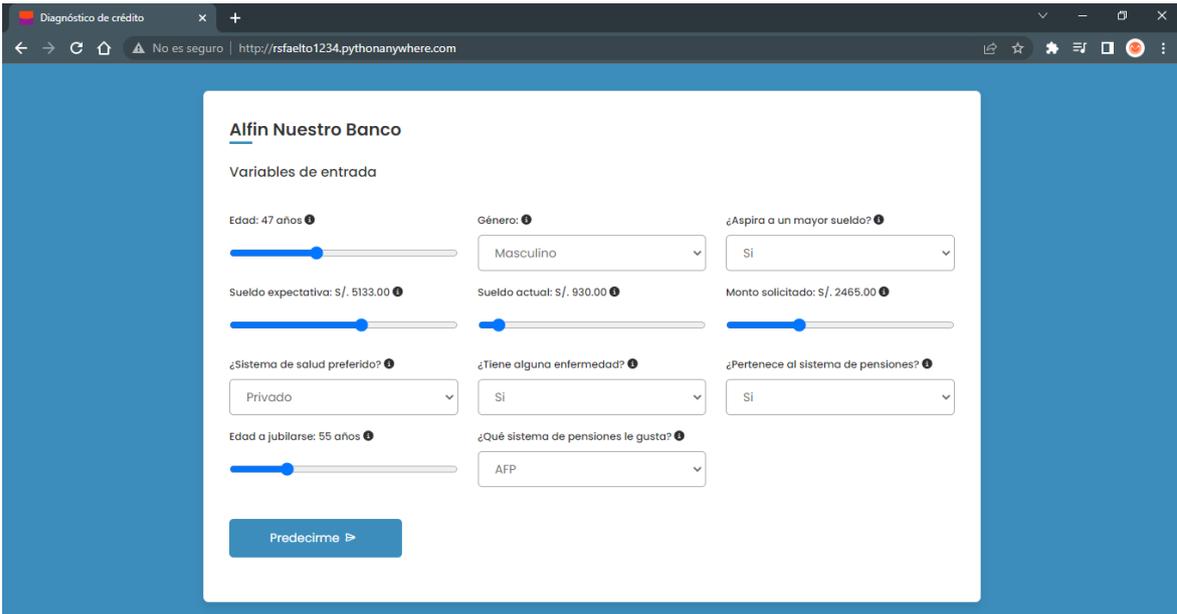


Fuente: Elaboración propia

Cuando se seleccionen los campos se podrá ver que todo está listo para mandar a predecir con los datos proporcionados por el usuario así como se visualizan en la siguiente figura:

Figura 83

Interfaz del aplicativo con datos

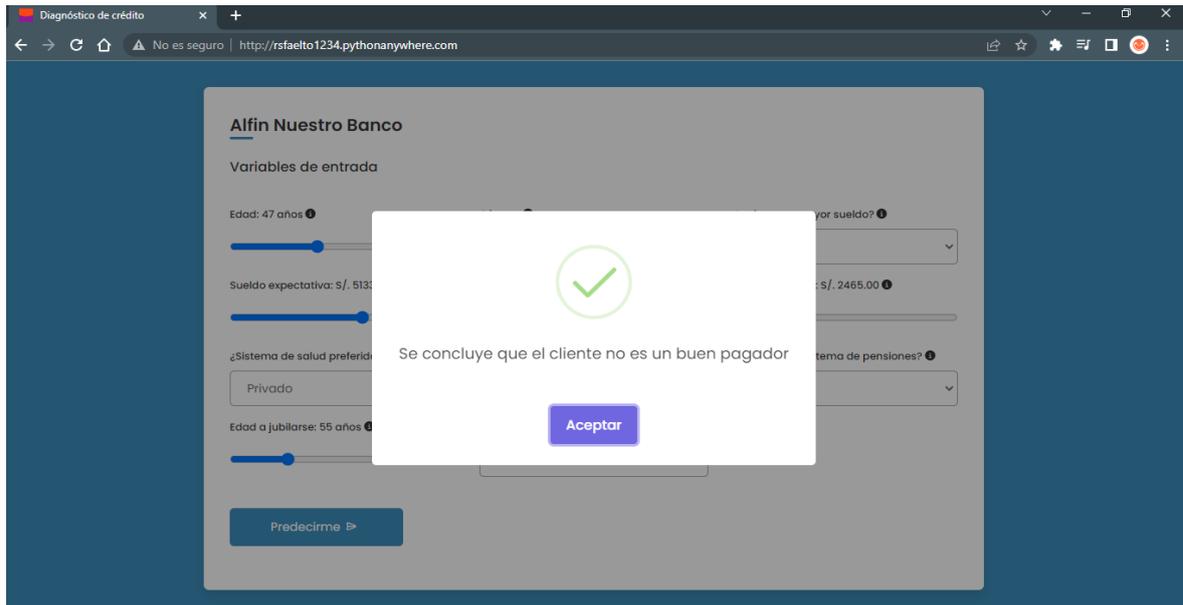
A screenshot of a web browser displaying a credit diagnosis form. The browser's address bar shows 'http://rsfaelto1234.pythonanywhere.com'. The form is titled 'Alfin Nuestro Banco' and 'Variables de entrada'. It contains several input fields: 'Edad: 47 años' with a slider; 'Género: Masculino' with a dropdown; '¿Aspira a un mayor sueldo?: Si' with a dropdown; 'Sueldo expectativa: S/. 5133.00' with a slider; 'Sueldo actual: S/. 930.00' with a slider; 'Monto solicitado: S/. 2465.00' with a slider; '¿Sistema de salud preferido?: Privado' with a dropdown; '¿Tiene alguna enfermedad?: Si' with a dropdown; '¿Pertenece al sistema de pensiones?: Si' with a dropdown; 'Edad a jubilarse: 55 años' with a slider; and '¿Qué sistema de pensiones le gusta?: AFP' with a dropdown. A blue 'Predecirme' button is at the bottom.

Fuente: Elaboración propia

Al usuario se le facilitará ingresar los datos con los Tootips explicando anteriormente, ya que la capacitación realizada en el pasado sirvió como reactivo para poder generar un contexto a las etiquetas mostradas, de modo que las dudas acerca del formulario sean absueltas en el sistema.

Figura 84

Interfaz del aplicativo con resultado

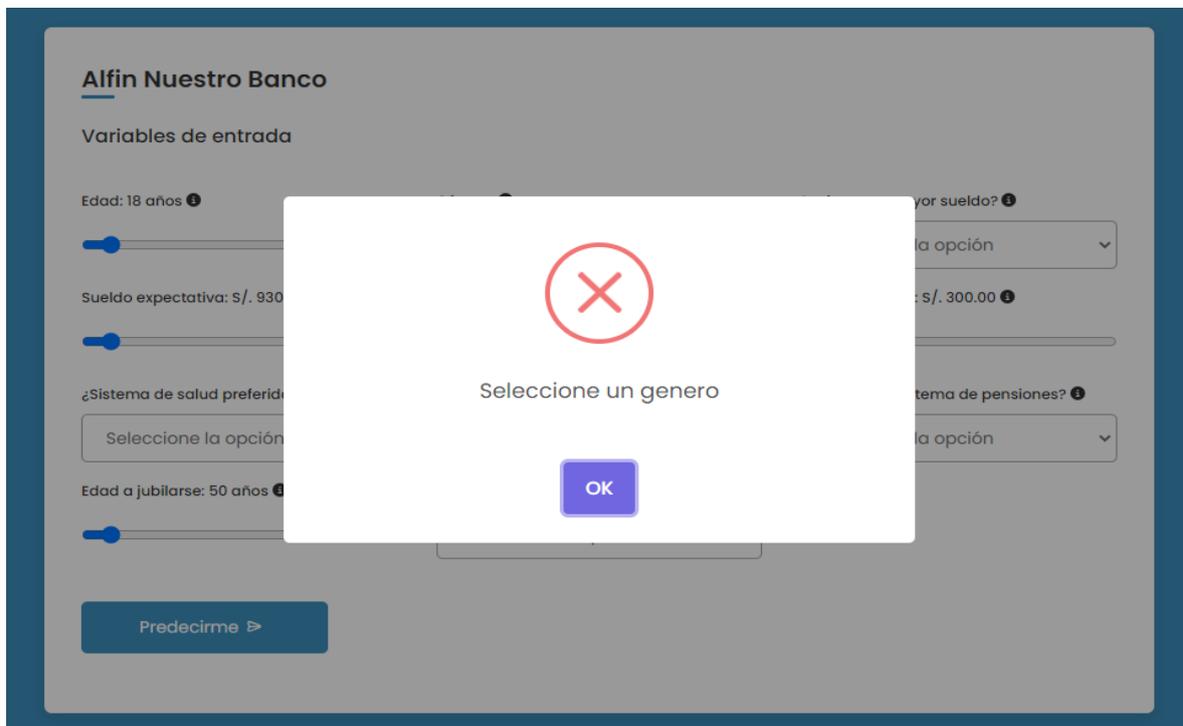


Fuente: Elaboración propia

La figura se mostrará siempre que el usuario haya ingresado todos sus campos, caso contrario se le mostrará una validación implementada con el SweetAlert:

Figura 85

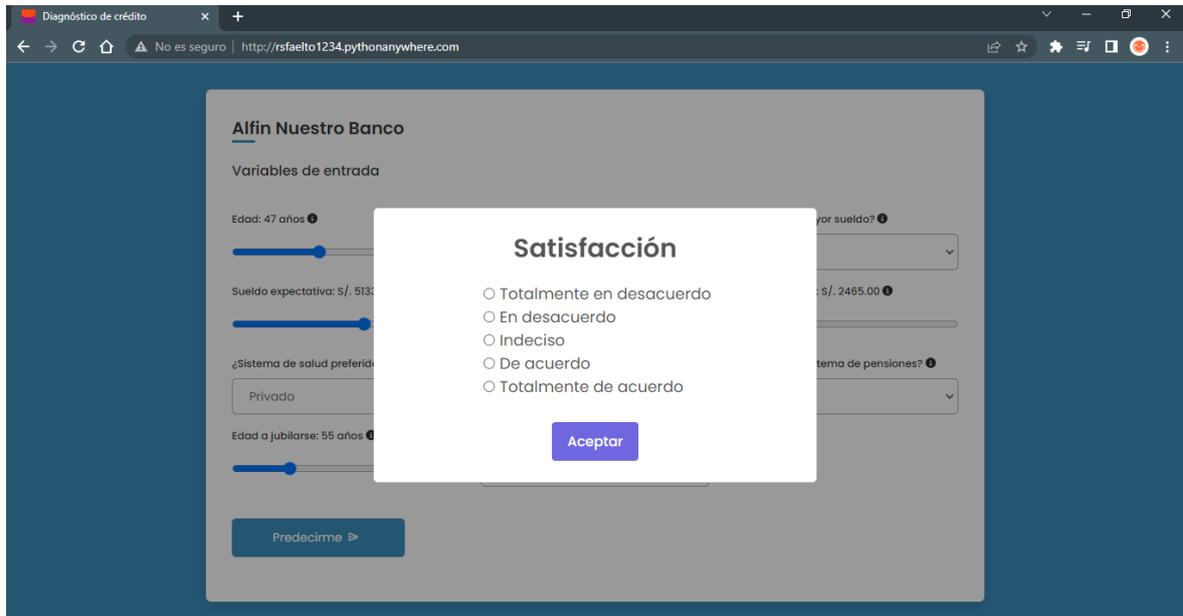
Validación de campos



Fuente: Elaboración propia

Figura 86

Interfaz del aplicativo para satisfacción

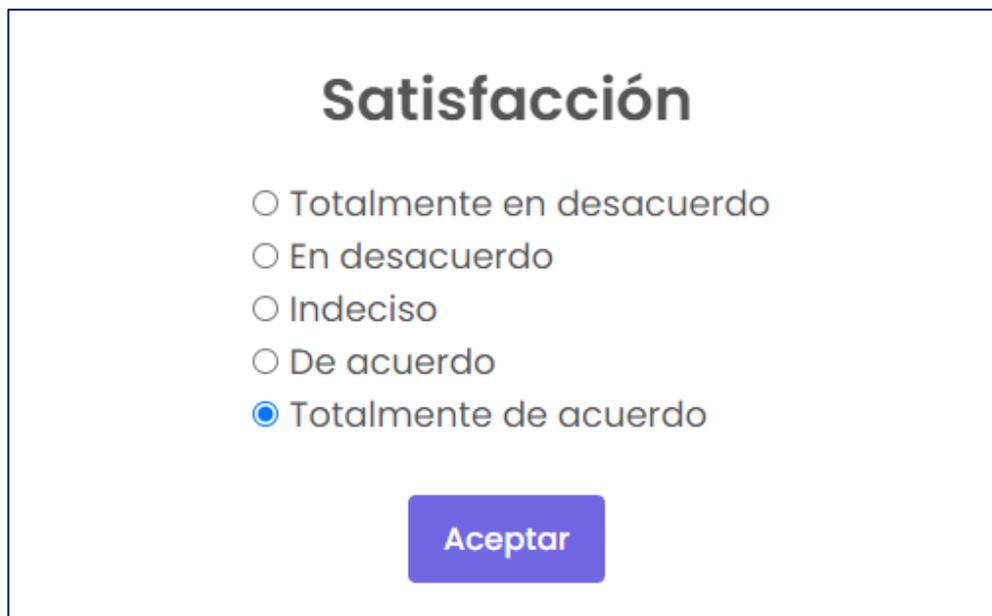


Fuente: Elaboración propia

La modal de satisfacción se muestra porque el usuario ya vio su resultado. En la escala de Likert la satisfacción que se propone se ha realizado con SweetAlert.

Figura 87

Modal de satisfacción



Fuente: Elaboración propia

La aplicación se ha programado para mostrar por console log la uID como se observa a continuación:

Figura 88

uID del documento generado

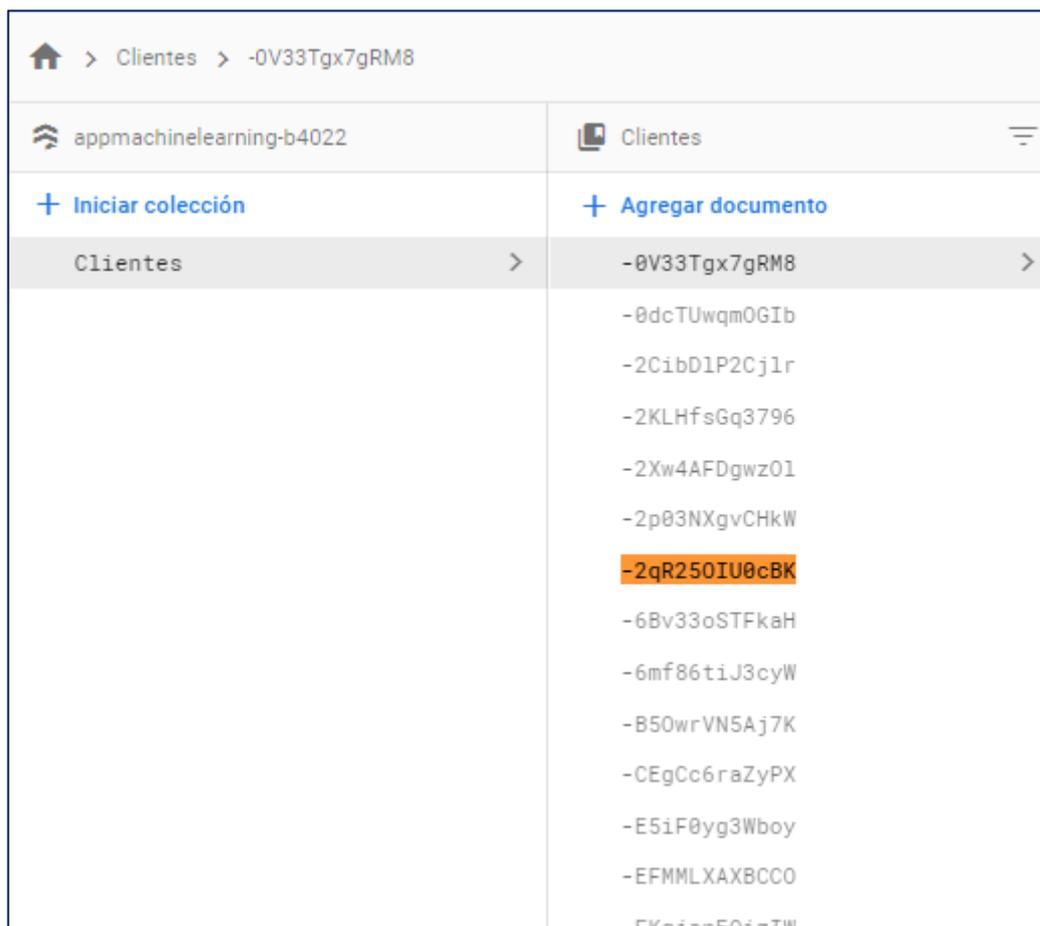
```
Se esta actualizando el id:  
-2qR250IU0cBK
```

Fuente: Elaboración propia

El documento generado tiene su uID, el cuál se encuentra en firebase en el nodo de “Clientes”.

Figura 89

uID del documento generado en Firestore



Fuente: Elaboración propia

Este uID nos trae los siguientes datos:

Figura 90

Indicadores en el sistema

The screenshot displays a web application interface for managing clients. On the left, a list of client IDs is shown, with '-2qR250IU0cBK' selected. On the right, the details for this client are displayed, including various attributes and their values. Three specific data points are highlighted with red boxes:

- resultado: "Se concluye que el cliente no es un buen pagador"
- satisfaccion: "Totalmente de acuerdo"
- tiemporeal: "9:36"

Fuente: Elaboración propia

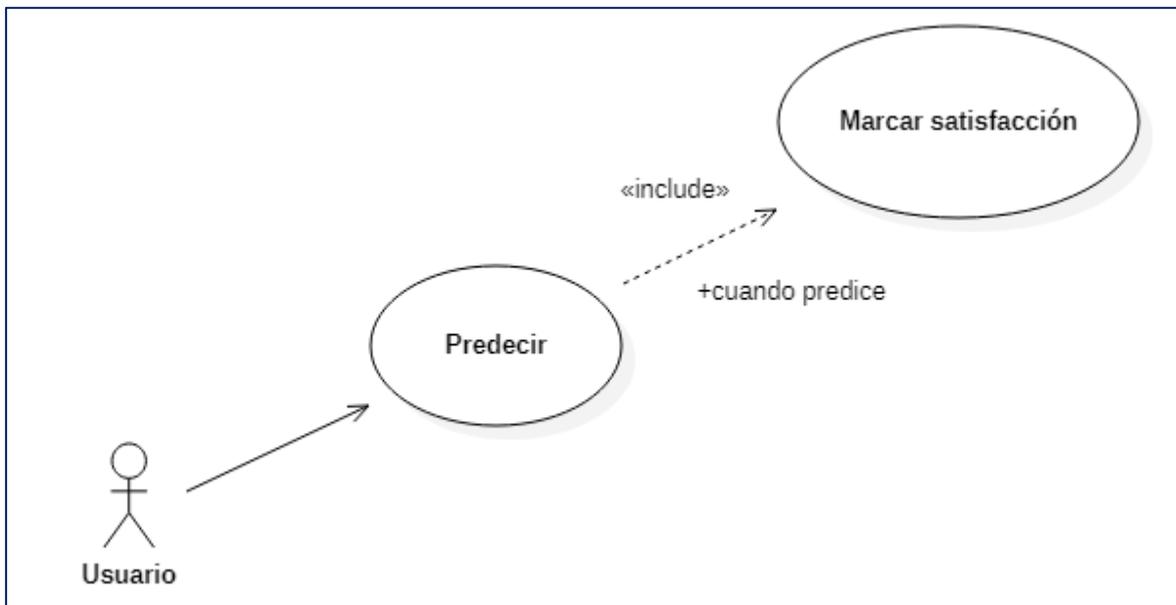
Esta figura hace énfasis en 3 datos capturados que son: resultado, el cual es el resultado de la predicción que se realizó en función de los datos ingresados por el usuario, también se tiene la satisfacción que es la modal que apareció después de aceptar los resultados para que el usuario indique su satisfacción y por último se muestra el tiempo real que es un cronómetro implementado para que se capture el tiempo que transcurrió el usuario desde que ingresó al sistema hasta que hizo clic en "Predecirme" y se le mostrará las respectivas modales.

De esta forma se hace cumplir todos los indicadores, ya que para el tiempo de espera se requiere este tiempo mostrado, para costo de personal, se requiere este tiempo real y una variable auxiliar que se conoce por la información del salario que hizo conocer el banco Alfin que es constante, la satisfacción del cliente que se cumple con la escala de Likert y la eficiencia de evaluación que está en función del tiempo real y un tiempo objetivo proporcionado por el banco Alfin en concordancia con sus necesidades.

Anexo 19: Ingeniería de software

Figura 91

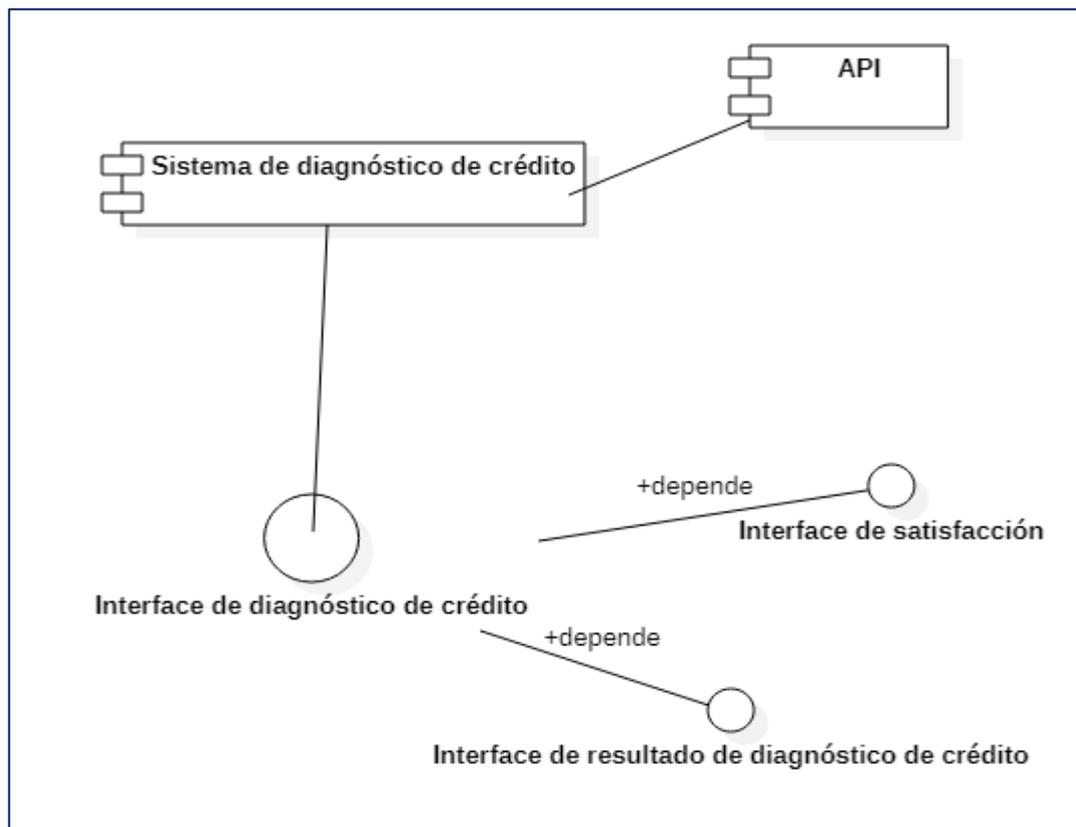
Diagrama de casos de uso



Fuente: Elaboración propia

Figura 92

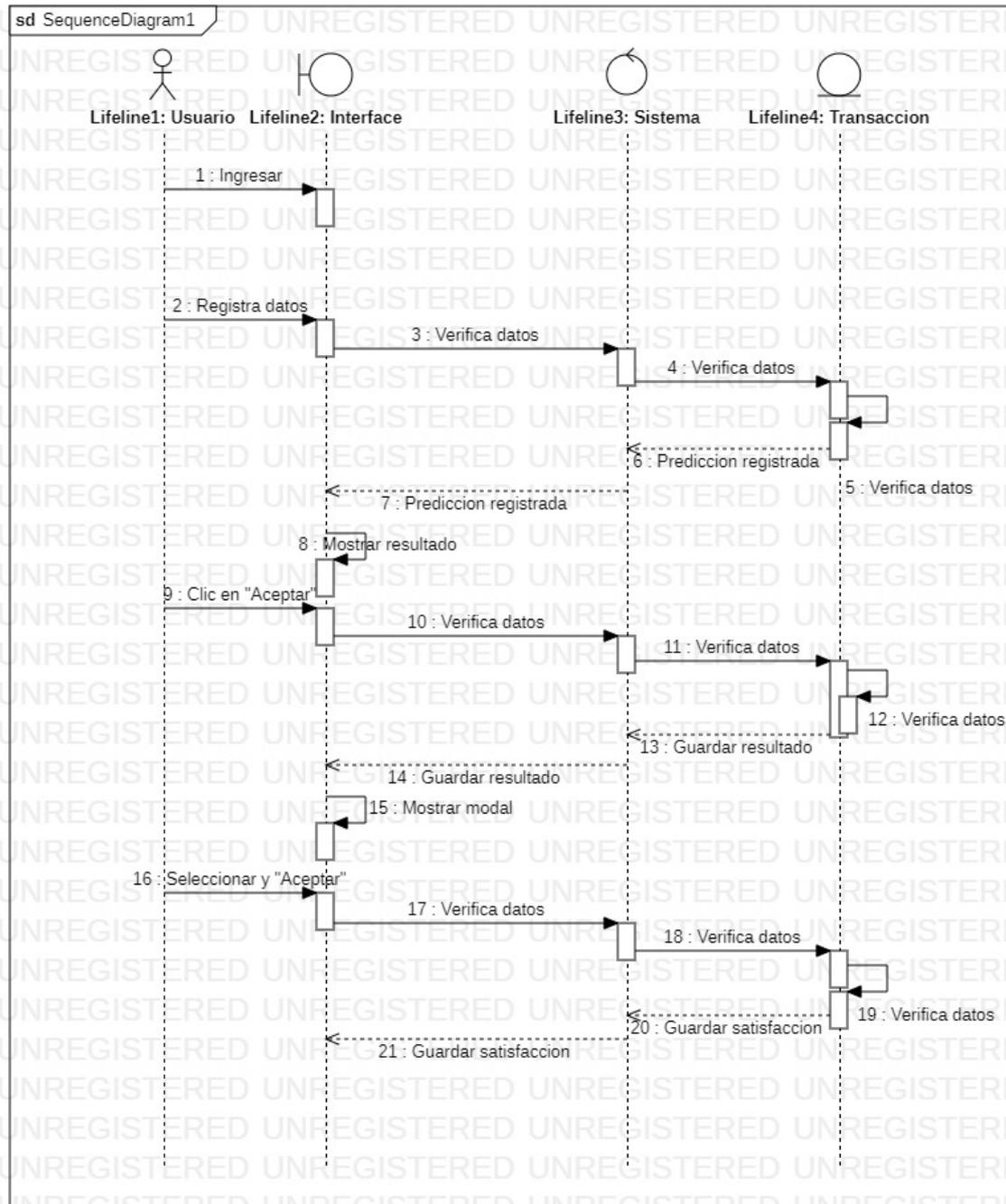
Diagrama de componentes



Fuente: Elaboración propia

Figura 93

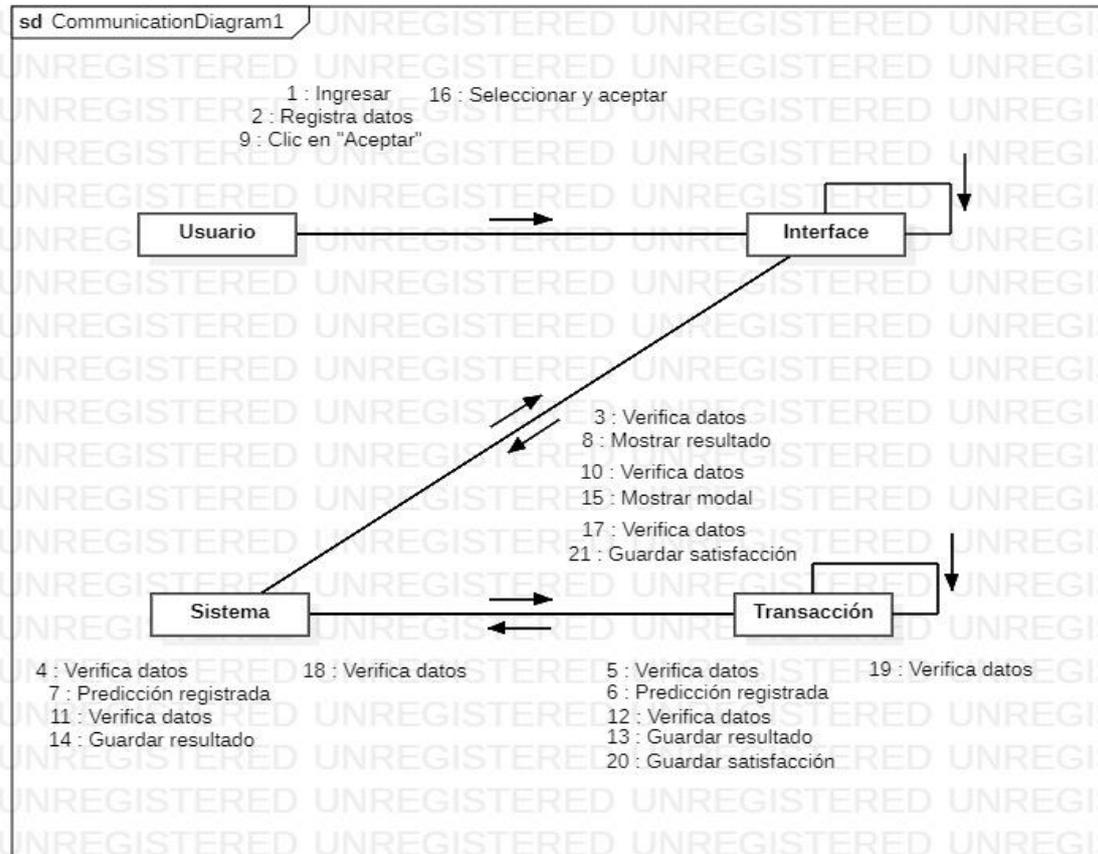
Diagrama de secuencia



Fuente: Elaboración propia

Figura 94

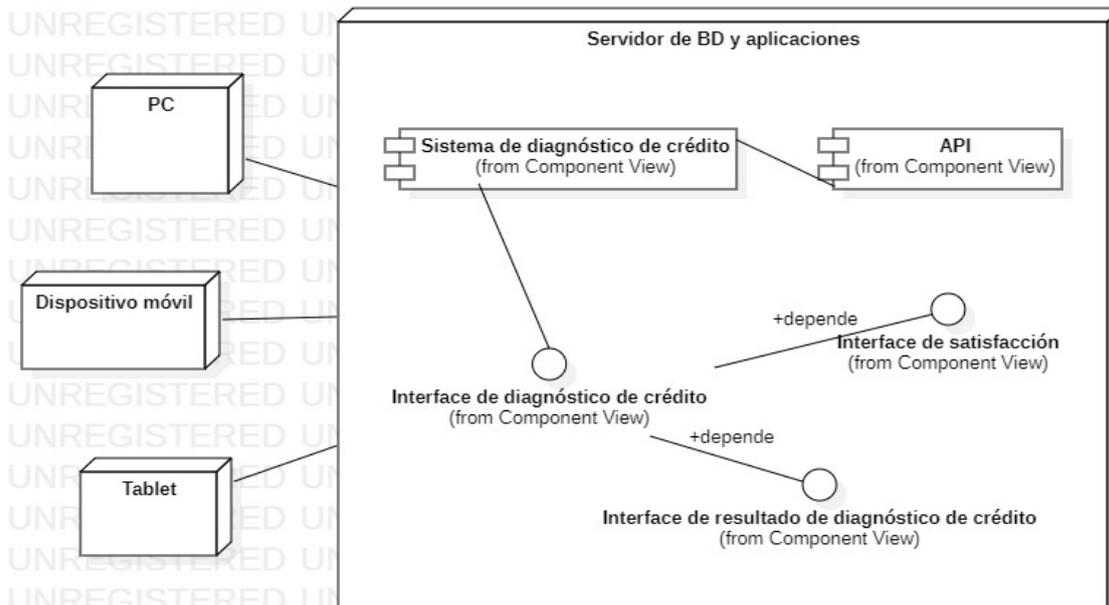
Diagrama de colaboración



Fuente: Elaboración propia

Figura 95

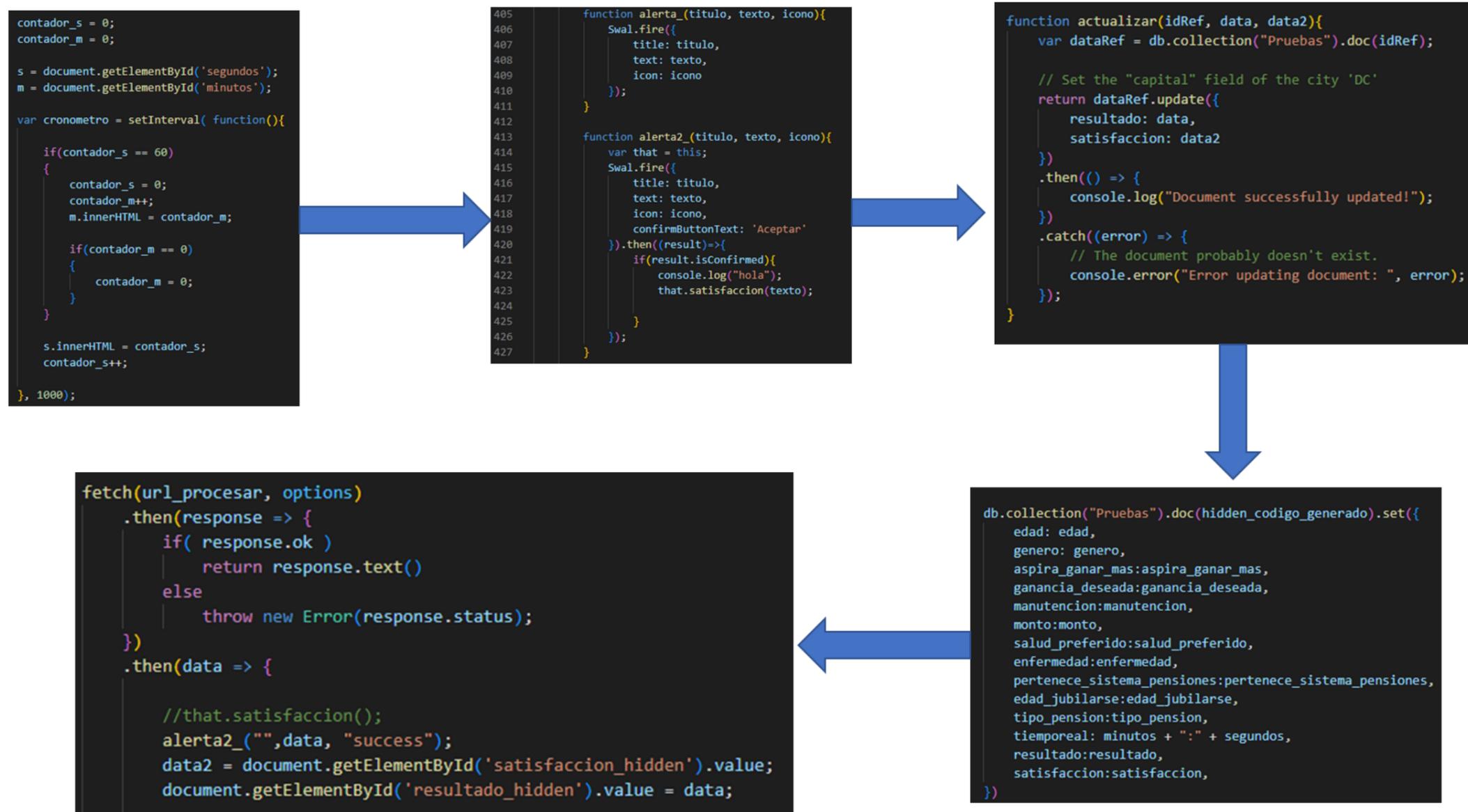
Diagrama de despliegue



Fuente: Elaboración propia

Mostrar la modal de satisfacción después de mostrarle al usuario la modal de su resultado de la predicción y que la opción que elija de su satisfacción se actualice en la base de datos para un formulario que se construye desde la lógica por temas de consulta de post a un api rest y que el resultado de su predicción, su satisfacción y el tiempo real hasta que se le mostró su resultado y los datos que registró en el formulario se guarden en la base de datos en el mismo doc sin realizar una recarga o llevarlo a otra página.

Figura96
Algoritmo innovador



Fuente: Elaboración propia

Figura 97

Arquitectura tecnológica

1. El usuario podrá ingresar al sistema desde su dispositivo móvil, tableta u ordenador de escritorio.

Alfin Nuestro Banco

Variables de entrada

Edad: 18 años

Género: Seleccione el género

¿Aspira a un mayor sueldo?: Seleccione la opción

Sueldo esperativa: \$/ 930.00

Sueldo actual: \$/ 930.00

Monto solicitado: \$/ 300.00

¿Sistema de salud preferido?: Seleccione la opción

¿Tiene alguna enfermedad?: Seleccione la opción

¿Pertenece al sistema de pensiones?: Seleccione la opción

Edad o jubilarse: 50 años

¿Qué sistema de pensiones le gusta?: Seleccione la opción

Predecir Cliente >>

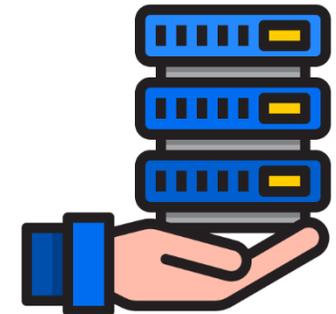
2. A través de la conexión a internet podrá predecir cliente.



3. Una vez presionado, el web service enviará la petición al servidor



4. El servidor comprobará como llegaron los datos y si están completos para enviar una respuesta



5. El web service envía la respuesta al sistema

6. El sistema muestra el resultado al usuario

Alfin Nuestro Banco

Variables de entrada

Edad: 18 años

Género: Seleccione el género

¿Aspira a un mayor sueldo?: Seleccione la opción

Sueldo esperativa: \$/ 930.00

Sueldo actual: \$/ 930.00

Monto solicitado: \$/ 300.00

¿Sistema de salud preferido?: Seleccione la opción

¿Tiene alguna enfermedad?: Seleccione la opción

¿Pertenece al sistema de pensiones?: Seleccione la opción

Edad o jubilarse: 50 años

¿Qué sistema de pensiones le gusta?: Seleccione la opción

Predecir Cliente >>

Resultado de su evaluación!

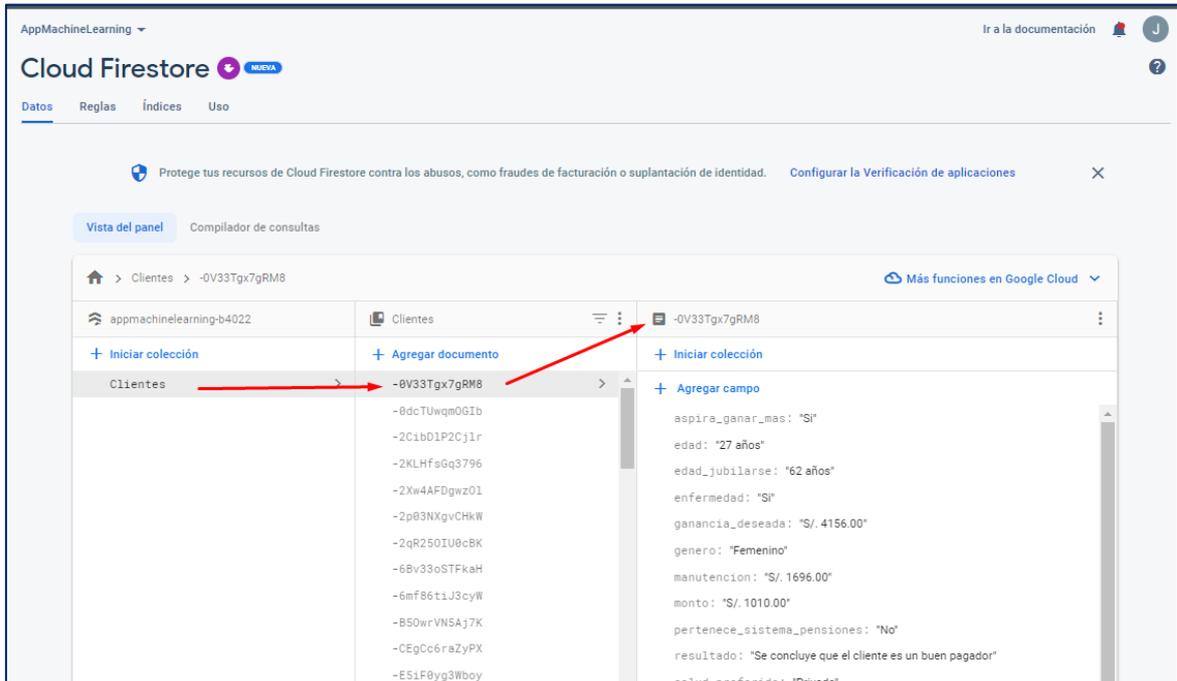
Se concluye que el cliente es buen pagador

OK

Fuente: Elaboración propia

Figura 98

Diseño de base de datos



Fuente: Elaboración propia

Tabla 31

Diccionario de datos

Campos	Tipo de dato
Aspira_ganar_mas	
Edad	
Edad_jubilarse	
Enfermedad	
Ganancia_deseada	
Genero	
manutencion	
Monto	String
Pertenece_sistema_pensiones	
Resultado	
Salud_preferido	
Satisfacción	
tiemporeal	
Tipo_pension	

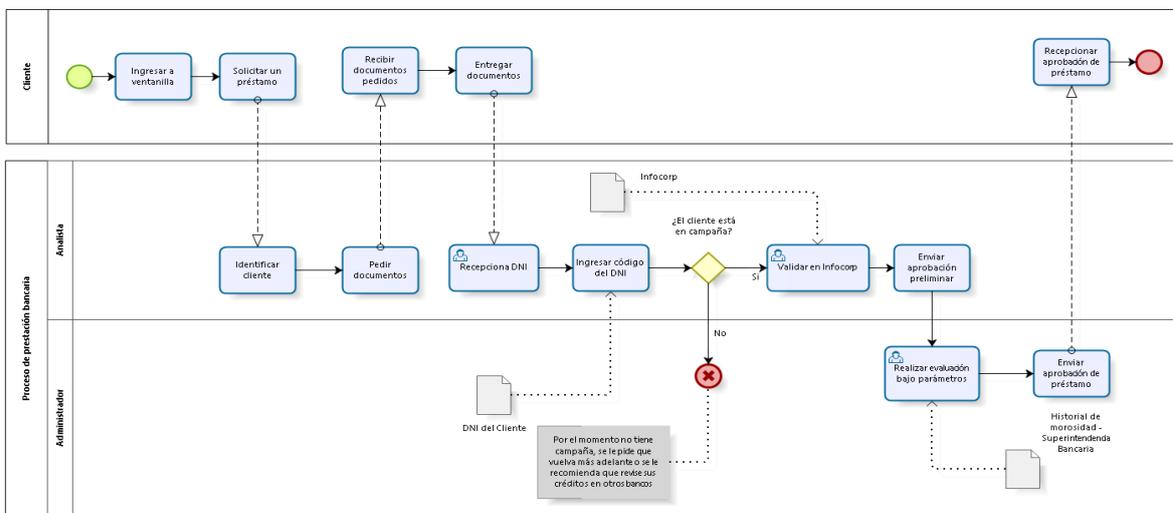
Fuente: Elaboración propia

Anexo 20: Proceso AS-IS y TO-BE

A continuación, se presenta el proceso de negocios, mostrando las actividades que posee la Figura 91, explicada desde el punto de vista que tienen muchos clientes bancarios al momento de su solicitud de un préstamo en el proceso de negocios llamado “Proceso de prestación bancaria de Banco Afil (AS-IS)” y con la mejora propuesta (TO-BE) mediante un sistema con Machine Learning en la Figura 92:

Figura 99

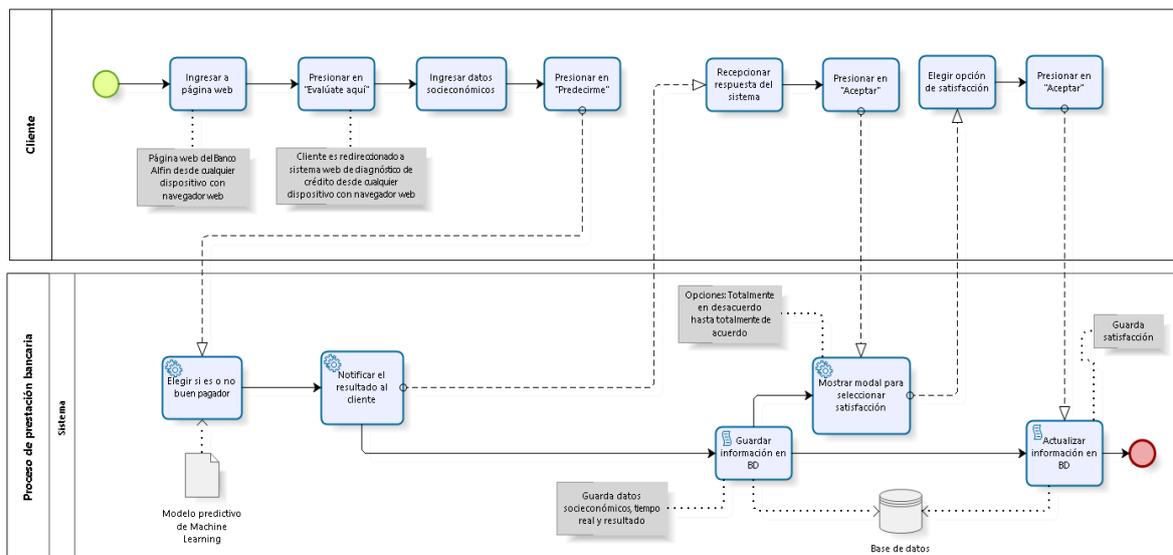
Proceso de prestación bancaria de Banco Afil AS-IS



Fuente: Elaboración propia

Figura 100

Proceso de prestación bancaria de Banco Afil TO-BE

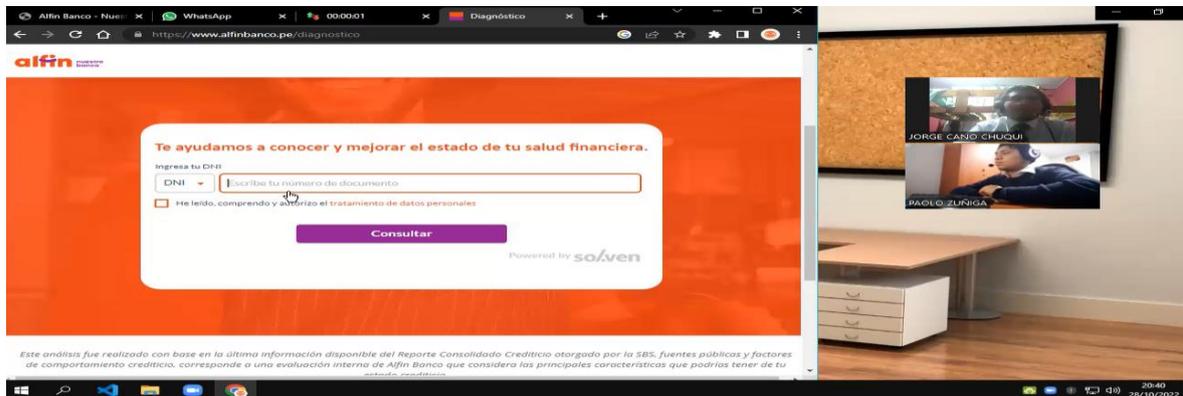


Fuente: Elaboración propia

Anexo 21: Capacitación antes de implementar retroalimentación

Figura 101

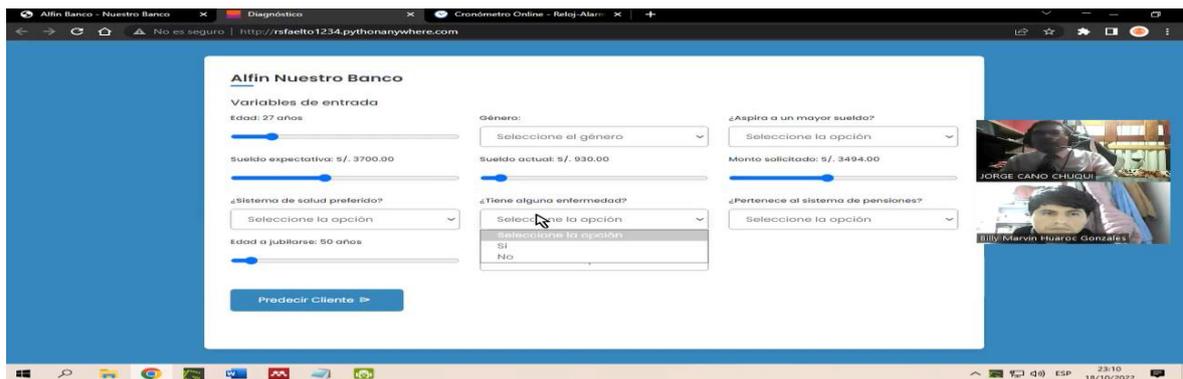
Evidencia de capacitación N°1



Fuente: Elaboración propia

Figura 102

Evidencia de capacitación N°2



Fuente: Elaboración propia

Figura 103

Evidencia de capacitación N°3



Fuente: Elaboración propia

https://drive.google.com/drive/folders/1uj-yXbE_29-rKAWcFbJgdXqXz2ItYMdr?usp=share_link

Anexo 22: Carta de aceptación



CARTA DE ACEPTACIÓN DEL BANCO ALFIN

Lima 28 de noviembre del 2022

Señores:

Universidad César Vallejo – UCV

De manera atenta manifestamos nuestro interés y conocimiento de la propuesta del proyecto de tesis titulada “Machine Learning para mejorar el proceso de prestación bancaria en una entidad financiera”. Elaborado por el estudiante de ingeniería **CANO CHUQUI JORGE** con DNI 77929589. En ese sentido, nos comprometemos a participar en estos procesos ofreciendo la información de apoyo necesario de la UCV como elemento de consulta para el público. Conocemos y aceptamos el reglamento y sus disposiciones sobre la realización de opciones de grado de la UCV.

Entregas aprobadas:

Sistema web con Machine Learning

Atentamente,

A handwritten signature in black ink is written over a circular stamp. The stamp contains the logo of Banco Azteca del Perú S.A. and the text 'Banco Azteca del Perú S.A.'.

Banco Azteca del Perú S.A.
Jesús Martín Ronceros Salhuana
Supervisor de Help Desk

Firma y Sello

Anexo 23: Artículo de investigación

El siguiente artículo fue publicado sobre el tema de la actual tesis.

WSEAS TRANSACTIONS on COMPUTER RESEARCH
DOI: 10.37394/232018.2022.10.9

Cano Chuqui Jorge, Ogosi Auqui José Antonio,
Guadalupe Mori Victor Hugo,
Obando Pacheco David Hugo

Machine Learning for Personal Credit Evaluation: A Systematic Review

CANO CHUQUI JORGE
Faculty of Engineering and Architecture
Universidad Privada César Vallejo
Av. Del Parque 640, San Juan de Lurigancho 15434
PERÚ

OGOSI AUQUI JOSÉ ANTONIO
Faculty of Engineering
Universidad Tecnológica del Perú
Av. Arequipa 265, Cercado de Lima 15046
PERÚ

GUADALUPE MORI, VICTOR HUGO
Faculty of Engineering
Universidad Privada San Juan Bautista
Ex Hacienda Villa, Av. José Antonio Lavalle s/n, Chorrillos 15067
PERÚ

OBANDO PACHECO, DAVID HUGO
Faculty of Engineering
Universidad Peruana de Ciencias Aplicadas
Prolongación Primavera 2390, Lima 15023
PERÚ

Abstract: - The importance of information in today's world as it is a key asset for business growth and innovation. The problem that arises is the lack of understanding of knowledge quality properties, which leads to the development of inefficient knowledge-intensive systems. But knowledge cannot be shared effectively without effective knowledge-intensive systems. Given this situation, the authors must analyze the benefits and believe that machine learning can benefit knowledge management and that machine learning algorithms can further improve knowledge-intensive systems. It also shows that machine learning is very helpful from a practical point of view. Machine learning not only improves knowledge-intensive systems but has powerful theoretical and practical implementations that can open up new areas of research. The objective set out is the comprehensive and systematic literature review of research published between 2018 and 2022, these studies were extracted from several critically important academic sources, with a total of 73 short articles selected. The findings also open up possible research areas for machine learning in knowledge management to generate a competitive advantage in financial institutions.

Key-Words: machine learning, credit scoring, risk assessment, algorithms, artificial intelligence.

Received: March 25, 2021. Revised: April 14, 2022. Accepted: May 12, 2022. Published: July 1, 2022.

1 Introduction

Time analysis is critical because financial institutions consistently implement the credit scoring model over time. Therefore, due to the complexity and flexibility of the training process, ML methods can be more sensitive to overtime

disturbances. It is susceptible to overfitting problems and may become unstable over time.

In addition, there is a lack of ability to account for time complexity in actual business operations. Everything depends on the model life cycle, from data collection, model development, and validation to the final model. This model can present new challenges and uncertainties to the generalization of

Anexo 24: Pantallazo software Turnitin

DESARROLLO DE PROYECTO DE INVESTIGACIÓN - TRILCE.docx

INFORME DE ORIGINALIDAD

13% INDICE DE SIMILITUD	12% FUENTES DE INTERNET	2% PUBLICACIONES	5% TRABAJOS DEL ESTUDIANTE
-----------------------------------	-----------------------------------	----------------------------	--------------------------------------

FUENTES PRIMARIAS

1	repositorio.ucv.edu.pe Fuente de Internet	3%
2	Submitted to Universidad Cesar Vallejo Trabajo del estudiante	2%
3	repositorio.autonoma.edu.pe Fuente de Internet	1%
4	cybertesis.unmsm.edu.pe Fuente de Internet	1%
5	Submitted to Universidad Técnica de Machala Trabajo del estudiante	1%
6	hdl.handle.net Fuente de Internet	1%
7	purl.org Fuente de Internet	<1%
8	repositorio.unap.edu.pe Fuente de Internet	<1%
9	support.minitab.com Fuente de Internet	<1%

Anexo 25: Actas del producto



UNIVERSIDAD CÉSAR VALLEJO

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

Autorización de Publicación en Repositorio Institucional

Yo, CANO CHUQUI JORGE identificado con N° de Documento N° 77929589 (respectivamente), estudiante de la FACULTAD DE INGENIERÍA Y ARQUITECTURA y de la escuela profesional de INGENIERÍA DE SISTEMAS de la UNIVERSIDAD CÉSAR VALLEJO SAC - LIMA ESTE, autorizo (X), no autorizo () la divulgación y comunicación pública de mi Tesis Completa: "MACHINE LEARNING PARA MEJORAR EL PROCESO DE PRESTACIÓN BANCARIA EN UNA ENTIDAD FINANCIERA".

En el Repositorio Institucional de la Universidad César Vallejo, según esta estipulado en el Decreto Legislativo 822, Ley sobre Derecho de Autor, Art. 23 y Art. 33.

Fundamentación en caso de NO autorización:

--

LIMA, 03 de Diciembre del 2022

Apellidos y Nombres del Autor	Firma
CANO CHUQUI JORGE DNI: 77929589 ORCID: 0000-0003-4809-6008	Firmado electrónicamente por: CCANOCHU el 03-12- 2022 18:42:05

Código documento Trilce: TRI - 0470029



UNIVERSIDAD CÉSAR VALLEJO

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

Declaratoria de Originalidad del Autor

Yo, CANO CHUQUI JORGE estudiante de la FACULTAD DE INGENIERÍA Y ARQUITECTURA de la escuela profesional de INGENIERÍA DE SISTEMAS de la UNIVERSIDAD CÉSAR VALLEJO SAC - LIMA ESTE, declaro bajo juramento que todos los datos e información que acompañan la Tesis titulada: "MACHINE LEARNING PARA MEJORAR EL PROCESO DE PRESTACIÓN BANCARIA EN UNA ENTIDAD FINANCIERA", es de mi autoría, por lo tanto, declaro que la Tesis:

1. No ha sido plagiada ni total, ni parcialmente.
2. He mencionado todas las fuentes empleadas, identificando correctamente toda cita textual o de paráfrasis proveniente de otras fuentes.
3. No ha sido publicada, ni presentada anteriormente para la obtención de otro grado académico o título profesional.
4. Los datos presentados en los resultados no han sido falseados, ni duplicados, ni copiados.

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

Nombres y Apellidos	Firma
JORGE CANO CHUQUI DNI: 77929589 ORCID: 0000-0003-4809-6008	Firmado electrónicamente por: CCANOCHU el 03-12- 2022 18:56:06

Código documento Trilce: TRI - 0470032





UNIVERSIDAD CÉSAR VALLEJO

FACULTAD DE INGENIERÍA Y ARQUITECTURA

ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS

ACTA DE SUSTENTACION DE TESIS COMPLETA

Siendo las 08:05 horas del 04/12/2022, el jurado evaluador se reunió para presenciar el acto de sustentación de Tesis Completa titulada: "MACHINE LEARNING PARA MEJORAR EL PROCESO DE PRESTACIÓN BANCARIA EN UNA ENTIDAD FINANCIERA", presentado por el autor CANO CHUQUI JORGE estudiante de la escuela profesional de INGENIERÍA DE SISTEMAS.

Concluido el acto de exposición y defensa de Tesis Completa, el jurado luego de la deliberación sobre la sustentación, dictaminó:

Autor	Dictamen
JORGE CANO CHUQUI	Unanimidad

Firmado electrónicamente por:
JCHUMPEA el 11 Dic 2022 10:06:33

JUAN BRUES LEE CHUMPE AGESTO
PRESIDENTE

Firmado electrónicamente por: DLAZOL el
09 Dic 2022 21:06:16

DAVID HUBER LAZO NEIRA
SECRETARIO

Firmado electrónicamente por:
EFLORESCH01 el 13 Dic 2022 21:06:20

ERICK GIOVANNY FLORES CHACÓN
VOCAL

Código documento Trilce: TRI - 0470028





UNIVERSIDAD CÉSAR VALLEJO

FACULTAD DE INGENIERÍA Y ARQUITECTURA
ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS
Dictamen para Sustentación

LIMA, 03 de Diciembre del 2022

El jurado encargado de evaluar la Tesis Completa presentado por el autor CANO CHUQUI JORGE de la escuela profesional de INGENIERÍA DE SISTEMAS, cuyo título es " MACHINE LEARNING PARA MEJORAR EL PROCESO DE PRESTACIÓN BANCARIA EN UNA ENTIDAD FINANCIERA", damos fe de que hemos revisado el documento antes mencionado, luego que el estudiante levantado todas las observaciones realizadas por el jurado, y por lo tanto está APTA para su defensa en la respectiva sustentación.

Firmado electrónicamente por: JCHUMPEA
el 05 Dic 2022 22:56:45

JUAN BRUES LEE CHUMPE AGESTO
PRESIDENTE

Firmado electrónicamente por: DLAZOL el
09 Dic 2022 21:04:16

DAVID HUBER LAZO NEIRA
SECRETARIO

Firmado electrónicamente por: EFLORESCH01 el
06 Dic 2022 20:29:44

ERICK GIOVANNY FLORES CHACÓN
VOCAL

Código documento Trilce: TRI - 0470030





UNIVERSIDAD CÉSAR VALLEJO

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

Declaratoria de Autenticidad del Asesor

Yo, FLORES CHACÓN ERICK GIOVANNY, docente de la FACULTAD DE INGENIERÍA Y ARQUITECTURA de la escuela profesional de INGENIERÍA DE SISTEMAS de la UNIVERSIDAD CÉSAR VALLEJO SAC - LIMA ESTE, asesor de Tesis titulada: "MACHINE LEARNING PARA MEJORAR EL PROCESO DE PRESTACIÓN BANCARIA EN UNA ENTIDAD FINANCIERA", cuyo autor es CANO CHUQUI JORGE, constato que la investigación tiene un índice de similitud de 13.00%, 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, 03 de Diciembre del 2022

Apellidos y Nombres del Asesor:	Firma
FLORES CHACÓN ERICK GIOVANNY DNI: 07964931 ORCID: 0000-0002-4028-8059	Firmado electrónicamente por: EFLORESCH01 el 04-12-2022 10:18:57

Código documento Trilce: TRI - 0470031





UNIVERSIDAD CÉSAR VALLEJO

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

Declaratoria de Autenticidad del Asesor

Yo, FLORES CHACÓN ERICK GIOVANNY, docente de la FACULTAD DE INGENIERÍA Y ARQUITECTURA de la escuela profesional de INGENIERÍA DE SISTEMAS de la UNIVERSIDAD CÉSAR VALLEJO SAC - LIMA ESTE, asesor de Tesis titulada: "MACHINE LEARNING PARA MEJORAR EL PROCESO DE PRESTACIÓN BANCARIA EN UNA ENTIDAD FINANCIERA", cuyo autor es CANO CHUQUI JORGE, constato que la investigación tiene un índice de similitud de 13.00%, 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, 03 de Diciembre del 2022

Apellidos y Nombres del Asesor:	Firma
FLORES CHACÓN ERICK GIOVANNY DNI: 07964931 ORCID: 0000-0002-4028-8059	Firmado electrónicamente por: EFLORESCH01 el 04-12-2022 10:18:57

Código documento Trilce: TRI - 0470031