



UNIVERSIDAD CÉSAR VALLEJO

FACULTAD DE INGENIERÍA Y ARQUITECTURA

ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS

Sistema de Pronóstico de la Calificación Crediticia de Clientes Basado en
Árbol de clasificación y Regresión lineal

TESIS PARA OBTENER EL TÍTULO PROFESIONAL DE:

Ingeniero de Sistemas

AUTORES:

Juan Erick López De la Cruz (0000-0001-7477-7435)

Juan José Pérez Alarcón (0000-0002-5642-5628)

ASESOR:

Dr. Francisco Manuel Hilario Falcón (0000-0003-3153-9343)

LÍNEA DE INVESTIGACIÓN:

Sistema de Información y Comunicaciones

LIMA – PERÚ

2019

Dedicatoria

Dedicados a nuestros padres que nos dieron el apoyo para llevar adelante esta investigación, a nuestros hermanos por su aliento. Asimismo, a nuestros docentes que confiaron en nosotros.

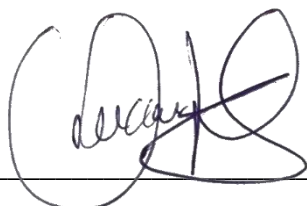
Agradecimientos

Agradecemos a dios por brindarnos salud, a nuestros padres, hermanos y amigos por apoyarnos en este último tramo de nuestras carreras. A nuestros Asesores por brindarnos su paciencia y dedicación.

Declaratoria de autenticidad

Nosotros Juan Erick López De la cruz con DNI N° 45989478, y Juan José Pérez Alarcón con DNI N° 45164035, en efecto de cumplir con las disposiciones vigentes consideradas en el Reglamento de Grados y Títulos de la Universidad César Vallejo, Facultad de Ingeniería, Escuela Profesional de Ingeniería de sistemas, declaramos bajo juramento que toda la documentación que acompaño es veraz y auténtica. Asimismo, declaramos también bajo juramento que toda la data e información que presentamos en la presente tesis es auténtica y veraz. En tal sentido, asumimos la responsabilidad que corresponda ante cualquier falsedad, ocultamiento u omisión tanto de los documentos como de información aportada por lo cual nos sometemos a lo dispuesto en las normas académicas de la Universidad César Vallejo.

Lima, 20 de octubre de 2019



Juan Erick López De la Cruz

DNI: 45989478

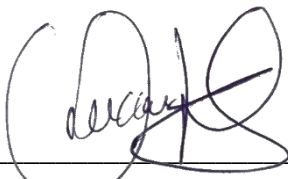


Juan José Pérez Alarcón

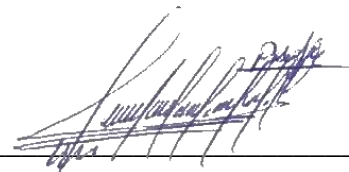
DNI: 45164035

Presentación

Señores miembros del jurado, cumplimiento el Reglamento de Grados y Títulos de la Universidad César Vallejo les presentamos ante ustedes la tesis titulada “Sistema de Pronóstico de la Calificación Crediticia de Clientes Basado en Árbol de clasificación y Regresión lineal”, que tiene como objetivo determinar el impacto del sistema de pronóstico de la calificación crediticia de clientes basado en árbol de clasificación y regresión lineal, esperando que cumpla con los requisitos de aprobación para obtener el título profesional de Ingeniero de Sistemas. Dicha investigación se distribuye en seis capítulos. El capítulo primero explicamos la realidad problemática que sirvió como base de nuestra investigación, de donde se desprenden la revisión de trabajos previos, teorías relacionadas, la formulación y planteamiento del problema, justificación del estudio, los objetivos y por último se plantea la hipótesis; en el segundo capítulo se muestra el diseño de la investigación, definición de las variables y Operacionalización seguidamente de define la población, muestra y muestreo, para después se mencionan las técnicas e instrumentos de recolección de datos, validez y confiabilidad, el método de análisis de datos y por último los aspectos éticos, el capítulo tercero se describen el resultado obtenido en la investigación. El capítulo cuarto presentamos discusión. El capítulo quinto describimos la conclusión de la investigación. El capítulo sexto realizamos la recomendación para futuras investigaciones.



Juan Erick López De la Cruz
DNI: 45989478



Juan José Pérez Alarcón
DNI: 45164035

Resumen

El presente proyecto muestra al desarrollar la comprensión del negocio, preparación, modelamiento y evaluación de los datos para un sistema de pronóstico de la calificación crediticia de clientes en una empresa financiera, en donde se viene utilizando una metodología Credit Scoring el cual se busca mejorar ya que no se cuenta con una precisión adecuada y existe un porcentaje de error y un tiempo de procesamiento elevados.

El objetivo general de la presente investigación es determinar el impacto del sistema de pronóstico de la calificación crediticia de clientes se basa al árbol de clasificación y regresión lineal. Al analizar la bibliografía correspondiente a las variables en estudio, se determina que el pronóstico de la calificación crediticia de clientes basado en árbol de clasificación y regresión lineal incrementa la precisión del pronóstico, reduce el error y reduce el tiempo de respuesta, para lograr este objetivo se presenta una propuesta de sistema híbrido, tomando como referencia los principales modelos y parámetros asociados a la calificación crediticia.

Palabras clave: Sistema, modelo de pronóstico, calificación crediticia, árbol de clasificación, regresión lineal, CRISP-DM

Abstract

This project shows when developing the understanding of the business, preparation, modeling and evaluation of the data for a forecasting system of the credit rating of clients in a financial company, where a Credit Scoring methodology is being used which seeks to improve already that adequate precision is not available and there is a high error rate and processing time

The general objective of this research is to determine the impact of the forecasting system on customer credit rating based on the classification tree and linear regression. When analyzing the bibliography corresponding to the variables under study, it is determined that the forecast of the credit rating of clients based on classification tree and linear regression increases the precision of the forecast, reduces the error and reduces the response time, to achieve this objective A proposal for a hybrid system is presented, taking as a reference the main models and parameters associated with the credit rating.

The general objective of the present investigation is to determine the impact of the forecasting system of the credit rating of clients based on classification tree and linear regression. When analyzing the bibliography corresponding to the variables under study, it is determined that the forecast of the credit rating of clients based on classification tree and linear regression increases the accuracy of the forecast, reduces the error and reduces the response time, to achieve this objective A proposal for a hybrid system is presented, taking as a reference the main models and parameters associated with the credit rating.

Keywords: System, forecast model, credit rating, classification tree, linear regression, CRISP-DM

Índice general

I. INTRODUCCIÓN	13
I.1 Realidad problemática	14
I.2 Trabajos previos	19
I.3 Teorías relacionadas	24
I.3.1 Minería de Datos (<i>Data Mining</i>)	24
I.3.2 Metodologías	25
I.3.2.1 Metodología KDD	25
I.3.2.2 Metodología SEMMA	26
I.3.2.3 Metodología CRISP – DM	27
I.3.3 Cuadro comparativo entre las metodologías presentadas	28
I.3.4 Modelo Sistémico	29
I.3.5 Modelo Lineal	29
I.3.6 Modelo No Lineal	29
I.3.7 Credit Scoring	29
I.3.8 Pronostico	30
I.3.9 Dimensiones	30
I.3.9.1 Precisión	30
I.3.9.2 Error y fiabilidad	31
I.3.9.3 Tiempos de respuesta	31
I.3.9.3.1 Pruebas de desempeño	31
I.3.9.3.2 Pruebas de carga	33
I.4 Formulación del problema	34
I.5 Justificación del estudio	34
I.6 Objetivos	36
I.7 Hipótesis	37
II. MÉTODO	40
II.1 Diseño de la investigación	41
II.1.1 <i>Enfoque de la investigación</i>	41
II.1.2 <i>Tipo de estudio</i>	41
II.1.3 <i>Diseño de investigación</i>	41
II.2 Variable, Operacionalización	42
II.2.1 <i>Variable</i>	42

II.2.2	Operacionalización de la variable	42
II.2.3	Matriz de Operacionalización de las variables ^o	44
II.3	Población y muestra	45
II.4	Técnicas e instrumentos de recolección de datos, validez y confiabilidad	46
II.5	Métodos de análisis de datos	47
II.6	Aspectos éticos	48
III.	RESULTADOS	49
3.1	Pruebas de normalidad	50
3.2	Dimensión 1: Precisión del pronóstico	51
3.2.1	Indicador 1: Incremento de la precisión del pronóstico respecto al modelo: Credit scoring	51
3.3	Dimensión 2: Error del pronóstico del Sistema	58
3.3.1	Indicador 2: Reducción de error del pronóstico del sistema respecto al modelo: Credit scoring	58
3.4	Dimensión 3: Tiempo de Respuesta del Sistema	65
3.4.1	Indicador 3: Reducción del tiempo de respuesta del sistema respecto al modelo: Credit scoring	65
IV.	DISCUSIÓN	71
V.	CONCLUSIONES	74
VI.	RECOMENDACIONES	76
VII.	REFERENCIAS	78
	ANEXOS	84

Índice de tablas

Tabla 1. Cuadro comparativo entre las metodologías presentadas	28
Tabla 2. Matriz de Operacionalización de las variables de la investigación	44
Tabla 3. Recolección de datos	47
Tabla 4: Cálculos Estadísticos Descriptivos- Precisión del Pronóstico	52
Tabla 5: Tabla de Frecuencia Pre Test – Precisión del Pronostico	52
Tabla 6: Tabla de Frecuencia Post Test – Precisión del Pronóstico.....	53
Tabla 7: Prueba de Kolmorov – Precisión del Pronóstico	56
Tabla 8: Aplicación de la muestra relacionadas T student – Precisión del Pronóstico.....	57
Tabla 9: Calculo de datos descriptivos - Error de Pronostico del Sistema	59
Tabla 10: Tabla de Frecuencia Pre Test - Error del Pronóstico del Sistema	59
Tabla 11 : Tabla de Frecuencia Post Test - Error del Pronostico del Sistema.....	60
Tabla 12: Prueba de Kolmorov- Error del Pronostico del Sistema	63
Tabla 13: Aplicación de la muestra relacionadas T Student – Error del Pronostico del Sistema.....	64
Tabla 14: Calculo Estadístico descriptivo – Tiempos de Respuesta del Sistema	65
Tabla 15: Tabla de frecuencia Pre Test- Tiempo.....	66
Tabla 16: Tabla de frecuencias Post -Test - Tiempo	66
Tabla 17: Prueba de Kolmogorov - Tiempo de Respuestadel Sistema.....	69
Tabla 18: Prueba de Hipótesis – Tiempo	70
Tabla 19: Estadística de Prueba Wilcoxon	70
Tabla 20. Matriz de consistencia	85
Tabla 21 Estadísticos descriptivos	106
Tabla 22 Correlaciones	107
Tabla 23 Variables de entrada/eliminadas	112
Tabla 24 Resumen del modelo	113
Tabla 25 Tabla ANOVA	114
Tabla 26 Matriz de coeficientes	115

Índice de figuras

Figura 1. Prototipo de interfaz de ingreso al sistema de pronostico	86
Figura 2. Prototipo de interfaz para seleccionar el ingreso de datos del cliente y datos de la solicitud de crédito	86
Figura 3. Prototipo de Listado de clientes.....	87
Figura 4. Prototipo para registrar o actualizar los datos generales de un cliente	88
Figura 5. Prototipo para registrar o actualizar los datos de SUNAT de un cliente	88
Figura 6. Prototipo para registrar o actualizar los datos de dirección y UBIGEO de un cliente.....	89
Figura 7. Prototipo para registrar o actualizar los datos de los socios de un cliente....	89
Figura 8. Prototipo para registrar o actualizar los datos del representante legal de un cliente.....	90
Figura 9. Prototipo para verificar o levantar deudas de un cliente	90
Figura 10. Prototipo para validación final de los datos de un cliente	91
Figura 11. Prototipo para registrar o actualizar los estados financieros de un cliente .	91
Figura 12. Prototipo para registrar o actualizar los datos de experiencia de un cliente	92
Figura 13. Prototipo para registrar los datos de una solicitud de crédito.....	92
Figura 14. Prototipo para procesar y generar el pronóstico de calificación crediticia de un cliente.....	93
Figura 15. Diagrama de estrella	94
Figura 16 Recolección de datos 1.....	104
Figura 17 Recolección de datos 2.....	104
Figura 18 Modelo de Pronostico en Microsoft Azure.....	116
Figura 19 Autorización para realizar la investigación	117

Índice de anexos

Anexo 1 Matriz de consistencia	85
Anexo 2 Prototipos	86
Anexo 3 Diagrama estrella	94

I. INTRODUCCIÓN

Este primer capítulo inicia definiendo la realidad problemática para posteriormente definir el planteamiento del problema. Seguidamente de la realidad problemática se procede a la selección y revisión de trabajos de investigación previos que nos servirán de referencia, así como la definición de términos y teorías que utilizaremos como sustento.

Planteamiento del problema según indicaron Hernández, Collado y Baptista (2014) es el corazón, el centro de la investigación: define o dicta los pasos. Ellos explicaron que las principales fuentes más consultadas y utilizadas para desarrollar el marco teórico son: artículos, libros, revistas científicas entre otras, porque sistematizan, profundizan y son especializadas en información además de su fácil acceso a ellas por internet. Señalaron el Por qué investigar indicando sus razones A través de la justificación vamos hacer la demostración que el estudio es importante y necesario. (p. 34).

Luego, la formulación de una hipótesis, que, de acuerdo con las declaraciones de Hernández, Collado y Baptista (2014), se define como una explicación de prueba del fenómeno en estudio, que se conoce como proposiciones o afirmaciones. Finalmente, los objetivos se describen para indicar que esto es una inspiración para la investigación y, como tal, debe establecerse claramente, ya que son guías de estudio (p. 102).

I.1 Realidad problemática

Los especialistas del Banco Central de Reserva del Perú (2019) indicaron que en la última década todos los países del mundo han considerado de vital importancia tener conocimiento del rating, o calificación de la deuda que posee su nación, ya que en general según este indicador permite medir la capacidad del país, empresas, gobierno de esta manera hace frente a su responsabilidad financiera o deuda, lo cual, lo riesgoso que es invertir en ella. Cuanto más grande sea el riesgo, tendrá una calificación negativa, la cual será por el riesgo del impago de esta. Entonces, entendiendo la calificación de la deuda, como un indicador ya sea para una nación o para potenciales inversores acerca del riesgo que asumirán cuando inviertan en los productos financieros en algún momento determinado resulta muy importante (p. 15).

Otro factor importante mencionado por especialistas del BCR del Perú (2019) viene a ser dinámica de las relaciones comerciales internacionales y la participación de los agentes financieros dentro de la Comunidad Internacional han acogido la aplicación de medidas en virtud de salvaguardar sus intereses. Se demuestran que en la actualidad existe una alta incidencia de riesgos en el otorgamiento de créditos a empresas o personas naturales con negocio, tanto es así que el riesgo de control de crédito radica básicamente en la falta de medidas preventivas y herramientas que permitan una mejor calificación al momento de otorgar un crédito, para estos créditos se consideran relevantes variables como políticas de control de crédito, capacidad de pago, nivel de garantía y sobre todo una mejor clasificación de clientes(p. 15).

En una investigación realizada en el sistema bancario europeo Wahrenburg y Frerichs (2003) señalaron que en el marco de la vigencia de las primeras recomendaciones del “Primer Convenio de Basilea”, los agentes reguladores y las entidades bancarias en función de los riesgos que afrontaba este sector, tenían la potestad de utilizar herramientas cuantitativas con el objeto de validar la utilidad e idoneidad de diversos sistemas de calificación crediticia, en algunos casos, basados en una metodología de “calificación interna”. En base a los datos del Deutsche Bundesbank y la aplicación de herramientas de simulación dirigida, pudieron identificar el uso masivo de sistemas de calificación crediticia muy débiles y desactualizados en el contexto financiero actual. Estos sistemas estaban fundamentados mayoritariamente, en base a estadísticas que median la calidad crediticia de los agentes prestatarios, en una escala dicotómica (bueno o malo) subjetiva o en cálculos probabilísticos sesgados, sobre la calidad crediticia de datos de previsión financiera (p. 45).

En una iniciativa por encontrar nuevas herramientas, Karminsky y Khromova (2016) construyeron un modelo confiable basado en información pública para el uso práctico de los agentes interesados (agentes reguladores y los propios bancos). Su investigación evidenció la construcción de un rango de calificaciones basado en una tabla de variables representativas. Asimismo, los datos utilizados en la investigación corresponden a la base de datos Bankscope que contiene información financiera sobre el

sistema bancario internacional, correspondiente al período de 1996- 2011. Como parte de sus resultados, al desarrollar un modelo de prueba, se obtuvieron las siguientes conclusiones: 1) las variables macro mejoran el poder explicativo del modelo y 2) el efecto de afiliación regional es significativo. Además, los grados de calificación se ajustan al ciclo comercial. En este propósito, determinaron que los Standard Poor's & Moody's son las agencias con calificaciones más conservadora y menos conservadora, respectivamente. El modelo planteado por los investigadores se clasificó como prácticamente útil, ya que la predicción de los grados de calificación señaló un 31% de resultados precisos y hasta un 70% de pronósticos con un error dentro de una calificación, mientras que la predicción de las clases de calificación resultó en 62% y 95% respectivamente (p 69).

Según indica Karminsky y Khromova (2016) en Latinoamérica, los sistemas de evaluación crediticia que utilizan los bancos son lo suficientemente particulares y apropiados para el logro de los propósitos empresariales. Por ejemplo, en Chile los sistemas de evaluación permiten reducir los riesgos, de una manera rápida y sencilla siendo los más apropiados en el mercado financiero, en un contexto de una economía de mercado tan variable. En el propósito de confirmar esto, Leal, Arangüiz y Gallegos (2018) demostraron que los modelos de credit scoring se sostienen de variables (cualitativas y cuantitativas) con crítica en la consideración al proceso concesión de crédito. Se demostró que la calidad del modelo en base a su análisis crediticia propuesta resulta en un 81,83% de los créditos brindados al cliente han sobrepasado el rango inferior del análisis o el máximo de aprobar por la entidad (p. 123).

En el caso Leal, Arangüiz y Gallegos (2018) demostraron que casi en todos los países latinoamericanos y en las demás regiones a desarrollar en los años noventa y setenta, con auge crediticio provocado principalmente por la apertura financiera finalizó con colapsar el sistema bancario y el colapso económico de muchos años. A principios de la década de 1990 después de la liberalización financiera el cual lo financiaron primordialmente en préstamos extranjeros y depósitos reembolsados. A inicios, el incremento del crédito se respaldó en la mayor expansión en el sector real; pero en el año 1996, comenzaron a surgir los primeros signos de la expansión crediticia carecía de soporte. 2 años después, los shocks externos de Rusia y Asia se convirtieron en los

impulsores del sistema ya vulnerablemente: ocasionan que el riesgo de liquidez se convirtiera rápidamente en riesgo solventica imparabile, intenso y prolongada.

La investigación de Hernández (2014) busca investigar lo que sucedió en el sistema financiero de Perú en las últimas décadas, explorando el factor que determinaran deriva crediticia en ese tiempo, en particular, intenta mostrar a los bancos han tenido diversas reacciones a los primeros signos de crisis, dependiendo de la magnitud de estos. Como principal objetivo del documento es demostrar, como diversas crisis financieras, las que ocurrieron en Perú en esta década tienen un fuerte componente especulativo, sobre- miopía teórica generalizada y optimismo.

Según Guillén (2002), el riesgo financiero en Perú en la década última representa a las características de quienes preceden al auge crediticio. Tras el rápido incremento del crédito a principios de los años 1990 después de la liberalización financiera, en 1996 se comenzó los primeros signos de la expansión que tiene una naturaleza especulativa, y a su vez deja un vulnerable sistema. Así el shock de 1998, manifestado por el problema de la falta de liquidez, convirtiéndose rápidamente en una crisis de insolvencia. Sin lugar a duda, uno de los primeros signos de vulnerabilidad fue el aumento de la delincuencia, como ha sucedido en otras crisis bancarias internacionales (p. 38).

Guillén (2002) también indico que la acumulación de bancos se vio afectada no solo en factor macroeconómico. En Perú, se ha observado, la sensibilidad de la delincuencia a estos factores depende de la medición de la organización. Es así que los bancos mayores tienen dependencia determinante externos que internos, probablemente esto se deba a un mayor uso de las economías de escala en información y recursos para evaluar y gestionar el riesgo crediticio. El banco pequeño y mediano tiene mucha dependencia de los factores internos y externos. Entre estos últimos se encuentran los relacionados con las políticas de aprobación de crédito y la eficacia de producción en el caso del banco mediano y el comportamiento de riesgo en el caso del banco pequeño (p.20).

Por lo tanto, según Guillén (2002) el funcionamiento adecuado de los sistemas de pronóstico de calificación crediticia es muy importante para las empresas. Hoy, Perú busca reducir el control interno de riesgos al otorgar préstamos o financiamiento, por lo que es necesario contar con un sistema favorable con métodos que ayuden a mejorar el pronóstico al cometer menos errores para lograr la economía y las finanzas de la organización de manera adecuada con un mejor control y análisis en la gestión de ingresos y gastos, lo que resulta en una buena rentabilidad (p.21).

Los especialistas de FOGAPI (FOGAPI,2019) indican que la Financiera es una entidad financiera regular la cual está en supervisión por la Superintendencia de Banca, AFP y Seguros, especializado en brindar servicio de garantía para micro y pequeñas empresas (MIPE), que proporciona los accesos al crédito a personas físicas o jurídicas a través de las actividades económicas. En este contexto, esta investigación buscará ajustar y mejorar su pronóstico de selección de clientes mediante el desarrollo de un sistema de pronóstico de crédito que tenga un impacto positivo en FOGAPI financieramente. El punto de desarrollo clave es tener métricas de gestión clave, herramientas de medios informáticos, programas eficientes y amigables que supervisen y ayuden a controlar y aplicar modelos de regresión lineal que tomaremos como base, así como extraer datos en análisis exploratorios y predictivos.

Los especialistas de FOGAPI (FOGAPI,2019) indicaron que actualmente en la financiera FOGAPI no se cuenta con una herramienta sistémica que permita realizar la evaluación y predicción de récord crediticio, ya que se vienen utilizando metodologías y técnicas empíricas basadas en otras financieras y por ello no se logra tener una mejor precisión ya que las realidades muchas veces son totalmente distintas, por otro lado el nivel de error al realizar las predicciones en su mayoría son relevantes y esto afecta a la liquidez y la diversificación de fondos de capital mal distribuidos.

De igual forma, según los especialistas de FOGAPI (FOGAPI, 2019) esto impacta en el nivel de tiempos de respuesta para una solicitud de aprobación o evaluación para nuestros

clientes incurriendo muchas veces en menor captación de clientes potenciales. Por último, la metodología que se utiliza para la evaluación es poco amigable ya que solo aquellos que tienen experiencia por varios años con el método son los que realizan dicha evaluación incurriendo en dependencia de personal que solo sabe utilizar dicha herramienta. En este contexto planteamos mejorar todos los puntos anteriormente mencionados para optimizar los procesos de selección y evaluación récord crediticio en la financiera FOGAPI.

I.2 Trabajos previos

En una publicación de Baesens, Van Gestel, Viaene, Stepanova, Suykens, y Vanthienen (2003) que tiene por título *Benchmarking state-of-the-art classification algorithms for credit scoring. Journal of the Operational Research Society* se muestra el rendimiento de varios algoritmos de clasificación de vanguardia aplicados a ocho conjuntos de datos de puntuación de crédito en la vida real. Algunos de los conjuntos de datos se originan en las principales instituciones financieras del Benelux y el Reino Unido. Se evalúan y comparan diferentes tipos de clasificadores. Además de los algoritmos de clasificación conocidos (p. Ej., Regresión logística, análisis discriminante, vecino más cercano a k, redes neuronales y árboles de decisión), este estudio también investiga la idoneidad y el rendimiento de algunos algoritmos de clasificación avanzados basados en el núcleo propuestos recientemente, como el vector de soporte Máquinas de vectores y máquinas de vectores de mínimos cuadrados (LS-SVM). El rendimiento se evalúa utilizando la precisión de clasificación y el área debajo de la línea de características operativas por el receptor. Las diferencias de rendimiento estadísticamente significativas se identifican utilizando las estadísticas de prueba apropiadas. Se encontró que tanto el LS-SVM como los clasificadores de redes neuronales producen un muy buen rendimiento, pero también los clasificadores simples como la regresión logística y el análisis discriminante lineal funcionan muy bien en la credit scoring (p. 66).

En un intento por explicar los credit scoring Reza (2018) en un artículo que tiene por título *Credit Scoring Using CART Algorithm and Binary Particle Swarm*

Optimization, International Journal of Electrical and Computer Engineering lo define como un procedimiento que existe en toda institución financiera. Una forma de predecir si el deudor estaba calificado para recibir el préstamo o no y ha sido una preocupación importante en los pasos generales del proceso de préstamo. Casi todos los bancos y diversas instituciones financieras tienen sus propios medios de calificación crediticia. Hoy en día, el enfoque de minería de datos ha sido aceptado como uno de los métodos más conocidos. Ciertamente, la precisión también fue un problema importante en este enfoque. Esta investigación propuso un método híbrido utilizando el algoritmo CART y la optimización de enjambre de partículas binarias. Los indicadores de rendimiento que se utilizan en esta investigación son la precisión de la clasificación, la tasa de error, la sensibilidad, la especificidad y la precisión. Los resultados experimentales basados en el conjunto de datos públicos mostraron que la precisión del método propuesto es del 78%. En comparación con varios algoritmos populares, como la red neuronal, la regresión logística y la máquina de vectores de soporte, el método propuesto mostró un rendimiento sobresaliente (p. 41).

Rogers y Securato (2015) indicaron en su investigación que tiene por título *ABOUT PSYCHOLOGICAL VARIABLES IN APPLICATION SCORING MODELS* tuvo como propósito el estudio de la variable psicológica y la escala sugerida por la Psicología Económica para predecir el incumplimiento de los individuos. Teniendo la muestra de 555 personas completó un formulario de auto-completado, que estaba realizado por la variable psicológica y de escala. Al recoger la metodología de la regresión logística, se encontraron las siguientes características psicológicas y de comportamiento relacionadas al conjunto de personas en incumplimiento: a) dimensión negativa en relación al dinero; b) puntuaciones superiores a la escala de autoeficacia, lo que nos indica un superior grado de optimismo y confianza; c) comprador clasificado como compulsivo; d) personas que consideran una necesidad de dar obsequios a niños y amistades en fechas festivas, hay diversas personas lo consideran un lujo; e) problema de auto controlarse identificado en individuos que beben un promedio de más de un vaso de bebidas alcohólicas diaria (p. 19).

Según manifiestan Vargas y Mostajo (2014), normalmente el agente económico acogen el perfil que sea adverso de los riesgos, por eso al orientarlo el accionar al mínimo, transferencia y/o mitigar el riesgo conducían al banco a denegar a la operación que no ofrezcan garantías plenas, actualmente la gestión de riesgo de crédito tiene como objetivo la gestión del riesgo de crédito y tener la rentabilidad de acuerdo al rango de asumir la pérdida esperada, compromete una parte del propio capital para cumplimiento de la normativa. Entendemos que la operación crediticia que tenga la más alta probabilidad de no pagar no específicamente se calificara como un negocio malo, se tiene que tener la mayor rentabilidad para compensar el riesgo de crédito que se está asumiendo.

Para tener una mejor gestión de cartera de créditos y de esta forma se realiza la adecuada colocación de recursos, las entidades financieras tienen la necesidad en medición de los riesgos de crédito, usando diversas metodologías y enfoques de este ámbito. En el proyecto se desarrolló la aplicación de Métodos Basándonos en la Calificación Interna iniciando en 2 modelos financieros los cuales se usaron para la calculación de la medida del riesgo crediticio, como resultado obtenido fue comparar con los requerimientos establecido en las normativas bolivianas para Entidades de Intermediación Financiera (p. 75).

En su trabajo “FACTORES DETERMINANTES PARA LA CONCESIÓN DE CRÉDITO DE PARTE DE LAS ENTIDADES FINANCIERAS A LAS MIPYMES” Rocca, García y Gómez (2018) analizaron el criterio los cuales consideraran como primordiales entidades financieras en el momento de dar un crédito MIPYMES, como objetos minimiza la probabilidad de impago. Como resultado se muestra que el criterio a considerar por el analista de riesgo que es importante para poder brindar el crédito son: integridad y honestidad del empresario, conocimiento en el sector del empresario, calificación del personal de la entidad y cartera de clientes de la empresa. El criterio relacionado con la información contable, el analista de riesgo verifica minuciosamente los ratios de liquidez y endeudamiento, que las entidades deben tener un sistema de costo fiable y no aparecer en registro de impago (INFOCORP). De este modo se tuvo como resultado una herramienta útil para las entidades financieras que ayudo en la mejora del sistema de calificación y a su vez beneficiaron a las MIPYMES con una mayor relación

en las entidades financieras (p. 85).

De acuerdo con Leal, Arangüiz y Gallegos (2018) en su trabajo de investigación denominado “ANÁLISIS DE RIESGO CREDITICIO, PROPUESTA DEL MODELO *CREDIT SCORING*” aplicado a una empresa que experimento en los últimos años un incremento en las ventas y con ello, una reducción de rango líquido y calidad en las cuentas por cobrar. En cambio, tal crecimiento en la cuenta por cobrar está asociado a grandes riesgos asumidos por los cobros, con esta política liberal de cuentas por cobrar. Pero, no dispuso del sistema de gestión de crédito con objetivos que permitirá la adecuada calificación de calidad y capacidad crediticia en los clientes. Así mismo, en el artículo se propuso a la empresa una muestra de evaluación crediticia ajustado y ponderado, así como a sus clientes, que permitió reducir el riesgo de crédito. El trabajo en consideró, una minuciosa descripción de las muestras de evaluar el crédito y en especial las muestras de *credit scoring*. Mediante entrevista a experto, se deducen variables cualitativas y cuantitativas críticas para consideración en el proceso de gestión de créditos. Con calidad de evaluación crediticia por la modelo propuesta, el resultado que se obtuvieron fueron 81,82% del crédito otorgado a los clientes superaron el rango mínimo al evaluar o límite de aprobación por la entidad (p. 34).

En el trabajo que realizaron Ochoa, Galeano y Agudelo (2010) para perfiles de clientes óptimos para conceder el crédito para COOFINEP Cooperativa Financiera Construir un modelo de calificación crediticia para las entidades financieras, se refiere a la Autoridad de Supervisión Financiera de Colombia de acuerdo con la Circular 31 y la Circular 11 de 2002, la institución financiera son sistemas de gestión de riesgo de crédito (SARC);

Tiene en cuenta la volatilidad de las variables financieras, todo esto debido a la globalización del mercado financiero mundiales y con mayor importancia en el sistema financiero sólido. Se implementa un estudio de la técnica de análisis discriminando y formar modelos de calificación crediticia. En el análisis estadístico de las variables

cuantitativas y cualitativas en la base de la data proporcionada por las cooperativas financieras del Valle de Aburrá, pretendiendo la definición del perfil de los prestatarios propensos en los incumplimientos de la obligación y el perfil del buen prestatario. (p. 72).

En el artículo con nombre “DESARROLLAR LA PROPIA METODOLOGÍA PARA ANÁLISIS DE CRÉDITO EMPRESARIAL EN UNA ENTIDAD FINANCIERA” los economistas Hernández y Benavides (2005) indicaron el factor relevante al momento de realizar la metodología para analizar el crédito empresarial. Primero, realizamos los estudios del estado financiero del cliente para el crédito. El estado financiero lo diagnosticamos y así realizar la evaluación de calidad del cliente. Posteriormente se llega a exponer 2 metodologías y evaluamos los riesgos: modelo de scoring y modelo interno de ponderación predeterminada (p. 22).

En la publicación que tiene como título CMIN - herramienta case que se basa en *CRISP-DM* en soporte para proyectos de minería de datos los autores Cobos, Guarín, León y Mendoza (2010) presentan la CMIN, una herramienta *CASE* basada en *CRISP-DM* 1.0 de esta forma soporta el desarrollo del proyecto de minería de datos. En primer lugar, exponemos las funcionalidades generales del CMIN, incluimos las gestiones de procesos, proyectos y plantillas, destacamos las capacidades del CMIN y realizamos los seguimientos del proyecto de una manera intuitiva, sencilla, CMIN hace posible al usuario incrementa los conocimientos al usar *CRISP-DM* o cualquier proceso que esté definido en la herramienta mediante la información y ayuda ofreciendo en cada paso a paso en el proceso. Posteriormente, detallamos cómo CMIN nos permita el enlace en tiempo de ejecución, nuevos algoritmos en minería de datos y apoyar en el trabajo de modelado en un proyecto de minería de datos. Para finalizar, ofrecieron las respuestas de 2 evaluaciones de la herramienta, trabajo futuro y conclusiones (p. 46).

Según indicaron Salcedo, Galena y Rodríguez (2010) en la publicación que tiene como título Metodología *CRISP* para la implementación *Data Warehouse*. Actualmente la generación de informe conciso, claro y sobre todo veraz, en base a la información de

las empresas, son elementos fundamentales en la toma de decisiones. Esto se debe a que necesita inminente surge *Data Warehouse* que es un recurso principal para realizar los procesos, cimentado primero por debajo de filosofía OLAP que utilizara el concepto de EIS y DSS y realizar el informe. Entre los procesos que se realiza en la construcción de las bodegas de datos destaca primordialmente transformación, extracción y manipular la información y posteriormente la definición de los metadatos, el cual es utilizado en la definición del *Data Warehouse* que está integrada.

La tendencia hacia lo que apunta la "inteligencia de negocios" es divulgar la información, para todos aquellos que necesiten en diferente dimensión y nivel asociado, de esta maneta tener los informes consolidados, detallada y facilitar la síntesis de determinado proceso empresarial y repercutirá al momento de hacer la toma de decisiones, viene a ser el mismo objetivo de los *Data Warehouses*.

Para implementar este proceso se necesita la disposición de la adecuada metodología, el proyecto se tiene que diseñar bajo las estructuraciones de internacionales estándares, el cual construirán la base y poder obtener un excelente resultado sobre la puesta en marcha del proyecto (p 55).

I.3 Teorías relacionadas

I.3.1 Minería de Datos (*Data Mining*)

Al respecto, Chitra y Subashini (2013) afirmaron que la minería de datos se entiende como la revelación de información valiosa o relevante para una organización, que se encuentra oculta dentro de enormes DB. La minería de datos es un método utilizada en la extracción de información vital de gran cantidad de datos existente y permitir una mejor toma de decisiones para el comercio minorista y la banca. Ellos usan el almacenamiento de datos para combinar diversos datos de bases de datos en un formato aceptable para que los datos pueden ser extraídos. Se analiza entonces la data y la información que captura se utiliza en todas las organizaciones y así ayudar en la toma de

decisiones. Las técnicas en minería de datos son muy útiles para el sector bancario para una mejor orientación y conseguir clientela nueva, la retención de clientes más valioso, la aprobación del crédito automático que se utiliza para la prevención del fraude, detección de fraudes en tiempo real, proporcionando productos a base de segmento, el análisis de los clientes, transacciones patrones en el tiempo para una mejor retención y la relación, la gestión de riesgos y la comercialización (p. 77).

I.3.2 Metodologías

I.3.2.1 Metodología KDD

Usama y Shapiro (1997) señalaron que según sus siglas Knowledge Discovery in Databases es una metodología que traducida al español sería “Descubrir el conocimiento en base de datos”, esta metodología propone cinco fases: pre procesamiento, selección, minería de datos, transformación, implantación y evaluación. Es un proceso interactivo e iterativo. Estas cinco fases se pueden explicar en nueve pasos:

- **Primer paso:** consiste en el desarrollo de entender a la app, tener conocimiento y dominio previos además identificar el final del proceso de KDD teniendo presente la vista del cliente.
- **Segundo paso:** consiste en la creación de los conjuntos de datos: seleccionando los conjuntos de los datos, o los subconjuntos de la variable o data de muestra, al descubrir que se lleva a cabo.
- **Tercer paso:** Limpiar y pre procesamiento de data. Operación básica incluyendo la ejecución de ruido campos de data sin rellenar, etc.
- **Cuarto paso:** reduciendo la data y las proyecciones: buscando la característica que utilizaremos para representar los datos en función del objetivo.
- **Quinto paso:** colocaremos los objetivos del KDD (primer paso) en el método de minería de data.
- **Sexto paso:** se realiza el análisis para explorar la hipótesis, la selección del modelo: utilizaremos en la búsqueda para patrón de data el algoritmo de minería de datos que hallamos escogido.
- **Séptimo paso:** Minería de datos: buscamos el patrón de intereses en una

determinada manera de representaciones.

- **Octavo paso:** Interpretamos los patrones minados, nos brinda el regreso a cualquiera de los anteriores pasos y tener mayores iteraciones. En este paso podemos realizar la implicación de la visualización del patrón y modelo extraído o visualizar los datos que figura extraído en el modelo.
- **Noveno paso:** Actúa sobre el conocimiento descubierto: utilizamos el conocimiento directamente, al incorporar los conocimientos en el sistema y adoptar la nueva medida o, sencillamente documentaciones y presentación del informe a la parte interesada. Aquí incluimos la comprobación y soluciones del posible conflicto con creían (o extrae) del conocimiento.

I.3.2.2 Metodología SEMMA

Los especialistas del *SAS Institute Inc* (2019) indicaron que SEMMA es un acrónimo que corresponde a 5 fase: (*Sample, Explore, Modify, Model, Assess*) Las metodologías están propuestas por *SAS*, y la puntualiza como: proceso de selección, exploraciones y modelar una gran cantidad de datos y descubrimos patrones desconocidos del negocio. Cada una de las iniciales hace referencia a las fases de minería de datos en los proyectos, teniendo diversos nodos para que al analizar logren escoger dependiendo del modelo que va llevar a cabo.

- **Muestra:** La primera fase del proyecto. Es aquí donde se prepara los datos para explorarla. En la fase el nodo de partición. Usualmente se utiliza el 70% en la muestra de entrenamiento y 3% en validar
- **Exploración:** Exploramos los datos. Esta es la etapa mayor trabajosa, y a la vez hermosa. Teniendo el nodo el cual ayuda al explorar los gráficos de los datos, también nos brinda la herramienta de exterminar los inputs que no tengan una relación con la variable, también se puede realizar “*clustering*” o una segmentación
- **Modificación:** Al llegar a esta etapa ya se puede comenzar a hablar en serio. En esta etapa se centra en seleccionar y transformar las variables y datos las que servirían al construir los modelos. También resalta: imputación de valores

"missing", "outliers", reducción de dimensiones, etc.

- **Modelo:** en esta etapa tenemos que elegir el modelo. Al elegir el modelo esto dependerá sobre todo de los datos que ya se tiene y las variables y de lograr el modelo sencillo de entender. Se puede escoger redes neuronales, arboles de decisión, regresión logística, regresión, análisis factorial discriminante. En esta etapa aplicaremos muchos a la vez, y posterior se comparará el resultado obtenido
- **Evaluación:** Al realizar el trabajo en esta etapa final se compara los modelos obtenidos. Es fácil usar el análisis del diagrama ROC. La curva ROC lo usaremos al comprar el comportamiento global de un modelo. El gráfico ROC enfrentara 2 variables: especificidad y sensibilidad. Siendo lo menos que las dos sean mayores.

I.3.2.3 Metodología CRISP – DM

Los especialistas del diario de almacenamiento de datos (2000) indicaron que esta metodología es nueva que inicia no mucho tiempo después de la Minería de Datos se forje como la mayor disciplina en el análisis de los Activos de Información que tiene la organización.

Esta herramienta de trabajo inicia de la necesidad de conocer técnicas nuevas y aplica en una mayor comprensión de Minería de Datos y el resultado que se basa en el proceso jerárquico esta se encuentra compuesta en diversos tareas o niveles que ofrecen resultado a un menor tiempo, es por eso que esta metodología está considerada como Minería de Datos para inexpertos según uno que otro texto que hablan sobre Business Intelligence. Para construirla no solo se basa en teoría sino también en experiencias prácticas, como es una metodología abierta esta puede ser utilizada con diversas herramientas de *Business Intelligence*. La principal característica es al comienzo de analizar lo hace desde una perspectiva global que enfatiza con el conocimiento del negocio, es así que se convierte en una metodología apegada al proyecto que se realiza en la Organización en primera fase.

I.3.3 Cuadro comparativo entre las metodologías presentadas

Tabla 1. Cuadro comparativo entre las metodologías presentadas

	KDD	SEMMA	CRISP-DM
FASES	-Selección -matriz Pre procesamiento -Transformación -Minería de datos -Evaluación e implementación (Usama y Shapiro, 1997, p. 33)	-Muestreo -Exploración -Modificación -Modelo -Evaluación (SAS, 2019, p. 42)	-Conocimiento del negocio -Conocimiento de los datos -Preparación de datos -Modelamiento -Evaluación -Despliegue (Galán, 2016, p. 94)
VENTAJAS	Se utiliza mayormente con modelos interpretativos (Usama y Shapiro, 1997, p. 33)	Al utilizar una muestra facilita la obtención de resultados (SAS, 2019, p. 42)	-Se utiliza con modelos interpretativos y predictivos. -Se enfoca en el conocimiento del negocio (Galán, 2016, p. 94)
DESVENTAJAS	Utiliza métodos de reducción de datos (Usama y Shapiro, 1997, p. 33)	Utiliza métodos de reducción de datos (SAS, 2019, p. 42)	Se requiere grandes dimensiones de base de datos para su mejor explicación. (Galán, 2016, p. 94)
ENFOQUE	Estadístico (Usama y Shapiro, 1997, p. 33)	Estadístico (SAS, 2019, p. 42)	Estadístico/Inteligencia artificial (Galán, 2016, p. 94)

Una vez revisada cada metodología KDD, CRISP-DM, se puede concluir que, si bien existen diferencias en la cantidad de fases para realizar determinada acción, éstas siguen una metodología principal que constaría de cinco pasos:

Selección de datos, procesado o limpieza, implementación de técnica elegida, evaluación de modelo o resultados y despliegue. Si bien es cierto cada metodología tiene sus ventajas hemos convenido en utilizar la metodología *CRISP – DM* ya que según la cronología es la que engloba y ha venido mejorando las limitaciones con respecto a las anteriores además de la sencillez en su uso y fácil complemento con diversos modelos para la minería de datos.

I.3.4 Modelo Sistémico

Según Bertalanffi (1989), el sistema puede estar abierto o cerrado (dependiendo de si intercambiarán energía, materia o información con el medio ambiente), actúan como un todo, tienen circularidad (esto se debe a la conexión de los elementos, la causal es redonda en lugar de recta) y equivalencia (esta causa suele tener diversos efectos). Podemos decir que el enfoque de sistemas se basa, en principio, en el entendimiento de que el sistema tiene propiedades diferentes en la sencilla adición en los elementos. El modelo del sistema pasa de estudiar al aislado individuo a estudiar los sistemas y la relación con sus elementos que lo hacen. Los sistemas es un todo organizado formado por componentes que se comunican de cierta manera (p. 55).

I.3.5 Modelo Lineal

Mckeenw (2000) señaló que este es un modelo en el que la mayoría de relaciones funcionales son tales que al sumar las variables independientes son proporcionales a la variable dependiente. En este modelo se tiene como características que para cada cambio la variable independiente, y los demás se quedan fijos, la modificación de la variable dependiente es proporcional directamente (p. 20).

I.3.6 Modelo No Lineal

Mckeenw (2000, p. 24), señaló que un modelo en el que hay ecuaciones o funciones que no son proporcionales. Estos modelos tienen variables elevadas a una potencia diferente de una y / o tienen productos de dos o más variables.

I.3.7 Credit Scoring

Hand y Henley (1997, p. 243) Señalaron que la calificación crediticia es un método estadístico secuencial que se utiliza en la clasificación u organizar a aquellos que

requieren un préstamo, incluidos aquellos que ya son clientes del prestamista, en los tipos de riesgos "buenos" y "malos". Inicialmente, durante la década de 1970, las aplicaciones de calificación crediticia se construyeron con técnicas estadísticas. Después, los métodos utilizados evolucionan a técnica de inteligencia matemática, econométrica y artificial. El nacimiento de cualquier solicitud de evaluación de crédito se realizará al tomar la data del cliente llenado en la solicitud de crédito de fuente interna hasta de fuentes de información externas para una evaluación posterior.

Medina y Selva (2013, p. 303) afirmaron que el *credit scoring* se aplica a cualquier sistema de evaluación crediticia que permita evaluar automáticamente los riesgos se asocian con cada una de las solicitudes de préstamo. El riesgo que dependerá de la calidad crediticia del prestatario, el tipo de crédito, las condiciones y otra característica del negocio y el cliente para definir cada observación o cada solicitud de crédito. No habrá riesgo en la operación de préstamo solo si la entidad que los realiza actúa como intermediario o intermediario, o cuando el préstamo se otorga con una garantía del gobierno.

I.3.8 Pronostico

Brockwell y Davis (2010) analizaron que el término predicción se debe referir al efecto y la acción de predecir algo; en este sentido, al predecir es dar un anuncio ya sea por revelaciones, conjeturar que algo sucederá o ciencia. En estadística los pronósticos son los procesos de estimaciones en la situación de incertidumbre (p. 28).

I.3.9 Dimensiones

I.3.9.1 Precisión

Armstrong (2001) Observo que hay unas medidas de precisión de un común cuya escala va a depender del volumen de datos. Son útiles al comparar el diferente método en el mismo conjunto de datos, pero no deben usarse, en comparación de diversos conjuntos de datos que tiene varias escalas. Las medidas que tienen mayor utilización según la escala están basadas en un error absoluto o cuadrado (p. 99).

I.3.9.2 Error y fiabilidad

Serna y Arango (2012) Señalaron que los modelos matemáticos de los escenarios de prueba y los datos de error estaban bien establecidos. Estos modelos tienen como objetivo predecir cuál será el comportamiento del software en un entorno funcional en función de comportamiento durante la prueba. Para lograrlo, mayormente requiere una especificación de perfil operativo: la descripción de esperar que apliquen las entradas los usuarios. Y realizar el cálculo de probabilidad de falla, los modelos nos muestran las suposiciones sobre la distribución de probabilidades fundamentales que rigen la aparición de anomalías (p. 219).

I.3.9.3 Tiempos de respuesta

De acuerdo a Pressman (2010, págs. 526-552), los tiempos de respuesta están íntimamente relacionado con la prueba de desempeño y la prueba de carga de todo software. A continuación, describiremos cada una de estas pruebas:

I.3.9.3.1 Pruebas de desempeño

Su objetivo es la validación del tiempo en dar las respuestas para la transacción o función de los negocios en:

- Volumen normal anticipado
- Volumen máximo anticipado

La prueba de desempeño mide el tiempo en dar respuesta, índice de procesamiento en transacción y requisito sensible al periodo. Como objetivo de la prueba de desempeño será validación y verificación del requisito de desempeño que está especificados.

La prueba de desempeño mayormente se tiene que ejecutar varias veces, usando diferentes cargas en el sistema. Las pruebas iniciales deben ser ejecutadas con similar carga por el sistema. La segunda se tiene que utilizar una carga máxima.

También la prueba de desempeño se puede utilizar en refinar y perfilar el cumplimiento del sistema en función de condición de las configuraciones o carga de HW.

Principales actividades

- Comparación de los requisitos con el desempeño en el sistema actual.
- La mejora de las métricas a desempeñar y proyectarnos la capacidad a futuro de la carga del sistema.

Los objetivos definidos por el nivel de los servicios tienen que guiar las pruebas de performance. La característica que afecta al desempeño:

- Errores lógicos
- Procesamiento ineficiente
- Diseño pobre: demasiadas interfaces, instrucción y entrada/salida
- Cuello de botella en CPU, disco, canal de entrada/salida
- Salida del sistema
- Tiempo en las respuestas
- Capacidad para almacenar
- entrada/salida de la data
- Número de transacciones a ser manejadas en simultáneo.

La prueba de desempeño usa técnicas de caja negra y caja blanca.

Técnica

La prueba a desempeñar usa técnicas de caja negra y blanca. Usando el procedimiento de pruebas para desarrollo en la prueba del modelo del negocio. Modifican archivos de data o scripts y así aumentar las veces que ocurre cada transacción. El script debe estar ejecutado en un ordenador y tienen que estar repetidas con clientes múltiples.

Criterio de completitud

Las pruebas se completarán sin error y en un espacio adecuado para las transacciones. En cuanto mayor sea el número de transacción, mayor será el número de usuarios. La prueba de script se completará sin error y en el tiempo

adecuado.

Consideraciones especiales

Una prueba de desempeño integral incluirá en una carga background para el servidor. Teniendo muchos métodos a utilizar y poder realizar esta transacción dirigida directa al servidor, normalmente se realiza de forma en sentencias SQL, creaciones del usuario virtual para la simulación de diversos clientes. Utilizaremos herramientas de emulación de terminal remota y llegar a la obtención de la carga. Dicha técnica a su vez se puede utilizar y de esta forma realizar la carga de tráfico de red.

Se utilizó muchos clientes, el cliente corrió los scripts en la prueba.

- La prueba de desempeño tienen que estar ejecutada en un ordenador en un tiempo dedicado. De esta forma permitirá un control total y la medida precisa.
- La DB que se utilizó en la prueba de desempeño tiene que tener un apropiado tamaño o superior a la diseñada.

I.3.9.3.2 Pruebas de carga

Tiene como objetivo la verificación del tiempo de respuesta del sistema en cada transacción, caso de uso de negocio, en diversa condición de carga.

- Descripción de la Prueba:

La prueba de carga calcula la capacidad del sistema en el apropiado funcionamiento en diversa condición de carga.

La finalidad de las pruebas en carga es la determinación y aseguramiento de la funcionalidad apropiada del sistema superando la carga de trabajo máximo esperado. Posteriormente, la prueba de carga evaluará la característica de desempeño.

- Criterio de completitud

Muchas transacciones. Concluye la prueba del script sin falla alguna y en el tiempo apropiado.

- Consideraciones especiales

La prueba de carga tiene que ser ejecutada en un ordenador en un tiempo dedicado. El cual permitirá controlar totalmente y tener precisas medidas.

- La DB que se utilizó en la prueba de desempeño tiene que tener un apropiado tamaño o superior a la diseñada.

I.4 Formulación del problema

Sobre la base de realidad problemática presentada se planteó los siguientes problemas de investigación:

1.4.1 Problema general

¿Cuál es el impacto de un sistema de pronóstico de calificación crediticia de clientes basado en árbol de clasificación y regresión lineal?

1.4.2 Problemas específicos

- ¿Cuál es el impacto de un sistema de pronóstico de calificación crediticia de clientes basado en árbol de clasificación y regresión lineal en la precisión del pronóstico?
- ¿Cuál es el impacto de un sistema de pronóstico de calificación crediticia de clientes basado en árbol de clasificación y regresión lineal en el error del pronóstico?
- ¿Cuál es el impacto de un sistema de pronóstico de calificación crediticia de clientes basado en árbol de clasificación y regresión lineal en el tiempo de respuesta del pronóstico?

I.5 Justificación del estudio

Las MYPES en nuestro país es un sector importante de la economía, por lo que estas

empresas que desean impulsar la economía con sus empresas y nuevos proyectos deben realizar una evaluación adecuada y detallada, incluidos otros factores que nos ayuden a conocer la realidad de cada cliente, esto sería propicio para brindar oportunidades nuevos clientes que no tienen o no tienen historial crediticio con opciones de pago y liquidez, y esta medida permitiría una mejor evaluación de riesgos y asignación de crédito a diferentes entidades para el beneficio de los clientes. Por otro lado, se beneficiaría financieramente en una mejor administración de sus recursos financieros.

1.5.1 Justificación teórica

Según Hernández (2014, págs. 39-49), cada estudio requiere justificación presentando sus razones en detalle, dejando en claro por qué y/o por qué el estudio tiene un propósito específico que es lo suficientemente significativo como para justificar su realización.

1.5.2 Justificación tecnológica

En el presente proyecto se han tomado diversas referencias y modelos ya utilizados de investigaciones en las principales fuentes y uso de tecnologías que se utilizan, para así lograr una mejor afinación de estimación de pronóstico de riesgo en la calificación crediticia o *credit scoring* a continuación citaremos las tecnologías que nos han servido de referencia para la implementación de nuestro sistema basadas en la selección de características de minería de datos para un modelo de calificación de crédito

Según Rayo, Lara y Camino (2010, págs. 89-124), estos son los procedimientos estadísticos utilizados para clasificar a aquellos que requieren un préstamo, incluidos aquellos que ya son clientes del prestamista, en los tipos de riesgos "buenos" y "malos". Estos sistemas operan sobre la base de un historial de cumplimiento e incumplimiento. En cada caso, la DB o equivalente historial deberá tener información de la característica del préstamo dispuesto para llegar a descifrar el formulario para poder predecir exitosamente si la transacción de crédito podría incumplir basado en experiencia similar. El pronóstico se expresa de 2 maneras: probabilidad de incumplimiento y resultado numérico.

Según Liu y Schulman (2005), El árbol modelo M5 es un algoritmo de predicción

numérica. El algoritmo M5 genera una estructura de árbol de decisión convencional con modelos regresiones lineal en las ramas, lo que permitiría una distribución más uniforme de los resultados en lugar de etiquetas de clase discretas. El árbol base se crea mediante un método de partición recursiva similar a C4.5. Luego se construye una función de regresión para cada nodo de árbol construido usando un algoritmo de regresión estándar. Las funciones de regresión se simplifican minimizando la estimación del error esperado multiplicado por el factor $(n + v) / (n-v)$.

1.5.3 Justificación económica

El presente trabajo tendría una sólida justificación económica, ya que mediante su implementación se pueden reducir costos en la gestión financiera de FOGAPI, que a su vez se verían reflejado en menores tasas y comisiones para sus clientes. Al respecto, los especialistas del BCR del Perú (2006) Afirmaron que los análisis en el costo del crédito estarán inconclusos si se enfoca en la TI que se cobró, debido a que hay costos adicionales tales como costo y comisión, que suelen aumentar rigurosamente los costos efectivos para los clientes. Para calcular esto último, se utiliza una metodología de tasa interna de rendimiento (TIR) para considerar el flujo de efectivo de un préstamo, encontrando el caso que hay una gran diferencia entre la TI primera y el precio efectivo en el préstamo, información que los deudores deben tener en cuenta para tomar la decisión correcta.

1.6 Objetivos

1.6.1 Objetivo general

El objetivo general es determinar el impacto del sistema de pronóstico de la calificación crediticia de clientes basado en árbol de clasificación y regresión lineal.

1.6.2 Objetivos específicos

OE1: Determinar el impacto del sistema de pronóstico de la calificación crediticia de clientes basado en árbol de clasificación y regresión lineal en la precisión del pronóstico.

OE2: Determinar el impacto del sistema de pronóstico de la calificación crediticia de clientes basado en árbol de clasificación y regresión lineal en el error del pronóstico.

OE3: Determinar el impacto del sistema de pronóstico de la calificación crediticia de clientes basado en árbol de clasificación y regresión lineal en el tiempo de respuesta del pronóstico.

I.7 Hipótesis

1.6.1 Hipótesis general

HG: El sistema de pronóstico de calificación crediticia de clientes basado en árbol de clasificación y regresión lineal incrementa la precisión del pronóstico, reduce el error y el tiempo de respuesta

En un intento por explicar los credit scoring Reza (2018) en un artículo titulado Credit Scoring usando el algoritmo CART y la optimización de enjambre de partículas binarias, el International Journal of Electrical and Computer Engineering lo define como un proceso que existe en todas las instituciones financieras. Una forma de predecir si un prestatario es elegible para recibir un préstamo o no es de gran preocupación en los pasos generales del proceso del préstamo. Casi todas las instituciones financieras tienen sus métodos de calificación crediticia. Hoy, el acceso a la minería de datos es aceptado como uno de los métodos más conocidos. Ciertamente, la precisión también fue un problema importante en este enfoque. Esta investigación propuso un método híbrido utilizando el algoritmo CART y la optimización de enjambres de partículas binarias. Los indicadores de rendimiento utilizados en esta investigación son precisión de clasificación, tasa de error, sensibilidad, especificidad y precisión. Los resultados experimentales basados en un conjunto de datos públicos mostraron que la precisión del procedimiento propuesto fue del 78%. En comparación con varios algoritmos populares, como la red neuronal, la regresión logística y el vector de soporte, el método propuesto ha mostrado un rendimiento sobresaliente (p. 41).

1.6.2 Hipótesis específicas

HE1: El sistema de pronóstico de calificación crediticia de clientes basado en árbol de clasificación y regresión lineal incrementa la precisión del pronóstico.

De acuerdo con Leal, Arangüiz y Gallegos (2018) trabajo de investigación titulado "ANÁLISIS DE RIESGO DE CRÉDITO, UNA PROPUESTA DE MODELO DE PUNTUACIÓN DE CRÉDITO" se aplicó a una empresa que experimento un importante incremento de ventas en el último año, y con ello una reducción en el nivel de calidad y liquidez de la cuenta. Sin embargo, dicho aumento en las cuentas por cobrar se asoció con un más alto riesgo de cobranza, en política liberalizador de cuentas por cobrar. Además, no tenía ni un solo sistema para la gestión de crédito que permitiera la evaluación acorde a la calidad y la solvencia en los clientes. Por lo tanto, el artículo nos brinda una muestra rigurosa y ponderada para evaluar la solvencia de la empresa y sus clientes, lo que ha permitido reducir el riesgo de crédito o deudas no cobradas. El documento consideró una descripción del modelo de evaluación crediticia y específicamente el modelo de calificación crediticia. Mediante entrevista con expertos, se han identificado variables cualitativas y cuantitativas críticas que se tendrán en cuenta en la gestión. Con respecto a la calidad de la muestra de evaluación crediticia propuesta, los resultados obtenidos fueron que un 81.82% de los préstamos que se otorgan a los clientes excedieron un menor nivel de evaluación o aprobación de la compañía (p. 34).

HE2: El sistema de pronóstico de calificación crediticia de clientes basado en árbol de clasificación y regresión lineal reduce el error del pronóstico.

En el propósito de confirmar esto, Leal, Arangüiz y Gallegos (2018) Han demostrado que los modelos de puntuación se basan en variable cuantitativa y cualitativa criticadas al considerar el proceso de gestión de crédito. Se ha demostrado que la calidad de la modelo basada en su evaluación crediticia propuesta observamos el 81.82% de préstamos brindados a clientes exceden el margen mínimo al evaluar o aprobación de la compañía (p. 123).

HE3: El sistema de pronóstico de calificación crediticia de clientes basado en árbol de clasificación y regresión lineal reduce el tiempo de respuesta del pronóstico.

En una iniciativa por encontrar nuevas herramientas, Karminsky y Khromova (2016) han construido un modelo confiable basado en información pública para el uso práctico de los agentes (agentes reguladores y los propios bancos). Su investigación demostró construir una serie de calificaciones basadas en una tabla de variables representativas. Del mismo modo, los datos utilizados en la investigación corresponden a una base de datos bancarios que contiene información financiera sobre el sistema bancario internacional, que corresponde al período 1996-2011. Como parte de sus resultados, se obtuvieron los siguientes datos al diseñar el modelo de prueba: Conclusiones: 1) las variables macro mejoran el poder explicativo del modelo y 2) el efecto de la afiliación regional es significativo. Además, los grados de calificación son consistentes con el ciclo comercial. En este sentido, descubrieron que Standard Poor's & Moodi's son las agencias con las calificaciones más conservadoras y menos conservadoras. El modelo propuesto por los investigadores se clasificó como prácticamente útil, ya que la predicción de calificación mostró un 31% de resultados correctos y hasta un 70% de predicción con error dentro de las calificaciones, mientras que la predicción del reloj resultó en 62% y 95% respectivamente (p 69).

II. MÉTODO

II.1 Diseño de la investigación

II.1.1 Enfoque de la investigación

El enfoque cuantitativo es un proceso secuencial y basado en evidencia. De las diversas etapas que están interconectadas y los pasos no pueden adelantarse o retrasarse, la secuencia es secuencial, por lo que algunas etapas pueden redefinirse. Una vez detallados y delineados, se elaboran objetivos o preguntas de investigación, se desarrollan revisiones de literatura y un conjunto de procedimientos o perspectivas teóricas. (Hernández, 2014, p.4).

El presente trabajo contiene el enfoque cuantitativo, debido a que obtiene data que será analizada.

II.1.2 Tipo de estudio

Utilizamos la investigación aplicada que consiste en conocer el estado de la sociedad y crear modificaciones para mejorar, generar innovación y actos concretos en el cambio (Vargas, 2009). La investigación aplicada aumenta la realidad, resolviendo problemas percibidos. Por lo tanto, podemos afirmar que esta investigación es de tipo aplicado porque proponemos una solución aplicando un sistema de predicción.

II.1.3 Diseño de investigación

Se define un proyecto que la realizamos sin la manipulación de las variables. Quiere decir, las variables no se cambian intencionalmente. Se hace en la investigación no experimental es la observación de lo que ocurren en su estado natural; sucesivamente estos cambios son analizados por expertos en el campo (Hernández, 2014 p.149)

Según Hernández, Fernández y Baptista (2014), Los diseños pre-experimentales manipulan intencionalmente al menos una variable independiente para notar su efecto en una o más variables dependientes, excepto que difieren de los experimentos "puros" en el grado de certeza que puede estar en la equivalencia inicial de los grupos. (p. 151).

El diseño de la investigación es pre experimental, debido a que se manipulara la variable independiente (sistema de pronóstico de la calificación crediticia basado en árbol de

clasificación y regresión lineal) para analizar el impacto que genera.

Utilizamos diseño pre experimental, para el cual aplicamos un pre test y post test al mismo grupo.

G: O1 X O2

Dónde:

G: Conjunto de estudio experimental

O1: Observación clasificación crediticia de clientes

X: Variables: Sistema de pronóstico de la clasificación crediticia.

O2: Observación de la clasificación crediticia de clientes después de la implementación sistema de pronóstico de la clasificación crediticia.

El diseño pre experimental Según Hernández, Collado y Baptista (2014, p. 141) viene a ser el diseño del grupo con mínimo de control. Normalmente se utiliza al comenzar un acercamiento hacia la problemática del proyecto en la realidad científica en los que se clasifique los diseños de investigación científica.

II.2 Variable, Operacionalización

II.2.1 Variable

Variable:

Impacto del sistema de pronóstico de la clasificación crediticia basado en árbol de clasificación y regresión lineal

II.2.2 Operacionalización de la variable

Variable: Impacto del sistema de pronóstico de la clasificación crediticia basado en árbol de clasificación y regresión lineal

El impacto del sistema de pronóstico que desarrollaremos se basa en utilizar técnicas de

Minería de datos utilizando algoritmos de árbol de clasificación y regresión lineal.

Definición conceptual

Enríquez (2016) nos dice que los SI son conjuntos de recurso tecnológico, humanos y económicos, interrelacionados entre sí, con único propósito de satisfacción en necesidad de información de la entidad, y tener una buena toma de decisiones y gestión.

En este proyecto de investigación el sistema informático se basa en:

1. Un árbol de clasificación: según Vizcaino (2008) es un conjunto de condiciones o reglas organizadas en una estructura jerárquica, de tal manera que la decisión final se puede determinar siguiendo las condiciones que se cumplen desde la raíz hasta alguna de sus hojas.
2. La regresión lineal: según Mckeonw (2000) es un modelo en el que todas las relaciones funcionales son de tal naturaleza que la variable dependiente es proporcional a la suma de las variables independientes. Estos modelos también tienen la característica de que, para cualquier cambio en una variable independiente, mientras se mantienen las demás fijas, el cambio en la variable dependiente es directamente proporcional.

Definición operacional

Utilización de grandes bases de datos para encontrar y generar patrones reiterativos que nos ayuden a obtener la información adecuada cuyas métricas se expresan en **funcionalidad** (Nivel de precisión del pronóstico, nivel de error del sistema), y **eficiencia** (Nivel de tiempos de respuesta).

II.2.3 Matriz de Operacionalización de las variables

Tabla 2. Matriz de Operacionalización de las variables de la investigación

Variable	Definición Conceptual	Definición Operacional	Dimensión	Indicador	Instrumento	Escala de Medición
Impacto del sistema de pronóstico de la calificación crediticia basado en árbol de clasificación y regresión lineal Brockwell & Davis, (2010). Bertalanffy, (1989).	Según Enríquez (2016) un sistema informático es un conjunto de recursos tecnológicos, humanos y económicos, interrelacionados entre sí, con el propósito de satisfacer las necesidades de información de una institución, para la correcta gestión y toma de decisiones. En este proyecto de investigación el sistema informático se basa en:	Utilización de grandes bases de datos para encontrar y generar patrones reiterativos que nos ayuden a obtener la información adecuada cuyas métricas se expresan en funcionalidad (Nivel de precisión del pronóstico, nivel de error del sistema), y eficiencia (Nivel de tiempos de respuesta).	Precisión del pronóstico	Incremento de la precisión del pronóstico respecto al modelo: Credit scoring	TS=Tiempo en segundos TMEF = TMDF+TMDR TMDF: tiempo medio de fallo TMDR: tiempo medio de reparación	nominal
	1. Un árbol de clasificación: según Vizcaino (2008) es un conjunto de condiciones o reglas organizadas en una estructura jerárquica, de tal manera que la decisión final se puede determinar siguiendo las condiciones que se cumplen desde la raíz hasta alguna de sus hojas.		Error de pronóstico del sistema	Reducción de error del pronóstico del sistema respecto al modelo: Credit scoring	CP = # Pruebas sin error / Total Población Registrado.	nominal
	2. La regresión lineal: según Mckeonw (2000) es un modelo en el que todas las relaciones funcionales son de tal naturaleza que la variable dependiente es proporcional a la suma de las variables independientes. Estos modelos también tienen la característica de que, para cualquier cambio en una variable independiente, mientras se mantienen las demás fijas, el cambio en la variable dependiente es directamente proporcional.		Tiempos de respuesta del sistema	Reducción del tiempo de respuesta del sistema respecto al modelo: Credit scoring	TPM = Técnica Modelo Anterior - Técnica del Modelo Propuesto	nominal

II.3 Población y muestra

Población

Arias (2012) nos dice que es conjunto infinito o finito de elementos de comunes características (p. 81).

La población en estudio está conformada por la información de solicitudes de crédito y clientes que se tiene recopilada en la DB desde 2014 hasta 2019, correspondiente a 50310 registros.

Muestra

Viene a ser parte de la población la cual queda seleccionada, de esta parte obtenemos la formación y así podemos desarrollar el proyecto, en el cual efectuaremos el objeto de estudio la medición y observación de variables (Bernal, 2010 p. 161).

La muestra de estudio está basada en la totalidad de la población, de los cuales se cuenta con información detallada para nuestros fines de desarrollo de extracción y minería de datos.

Muestreo

Los criterios inclusión y exclusión se detallan a continuación:

Criterios de inclusión de la base de datos de clientes:

- Clientes con más de seis meses de antigüedad.
- Clientes que se encontraban la totalidad de los datos a utilizar en el modelo.

Criterios de exclusión la base de datos de clientes:

- Clientes que cuentan con menos de seis meses de antigüedad.
- Clientes que no se encontraban la totalidad de los datos a utilizar en el modelo.

II.4 Técnicas e instrumentos de recolección de datos, validez y confiabilidad

Registro

Según Hernández, (2010) quienes sostuvieron que observación es un confiable registro, también válido en comportamientos y observables situaciones, comenzando por categoría de conjunto (p.260)

Ficha de registro

Es el conjunto de registros tomadas de acuerdo a los indicadores que presenta cada objetivo, y posteriormente pasa por la fase de análisis para realizar una retroalimentación o verificar el estado actual con los cambios posteriores que se realizan de acuerdo al modelo de análisis que se presentan en cada indicador propuesto en el caso de estudio.

Validez

Esta referido a nivel de un instrumento que se encarga de medir las variables las cuales se deben medir, para lo cual se debe conocer cuáles son los rasgos o características que deseamos estudiar (Hernández, 2017)

Para validar el proyecto se utilizará el Juicio de expertos.

Confiabilidad

Instrumento de medición que hace referencia al grado en que su reiterativa aplicación al mismo objeto o individuo origina resultados iguales, veraces y coherentes basados en su validez y cumplimiento de objetivos, por ejemplo, un test de inteligencia estará aplicada a un grupo de individuos obteniendo la inteligencia actual, después de un mes se aplica el mismo test, pero se incorpora un valor diferente, similar a subsecuentes mediciones, este test no estaría

confiable. (Hernández, Fernández y Baptista, 2010).

El registro de observación usado como instrumento, no requiere de los cálculos de confiabilidad, Ya que la información es extraída directamente de la DB de la entidad.

II.5 Métodos de análisis de datos

Usama (1997) comento que para este proyecto se empleará la minería de datos con la metodología CRISP – DM apoyada de estadística descriptiva con inferencial y comprobación de la hipótesis sea correcta, teniendo presente el resultado que nos proporcione el instrumento de recolección de información en este caso análisis documental. Se ejecuta:

- Recolección de datos

Tabla 3. Recolección de datos

Técnica	Instrumento	Descripción
Observación	Registro de Observación de Datos	Se revisará los registros de datos de las transacciones y operaciones del periodo 2010 hasta 2018, correspondiente a 25000 clientes.

- Analizar el entorno de la fuente de datos
- Procesar y analizar la información
- Tabulación y graficas de datos
- Analizar descriptivos
- Analizar inferencial
- Contrastar la hipótesis

II.6 Aspectos éticos

Acatando la disposición vigente incluidas en el reglamento del grado y título, y como estudiante de la Universidad César Vallejo en la Facultad de Ingeniería de Sistemas; asevero que durante los desarrollos de este proyecto de investigación se empleó información auténtica, de manera legítima, fundada y respetando las disposiciones legales.

Según los especialistas de la Universidad de Celaya (2011) la realización del análisis crítico del objeto de estudio estará sujeto a alineamientos éticos básicos. En esta investigación se tuvo siempre cuenta de los aspectos éticos tales como el respeto en la propiedad intelectual, uso de fuentes fidedignas debidamente corroboradas, la protección a la identidad del participante en el presente proyecto, la confidencialidad de la data otorgados por la organización seleccionada, la honestidad al desarrollar el tema y circunspección en la obtención de información.

Según Abad y Morales (2005) indicaron que El investigador asume estos principios, desde el inicio, durante y después del proceso; a efectos de cumplir el principio de reserva, el respeto a la dignidad humana y el derecho a la intimidad. Así mismo para la disposición de datos y acceso a la información de la Financiera FOGAPI anteriormente, presentamos el proyecto al Gerente y autoridades de la entidad, así mismo a la escuela de Ingeniería de Sistemas para las respectivas aprobaciones. Las personas, que tuvieron una participación en el proyecto, se le informo con anticipación, sobre el beneficio, y el riesgo que predispone, respetando sus principios bioéticos de su persona como la autonomía, respetar, la no maleficencia y las beneficencias, también se pidió a todos los colaboradores, que participen de forma voluntaria y con respectiva veracidad del proceso de Investigación, la cual se encuentra respaldada por un documento el cual se encuentra en los anexos del presente proyecto.

III. RESULTADOS

3.1 Pruebas de normalidad

Prueba de normalidad que utilizado fue el método Kolmogorov aplicada al indicador. Ya que la muestra es mayor a 50, Indicaciones:

$N \geq 50$, empleamos el método de Kolmogorov-Smirnov

$N < 50$, empleamos el método de Shapiro-Wilk

Entonces considerando las siguientes medidas y obteniendo la prueba de normalidad se puede determinar que:

Sig. < 0.05 , Adopta una distribución no normal – Wilcoxon

Sig. ≥ 0.05 , Adopta una distribución normal – T Student

Donde Sig. Es el nivel crítico de contraste. Por lo cual basados en la definición de comprobación de hipótesis tenemos.

3.2 Dimensión 1: Precisión del pronóstico

3.2.1 Indicador 1: Incremento de la precisión del pronóstico respecto al modelo: Credit scoring

El puntaje obtenido del PreTest y PosTest, añadimos diferencia.

	PRESICIONP OST	PRESICIONP RE	DIFERENCIA	vs
1	93,00	83,00	10,00	
2	92,00	82,00	10,00	
3	95,00	70,00	26,00	
4	97,00	78,00	19,00	
5	98,00	66,00	33,00	
6	96,00	70,00	26,00	
7	96,00	79,00	17,00	
8	99,00	73,00	26,00	
9	99,00	73,00	26,00	
10	93,00	83,00	10,00	
11	92,00	81,00	11,00	
12	95,00	84,00	11,00	
13	94,00	59,00	35,00	
14	92,00	77,00	15,00	
15	90,00	76,00	15,00	
16	98,00	74,00	24,00	
17	93,00	61,00	32,00	
18	96,00	67,00	29,00	
19	98,00	83,00	14,00	
20	90,00	83,00	7,00	
21	93,00	73,00	20,00	

Figura Nº 1: Puntaje obtenidos Pre Test, Post Test y Diferencia - Precisión del Pronostico

Calculo de Datos Descriptivos

Tabla 4: Cálculos Estadísticos Descriptivos- Precisión del Pronóstico

Estadísticos descriptivos						
	N	Mínimo	Máximo	Media	Desviación estándar	Varianza
PRESICIONPOST	650	90,00	99,00	94,3892	2,63277	6,931
PRESICIONPRE	650	56,00	86,00	71,0446	8,43827	71,204
DIFERENCIA	650	5,00	42,00	23,3492	8,74724	76,514
N válido (por lista)	650					

Tabla de Frecuencias

Tabla 5: Tabla de Frecuencia Pre Test – Precisión del Pronóstico

PRESICIONPRE					
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	56,00	1	,2	,2	,2
	57,00	15	2,3	2,3	2,5
	58,00	18	2,8	2,8	5,2
	59,00	34	5,2	5,2	10,5
	60,00	27	4,2	4,2	14,6
	61,00	20	3,1	3,1	17,7
	62,00	25	3,8	3,8	21,5
	63,00	21	3,2	3,2	24,8
	64,00	23	3,5	3,5	28,3
	65,00	20	3,1	3,1	31,4
	66,00	27	4,2	4,2	35,5
	67,00	17	2,6	2,6	38,2
	68,00	17	2,6	2,6	40,8
	69,00	18	2,8	2,8	43,5
	70,00	27	4,2	4,2	47,7
	71,00	22	3,4	3,4	51,1
	72,00	24	3,7	3,7	54,8
	73,00	32	4,9	4,9	59,7
	74,00	20	3,1	3,1	62,8
	75,00	13	2,0	2,0	64,8

76,00	29	4,5	4,5	69,2
77,00	23	3,5	3,5	72,8
78,00	17	2,6	2,6	75,4
79,00	20	3,1	3,1	78,5
80,00	27	4,2	4,2	82,6
81,00	13	2,0	2,0	84,6
82,00	26	4,0	4,0	88,6
83,00	27	4,2	4,2	92,8
84,00	20	3,1	3,1	95,8
85,00	26	4,0	4,0	99,8
86,00	1	,2	,2	100,0
Total	650	100,0	100,0	

Tabla de Frecuencia Post Test – Precisión del Pronóstico

PRESICIONPRE

	Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
90,00	40	6,2	6,2	6,2
91,00	70	10,8	10,8	16,9
92,00	72	11,1	11,1	28,0
93,00	88	13,5	13,5	41,5
94,00	75	11,5	11,5	53,1
95,00	64	9,8	9,8	62,9
96,00	80	12,3	12,3	75,2
97,00	54	8,3	8,3	83,5
98,00	66	10,2	10,2	93,7
99,00	41	6,3	6,3	100,0
Total	650	100,0	100,0	

HISTOGRAMA

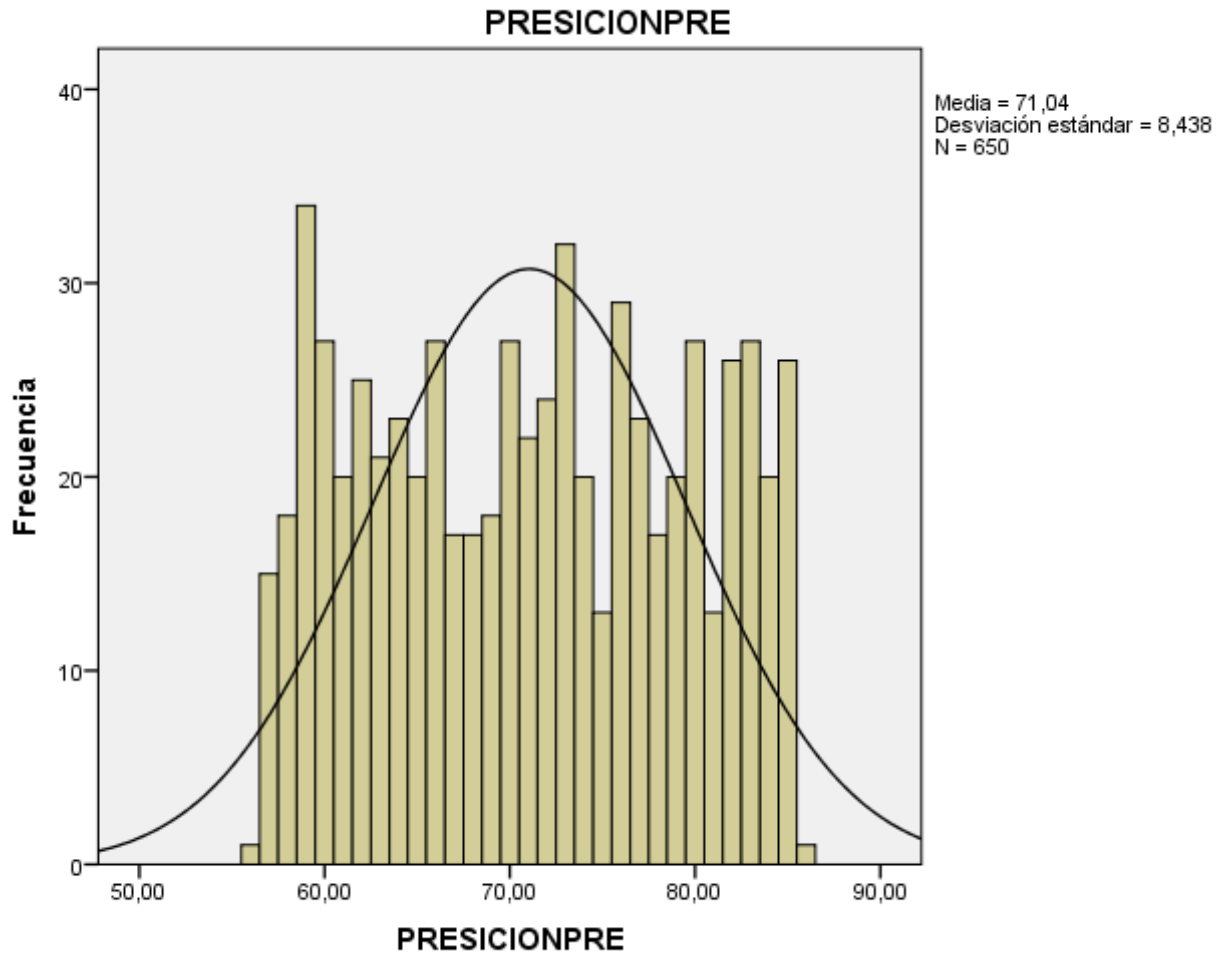


Figura Nº 2: Grafico de puntajes obtenidos en el Pre Test - Precisión del Pronostico

Resultado del análisis del PreTest en el gráfico, representamos el histograma obtenido de los puntajes del indicador Precisión del Pronóstico, obtuvimos una media de 71.04, con una desviación estándar de 8.438.

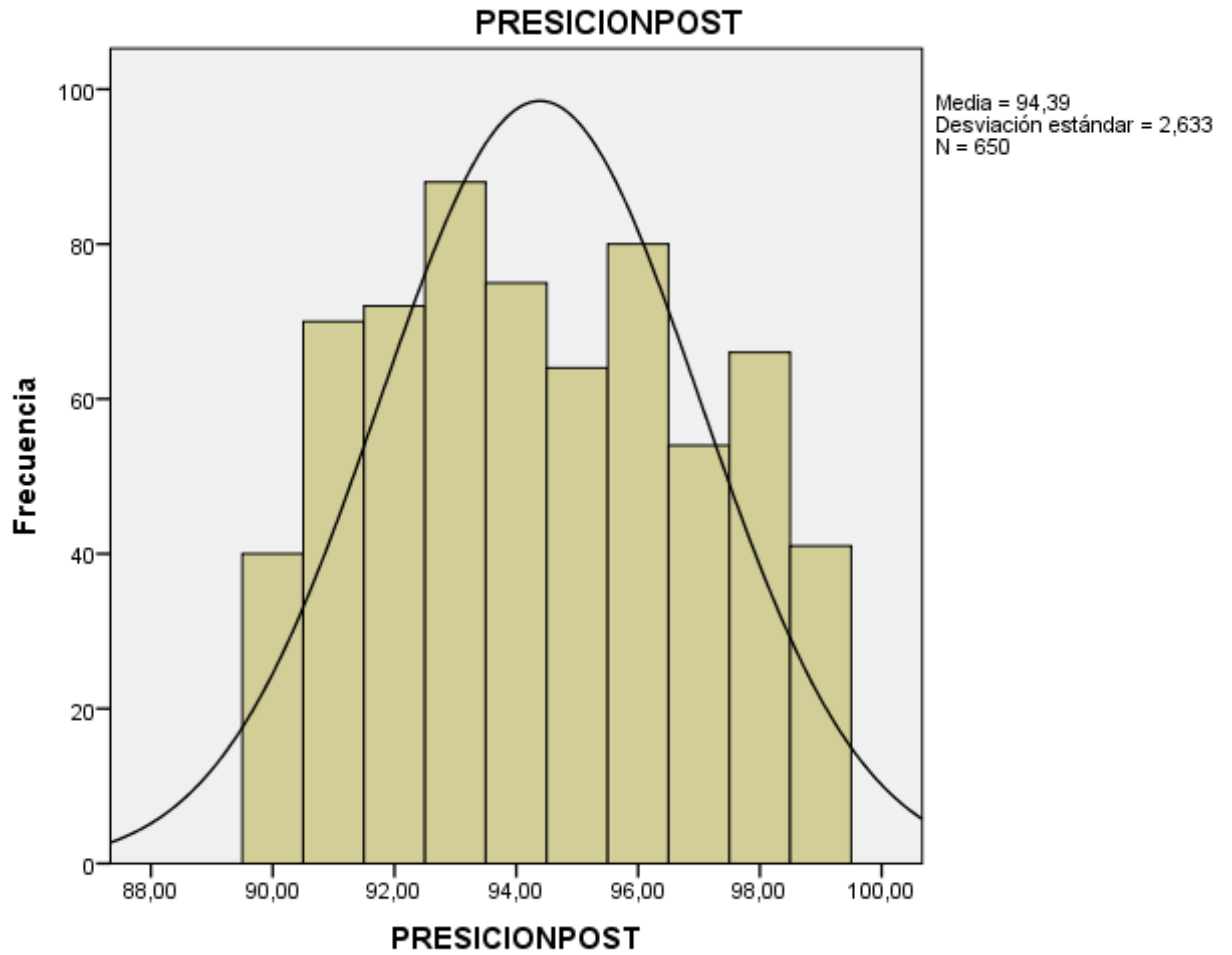


Figura Nº 3 : Grafico de puntajes obtenidos en el Post Test – Precisión del Pronostico

Resultado de análisis del PosTest en el gráfico, representamos el histograma obtenido de los puntajes del indicador Precisión del Pronóstico, obtuvimos una media de 94.39, con una desviación estándar de 2.633

Prueba de Normalidad

Determinamos que la distribución es normal o no, utilizamos la prueba de Kolmorov-Smirnov en la muestra.

Tabla 7: Prueba de Kolmorov – Precisión del Pronóstico

		Prueba de Kolmogorov-Smirnov para una muestra		
		PRESICIONPR	PRESICIONP	
		E	OST	DIFERENCIA
N		650	650	650
Parámetros normales	Media	71,0446	94,3892	23,3492
	Desviación estándar	8,43827	2,63277	8,74724
Máximas diferencias extremas	Absoluta	,081	,117	,081
	Positivo	,081	,117	,064
	Negativo	-,074	-,100	-,081
Estadístico de prueba		,081	,117	,081
Sig. asintótica (bilateral)		,007 ^c	,023 ^c	,148 ^c

Se observa en la tabla, el valor de significancia (Sig) de la columna diferencia es mayor a 0.05, por tal motivo podemos afirmar que el indicador Precisión del Pronostico sigue en una distribución normal.

Prueba de hipótesis

La distribución es normal, aplicamos la prueba Estadística comparación de medias. Se aplica la prueba de muestra relacionadas de T student. La H0 y HA:

- Hipótesis Nula (H0): El sistema de pronóstico de calificación crediticia de clientes basado en árbol de clasificación y regresión lineal no incrementa la precisión del pronóstico.
- Hipótesis Alterna (H1): El sistema de pronóstico de calificación crediticia de clientes basado en árbol de clasificación y regresión lineal incrementa la precisión del pronóstico.

Tabla 8: Aplicación de la muestra relacionadas T student – Precisión del Pronóstico

		Media	N	Desviación estándar	Media de error estándar
Par 1	PRESICIONPOST	94,3892	650	2,63277	,10327
	PRESICIONPRE	71,0446	650	8,43827	,33098

		Diferencias emparejadas				t	gl	Sig. (bilateral)	
		Media	Desviación estándar	Media de error estándar	95% de intervalo de confianza de la diferencia				
					Inferior				Superior
Par 1	PRESICIONPOST - PRESICIONPRE	23,34462	8,75614	,34344	22,67022	24,01901	67,972	650	,000

El valor de Sig. $0 < \alpha < 0.05$, se niega la H_0 y se acepta la H_A : El sistema de pronóstico de calificación crediticia de clientes basado en árbol de clasificación y regresión lineal incrementa la precisión del pronóstico.

3.3 Dimensión 2: Error del pronóstico del Sistema

3.3.1 Indicador 2: Reducción de error del pronóstico del sistema respecto al modelo: Credit scoring

El puntaje obtenido en el PreTest y PosTest añadimos diferencia.

	ERRORPOST	ERRORPRE	DIFERENCIA	var
1	7,00	17,00	10,00	
2	8,00	18,00	10,00	
3	4,00	21,00	17,00	
4	8,00	26,00	18,00	
5	5,00	37,00	31,00	
6	2,00	21,00	19,00	
7	10,00	32,00	22,00	
8	5,00	30,00	25,00	
9	3,00	39,00	36,00	
10	5,00	32,00	27,00	
11	4,00	22,00	17,00	
12	7,00	18,00	11,00	
13	4,00	19,00	15,00	
14	7,00	17,00	10,00	
15	2,00	16,00	14,00	
16	2,00	26,00	25,00	
17	7,00	34,00	27,00	
18	3,00	30,00	27,00	
19	7,00	29,00	22,00	
20	6,00	18,00	11,00	
21	1,00	16,00	14,00	
--				

Figura Nº 4: puntajes obtenidos Pre Test, Post Test y Diferencia – Error del Pronostico del Sistema

Cálculo de Datos Descriptivos

Tabla 9: Calculo de datos descriptivos - Error de Pronostico del Sistema

Estadísticos descriptivos						
	N	Mínimo	Máximo	Media	Desviación estándar	Varianza
ERRORPOST	650	1,00	10,00	5,7569	2,60543	6,788
ERRORPRE	650	14,00	44,00	28,3646	8,48515	71,998
DIFERENCIA	650	5,00	41,00	22,6277	8,74408	76,459
N válido (por lista)	650					

Tabla de Frecuencia

Tabla 10: Tabla de Frecuencia Pre Test - Error del Pronóstico del Sistema

ERRORPRE					
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	14,00	3	,5	,5	,5
	15,00	26	4,0	4,0	4,5
	16,00	32	4,9	4,9	9,4
	17,00	27	4,2	4,2	13,5
	18,00	26	4,0	4,0	17,5
	19,00	20	3,1	3,1	20,6
	20,00	27	4,2	4,2	24,8
	21,00	19	2,9	2,9	27,7
	22,00	20	3,1	3,1	30,8
	23,00	21	3,2	3,2	34,0
	24,00	28	4,3	4,3	38,3
	25,00	10	1,5	1,5	39,8
	26,00	30	4,6	4,6	44,5
	27,00	19	2,9	2,9	47,4
	28,00	14	2,2	2,2	49,5
	29,00	24	3,7	3,7	53,2
	30,00	18	2,8	2,8	56,0
	31,00	26	4,0	4,0	60,0
	32,00	17	2,6	2,6	62,6
	33,00	23	3,5	3,5	66,2
	34,00	31	4,8	4,8	70,9

35,00	24	3,7	3,7	74,6
36,00	17	2,6	2,6	77,2
37,00	27	4,2	4,2	81,4
38,00	19	2,9	2,9	84,3
39,00	28	4,3	4,3	88,6
40,00	31	4,8	4,8	93,4
41,00	12	1,8	1,8	95,2
42,00	14	2,2	2,2	97,4
43,00	15	2,3	2,3	99,7
44,00	2	,3	,3	100,0
Total	650	100,0	100,0	

Tabla 11 : Tabla de Frecuencia Post Test - Error del Pronostico del Sistema

ERRORPOST

	Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido 1,00	30	4,6	4,6	4,6
2,00	61	9,4	9,4	14,0
3,00	71	10,9	10,9	24,9
4,00	57	8,8	8,8	33,7
5,00	71	10,9	10,9	44,6
6,00	86	13,2	13,2	57,8
7,00	83	12,8	12,8	70,6
8,00	66	10,2	10,2	80,8
9,00	81	12,5	12,5	93,2
10,00	44	6,8	6,8	100,0
Total	650	100,0	100,0	

HISTOGRAMA

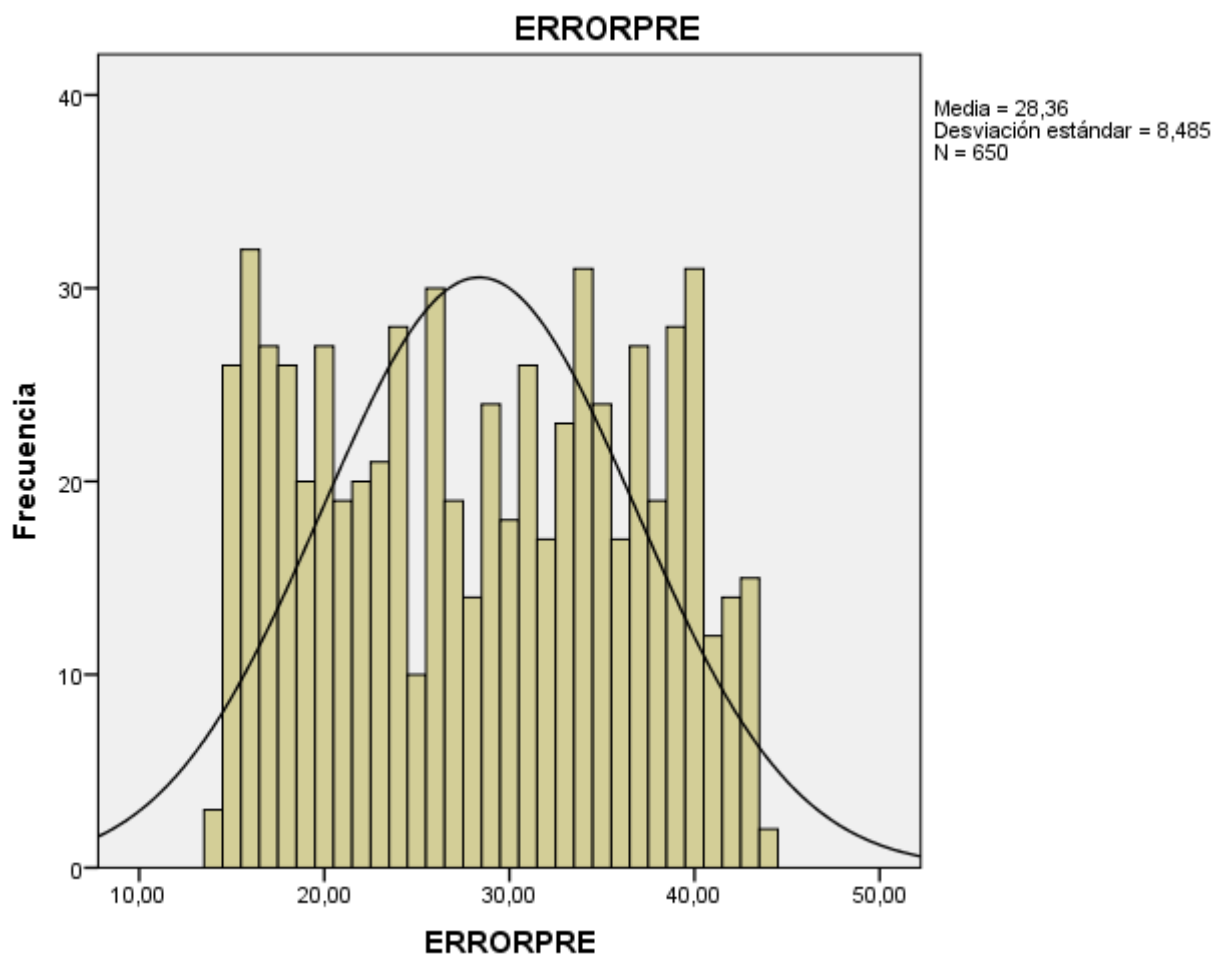


Figura N° 5: Grafico de puntajes obtenidos en el Pre Test - Error del Pronostico del Sistema

Resultado de análisis de PreTest en el gráfico, representamos el histograma obtenido de los puntajes del indicador Error del Pronostico del Sistema, obtuvimos una media de 28.36, con una desviación estándar de 8.485

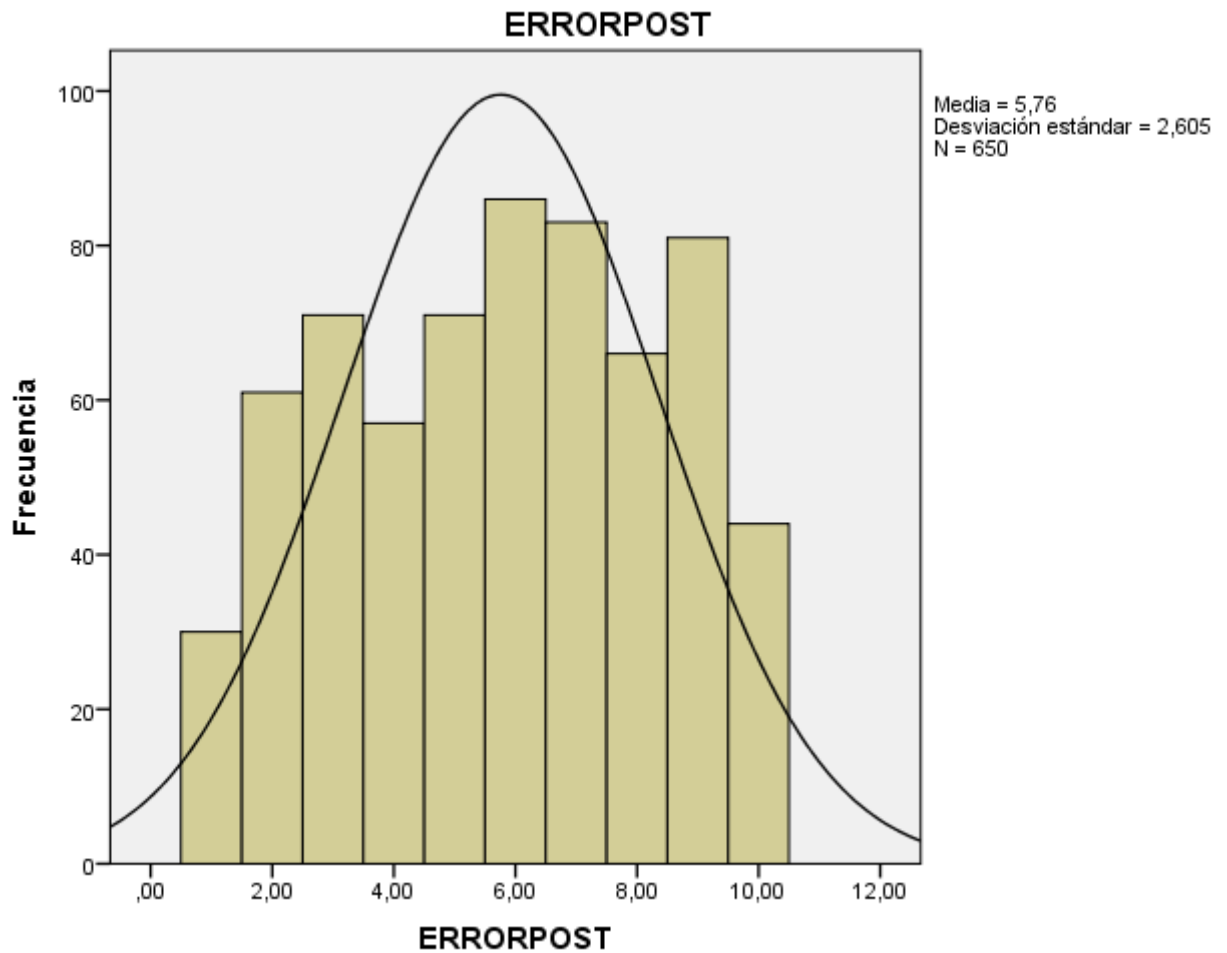


Figura Nº 6: Grafico de puntajes obtenidos en el Post Test - Error

Resultado de análisis del PosTest en el gráfico, representamos el histograma obtenido de los puntajes del indicador Error del Pronostico del Sistema, obtuvimos una media de 5.76, con una desviación estándar de 2.605

Prueba de normalidad

Determinamos que la distribución es normal o no, utilizamos la prueba de Kolmorov-Smirnov en la muestra.

Tabla 12: Prueba de Kolmorov- Error del Pronostico del Sistema

Prueba de Kolmogorov-Smirnov para una muestra

		ERRORPOST	ERRORPRE	DIFERENCIA
N		650	650	650
Parámetros normales	Media	5,7569	28,3646	22,6277
	Desviación estándar	2,60543	8,48515	8,74408
Máximas diferencias extremas	Absoluta	,105	,086	,082
	Positivo	,104	,086	,082
	Negativo	-,105	-,085	-,078
Estadístico de prueba		,105	,086	,082
Sig. asintótica (bilateral)		,000 ^c	,000 ^c	,038 ^c

Observamos en la tabla, el valor de Sig en la diferencia siendo mayor a 0.05, en tanto se puede afirmar en el indicador Error del Pronostico del Sistema, sigue en una distribución normal.

Prueba de hipótesis

Siendo distribución normal, aplicamos la prueba Estadística comparación de medias. Se aplica la prueba de muestra relacionadas de T student. La H0 y HA:

- Hipótesis nula (H0): El sistema de pronóstico de calificación crediticia de clientes basado en árbol de clasificación y regresión lineal no reduce el error del pronóstico.
- Hipótesis Alternativa (H1): El sistema de pronóstico de calificación crediticia de clientes basado en árbol de clasificación y regresión lineal reduce el error del pronóstico.

Tabla 13: Aplicación de la muestra relacionadas T Student – Error del Pronostico del Sistema

		Media	N	Desviación estándar	Media de error estándar
Par 1	ERRORPOST	5,7569	650	2,60543	,10219
	ERRORPRE	28,3646	650	8,48515	,33282

	Diferencias emparejadas					t	gl	Sig. (bilateral)
	Media	Desviación estándar	Media de error estándar	95% de intervalo de confianza de la diferencia				
				Inferior	Superior			
Par 1	ERRORPOST – ERRORPRE	-22,60769	8,73960	,34280	-23,28081	-21,93457	- 65,951	649 ,000

El valor de Sig. Es menor a 0.05, se rechazó H0 y aceptamos la HA: El sistema de pronóstico de calificación crediticia de clientes basado en árbol de clasificación y regresión lineal reduce el error del pronóstico.

3.4 Dimensión 3: Tiempo de Respuesta del Sistema

3.4.1 Indicador 3: Reducción del tiempo de respuesta del sistema respecto al modelo: Credit scoring

El puntaje obtenido del PreTest y PosTest, con la diferencia.

	TIEMPOPOST T	TIEMPOPRE	DIFERENCIA	var
1	2,00	6,00	4,00	
2	4,00	5,00	1,00	
3	3,00	5,00	2,00	
4	2,00	8,00	6,00	
5	4,00	6,00	2,00	
6	3,00	5,00	2,00	
7	3,00	7,00	4,00	
8	3,00	7,00	4,00	
9	4,00	8,00	4,00	
10	2,00	7,00	5,00	
11	3,00	8,00	5,00	
12	2,00	8,00	6,00	
13	3,00	5,00	2,00	
14	4,00	7,00	3,00	
15	2,00	8,00	6,00	
16	2,00	7,00	5,00	
17	3,00	7,00	4,00	
18	3,00	6,00	3,00	
19	3,00	8,00	5,00	
20	3,00	8,00	5,00	
21	2,00	5,00	3,00	

Figura N° 7: Puntajes obtenidos Pre Test, Post Test y Diferencia – Tiempos de Respuesta del Sistema

Cálculo de Datos Descriptivos

Tabla 14: Calculo Estadístico descriptivo – Tiempos de Respuesta del Sistema

Estadísticos descriptivos						
	N	Mínimo	Máximo	Media	Desviación estándar	Varianza
TIEMPOPOST	650	2,00	4,00	2,5677	,53177	,283
TIEMPOPRE	650	4,00	14,00	7,0154	1,69648	2,878
DIFERENCIA	650	1,00	12,00	4,4477	1,80426	3,255
N válido (por lista)	650					

Tabla de Frecuencia

Tabla 15: Tabla de frecuencia Pre Test- Tiempo

TIEMPOPRE					
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	4,00	21	3,2	3,2	3,2
	5,00	101	15,5	15,5	18,8
	6,00	125	19,2	19,2	38,0
	7,00	175	26,9	26,9	64,9
	8,00	165	25,4	25,4	90,3
	9,00	9	1,4	1,4	91,7
	10,00	26	4,0	4,0	95,7
	11,00	10	1,5	1,5	97,2
	12,00	12	1,8	1,8	99,1
	13,00	3	,5	,5	99,5
	14,00	3	,5	,5	100,0
	Total	650	100,0	100,0	

Tabla 16: Tabla de frecuencias Post -Test – Tiempo

TIEMPOPOST					
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	2,00	293	45,1	45,1	45,1
	3,00	345	53,1	53,1	98,2
	4,00	12	1,8	1,8	100,0
	Total	650	100,0	100,0	

Histograma

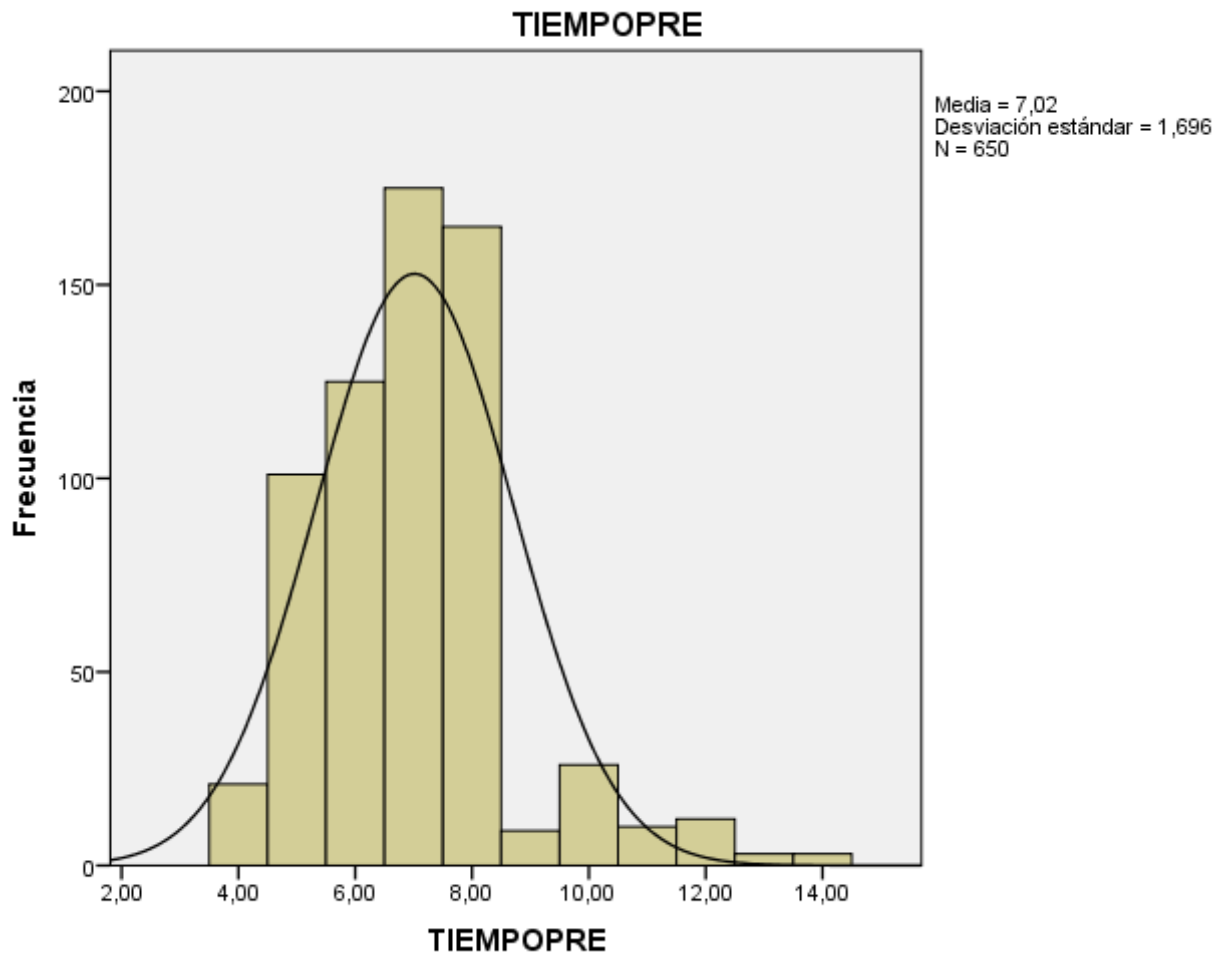


Figura Nº 8: Gráfico de puntajes obtenidos en el Pre Test- Tiempo de Respuesta del Sistema

Resultado del análisis del PreTest en el gráfico, representamos un histograma obtenido de los puntajes del indicador Tiempo de Respuesta del Sistema, obtuvimos una media de 7.02, con desviación estándar de 1.696

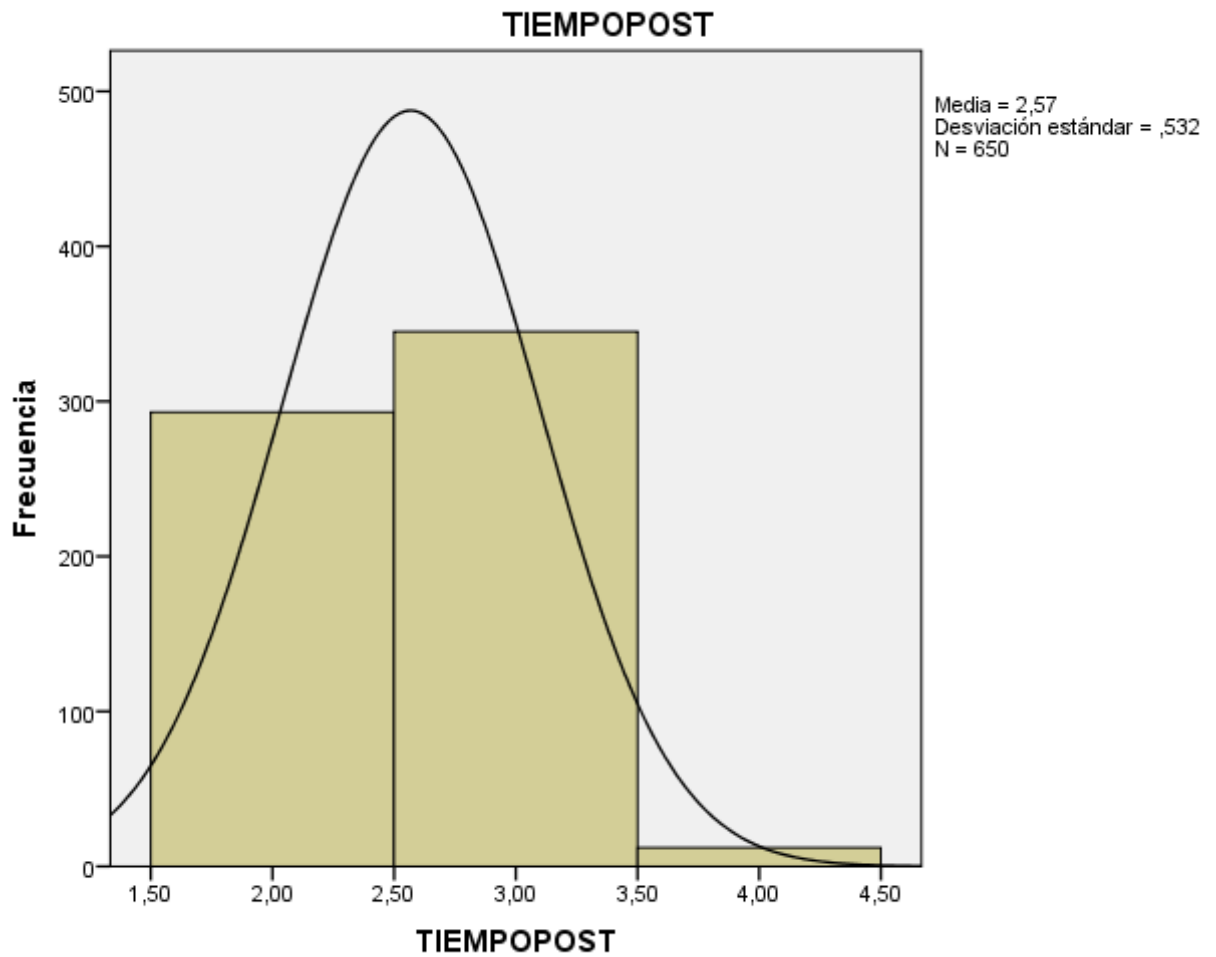


Figura Nº 9: Gráfico de puntajes obtenidos en el Post Test- Tiempo de Respuesta del Sistema

Resultado de análisis del PosTest en el gráfico, representamos el histograma obtenido de los puntajes del indicador Tiempo de Respuesta del Sistema, obtuvimos una media de 2.57, con desviación estándar de ,532

Prueba de Normalidad

Determinamos que la distribución es normal o no, utilizamos la prueba de Kolmorov-Smirnov en la muestra.

Tabla 17: Prueba de Kolmogorov - Tiempo de Respuesta del Sistema

		TIEMPOPRE	TIEMPOPOST	DIFERENCIA
N		650	650	650
Parámetros normales ^{a,b}	Media	7,0154	2,5677	4,4477
	Desviación estándar	1,69648	,53177	1,80426
Máximas diferencias extremas	Absoluta	,184	,341	,160
	Positivo	,184	,308	,160
	Negativo	-,116	-,341	-,096
Estadístico de prueba		,184	,341	,160
Sig. asintótica (bilateral)		,000 ^c	,000 ^c	,000 ^c

La Sig. En la columna diferencia > 0.05 , en tanto se puede afirmar en el indicador Tiempo de respuesta del sistema sigue en distribución no normal.

Prueba de hipótesis

Siendo la distribución de muestra no es normal, aplicamos la prueba Estadísticas no paramétricas, la cual es prueba de Rangos de Wisconsin. La H0 y HA son:

- Hipótesis nula (H0): El sistema de pronóstico de calificación crediticia de clientes basado en árbol de clasificación y regresión lineal no reduce el tiempo de respuesta del pronóstico.
- Hipótesis Alterna (H1): El sistema de pronóstico de calificación crediticia de clientes basado en árbol de clasificación y regresión lineal reduce el tiempo de respuesta del pronóstico.

Tabla 18: Prueba de Hipótesis – Tiempo

Rangos		N	Rango promedio	Suma de rangos
TIEMPOPRES –	Rangos negativos	0 ^a	,00	,00
TIEMPOPOST	Rangos positivos	650 ^b	325,50	211575,00
	Empates	0 ^c		
	Total	650		

- a. TIEMPOPRES < TIEMPOPOST
- b. TIEMPOPRES > TIEMPOPOST
- c. TIEMPOPRES = TIEMPOPOST

Tabla 19: Estadística de Prueba Wilcoxon

Estadísticos de prueba	
	TIEMPOPRES – TIEMPOPOST
Z	-22,188 ^b
Sig. asintótica (bilateral)	,000

- a. Prueba de rangos con signo de Wilcoxon
- b. Se basa en rangos negativos.

El valor de Sig. $0 < \alpha 0.05$, se niega la H_0 y se acepta la H_A : El sistema de pronóstico de calificación crediticia de clientes basado en árbol de clasificación y regresión lineal reduce el tiempo de respuesta del pronóstico.

IV. DISCUSIÓN

En consecuencia, se describen a detalle los resultados que obtuvimos en el proyecto al realizar el análisis y cuando comparamos los comportamientos de la media del indicador del aumento la precisión del pronóstico, reducción del error del pronóstico y reducción, tales como después del Sistema de Pronóstico de la Calificación Crediticia de Clientes Basado en Árbol de clasificación y Regresión lineal.

Partiendo desde la HE 1 que hemos planteado para posteriormente evidenciar el cálculo necesario, se determinó que la media de Precisión del pronóstico con el Credit scoring fue de 71.04 (que representara el 82.60% un puntaje máximo) y con el sistema de pronóstico dio como resultado 94.38 (representara el 95.33% un puntaje máximo). Entonces partiendo de los resultados s podemos afirmar que hay un aumento en 12.72% mediante los puntajes. Mediante la prueba de hipótesis realizada se niega la H₀, concluimos que el Sistema de Pronóstico de la Calificación Crediticia de Clientes Basado en Árbol de clasificación y Regresión lineal incrementa la precisión del pronóstico respecto al modelo: Credit scoring.

Partiendo desde la HE 2 que hemos planteado para posteriormente evidenciar el cálculo necesario se determinó que la media de Error del pronóstico del sistema con el Credit scoring fue de 28.36 (que representa el 64.45% del puntaje máximo) y con el sistema de pronóstico dio como resultado 5.75 (que representara el 57.5% un puntaje máximo). Entonces partiendo de los resultados podemos afirmar que hay una reducción de 6.95% mediante los puntajes. Mediante la prueba de hipótesis realizada se niega la H₀, concluimos que el Sistema de Pronóstico de la Calificación Crediticia de Clientes Basado en Árbol de clasificación y Regresión lineal reduce el error del pronóstico del sistema respecto al modelo: Credit scoring.

Partiendo desde la HE 3 que hemos planteado para posteriormente evidenciar el cálculo necesario, se determinó que la media de Tiempos de respuesta del sistema con el Credit scoring fue de 7.0154 (que representara el 50.11% un puntaje máximo) y con el sistema de pronóstico dio como resultado 2.5677 (que representara el 64.1925% un puntaje máximo). Entonces partiendo de los resultados podemos afirmar que hay una reducción de 14.0825% mediante los puntajes de ambos. Mediante la prueba de hipótesis realizada se niega la H₀, concluimos que el Sistema de Pronóstico de la Calificación Crediticia de Clientes Basado en Árbol de clasificación y Regresión lineal reduce el tiempo de respuesta del sistema respecto al modelo: Credit scoring.

Este incremento en la precisión del pronóstico, reducción del error y reducción del tiempo de respuesta reflejan el impacto del Sistema de pronóstico de calificación crediticia de clientes basado en árbol de clasificación y regresión lineal. De la misma forma, Reza (2018) en un artículo propuso un método híbrido utilizando el algoritmo CART y la optimización de enjambre de partículas binarias. Los indicadores de rendimiento que se utilizan en esta investigación son la precisión de la clasificación, la tasa de error, la sensibilidad, la especificidad y la precisión. Los resultados experimentales basados en el conjunto de datos públicos mostraron que la precisión del método propuesto es del 78%. Por otro lado, Leal, Arangüiz y Gallegos (2018) en el artículo se propuso a la empresa un modelo de evaluación crediticia ajustado y ponderado, así como a sus clientes, que permitió reducir el temor de crédito o incobrables. Con respecto a la evaluación crediticia que ya esta propuesta, los resultados que se obtuvieron fueron un 81,82% de los créditos brindados a sus clientes superaron el rango mínimo de evaluar o máximo para aprobar a la entidad. (p. 34).

Por lo tanto, concluimos el Sistema de pronóstico de calificación crediticia de clientes basado en árbol de clasificación y regresión lineal impacta en el incremento de la precisión del pronóstico, reducción del error y reducción del tiempo de respuesta

V. CONCLUSIONES

Las conclusiones de la investigación fueron las siguientes:

1. Con los resultados obtenidos, podemos concluir que el Sistema de pronóstico de calificación crediticia de clientes basado en árbol de clasificación y regresión lineal tuvo un impacto positivo ya que incrementó la precisión del pronóstico respecto al modelo: Credit scoring, esto lo podemos ver reflejado en la estadística, en donde la media de precisión del pronóstico con el Credit scoring fue de 71.04 y con el Sistema de pronóstico dio como resultado 94.38. Entonces partiendo de los resultados se puede afirmar que existe un incremento de 12.72% de la precisión del pronóstico respecto al modelo: Credit scoring.
2. También podemos concluir que el Sistema de pronóstico de calificación crediticia de clientes basado en árbol de clasificación y regresión lineal redujo el error del pronóstico del sistema respecto al modelo: Credit scoring, esto lo podemos ver reflejado en la estadística, en donde la media de error del pronóstico con el Credit scoring fue de 28.36 y con el Sistema de pronóstico dio como resultado 5.75. Entonces partiendo de los resultados se puede afirmar que existe una reducción de 6.95% del error del pronóstico respecto al modelo: Credit scoring.
3. Por último, podemos concluir que el Sistema de pronóstico de calificación crediticia de clientes basado en árbol de clasificación y regresión lineal redujo el tiempo de respuesta del sistema respecto al modelo: Credit scoring, esto lo podemos ver reflejado en la estadística, en donde la media de tiempo de respuesta del sistema con el Credit scoring fue de 7.0154 y con el Sistema de pronóstico dio como resultado 2.5677. Entonces partiendo de los resultados se puede afirmar que existe una reducción de 14.0825% del tiempo de respuesta del sistema respecto al modelo: Credit scoring.

En consecuencia, se concluye que el Sistema de pronóstico de calificación crediticia de clientes basado en árbol de clasificación y regresión lineal tiene un impacto significativo ya que incrementa la precisión del pronóstico, reduce el error y reduce el tiempo de respuesta.

VI. RECOMENDACIONES

Las recomendaciones para futuras investigaciones son las siguientes:

1. Se recomienda la definición de la estrategia de predicción basándose de los resultados ya obtenidos del análisis de predicción de datos a mayor escala, hacer promoción de servicios en relación por sucursal y estrategias relacionadas al ofrecimiento de servicios adicionales en base a los patrones de comportamiento que devuelva el modelo utilizado en el sistema de pronóstico.
2. Se recomienda tener una constante actualización y realizar mantenimiento, ya que es punto clave para la confiabilidad y generación del resultado que deseamos obtener.
3. Es recomendable que todo el personal de la empresa tenga capacitaciones continuas, de esta manera el personal estará apto para que lleve un buen funcionamiento del sistema de una forma rápida y eficaz, también estará familiarizado con el sistema que se ha implementado.
4. Incremento en las cantidades de la variable en el impacto de la medición para el sistema el cual será: costos, tiempos, cantidad de transacciones, y muchas más que enriquezcan más el resultado de la investigación.
5. Se recomienda implementar esta tecnología en varios procesos de la empresa a fin de determinar el impacto que tendría en el servicio que brinda.

VII. REFERENCIAS

- Armstrong, J. (2001). *Evaluating forecasting methods*. University of Pennsylvania. Recuperado de https://repository.upenn.edu/cgi/viewcontent.cgi?article=1182&context=marketing_papers
- Baesens, B., Van Gestel, T., Viaene, S., Stepanova, M., Suykens, J., & Vanthienen, J. (2003). *Benchmarking state-of-the-art classification*. *Journal of the Operational Research Society*, 627-635. Doi: 10.1057 / palgrave. jors.2601545
- Banco Central de Reserva del Perú. (2006). <http://www.bcrp.gob.pe/docs/Publicaciones>. Recuperado de <http://www.bcrp.gob.pe/docs/Publicaciones/Documentos-de-Trabajo/2006/Documento-Trabajo-04-2006.pdf>.
- Banco Central de Reserva del Perú. (2019). <http://www.bcrp.gob.pe>. Obtenido de <http://www.bcrp.gob.pe/publicaciones/glosario/r.html>.
- Bertalanffy, L. v. (1989). *TEORÍA GENERAL DE LOS SISTEMAS Fundamentos, desarrollo y aplicaciones*. México: FONDO DE CULTURA ECONÓMICA. Obtenido de https://cienciasyparadigmas.files.wordpress.com/2012/06/teoria-general-de-los-sistemas-_fundamentos-desarrollo-aplicacionesludwig-von-bertalanffy.pdf.
- Brockwell, P., & Davis, R. (2010). *Introduction to Time Series and Forecasting*. New York. Recuperado de https://www.researchgate.net/profile/Paul_Louangrath/post/Are_there_good_books_resources_on_stochastic_volatility/attachment/59d6387cc49f478072ea56e6/AS%3A273701147217926%401442266846239/download/2002-Brockwell-Introduction+Time+Series+and+Forecasting.pdf
- Chitra, K., & Subashini, B. (2013). Sector, Data Mining Techniques and its Applications in Banking. *International Journal of Emerging Technology and Advanced Engineering*. Recuperado de <https://pdfs.semanticscholar.org/87da/9ade5ca2f0547c86e0d29f65ede5f544de90.pdf>
- Cobos, C., Zúñiga, J., Guarín, J., León, E., & Mendoza, M. (2010). CMIN - herramienta case basada en CRISP-DM para el soporte de proyectos de minería de datos. *Ingeniería e Investigación*. Recuperado de

- http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S0120-56092010000300004&lang=es
- Comité de Basilea (1999). Credit Risk Modelling: Current practices and applications. Basle, abril. Recuperado de <https://www.bis.org/publ/bcbs49.pdf>
- Davila Hernández, F., & Sanchez Corrales, Y. (2012). Técnicas de minería de datos aplicadas al diagnóstico de entidades clínicas. Revista Cubana de Informática Médica. Recuperado de http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S1684-18592012000200007&lang=es
- Galán Cortina, Víctor (2016). Aplicar la metodología CRISP-DM al proyecto de minería de datos en el entorno universitario.
- Guillen Uyen, J. (2002). Morosidad crediticia y tamaño: Un análisis de la crisis bancaria peruana. Recuperado de <http://www.bcrp.gob.pe/docs/Publicaciones/Documentos-de-Trabajo/2001/Documento-Trabajo-05-2001.pdf>
- Hand, D., & Henley, W. (1997). Statistical Classification Methods in Consumer Credit Scoring: a Review. Journal of the Royal Statistical Society Series A, 160(3), 160. Retrieved from https://econpapers.repec.org/article/blajorssa/v_3a160_3ay_3a1997_3ai_3a3_3ap_3a523-541.htm
- Hernández Corrales, L., Meneses Ceron, L., & Benavides, J. (2005). DESARROLLO DE UNA METODOLOGÍA PROPIA DE ANÁLISIS DE CRÉDITO EMPRESARIAL EN UNA ENTIDAD FINANCIERA. Estudios Gerenciales, 21, 129-165. Recuperado de http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S0123-59232005000400007&lang=es
- Karminsky, A. y Khromova, E. (2016). Extended modelling of banks credit ratings. Retrieved from <https://core.ac.uk/download/pdf/82374668.pdf>.
- Leal, A. Arangüiz, M. y Gallegos, J. (2018). Análisis de riesgo crediticio, propuesta del modelo credit scoring. Revista de la Facultad de Ciencias Económicas. Vol. XXVI. Recuperado de <https://www.redalyc.org/pdf/909/90953767010.pdf>.
- Leal Fica, A., Aranguis Casanova, M., & Gallegos Mardones, J. (2018).

ANÁLISIS DE RIESGO CREDITICIO, PROPUESTA DEL MODELO CREDIT SCORING. Revista Facultad de Ciencias Económicas: Investigación y Reflexión, 26. Recuperado de http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S0121-68052018000100181&lang=es

Liu, Y., & Schulman, M. (2005). Data mining feature selection for credit scoring models. Journal of the Operational Research Society. Doi: 10.1057 / Palgrave. jors.2601976

Macias Rivero, Y., Guzmán Sanchez, M., & Martínez Suarez, Y. (2009). Modelo de evaluación para software que emplean indicadores métricos en la vigilancia científico-tecnológica. ACIMED. Recuperado de http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S1024-94352009001200003

Mariano Rodriguez, J., Frias C, S., Souquet C., G., & Marino Rodriguez, R. (2002). Dialnet. Recuperado de <https://dialnet.unirioja.es/descarga/articulo/4004764.pdf>

Mckeonw, D. (2000). Modelos Cuantitativos Para Administracion. (D. d. santos, Ed.) Grupo Editorial Iberoamerica. Recuperado de <https://es.scribd.com/document/244053395/Davis-McKeown-Capitulo-1-pdf>

Medina, R. P., & Selva, M. L. (2013). Análisis del credit scoring. Revista de Administração de Empresas. Recuperado de http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0034-75902013000300007&lang=es

MINISTERIO DE ECONOMIA Y FINANZAS DEL PERU. (2019). Clasificadores presupuestarios. Recuperado de https://www.mef.gob.pe:https://www.mef.gob.pe/contenidos/presu_public/anexos/Anexo_2_clasificador_Gastos_RD003_2019EF5001.pdf

Ochoa P., J., Galeano M., W., & Agudelo V., L. (2010). Construcción de un modelo de scoring para el otorgamiento de crédito en una entidad financiera. Perfil de Coyuntura Económica, 16. Recuperado de http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S1657-42142010000200010&lang=es

- Pressman, R. (2010). Ingeniería del software. Un enfoque práctico (séptima ed.). México: Mc Graw Hill. Recuperado de <http://cotana.informatica.edu.bo/downloads/Id-Ingenieria.de.software.enfoque.practico.7ed.Pressman.PDF>
- Pretel Ruiz, N. (2014). Repositorio Universidad Nacional de Trujillo. Recuperado de <http://dspace.unitru.edu.pe/handle/UNITRU/617>
- Rayo Canton, S., Lara Rubio, J., & Camino Blasco, D. (2010). Un Modelo de Credit Scoring para instituciones de microfinanzas en el marco de Basilea II. Journal of Economics, Finance and Administrative Scienc. Recuperado de http://www.scielo.org.pe/scielo.php?script=sci_arttext&pid=S2077-18862010000100005&lang=es
- Reza Firsandaya Malik, H. (2018). Credit Scoring Using Classification and Regression Tree (CART) Algorithm and Binary Particle Swarm Optimization. International Journal of Electrical and Computer Engineering (IJECE). doi:10.11591/ijece.v8i6.pp5425-5431
- Rocca Espinoza, E., Garcia Perz de Lema, D., & Gómez Guillamon, A. (2018). Factores determinantes para la concesión de crédito por parte de las entidades financieras a las Mipymes. Tec Empresarial, 19-30. Recuperado de https://www.scielo.sa.cr/scielo.php?script=sci_arttext&pid=S1659-33592018000100019&lang=es
- Rogers, P., Rogers, D., & Securato, J. (2015). ABOUT PSYCHOLOGICAL VARIABLES IN APPLICATION SCORING MODELS. Revista de Administração de Empresas, 55. Retrieved from http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0034-75902015000100038&lang=es
- SAS. (2019). <http://support.sas.com>. Recuperado de [http://support.sas.com: http://support.sas.com/documentation/cdl/en/emcs/66392/HTML/default/viewer.htm#n0pejm83csbj4n1xueveo2uoujy.htm](http://support.sas.com/documentation/cdl/en/emcs/66392/HTML/default/viewer.htm#n0pejm83csbj4n1xueveo2uoujy.htm)
- Salcedo Parra, O., Galena, R., & Rodriguez, L. (2010). Metodología crisp para la implementación Data Warehouse. Tecnura vol.14 no.26. Recuperado de http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S0123-921X2010000100005&lang=es
- Sanchez Martorelli, J. R. (2013). Indicadores de Gestion Empresarial. Estados

- Unidos de Norteamérica: Editorial Reviews. Recuperado de <https://es.scribd.com/read/387424265/Indicadores-De-Gestion-Empresarial-De-La-Estrategia-a-Los-Resultados>
- Serna M., E., & Arango I., F. (2012). Prueba del software: más que una fase en el ciclo de vida. Revista de Ingeniería. Recuperado de <http://www.scielo.org.co/pdf/ring/n35/n35a06.pdf>
- The Journal of Data Warehousing. (2000). the CRISP-DM Model: The New Blueprint for Data Mining. The Journal of Data Warehousing. Recuperado de <https://mineracaodedatos.files.wordpress.com/2012/04/the-crisp-dm-model-the-new-blueprint-for-data-mining-shearer-colin.pdf>
- Usama, F., Shapiro, P., & Padhraic, S. (1997). From Data Mining to Knowledge Discovery in Databases. AI Magazine, 17(3). Recuperado de <https://www.aaai.org/ojs/index.php/aimagazine/article/view/1230/1131>
- Vargas Sanchez, A., & Mostajo Castelu, S. (2014). MEDICIÓN DEL RIESGO CREDITICIO MEDIANTE LA APLICACIÓN DE MÉTODOS BASADOS EN CALIFICACIONES INTERNAS. Investigación & Desarrollo, Recuperado de http://www.scielo.org.bo/scielo.php?script=sci_arttext&pid=S2518-44312014000200002&lang=es#_ftnref2
- Wahrenburg, M. y Frerichs, H. (2003). Evaluating internal credit systems depending on bank size. Retrieved from https://www.researchgate.net/publication/5057245_Evaluating_internal_credit_rating_systems_dependent_on_bank_size
- Vizcaino, Paula Andrea (2008). Aplicar técnicas de inducción de árboles de decisión a problemas de clasificación mediante el uso de WEKA. Facultad de ingeniería de sistemas, Bogotá,

ANEXOS

Anexo 1 Matriz de consistencia

Tabla 20. Matriz de consistencia

PROBLEMAS	OBJETIVOS	HIPÓTESIS	VARIABLE	DIMENSIONES	INDICADORES
GENERAL	GENERAL	GENERAL			
¿Cuál es el impacto de un sistema de pronóstico de calificación crediticia de clientes basado en árbol de clasificación y regresión lineal?	Determinar el impacto del sistema de pronóstico de la calificación crediticia de clientes basado en árbol de clasificación y regresión lineal.	El sistema de pronóstico de calificación crediticia de clientes basado en árbol de clasificación y regresión lineal incrementa la precisión del pronóstico, reduce el error y reduce el tiempo de respuesta	Impacto del sistema de pronóstico de la calificación crediticia basado en árbol de clasificación y regresión lineal Brockwell & Davis, (2010). Bertalanffy, (1989).	X.1. Precisión del pronóstico (Armstrong, 2001)	X1.1. Incremento de la precisión del pronóstico respecto al modelo: <i>Credit scoring</i> Hand & Henley, (1997).
ESPECÍFICOS	ESPECÍFICOS	ESPECÍFICOS			
¿Cuál es el impacto de un sistema de pronóstico de calificación crediticia de clientes basado en árbol de clasificación y regresión lineal en la precisión del pronóstico?	Determinar el impacto del sistema de pronóstico de la calificación crediticia de clientes basado en árbol de clasificación y regresión lineal en la precisión del pronóstico.	El sistema de pronóstico de calificación crediticia de clientes basado en árbol de clasificación y regresión lineal incrementa la precisión del pronóstico. Armstrong, J. (2001).		X.2. Error del pronóstico del sistema	X2.1 Reducción de error del pronóstico del sistema respecto al modelo: <i>Credit scoring</i> Hand & Henley, (1997).
¿Cuál es el impacto de un sistema de pronóstico de calificación crediticia de clientes basado en árbol de clasificación y regresión lineal en el error del pronóstico?	Determinar el impacto del sistema de pronóstico de la calificación crediticia de clientes basado en árbol de clasificación y regresión lineal en el error del pronóstico.	El sistema de pronóstico de calificación crediticia de clientes basado en árbol de clasificación y regresión lineal reduce el error del pronóstico. Serna M., E., & Arango I., F. (2012).		X.3. Tiempos de respuesta del sistema	X3.1 Reducción del tiempo de respuesta del sistema respecto al modelo: <i>Credit scoring</i> Medina & Selva, (2013, págs. 303-315).
¿Cuál es el impacto de un sistema de pronóstico de calificación crediticia de clientes basado en árbol de clasificación y regresión lineal en el tiempo de respuesta del pronóstico?	Determinar el impacto del sistema de pronóstico de la calificación crediticia de clientes basado en árbol de clasificación y regresión lineal en el tiempo de respuesta del pronóstico.	El sistema de pronóstico de calificación crediticia de clientes basado en árbol de clasificación y regresión lineal reduce el tiempo de respuesta del pronóstico. (Pressman, (2010, págs. 526-552).			

Anexo 2 Prototipos

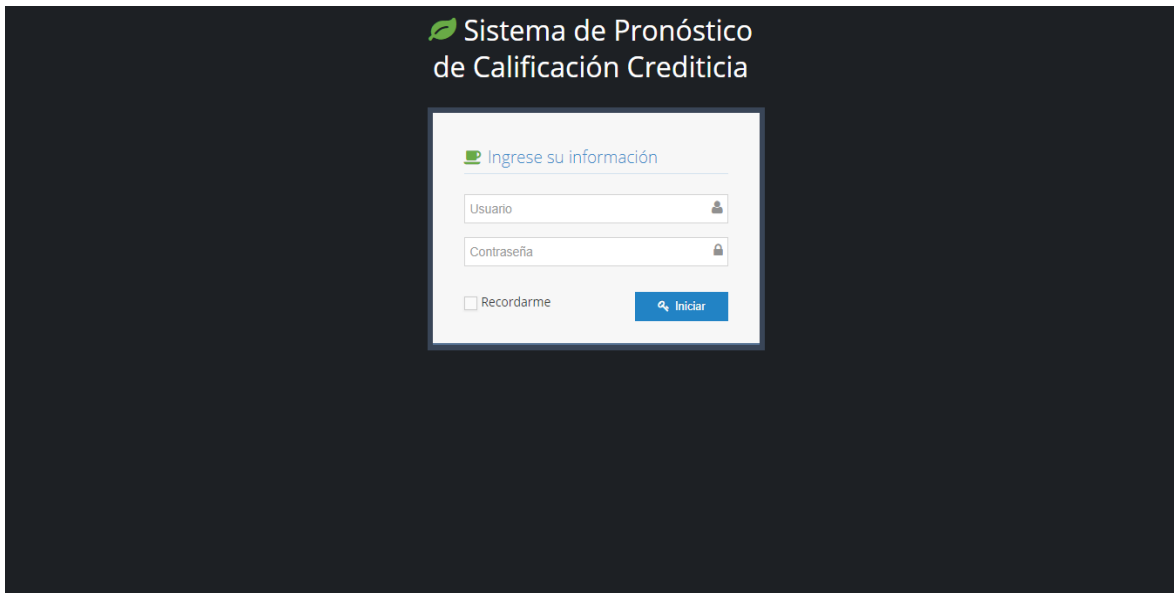


Figura 1. Prototipo de interfaz de ingreso al sistema de pronóstico

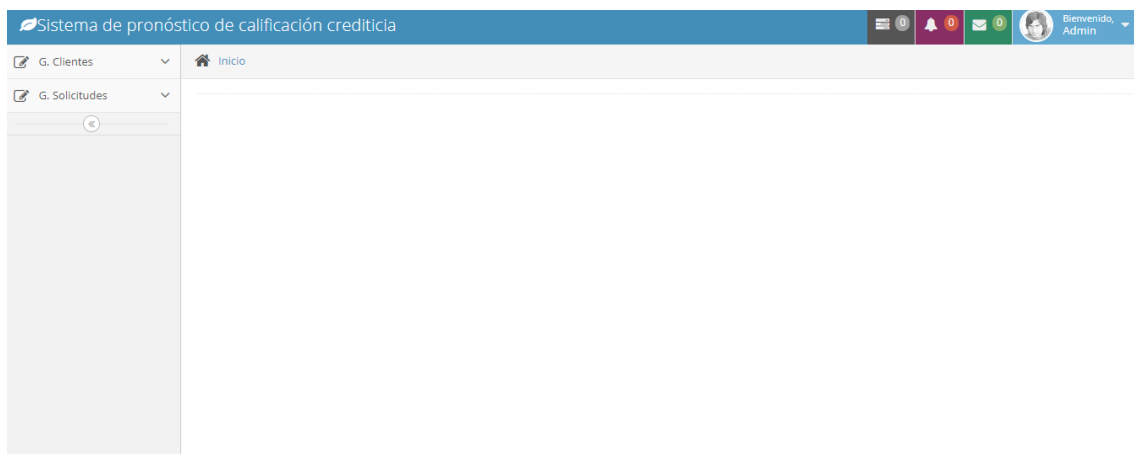


Figura 2. Prototipo de interfaz para seleccionar el ingreso de datos del cliente y datos de la solicitud de crédito

Sistema de pronóstico de calificación crediticia

G. Clientes Inicio

G. Solicitudes LISTA DE CLIENTES » Relación de Personas Jurídicas y Personas Naturales

RUC BUSCAR

Mostrar 10 registros Exportar

















<input type="checkbox"/>	RUC	RAZON SOCIAL o NOMBRES	DIRECCION	TELEFONO	ACCIONES
<input checked="" type="checkbox"/>	10198321937	MARIA CRISTINA FERNANDEZ MENDOZA	REAL NRO. 1786	064236665 / 920203120 / 914067858 / arqmariafernandez372@gmail.com	   
<input checked="" type="checkbox"/>	20559567150	CORPORACION ALDERU PERU SAC	ALTO TRUJILLO MZA. A 2 LOTE 17 DEP. 201 BARRIO 5 A	044573616 / 925224048 // ALDERUPERUSAC@HOTMAIL.COM	   
<input checked="" type="checkbox"/>	20552152515	SEMMAQ SAC	1 MZA. J LOTE 308 URB. SAN PEDRO DE GARAY	975028951 / 975028951 // ADMINISTRACION@SEMMAQ.COM.PE	   
<input checked="" type="checkbox"/>	20487568326	JC EJECUTORES SAC	DANIEL ALCIDES CARRION NRO. 196 INT. 507 CIUDAD CHICLAYO	074228227 / 969775977 / 951478818 / JCEJECUTORES@GMAIL.COM	   

Figura 3. Prototipo de Listado de clientes

Opción buscar un cliente por su número de RUC

RUC

Opción para registrar o actualizar información de un cliente



Opción para registrar o actualizar la información de experiencia del cliente



Opción para eliminar lógicamente a un cliente



Opción para registrar o actualizar los estados financieros de un cliente



Opción para registrar una solicitud de crédito



Validación de datos

DAT.GENERALES | DAT.SUNAT | DIRECCIONES | SOCIOS | REPRESENT. | CLIENTES | PROVEEDORES | VENT.BRUTAS | INF. CREDITICIA | VALIDACION FINAL

Datos Generales

Tipo Persona : JURIDICA
 Numero Ruc: 20559567150
 Perfil Riesgo: BAJO

Persona Juridica

Razón social: CORPORACION ALDERU PERU SAC
 Oficina Registral: ZONA V - SEDE TRUJILLO
 Documento En RR.PP.: FICHA
 Sector Económico: SERVICIOS

Tipo Sociedad: SAC
 Sub Sede Registral: OR TRUJILLO (SEDE CENTRAL)
 N° DE DOCUMENTO EN RRPP : 0011215582

Tipo Crédito: MEDIANA EMPRESA - DEUDAS M
 Monto Máximo De Contratación: 900,000.00
 Calificación: NORMAL
 Calificación SBS: NORMAL

Figura 4. Prototipo para registrar o actualizar los datos generales de un cliente

Validación de datos

DAT.GENERALES | DAT.SUNAT | DIRECCIONES | SOCIOS | REPRESENT. | CLIENTES | PROVEEDORES | VENT.BRUTAS | INF. CREDITICIA | VALIDACION FINAL

Datos Sunat

Estado: ACTIVO
 CIU: 4100 - CONSTRUCCION DE EDIFICIOS

Fecha de Inicio De Actividades: 04/06/2013
 Regimen Tributario: REGIMEN GENERAL

Condición: HABIDO



✓ GUARDAR CAMBIOS

Figura 5. Prototipo para registrar o actualizar los datos de SUNAT de un cliente

Validación de datos

DAT.GENERALES	DAT.SUNAT	DIRECCIONES	SOCIOS	REPRESENT.	CLIENTES	PROVEEDORES	VENT.BRUTAS	INF. CREDITICIA	VALIDACION FINAL
---------------	-----------	-------------	--------	------------	----------	-------------	-------------	-----------------	------------------

Datos de Direccion

Sec.	Tipo Direccion	Departamento	Provincia	Distrito	Tipo de via	Nombre de via	Nº / Mz	Lote	Int. / Dep.	
+ 01	DIRECCION LEGAL	LA LIBERTAD	TRUJILLO	EL PORVENIR	(OTROS)	ALTO TRUJILLO	A 2	17	201	 







[Agregar Direccion](#)

Figura 6. Prototipo para registrar o actualizar los datos de dirección y UBIGEO de un cliente

Validación de datos

DAT.GENERALES	DAT.SUNAT	DIRECCIONES	SOCIOS	REPRESENT.	CLIENTES	PROVEEDORES	VENT.BRUTAS	INF. CREDITICIA	VALIDACION FINAL
---------------	-----------	-------------	--------	------------	----------	-------------	-------------	-----------------	------------------

Socios

Sec.	Tip. Per.	Nombre o Razon social	Num. Doc	% Participación	Ruc	Cod. SBS	Cargo	
1	NATURAL	WILFREDO ESCOBEDO EUSTAQUIO	19053173	50		60698309	NO OSTENTE OTRO	 
2	NATURAL	DIANA MARISOL ESCOBEDO ESPINOLA	70268632	40		162672703	GERENTE	 
3	NATURAL	DANIEL DE LA CRUZ RUIZ	46859581	10		151164943	NO OSTENTE OTRO	 

[Agregar Socio](#)



[✓ GUARDAR CAMBIOS](#)

Figura 7. Prototipo para registrar o actualizar los datos de los socios de un cliente

Validación de datos

DAT.GENERALES DAT.SUNAT DIRECCIONES SOCIOS REPRESENT. CLIENTES PROVEEDORES VENT.BRUTAS INF. CREDITICIA VALIDACION FINAL

Representantes

Sec.	Nombre	Tip. Doc	Num. Doc	Num. Ruc	Esta. Civil	Cargo	
+ 1	DIANA MARISOL ESCOBEDO ESPINOLA	DNI	70268632	10702686321	SOLTERO	GERENTE GENERAL	 

[Agregar Representante](#)

Figura 8. Prototipo para registrar o actualizar los datos del representante legal de un cliente

Validación de datos

DAT.GENERALES DAT.SUNAT DIRECCIONES SOCIOS REPRESENT. CLIENTES PROVEEDORES VENT.BRUTAS INF. CREDITICIA VALIDACION FINAL

Información Crediticia

Fecha	Institucion	Calif.	Cred. Directo	Cred. Indirecto
31/05/2019	CAJA MUNICIPAL DE AHORRO Y CREDITO DE SULLANA S A	0	0	14013

Fecha de Consulta:

Deuda Coactiva:

Monto De Protestos:

[CARGAR SUSTENTO DE NO ADEUDO](#) [CARGAR SUSTENTO DE PROTESTOS](#)

[✓ GUARDAR CAMBIOS](#)

Figura 9. Prototipo para verificar o levantar deudas de un cliente

Validación de datos

DAT.GENERALES DAT.SUNAT DIRECCIONES SOCIOS REPRESENT. CLIENTES PROVEEDORES VENT.BRUTAS INF. CREDITICIA **VALIDACION FINAL**

Validacion Final Datos Cliente

Yo confirmo la veracidad y consistencia de esta información

✓ VALIDAR DATOS CLIENTE

Figura 10. Prototipo para validación final de los datos de un cliente

Tipo de EEFF: Año base: Mes base: Fecha de EEFF:

BALANCE	31/03/2019
ACTIVO	
ACTIVO CORRIENTE	
DISPONIBLE	
CAJA Y BANCOS (EFECTIVO)	0.00
TOTAL DISPONIBLE	
EXIGIBLE	
CLIENTES	0.00
LETRAS POR COBRAR	0.00
CUENTAS POR COBRAR DIVERSAS	0.00
CUENTAS POR COBRAR ACCIONISTAS	0.00
CARGAS DIFERIDAS	0.00
GASTOS PAGADOS POR ANTICIPADO	0.00
PROVISION DE COBRANZAS DUDOSAS	0.00
TRIBUTOS POR PAGAR(CREDITO FISCAL)	0.00
OTRAS CUENTAS DEL ACTIVO	0.00

VALIDAR **GUARDAR** **VOLVER**

Figura 11. Prototipo para registrar o actualizar los estados financieros de un cliente

Experiencia en obras

CONTRATOS EJECUTADOS CONTRATOS EN EJECUCION EXPERIENCIA DEL PERSONAL

Principales Contratos Ejecutados En Los Últimos 5 Años

Sec.	Entidad	Objetivo	Moneda	Monto	Plazo días	Ampliación Del Plazo	Fec. Inicio	Fec. Recep. Final	Consortiado
No hay obras ejecutadas									

Agregar Experiencia en obras

Figura 12. Prototipo para registrar o actualizar los datos de experiencia de un cliente

Datos la solicitud de crédito

Datos de la Solicitud

Tipo de Crédito:

Tipo de Garantía:

Plazo en días:

Moneda:

Monto del Crédito:

Porcentaje de la cobertura: %

Porcentaje Garantía:

Fecha de entrega solicitada:

Descripción:

Datos del proyecto

Actividad económica del proyecto:

Departamento:

Provincia:

Distrito:

Fecha de Inicio del Proyecto:

Fecha de Fin del Proyecto:

Tipo de Servicio:

Guardar Limpiar Imprimir

Figura 13. Prototipo para registrar los datos de una solicitud de crédito

Procesar una solicitud de Crédito

<input type="checkbox"/>	RUC	NOMBRE O RAZON SOCIAL	TIPO DE CRÉDITO	MONTO	FECHA DE SOLICITUD	⌚ TIEMPO TRANSCURRIDO
<input type="checkbox"/>	20452145214	EMPRESA CINCO ESTRELLAS # &	CAPITAL DE TRABAJO	5,689.00	21 julio 2019	0 dias 0 hrs 0 min

✓ PRONOSTICAR CALIFICACIÓN CREDITICIA

Figura 14. Prototipo para procesar y generar el pronóstico de calificación crediticia de un cliente

Anexo 3 Diagrama estrella

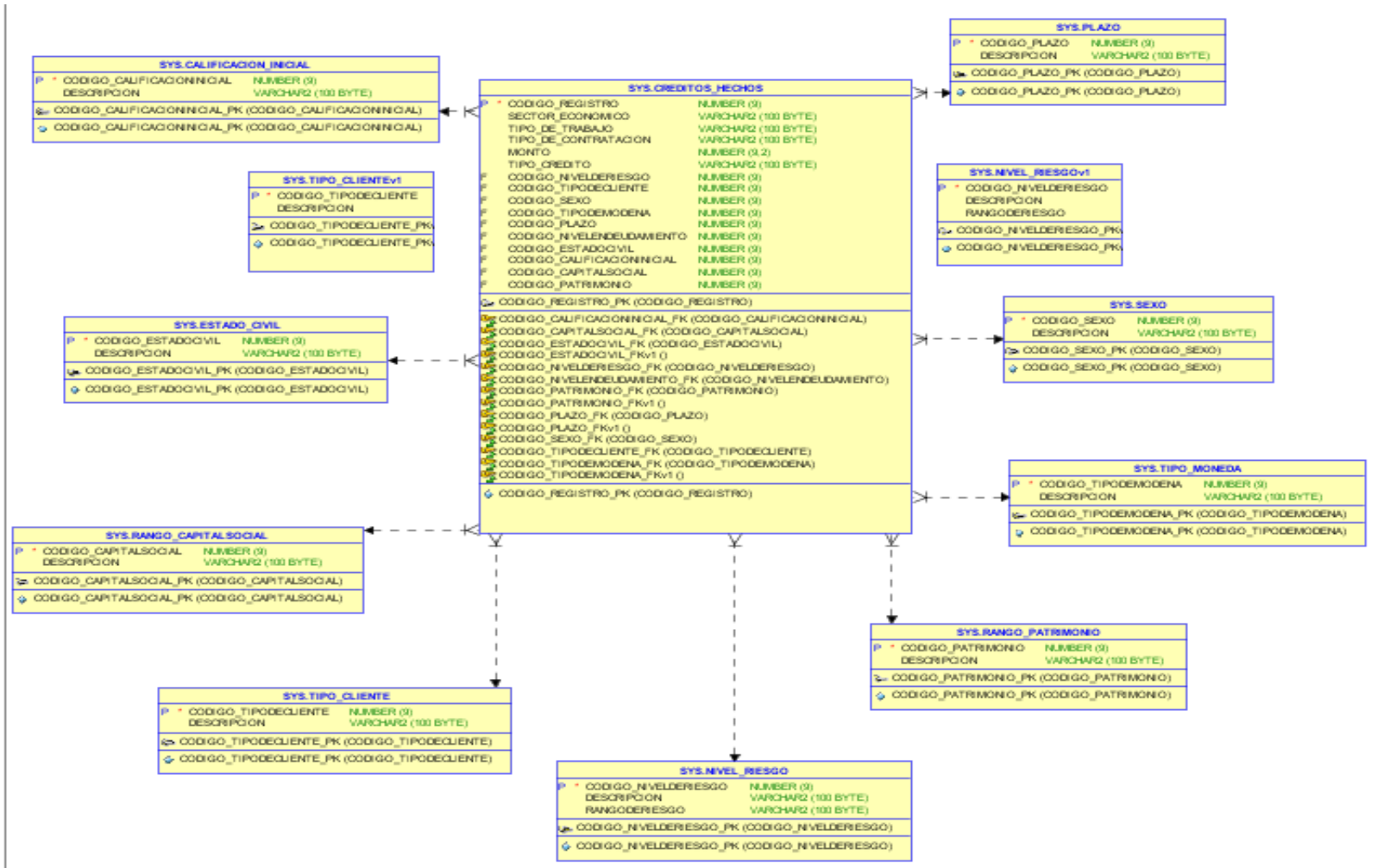


Figura 15. Diagrama de estrella

Anexo 4 Ficha de observación tiempo de respuesta

FICHA DE OBSERVACIÓN TIEMPO DE RESPUESTA DEL SISTEMA CREDIT SCORING

N.º DE FICHA DE OBSERVACIÓN	1
Observadores:	Juan López y Juan Pérez
Institución donde se investiga:	Institución Financiera
Indicador observado:	Av. Camino Real 157 San Isidro
Periodo de observación:	01/08/2019 al 19/08/2019

N.º Registro	Fecha	Tiempo Inicio TI	Tiempo Final TF	Resultado en Tiempo RT= TF - TI
C00001	01-08	09:15	09:21	6:00
C00002	01-08	09:34	09:39	5:00
C00003	02-08	09:21	09:26	5:00
C00004	05-08	09:07	09:15	8:00
C00005	05-08	10:13	10:19	7:00
C00006	05-08	11:09	11:14	5:00
C00007	06-08	09:25	09:32	7:00
C00008	07-08	09:35	09:42	7:00
C00009	09-08	09:11	09:15	4:00
C00010	12-08	09:15	09:22	7:00
C00011	12-08	09:45	09:53	8:00
C00012	13-08	10:15	10:23	8:00
C00013	14-08	09:18	09:23	5:00
C00014	14-08	09:47	09:54	7:00
C00015	15-08	10:15	10:23	8:00
C00016	15-08	10:49	10:56	7:00
C00017	15-08	11:46	11:53	7:00
C00018	16-08	09:34	09:40	6:00
C00019	16-08	10:15	10:23	8:00
C00020	17-08	11:15	11:23	8:00
C00021	17-08	12:02	12:07	5:00
C00022	17-08	12:22	12:28	6:00
C00023	18-08	10:13	10:18	5:00
C00024	18-08	10:56	11:02	6:00
C00025	19-08	09:07	09:15	8:00
				163 =2h.43 % (1.63)

FICHA DE OBSERVACIÓN TIEMPO DE RESPUESTA DEL SISTEMA DE PRONOSTICO DE CALIFICACIÓN CREDITICIA DE CLIENTES

N.º DE FICHA DE OBSERVACIÓN	2
Observadores:	Juan López y Juan Pérez
Institución donde se investiga:	Institución Financiera
Indicador observado:	Av. Camino Real 157 San Isidro
Periodo de observación:	02/09/2019 al 23/09/2019

N.º Registro	Fecha	Tiempo Inicio TI	Tiempo Final TF	Resultado en Tiempo RT= TF - TI
C00001	02-09	09:23	09:25	2:00
C00002	02-09	09:43	09:47	4:00
C00003	03-09	09:11	09:14	3:00
C00004	03-09	10:07	10:09	2:00
C00005	04-09	10:01	10:05	4:00
C00006	05-09	09:09	09:12	3:00
C00007	05-09	10:23	10:26	3:00
C00008	09-09	09:23	09:26	3:00
C00009	09-09	10:11	10:15	4:00
C00010	10-09	09:15	09:17	2:00
C00011	11-09	09:50	09:53	3:00
C00012	12-09	10:21	10:23	2:00
C00013	12-09	10:32	10:35	3:00
C00014	13-09	09:17	09:21	4:00
C00015	16-09	10:22	10:24	2:00
C00016	16-09	10:49	10:51	2:00
C00017	17-09	09:46	09:49	3:00
C00018	17-09	10:34	10:37	3:00
C00019	18-09	10:15	10:18	3:00
C00020	18-09	11:05	11:08	3:00
C00021	19-09	09:02	09:05	3:00
C00022	19-09	10:22	10:25	3:00
C00023	20-09	10:13	10:15	2:00
C00024	23-09	09:56	09:59	3:00
C00025	23-09	10:07	10:10	3:00
				72 = 1h:12 % (0.72)

FICHA DE OBSERVACIÓN PRECISIÓN DEL PRONOSTICO DE CREDIT SCORING

N.º DE FICHA DE OBSERVACIÓN	3
Observadores:	Juan López y Juan Pérez
Institución donde se investiga:	Institución Financiera
Indicador observado:	Av. Camino Real 157 San Isidro
Periodo de observación:	01/08/2019 al 19/08/2019

N.º Registro	Fecha	Precisión
C00001	01-08	83.00%
C00002	01-08	82.00%
C00003	02-08	82.5%
C00004	05-08	81.23%
C00005	05-08	83.94%
C00006	05-08	83.65%
C00007	06-08	80.19%
C00008	07-08	84.92%
C00009	09-08	82.53%
C00010	12-08	81.62%
C00011	12-08	85.35%
C00012	13-08	83.74%
C00013	14-08	85.53%
C00014	14-08	83.00%
C00015	15-08	83.84%
C00016	15-08	82.50%
C00017	15-08	83.72%
C00018	16-08	85.34%
C00019	16-08	83.27%
C00020	17-08	84.54%
C00021	17-08	82.83%
C00022	17-08	80.00%
C00023	18-08	82.00%
C00024	18-08	56.45%
C00025	19-08	79.03%

FICHA DE OBSERVACIÓN PRECISIÓN DE PRONOSTICO DEL SISTEMA DE PRONOSTICO DE CALIFICACIÓN CREDITICIA DE CLIENTES

N.º DE FICHA DE OBSERVACIÓN	4
Observadores:	Juan López y Juan Pérez
Institución donde se investiga:	Institución Financiera
Indicador observado:	Av. Camino Real 157 San Isidro
Periodo de observación:	02/09/2019 al 23/09/2019

N.º Registro	Fecha	Precisión
C00001	02-09	93.00%
C00002	02-09	92.00%
C00003	03-09	92.5%
C00004	03-09	91.23%
C00005	04-09	93.94%
C00006	05-09	93.65%
C00007	05-09	90.19%
C00008	09-09	94.92%
C00009	09-09	92.53%
C00010	10-09	91.62%
C00011	11-09	95.35%
C00012	12-09	93.74%
C00013	12-09	95.53%
C00014	13-09	93.00%
C00015	16-09	93.84%
C00016	16-09	92.50%
C00017	17-09	93.72%
C00018	17-09	95.34%
C00019	18-09	93.27%
C00020	18-09	94.54%
C00021	19-09	92.83%
C00022	19-09	90.00%
C00023	20-09	92.00%
C00024	23-09	96.45%
C00025	23-09	99.03%

FICHA DE OBSERVACIÓN ERROR DEL PRONOSTICO DE CREDIT SCORING

N.º DE FICHA DE OBSERVACIÓN	5
Observadores:	Juan López y Juan Pérez
Institución donde se investiga:	Institución Financiera
Indicador observado:	Av. Camino Real 157 San Isidro
Periodo de observación:	01/08/2019 al 19/08/2019

N.º Registro	Fecha	Error
C00001	01-08	17.00%
C00002	01-08	18.00%
C00003	02-08	17.50%
C00004	05-08	18.77%
C00005	05-08	16.06%
C00006	05-08	16.35%
C00007	06-08	19.81%
C00008	07-08	15.08%
C00009	09-08	17.47%
C00010	12-08	18.38%
C00011	12-08	14.65%
C00012	13-08	16.26%
C00013	14-08	14.47%
C00014	14-08	17.00%
C00015	15-08	16.16%
C00016	15-08	17.50%
C00017	15-08	16.28%
C00018	16-08	14.66%
C00019	16-08	16.73%
C00020	17-08	15.46%
C00021	17-08	17.17%
C00022	17-08	20.00%
C00023	18-08	18.00%
C00024	18-08	43.55%
C00025	19-08	20.97%

FICHA DE OBSERVACIÓN ERROR DEL PRONÓSTICO DEL SISTEMA DE PRONOSTICO DE CALIFICACIÓN CREDITICIA DE CLIENTES

N.º DE FICHA DE OBSERVACIÓN	6
Observadores:	Juan López y Juan Pérez
Institución donde se investiga:	Institución Financiera
Indicador observado:	Av. Camino Real 157 San Isidro
Periodo de observación:	02/09/2019 al 23/09/2019

N.º Registro	Fecha	Precisión
C00001	02-09	7.00%
C00002	02-09	8.00%
C00003	03-09	7.50%
C00004	03-09	8.77%
C00005	04-09	6.06%
C00006	05-09	6.35%
C00007	05-09	9.81%
C00008	09-09	5.08%
C00009	09-09	7.47%
C00010	10-09	8.38%
C00011	11-09	4.65%
C00012	12-09	6.26%
C00013	12-09	4.47%
C00014	13-09	7.00%
C00015	16-09	6.16%
C00016	16-09	7.50%
C00017	17-09	6.28%
C00018	17-09	4.66%
C00019	18-09	6.73%
C00020	18-09	5.46%
C00021	19-09	7.17%
C00022	19-09	10.00%
C00023	20-09	8.00%
C00024	23-09	3.55%
C00025	23-09	0.97%

Anexo 5 Desarrollo Comprensión del Negocio

1.1 Determinar los objetivos de Negocio

La institución financiera tiene como principales objetivos poder contar con herramientas que le ayuden con la evaluación crediticia a fin de tener el índice de morosidad bajos y mitigar el riesgo en los créditos otorgados. A su vez también requiere que estas herramientas sean de bajo costo de implementación y mantenimiento. Ellos cuentan con una DB en la cual estará la información de los clientes y de las solicitudes de crédito otorgadas.

Objetivos de negocio

- Predecir que clientes serán buenos pagadores en base a la información histórica con una mayor precisión.
- Predecir que clientes serán buenos pagadores en base a la información histórica en menos tiempo.
- Contar con herramientas a la evaluación de créditos que sea de bajo costo.

1.2 Evaluación de la situación

Se cuenta con un servidor Proliant DL580 G5, con un sistema operativo Windows server 2012 R2, con un gestor de base de datos Oracle Server 2012 Enterprise que contiene información transaccional desde el año 2014, para solución de la problemática nosotros podemos dar firmeza que disponemos de los datos con una buena cantidad.

Para implementar el sistema de información, se propone los siguientes requerimientos de Software.

- Servidor web Apache (Xampp v3.3.3)
- Editor de texto Sublime Text v3
- Extensiones dll para conectar con Oracle Server 2012
- Microsoft Azure Para Machine Learning

Evaluación inicial de Herramientas y técnicas

Las herramientas que vamos a usar en la realización para el proyecto de minería de datos es Microsoft Azure, la cual se encuentra acoplada en la metodología que utilizaremos. La ventaja de esta herramienta es que nos permite contar con un servicio web que se puede

invocar desde nuestro aplicativo y así contar con la funcionalidad del modelo de pronóstico.

En cuanto las técnicas que vamos a implementar para extraer la sabiduría, de las que proporciona Microsoft Azure, se usaran las siguientes:

Predicción:

- Clasificación
- Regresión

Para dar un mayor aporte de conocimiento las técnicas de predicción que se utilizarán en este proyecto de investigación serán la Regresión Lineal y Árbol de clasificación.

Algoritmo: regresión lineal

```
public class Regresion {
    private double[] x;
    private double[] y;
    private int n; //número de datos
    public double a, b; //pendiente y ordenada en el origen
    public Regresion(double[] x, double[] y) {
        this.x=x;
        this.y=y;
        n=x.length; //número de datos
    }
    public void lineal(){
        double pxy, sx, sy, sx2, sy2;
        pxy=sx=sy=sx2=sy2=0.0;
        for(int i=0; i<n; i++){
            sx+=x[i];
            sy+=y[i];
            sx2+=x[i]*x[i];
            sy2+=y[i]*y[i];
            pxy+=x[i]*y[i];
        }
        a=(n*pxy-sx*sy)/(n*sx2-sx*sx);
        b=(sy-b*sx)/n;
    }
    public double correlacion(){
        //valores medios
        double suma=0.0;
        for(int i=0; i<n; i++){
            suma+=x[i];
        }
        double mediaX=suma/n;

        suma=0.0;
        for(int i=0; i<n; i++){
            suma+=y[i];
        }
        double mediaY=suma/n;
        //coeficiente de correlación
        double pxy, sx2, sy2;
        pxy=sx2=sy2=0.0;
        for(int i=0; i<n; i++){
            pxy+=(x[i]-mediaX)*(y[i]-mediaY);
            sx2+=(x[i]-mediaX)*(x[i]-mediaX);
            sy2+=(y[i]-mediaY)*(y[i]-mediaY);
        }
        return pxy/Math.sqrt(sx2*sy2);
    }
}
```

Algoritmo:

árbol de clasificación

```

crear_sets <- function(datos, proporcion = .7) {
  sets <- list()

  sets[["entrenamiento"]] <- sample_frac(datos, proporcion)
  sets[["prueba"]] <- setdiff(datos, sets[["entrenamiento"]])

  sets
}

entrenar_arbol <- function(sets, objetivo, predictores = ".", mi_cp =
.01) {
  if(length(predictores) > 1) {
    predictores <- paste0(predictores, collapse = "+")
  }
  mi_formula <- paste0(objetivo, " ~ ", predictores) %>% as.formula()

  arbol <- list()
  arbol[["modelo"]] <-
    rpart(data = sets[["entrenamiento"]], formula = mi_formula,
          control = rpart.control(cp = mi_cp, xval = 35, minsplit =
5))
  arbol[["prediccion"]] <- predict(arbol[["modelo"]],
sets[["prueba"]], type = "class")
  arbol[["referencia"]] <- sets[["prueba"]][[objetivo]]

  arbol
}

obtener_diagnostico <- function(arbol, objetivo, mi_cp = 0.01) {
  diagnostico <- list()
  diagnostico[["matriz"]] <- confusionMatrix(data =
arbol[["prediccion"]],
                                             reference =
arbol[["referencia"]])

  cp <- with(arbol[["modelo"]], cptable[which.min(cptable[,
"xerror"]), "CP"])
  cp_original <- mi_cp
  podar <- if(cp < mi_cp) "SI" else "NO"
  diagnostico[["mincp"]] <- data.frame("CP mínimo" = cp, "CP original"
= cp_original, "Podar" = podar)

  diagnostico
}

crear_arbol <- function(datos, objetivo, predictores = ".", mi_cp =
0.01) {
  resultado <- list()
  resultado[["sets"]] <- crear_sets(datos)
  resultado[["arbol"]] <- entrenar_arbol(resultado[["sets"]],
objetivo, predictores, mi_cp)
  resultado[["diagnostico"]] <-
obtener_diagnostico(resultado[["arbol"]], objetivo, mi_cp)

  resultado
}

```

2. Comprensión de los Datos

2.1 Recolectar los Datos Iniciales

Los Datos utilizados comprenden de todos registros históricos de solicitudes de crédito desde el año 2014 al 2019.

```
1 SELECT S.AA_SOLI_SOL || '-' || S.NU_SOLI_SOL || '/' ||
2 FOGADBA.PKG_MISCELANEOS.F_OBT_DESC_SUCURSAL(S.CO_COMP_CIA,
3 S.CO_SUCU_SUC) AS SOLICITUD,
4 S.MO_TOTA_SPR AS MONTO,
5 (SELECT DECODE(L.MO_LIQU_LOP,NULL,L.MO_LIQU_DOLA_LOP,0,L.MO_LIQU_DOLA_LOP,L.MO_LIQU_LOP)
6 FROM FOGADBA.TA_LIQU_OPER_PROD_LOP L
7 WHERE L.AA_SOLI_SOL = S.AA_SOLI_SOL
8 AND L.NU_SOLI_SOL = S.NU_SOLI_SOL
9 AND L.CO_SUCU_SUC = S.CO_SUCU_SUC
10 AND L.NS_SOPR_SPR = S.NS_SOPR_SPR
11 AND L.CO_CPTO_CPT = '006'
12 AND ROWNUM = 1
13 ) AS PLAZO,
14 FOGADBA.PKG_MISCELANEOS.F_OBT_DESC_MONEDA(S.CO_TMON_TMO) AS MONEDA,
15 FOGADBA.PKG_MISCELANEOS.F_OBT_DESC_TIPOPRODUCTO(S.CO_TPRD_TPR) AS TIPO,
16 S.FE_SOLI_SPR AS FECHA,
17 S.CO_CALF_CAL ,
18 FOGADBA.PKG_MISCELANEOS.F_OBT_CALF_CAL(S.CO_CALF_CAL) AS CALIFICACION,
19 C.TI_PERS_CLI AS TIPO_PERSONA,
20 FOGADBA.PKG_MISCELANEOS.F_OBT_DESC_SECTOR_ECONOMICO(FOGADBA.PKG_MISCELANEOS.F_OBT_SEE_CIIU(C.CO_CIIU_CIU)) AS SECTOR_ECON,
21 FOGADBA.PKG_MISCELANEOS.F_OBT_DESC_CIIU(C.CO_CIIU_CIU) AS CIIU,
22 C.NU_RUC_CLI AS RUC,
23 CARGA.PKG_CRUCE_RCC.F_OBT_CAL_RCC@DBL_BIRCC(NVL(FOGADBA.PKG_CLIENTE.F_OBT_COD_SBS_XCLIE(C.CO_CLIE_CLI),FOGADBA.PKG_ALINEAMIENTO
.F_OBT_COD_SBS_RCC( FOGADBA.PKG_CLIENTE.F_OBT_RUCCLIENTE(C.CO_CLIE_CLI), DECODE(SUBSTR(FOGADBA.PKG_CLIENTE.F_OBT_RUCCLIENTE(C
.CO_CLIE_CLI),1,1), '1', SUBSTR(FOGADBA.PKG_CLIENTE.F_OBT_RUCCLIENTE(C.CO_CLIE_CLI),3,8), NULL))), '2019', '08') AS "CALF_201908",
24 C.DE_EST_SUNAT AS ESTADO_SUNAT,
25 C.DE_CON_CONDIC AS CONDICION,
26 C.FE_INI_ACT AS FE_INICIO_ACTIV,
27 FOGADBA.PKG_CLIENTE.F_OBT_DEPARTAMENTO(C.CO_CLIE_CLI) AS DEPARTAMENTO,
28 FOGADBA.PKG_CLIENTE.F_OBT_PROVINCIA(C.CO_CLIE_CLI) AS PROVINCIA,
29 FOGADBA.PKG_CLIENTE.F_OBT_DISTRITO(C.CO_CLIE_CLI) AS DISTRITO,
30 CARGA.PKG_CRUCE_RCC.F_OBT_CAL_RCC@DBL_BIRCC
```

Figura 16 Recolección de datos 1

```
31 CARGA.PKG_CRUCE_RCC.F_OBT_COD_RCC_DNI@DBL_BIRCC(FOGADBA.PKG_CLIENTE.F_OBT_DNI_REPRESENTANTE(C.CO_CLIE_CLI)),
32 '2018',
33 '09'
34 ) CALF_201908_REP,
35 FOGADBA.PKG_EEFF.F_SUM_MON_EEFF(C.CO_CLIE_CLI, '01', FOGADBA.PKG_EEFF.F_OBT_ULT_BALG(C.CO_CLIE_CLI), '0203', 1, 0) AS PATRIMONIO,
36 FOGADBA.PKG_EEFF.F_SUM_MON_EEFF(C.CO_CLIE_CLI, '01', FOGADBA.PKG_EEFF.F_OBT_ULT_BALG(C.CO_CLIE_CLI), '020301', 1, 0) AS VENTAS,
37 FOGADBA.PKG_EEFF.F_SUM_MON_EEFF(C.CO_CLIE_CLI, '01', FOGADBA.PKG_EEFF.F_OBT_ULT_BALG(C.CO_CLIE_CLI), '0301010101', 1, 0) AS CAPITAL_SOCIAL
38
39
40 FROM FOGADBA.TR_SOLI_PROD_SPR S
41 INNER JOIN FOGADBA.TR_CLIE_SOLI_TCS T
42 ON T.AA_SOLI_SOL = S.AA_SOLI_SOL
43 AND T.NU_SOLI_SOL = S.NU_SOLI_SOL
44 AND T.CO_SUCU_SUC = S.CO_SUCU_SUC
45 INNER JOIN FOGADBA.TP_CLIE_CLI C
46 ON C.CO_CLIE_CLI = T.Co_Clie_Cli
47
48 WHERE CO_PROD_PRD IN ('01', '02')
49 AND FE_SOLI_SPR >= '01/01/2014'
50
51 */
```

Figura 17 Recolección de datos 2

La consulta nos devolvió un total de 50310 con los cuales podremos entrenar y probar nuestro modelo.

A continuación, listamos los datos adquiridos:

- SOLICITUD
- MONTO
- PLAZO
- TIPO DE MONEDA
- TIPO DE CRÉDITO
- FECHA
- CALIFICACIÓN
- TIPO DE PERSONA
- SECTOR ECONÓMICO
- CIU
- RUC
- CALIFICACIÓN CREDITICIA
- ESTADO_SUNAT
- CONDICION
- FECHA DE INICIO DE ACTIVIDAD
- DEPARTAMENTO
- PROVINCIA
- DISTRITO
- CALIFICACION_CREDITICIA 2
- PATRIMONIO
- VENTA ANUAL
- CAPITAL SOCIAL

3. Preparación de los Datos

3.1 Datos seleccionados

En esta fase se basa en preparar las datas obtenidas para adecuarlos a las técnicas de minería de datos. Al extraer los datos obtenidas desde la DB de la Institución Financiera, se obtuvieron diferentes tipos de datos, para ello se seleccionó los atributos necesarios para el correcto funcionamiento del modelo.

4. Modelamiento

En fase de la metodología escogemos la técnica de Data Mining para poder predecir. Ya lo mencionado en la fase 1, la técnica que se utilizará en este proyecto de investigación es Regresión Lineal y Árbol de clasificación. En la Figura N° 18, se aplica el entrenamiento para obtener el pronóstico de clasificación crediticia.

5. Evaluación

La fase evalúa el modelo generado, utilizando los objetivos de negocio.

Tabla 21 Estadísticos descriptivos

	Estadísticos descriptivos		N
	Media	Desv. Desviación	
CALIFICA	1,25	,449	26869
MONTO_CREDITO	195476,688935	429354,218977	26869
	577460000	646400000	
PLAZO	121,4590	99,50092	26869
MONEDA	1,01	,099	26869
PRODUCTO	1,32	,472	26869
AÑO	2016,70	1,720	26869
MES	6,60	3,434	26869
TIPO_PERSONA	2,00	,000	26869
SECTOR_ECONOMICO	1,89	,718	26869
CALF_201908	,65	1,406	26869
ANTIGÜEDAD	10,94	7,183	26869
CALF_201908_REP	,49	1,208	26869
PATRIMONIO	451070,6768	1056527,35564	26869
VENTAS	221731,2690	638356,85391	26869
CAPITAL_SOCIAL	996766,3400	1889076,04602	26869

Resultado del análisis que se evidencia en la figura, se la media, la desviación y el número de datos analizados, teniendo una media de 195476.68 para el monto de crédito, 121.590 para el plazo, 10.01 para el tipo de moneda, 1.32 para el tipo de producto, 2016 el año, 06 para el mes, 2 para el tipo de persona, 1.89 para el sector económico, 0.65 para la calificación, 451070.6768 para el patrimonio, 221731.2690 para las ventas y 996766.34 para el capital social.

Tabla 22 Correlaciones

Correlaciones

		CALIFICA	MONTO_CREDITO	PLAZO	MONEDA	PRODUCTO	AÑO	MES	TIPO_PERSONA	SECTOR_ECONOMICO	CALF_201908	ANTIGÜEDAD	CALF_201908_REP	PATRIMONIO	VENTAS	CAPITAL_SOCIAL
Correlación de	CALIFICA	1,000	,019	,015	,010	,020	-,030	-,007	.	-,042	-,051	,058	-,031	,001	,014	-,003
Pearson	MONTO_CREDITO	,019	1,000	,000	-,025	,103	,085	,016	.	-,003	-,061	-,039	-,058	,100	,117	,090
	PLAZO	,015	,000	1,000	,062	-,043	,062	,027	.	-,082	-,055	,041	-,040	-,033	-,023	-,029
	MONEDA	,010	-,025	,062	1,000	-,058	,007	,003	.	,030	,014	,035	-,017	-,029	-,023	,019
	PRODUCTO	,020	,103	-,043	-,058	1,000	,085	-,003	.	-,023	-,151	-,221	-,069	-,085	-,061	-,085
	AÑO	-,030	,085	,062	,007	,085	1,000	-,001	.	-,067	-,262	-,229	-,219	,082	,058	,065
	MES	-,007	,016	,027	,003	-,034	-,001	1,000	.	,001	-,024	-,036	-,026	,024	,021	,021
	TIPO_PERSONA	1,000

	SECTOR_ECONOMICO	-,042	-,003	-,082	-,030	-,023	-,067	-,001	.	1,000	,065	-,016	,040	,046	,019	,113
	CALF_201908	-,051	-,061	-,055	-,014	-,151	-,262	-,024	.	,065	1,000	-,040	,587	-,067	-,036	-,068
	ANTIGÜEDAD	,058	-,039	,041	,035	-,221	-,229	-,036	.	-,016	-,040	1,000	,017	,053	,018	,047
	CALF_201908_REP	-,031	-,058	-,040	-,017	-,069	-,219	-,026	.	,040	,587	,017	1,000	-,073	-,045	-,068
	PATRIMONIO	,001	,100	,033	-,029	-,085	,082	,024	.	,046	-,067	,053	-,073	1,000	,834	,699
	VENTAS	,014	,117	,023	-,023	-,061	,058	,021	.	,019	-,036	,018	-,045	,834	1,000	,525
	CAPITAL_SOCIAL	-,003	,090	,029	-,019	-,085	,065	,021	.	,113	-,068	,047	-,068	,699	,525	1,000
Sig. (unilateral)	CALIFICA	.	,001	,008	,046	,000	,000	,134	,000	,000	,000	,000	,000	,430	,012	,297
	MONTO_CREDITO	,001	.	,473	,000	,000	,000	,004	,000	,295	,000	,000	,000	,000	,000	,000

PLAZO	,008	,473	.	,000	,000	,00	,00	,000	,000	,000	,000	,000	,000	,000	,000
						0	0								
MONEDA	,046	,000	,00	.	,000	,12	,32	,000	,000	,010	,000	,003	,000	,000	,001
			0			1	8								
PRODUCTO	,000	,000	,00	,000	.	,00	,00	,000	,000	,000	,000	,000	,000	,000	,000
			0			0	0								
AÑO	,000	,000	,00	,121	,000	.	,00	,000	,000	,000	,000	,000	,000	,000	,000
			0				0								
MES	,134	,004	,00	,328	,000	,00	.	,000	,431	,000	,000	,000	,000	,000	,000
			0			0									
TIPO_PERSONA	,000	,000	,00	,000	,000	,00	,00	.	,000	,000	,000	,000	,000	,000	,000
			0			0	0								
SECTOR_ECONOMICO	,000	,295	,00	,000	,000	,00	,43	,000	.	,000	,004	,000	,000	,001	,000
			0			0	1								
CALF_201908	,000	,000	,00	,010	,000	,00	,00	,000	,000	.	,000	,000	,000	,000	,000
			0			0	0								
ANTIGÜEDAD	,000	,000	,00	,000	,000	,00	,00	,000	,004	,000	.	,003	,000	,002	,000
			0			0	0								
CALF_201908_REP	,000	,000	,00	,003	,000	,00	,00	,000	,000	,000	,003	.	,000	,000	,000
			0			0	0								
PATRIMONIO	,430	,000	,00	,000	,000	,00	,00	,000	,000	,000	,000	,000	.	,000	,000
			0			0	0								
VENTAS	,012	,000	,00	,000	,000	,00	,00	,000	,001	,000	,002	,000	,000	.	,000
			0			0	0								

	CAPITAL_SO CIAL	,297	,000	,000	,001	,000	,000	,000	,000	,000	,000	,000	,000	,000	.
N	CALIFICA	26869	26869	26869	26869	26869	26869	26869	26869	26869	26869	26869	26869	26869	26869
	MONTO_CRE DITO	26869	26869	26869	26869	26869	26869	26869	26869	26869	26869	26869	26869	26869	26869
	PLAZO	26869	26869	26869	26869	26869	26869	26869	26869	26869	26869	26869	26869	26869	26869
	MONEDA	26869	26869	26869	26869	26869	26869	26869	26869	26869	26869	26869	26869	26869	26869
	PRODUCTO	26869	26869	26869	26869	26869	26869	26869	26869	26869	26869	26869	26869	26869	26869
	AÑO	26869	26869	26869	26869	26869	26869	26869	26869	26869	26869	26869	26869	26869	26869
	MES	26869	26869	26869	26869	26869	26869	26869	26869	26869	26869	26869	26869	26869	26869

TIPO_PERSONA	26869	26869	26869	26869	26869	26869	26869	26869	26869	26869	26869	26869	26869	26869	26869
SECTOR_ECONOMICO	26869	26869	26869	26869	26869	26869	26869	26869	26869	26869	26869	26869	26869	26869	26869
CALF_201908	26869	26869	26869	26869	26869	26869	26869	26869	26869	26869	26869	26869	26869	26869	26869
ANTIGÜEDAD	26869	26869	26869	26869	26869	26869	26869	26869	26869	26869	26869	26869	26869	26869	26869
CALF_201908_REP	26869	26869	26869	26869	26869	26869	26869	26869	26869	26869	26869	26869	26869	26869	26869
PATRIMONIO	26869	26869	26869	26869	26869	26869	26869	26869	26869	26869	26869	26869	26869	26869	26869
VENTAS	26869	26869	26869	26869	26869	26869	26869	26869	26869	26869	26869	26869	26869	26869	26869
CAPITAL_SOCIAL	26869	26869	26869	26869	26869	26869	26869	26869	26869	26869	26869	26869	26869	26869	26869

Tabla 23 Variables de entrada/eliminadas

Variables entradas/eliminada			
Modelo	Variables entradas	Variables eliminadas	Método
1	CAPITAL_SOCI AL, MONEDA, MES, ANTIGÜEDAD, CALF_201908_ REP, PLAZO, MONTO_CREDI TO, SECTOR_ECO NOMICO, PRODUCTO, AÑO, VENTAS, CALF_201908, PATRIMONIO ^b	.	Introducir

a. Variable dependiente: CALIFICA

b. Todas las variables solicitadas introducidas.

Tabla 24 Resumen del modelo

Resumen del modelo

Modelo	R	R cuadrado	R cuadrado ajustado	Error estándar de la estimación	Cambio en R cuadrado	Estadísticos de cambio			Sig. Cambio en F	Durbin-Watson
						Cambio en F	gl1	gl2		
1	,951 ^a	,904	,903	,447	,010	21,132	13	26855	,000	1,425

a. Predictores: (Constante), CAPITAL_SOCIAL, MONEDA, MES, ANTIGÜEDAD, CALF_201908_REP, PLAZO, MONTO_CREDITO, SECTOR_ECONOMICO, PRODUCTO, AÑO, VENTAS, CALF_201908, PATRIMONIO

b. Variable dependiente: CALIFICA

Del resumen del modelo podemos apreciar que el R es igual a 0.951, el R cuadrado es igual a 0.904 y el R cuadrado ajustado es igual a 0.903, lo que quiere decir hay un alto grado de correlación en la variable explicativa y la variable a predecir.

Tabla 25 Tabla ANOVA

		ANOVA ^a				
Modelo		Suma de cuadrados	gl	Media cuadrática	F	Sig.
1	Regresión	54,884	13	4,222	21,132	,000 ^b
	Residuo	5365,272	26855	,200		
	Total	5420,156	26868			

a. Variable dependiente: CALIFICA

b. Predictores: (Constante), CAPITAL_SOCIAL, MONEDA, MES, ANTIGÜEDAD, CALF_201908_REP, PLAZO, MONTO_CREDITO, SECTOR_ECONOMICO, PRODUCTO, AÑO, VENTAS, CALF_201908, PATRIMONIO

De la tabla podemos decir la sig. Es inferior 0.05 lo cual indica que nuestras variables explicativas, si son correlacionadas con nuestra variable a predecir

Tabla 26 Matriz de coeficientes

		Coeficientes ^a											
		Coeficientes no estandarizados		Coeficientes estandarizados			95,0% intervalo de confianza para B		Correlaciones			Estadísticas de colinealidad	
Modelo		B	Desv. Error	Beta	t	Sig.	Límite inferior	Límite superior	Orden cero	Parcial	Parte	Tolerancia	VIF
1	(Constante)	21,658	3,482		6,221	,000	14,834	28,482					
	MONTO_CREDITO	1,779E-8	,000	,017	2,753	,006	,000	,000	,019	,017	,017	,966	1,035
	PLAZO	4,140E-5	,000	,009	1,494	,135	,000	,000	,015	,009	,009	,978	1,023
	MONEDA	,053	,028	,012	1,921	,055	-,001	,107	,010	,012	,012	,986	1,014
	PRODUCTO	,024	,006	,025	3,984	,000	,012	,036	,020	,024	,024	,901	1,110
	AÑO	-,010	,002	-,039	-5,883	,000	-,014	-,007	-,030	-,036	-,036	,844	1,185
	MES	-,001	,001	-,009	-1,475	,140	-,003	,000	-,007	-,009	-,009	,983	1,018
	SECTOR_ECONOMICO	-,024	,004	-,038	-6,178	,000	-,031	-,016	-,042	-,038	-,038	,969	1,032
	CALF_201908	-,015	,002	-,047	-6,028	,000	-,020	-,010	-,051	-,037	-,037	,611	1,638
	ANTIGÜEDAD	,003	,000	,053	8,106	,000	,002	,004	,058	,049	,049	,876	1,142
	CALF_201908_REP	-,003	,003	-,009	-1,180	,238	-,009	,002	-,031	-,007	-,007	,647	1,546
	PATRIMONIO	-1,517E-8	,000	-,036	-2,674	,008	,000	,000	,001	-,016	-,016	,207	4,832
	VENTAS	3,054E-8	,000	,043	3,880	,000	,000	,000	,014	,024	,024	,294	3,397
	CAPITAL_SOCIAL	9,045E-11	,000	,000	,044	,965	,000	,000	-,003	,000	,000	,490	2,040

a. Variable dependiente: CALIFICA

De la matriz de coeficientes podemos obtener el grado de correlación que tienen las variables explicativas.

6. Implementación

En esta fase se muestra el modelo de pronóstico desarrollado con la herramienta Microsoft Azure.

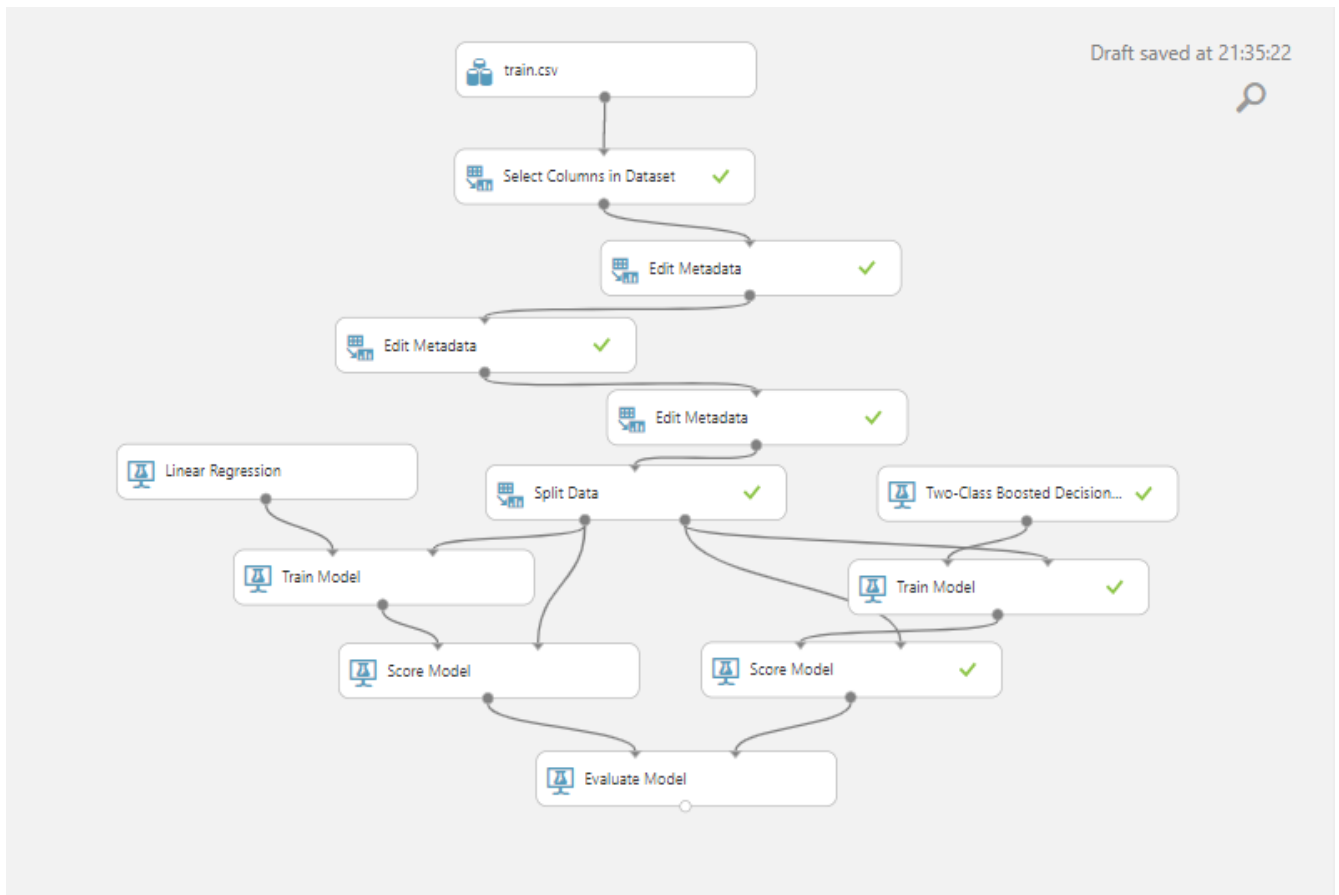


Figura 18 Modelo de Pronostico en Microsoft Azure

Anexo 6 Autorización para realizar la investigación



AUTORIZACIÓN PARA LA ELABORACIÓN DE TRABAJO DE INVESTIGACIÓN

Por medio del presente documento, Yo OSCAR ORLANDO PORTOCARRERO QUEVEDO, identificado con DNI N° 08866447 representante legal de La Fundación FONDO DE GARANTÍA PARA PRESTAMOS A LA PEQUEÑA INDUSTRIA - FOGAPI autorizo a JUAN ERICK LÓPEZ DE LA CRUZ identificado con DNI N° 45989478 a realizar el trabajo de investigación titulado: "SISTEMA DE PRONÓSTICO DE LA CALIFICACIÓN CREDITICIA DE CLIENTES BASADO EN ÁRBOL DE CLASIFICACIÓN Y REGRESIÓN LINEAL" que tiene como punto principal de referencia a FOGAPI.

La difusión de la información del trabajo de investigación, la misma que esta referenciada y citada de forma correcta en el documento y fuente correspondiente, es estrictamente con fines académicos.

Lima, 3 de DICIEMBRE de 2019

OSCAR ORLANDO PORTOCARRERO QUEVEDO
DNI N° 08866447
GERENTE GENERAL
FOGAPI

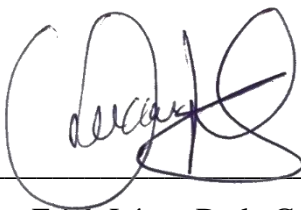
Av. Camino Real N° 157
San Isidro - Lima 27 - Perú
Teléfono: (511) 700 0100
Web: www.fogapi.com.pe

Figura 19 Autorización para realizar la investigación

Yo **LOPEZ DE LA CRUZ JUAN ERICK**, identificado con DNI N° **45989478** y **PÉREZ ALARCÓN JUAN JOSÉ**, identificado con DNI N° **45164035**, egresados de la Facultad de Ingeniería y carrera Profesional de Ingeniería de Sistemas de la Universidad César Vallejo campus Lima Este, declaramos bajo juramento que la tesis titulada:

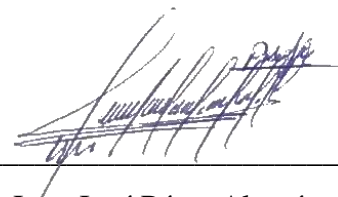
“Sistema de Pronóstico de la Calificación Crediticia de Clientes Basado en Árbol de clasificación y Regresión lineal”, que toda la documentación es veraz y autentica.

San Juan de Lurigancho, 21 de diciembre del 2019



Juan Erick López De la Cruz

DNI: 45989478



Juan José Pérez Alarcón

DNI: 45164035

Elaboró	Dirección de Investigación	Revisó	Responsable del SGC	Aprobó	Vicerrectorado de Investigación
---------	----------------------------	--------	---------------------	--------	---------------------------------