

**UNIVERSIDAD CATÓLICA SANTO TORIBIO DE MOGROVEJO**  
**FACULTAD DE INGENIERÍA**  
**ESCUELA DE INGENIERÍA DE SISTEMAS Y COMPUTACIÓN**



**Implementación de un modelo de Credit Scoring para la predicción de riesgo en el otorgamiento de crédito en la financiera Efectiva del grupo EFE**

**TESIS PARA OPTAR EL TÍTULO DE  
INGENIERO DE SISTEMAS Y COMPUTACIÓN**

**AUTOR**

**Carlos Fabricio Olivos Seclen**

**ASESOR**

**Maria Ysabel Aranguri Garcia**

<https://orcid.org/0000-0001-9220-5801>

**Chiclayo, 2025**

**Implementación de un modelo de Credit Scoring para la predicción de riesgo en el otorgamiento de crédito en la financiera Efectiva del grupo EFE**

PRESENTADA POR

**Carlos Fabricio Olivos Seclen**

A la Facultad de Ingeniería de la  
Universidad Católica Santo Toribio de Mogrovejo  
para optar el título de

**INGENIERO DE SISTEMAS Y COMPUTACIÓN**

APROBADA POR

Hector Miguel Zelada Vldivieso

PRESIDENTE

Ricardo David Iman Espinoza

SECRETARIO

Maria Ysabel Aranguri Garcia

VOCAL

## **Dedicatoria**

A DIOS por su guía en este camino de aprendizaje, a mis padres C. Olivos y M. Seclén, a mi segunda madre E. Seclén, a mis abuelitos que me cuidan desde el cielo y a toda mi familia en general por el apoyo desde el inicio de la formación académica.

## **Agradecimientos**

A mi asesora María Ysabel Arangurí García, por apoyarme en el transcurso del presente proyecto. A la vez agradecer a la empresa que permitió implementar la tesis y aquella que permitió el desarrollo de las practicas preprofesionales.

# TESIS\_OLIVOS\_CARLOS

## INFORME DE ORIGINALIDAD

<b>17%</b> INDICE DE SIMILITUD	<b>17%</b> FUENTES DE INTERNET	<b>1%</b> PUBLICACIONES	<b>3%</b> TRABAJOS DEL ESTUDIANTE
-----------------------------------	-----------------------------------	----------------------------	--------------------------------------

## FUENTES PRIMARIAS

<b>1</b>	<b>hdl.handle.net</b> Fuente de Internet	<b>10%</b>
<b>2</b>	<b>tesis.usat.edu.pe</b> Fuente de Internet	<b>1%</b>
<b>3</b>	<b>dspace.uniandes.edu.ec</b> Fuente de Internet	<b>1%</b>
<b>4</b>	<b>www.sbs.gob.pe</b> Fuente de Internet	<b>&lt;1%</b>
<b>5</b>	<b>repositorio.ucv.edu.pe</b> Fuente de Internet	<b>&lt;1%</b>
<b>6</b>	<b>www.slideshare.net</b> Fuente de Internet	<b>&lt;1%</b>
<b>7</b>	<b>www.coursehero.com</b> Fuente de Internet	<b>&lt;1%</b>
<b>8</b>	<b>Submitted to Universidad Nacional del Centro del Peru</b> Trabajo del estudiante	<b>&lt;1%</b>

## ÍNDICE

<b>RESUMEN</b> .....	11
<b>ABSTRACT</b> .....	12
<b>I. Introducción</b> .....	13
<b>II. Revisión de la literatura</b> .....	15
<b>2.1. Antecedentes</b> .....	15
<b>2.1.1. Antecedentes internacionales</b> .....	15
<b>2.1.2. Antecedentes nacionales</b> .....	15
<b>2.1.3. Antecedentes locales</b> .....	16
<b>2.2. Bases teóricas científicas</b> .....	16
<b>2.2.1. Credit scoring (CS)</b> .....	16
<b>2.2.2. Minería de datos (MD)</b> .....	22
<b>2.2.3. Términos financieros</b> .....	24
<b>III. Materiales y métodos</b> .....	24
<b>3.1. Tipo de investigación</b> .....	24
<b>3.2. Diseño de investigación</b> .....	25
<b>3.3. Población y muestra</b> .....	25
<b>3.4. Variables</b> .....	25
<b>3.5. Métodos de investigación</b> .....	25
<b>3.6. Técnicas e instrumentos de recolección de datos</b> .....	26
<b>3.7. Procedimientos</b> .....	26
<b>3.7.1. Metodología de desarrollo</b> .....	26
<b>3.7.2. Producto acreditable</b> .....	30
<b>3.7.3. Manual de usuario</b> .....	30
<b>3.8. Matriz de consistencia</b> .....	31
<b>3.9. Consideraciones éticas</b> .....	33
<b>IV. Resultados y discusión</b> .....	33

<b>4.1. En base a la metodología utilizada</b> .....	33
<b>4.1.1. Fase 1: Comprensión del negocio</b> .....	33
<b>4.1.2. Fase 2: Comprensión de los datos</b> .....	47
<b>4.1.3. Fase 3: Preparación de los datos</b> .....	55
<b>4.1.4. Fase 4: Modelamiento</b> .....	60
<b>4.1.5. Fase 5: Evaluación</b> .....	75
<b>4.1.6. Fase 6: Despliegue</b> .....	78
<b>4.2. En base a los objetivos de la investigación</b> .....	80
<b>4.2.1. Determinar la técnica Credit Scoring que reúna los requisitos de predictibilidad del sistema, para la toma de decisiones en la asignación de créditos</b> .....	80
<b>4.2.2. Implementar escenarios de predicción aplicando la técnica de Credit Scoring seleccionada para predecir el nivel de riesgo en el otorgamiento de crédito.</b> ...83	
<b>4.2.3. Medir el nivel de precisión del sistema para determinar la confiabilidad de los resultados del proceso.</b> ....	85
<b>4.2.4. Validar el grado de usabilidad del modelo Credit Scoring por parte de los usuarios finales</b> .....	86
<b>4.3. Impactos esperados</b> .....	89
<b>4.3.1. Impactos económicos</b> .....	89
<b>4.3.2. Impactos sociales</b> .....	89
<b>4.3.3. Impactos en tecnología</b> .....	89
<b>4.3.4. Impactos ambientales</b> .....	89
<b>V. Conclusiones</b> .....	92
<b>VI. Recomendaciones</b> .....	93
<b>VII. Referencias</b> .....	94

## Lista de tablas

<b>TABLA I. Métodos de investigación.....</b>	<b>25</b>
<b>TABLA II. Técnicas e instrumentos .....</b>	<b>26</b>
<b>TABLA III. Matriz de consistencia .....</b>	<b>31</b>
<b>TABLA IV. Recursos humanos.....</b>	<b>34</b>
<b>TABLA V. Fuente de datos .....</b>	<b>35</b>
<b>TABLA VI. Requisitos para un crédito .....</b>	<b>35</b>
<b>TABLA VII. Restricciones de los involucrados.....</b>	<b>35</b>
<b>TABLA VIII. Terminología financiera .....</b>	<b>36</b>
<b>TABLA IX. Equipo de trabajo.....</b>	<b>37</b>
<b>TABLA X. R.F. 01 .....</b>	<b>38</b>
<b>TABLA XI. R.F. 02 .....</b>	<b>38</b>
<b>TABLA XII. R.F. 03.....</b>	<b>38</b>
<b>TABLA XIII. R.F. 04 .....</b>	<b>39</b>
<b>TABLA XIV. R.F. 05 .....</b>	<b>39</b>
<b>TABLA XV. R.F. 06.....</b>	<b>39</b>
<b>TABLA XVI. R.F. 07 .....</b>	<b>40</b>
<b>TABLA XVII. R.F. 08.....</b>	<b>40</b>
<b>TABLA XVIII. R.F. 09 .....</b>	<b>40</b>
<b>TABLA XIX. R.F. 10 .....</b>	<b>41</b>
<b>TABLA XX. R.F. 11.....</b>	<b>41</b>
<b>TABLA XXI. R.N.F. 01.....</b>	<b>41</b>
<b>TABLA XXII. R.N.F.02 .....</b>	<b>42</b>
<b>TABLA XXIII. R.N.F.03 .....</b>	<b>42</b>
<b>TABLA XXIV. R.N.F.04.....</b>	<b>42</b>
<b>TABLA XXV. R.N.F.05 .....</b>	<b>43</b>
<b>TABLA XXVI. R.N.F.06.....</b>	<b>43</b>
<b>TABLA XXVII. Fase 1.....</b>	<b>44</b>
<b>TABLA XXVIII. Fase 2 .....</b>	<b>44</b>
<b>TABLA XXIX. Fase 3 .....</b>	<b>45</b>
<b>TABLA XXX. Fase 4.....</b>	<b>45</b>
<b>TABLA XXXI. Fase 5 .....</b>	<b>46</b>
<b>TABLA XXXII. Fase 6.....</b>	<b>46</b>

<b>TABLA XXXIII. Nivel de riesgo de las fases .....</b>	<b>47</b>
<b>TABLA XXXIV. Diccionario tabla ‘usuario’ .....</b>	<b>48</b>
<b>TABLA XXXV. Diccionario tabla ‘permisos’ .....</b>	<b>49</b>
<b>TABLA XXXVI. Diccionario tabla ‘usuariopermiso’ .....</b>	<b>49</b>
<b>TABLA XXXVII. Diccionario tabla ‘clientes’ .....</b>	<b>49</b>
<b>TABLA XXXVIII. Diccionario tabla ‘prestamos’ .....</b>	<b>50</b>
<b>TABLA XXXIX. Diccionario tabla ‘pagos’ .....</b>	<b>52</b>
<b>TABLA XL. Selección de datos.....</b>	<b>55</b>
<b>TABLA XLI. Variables BD original.....</b>	<b>57</b>
<b>TABLA XLII. Variables BD prestamos_modelo .....</b>	<b>58</b>
<b>TABLA XLIII. Formateo columna ‘situacion_laboral’ .....</b>	<b>59</b>
<b>TABLA XLIV. Formateo columna ‘formalidad_laboral’ .....</b>	<b>59</b>
<b>TABLA XLV. Formateo columna ‘casa_propia’ .....</b>	<b>60</b>
<b>TABLA XLVI. Comparación de algoritmos empleados.....</b>	<b>61</b>
<b>TABLA XLVII. Plan de pruebas .....</b>	<b>61</b>
<b>TABLA XLVIII. Matriz de confusión.....</b>	<b>61</b>
<b>TABLA XLIX. Métricas de evaluación.....</b>	<b>71</b>
<b>TABLA L. Comparativa de precisión de los tres algoritmos .....</b>	<b>75</b>
<b>TABLA LI. Matriz de confusión del algoritmo .....</b>	<b>76</b>
<b>TABLA LII. Comparativa de técnicas CS .....</b>	<b>81</b>
<b>TABLA LIII. Indicadores del Obj. Esp. 01 .....</b>	<b>82</b>
<b>TABLA LIV. Escenarios de predicción.....</b>	<b>83</b>
<b>TABLA 55. Matriz de confusión del modelo .....</b>	<b>85</b>
<b>TABLA LVI. Resultado de sensibilidad del modelo .....</b>	<b>85</b>
<b>TABLA LVII. Resultado de especificidad del modelo .....</b>	<b>85</b>
<b>TABLA LVIII. Resultado de eficacia del modelo .....</b>	<b>86</b>
<b>TABLA LIX. Ponderación de las respuestas (SUS) .....</b>	<b>86</b>
<b>TABLA LX. Resultados de la herramienta SUS .....</b>	<b>88</b>
<b>TABLA LXI. Comparativa antes/después de aplicar el modelo CS.....</b>	<b>91</b>

## Lista de figuras

Fig. 1. Proceso de crédito .....	15
Fig. 2. Regresión Logística.....	18
Fig. 3. Estructura de un árbol .....	19
Fig. 4. Componentes de una red neuronal .....	20
Fig. 5. Red con "n" entradas y una salida.....	21
Fig. 6. Visión de MD.....	22
Fig. 7. Tipos de datos.....	23
Fig. 8. Ciclo de vida – CRISP-DM .....	26
Fig. 9. Fase 1: Comprensión del negocio .....	27
Fig. 10. Fase 2: Comprensión de los datos .....	27
Fig. 11. Fase 3: Preparación de los datos .....	28
Fig. 12. Fase 4: Modelado .....	28
Fig. 13. Fase 5: Evaluación .....	29
Fig. 14. Fase 6: Implantación .....	29
Fig. 15. Proceso de otorgamiento de créditos.....	33
Fig. 16. Tabla ‘usuario’ .....	53
Fig. 17. Tabla ‘permisos’ .....	53
Fig. 18. Tabla ‘usuariopermiso’ .....	53
Fig. 19. Tabla ‘clientes’ .....	53
Fig. 20. Tabla ‘prestamos’ .....	54
Fig. 21. Tabla ‘pagos’ .....	54
Fig. 22. Modelo físico transaccional.....	54
Fig. 23. Limpieza de datos nulos .....	55
Fig. 24. Limpieza de datos con monto ‘0.00’ .....	56
Fig. 25. Limpieza de datos duplicados - Ej.1 .....	56
Fig. 26. Limpieza de datos duplicados - Ej.2 .....	56
Fig. 27. BD transaccional y BD de predicción .....	59
Fig. 28. Conexión a la BD .....	62
Fig. 29. Función inicial.....	62
Fig. 30. Parámetros de entrada de un nuevo cliente .....	62
Fig. 31. Parámetros de entrada de un préstamo.....	63
Fig. 32. Librería Tau Prolog.....	63

<b>Fig. 33. Tau Prolog .....</b>	<b>64</b>
<b>Fig. 34. Reglas de predicción .....</b>	<b>64</b>
<b>Fig. 35. Func. Calcular_nivel.....</b>	<b>65</b>
<b>Fig. 36. Componentes .....</b>	<b>65</b>
<b>Fig. 37. Interfaz Login.....</b>	<b>66</b>
<b>Fig. 38. Interfaz gestión de usuarios .....</b>	<b>66</b>
<b>Fig. 39. Interfaz de registro de usuario .....</b>	<b>67</b>
<b>Fig. 40. Interfaz gestionar clientes .....</b>	<b>67</b>
<b>Fig. 41. Interfaz registro de clientes.....</b>	<b>68</b>
<b>Fig. 42. Interfaz gestionar prestamos .....</b>	<b>68</b>
<b>Fig. 43. Interfaz registro de prestamos.....</b>	<b>69</b>
<b>Fig. 44. Interfaz resultado de predicción.....</b>	<b>69</b>
<b>Fig. 45. Interfaz de recomendación.....</b>	<b>70</b>
<b>Fig. 46. Selección de archivo -WEKA.....</b>	<b>72</b>
<b>Fig. 47. Selección de algoritmo .....</b>	<b>72</b>
<b>Fig. 48. Indicamos el porcentaje .....</b>	<b>73</b>
<b>Fig. 49. Prueba del algoritmo J48 .....</b>	<b>73</b>
<b>Fig. 50. Prueba del algoritmo Naive Bayes .....</b>	<b>74</b>
<b>Fig. 51. Prueba del algoritmo OneR .....</b>	<b>74</b>
<b>Fig. 52. Escala de aceptación de SUS.....</b>	<b>77</b>
<b>Fig. 53. Archivo .rar .....</b>	<b>78</b>
<b>Fig. 54. Petición POST.....</b>	<b>79</b>
<b>Fig. 55. Reglas de predicción .....</b>	<b>84</b>
<b>Fig. 56. Función del cálculo del nivel.....</b>	<b>84</b>
<b>Fig. 57. Comparativa antes/después de aplicar el modelo CS .....</b>	<b>92</b>

## RESUMEN

La presente investigación describe la problemática de una financiera en la ciudad de Chiclayo durante el proceso de otorgamiento de créditos, por ello se desarrolló un modelo Credit Scoring haciendo uso de las técnicas de modelamiento para analizar el nivel de riesgo de un cliente, así poder apoyar en la toma de decisiones en el proceso de otorgamiento de créditos. Los principales problemas durante la fase de la comprensión del negocio son: la demora en el proceso de créditos, el proceso es obsoleto y por último su índice de morosidad es del 7.4%. Por lo tanto, el objetivo general del estudio fue implementar un modelo Credit Scoring, para mejorar la toma de decisiones en el proceso de otorgamiento de créditos. El estudio es de tipo tecnológica aplicada de nivel cuantitativa cuasiexperimental. Para el desarrollo del módulo de predicción en base al algoritmo seleccionado, se empleó la metodología CRISP-DM para la construcción, evaluación y despliegue. El sistema se construyó utilizando lenguajes de programación JavaScript, HTML y el gestor de base de datos MySQL. Finalmente, se utilizó System Usability Scale la que determinó que el grado de usabilidad es aceptable con una puntuación promedio de 85.12 puntos, los resultados en las métricas de evaluación y el porcentaje de asertividad del 86% y un 81% de sensibilidad, por último, se realizaron las pruebas de caja blanca y las pruebas de caja negra para conocer el desempeño funcional del sistema.

**Palabras clave:** Credit Scoring, modelos supervisados, créditos. nivel de riesgo.

## ABSTRACT

The present investigation describes the problems of a financial company in the city of Chiclayo during the credit granting process, for this reason a credit scoring model was developed using modeling techniques to analyze the level of risk of a client, thus being able to support in decision-making in the credit granting process. The main problems during the business understanding phase are: the delay in the credit process, the process is obsolete and finally its delinquency rate is 7.4%. Therefore, the general objective of the study was to implement a Credit Scoring model, to improve decision-making in the credit granting process. The study is of an applied technological type of quasi-experimental quantitative level. For the development of the prediction module based on the selected algorithm, the CRISP-DM methodology was used for construction, evaluation and deployment. The system was built using the programming languages JavaScript, HTML and the MySQL database manager. Finally, the System Usability Scale was used, which determined that the degree of usability is acceptable with an average score of 85.12 points, the results in the evaluation metrics and the assertiveness percentage of 86% and 81% sensitivity, finally, White box tests and black box tests were carried out to know the functional performance of the system.

**Keywords:** Credit Scoring, supervised models, credits. risk level.

## I. Introducción

El presente trabajo de investigación tiene como finalidad analizar los principales factores que conllevan a un alto índice de morosidad en instituciones financieras; la situación de incertidumbre en la actualidad con respecto al sector económico del país y del mundo trae como consecuencia el aumento de precios de todo tipo de productos, afectando mucho a los bolsillos de los peruanos. [1] La región que sufrió mayores estragos durante la pandemia del COVID-19 fue América Latina y el Caribe (ALC).

En el 2021 la región ALC se recuperó económicamente en un 6,3%, para el año 2023 las perspectivas de crecimiento económico han ido disminuyendo en los últimos meses en un 1.4%. con respecto al Perú, según la Superintendencia de Banca y Seguros (SBS) [2], la economía peruana crecería en un 3.4% en 2022, impulsada por la normalización de las actividades de los sectores más rezagados, y 3.2% en 2023, debido a la reactivación de las actividades económicas lo cual permitió que el Perú pueda lograr los niveles previos a la pandemia. Esto se da en un avance significativo en la vacunación de la población, se mantiene el impulso fiscal y las medidas tomadas por el gobierno, todas ellas fueron variables de influencia para la recuperación del crecimiento económico. El Ministerio de Economía y Finanzas (MEF) [3], tomó diferentes medidas, siendo una de ellas el Programa Reactiva Perú que incluye prestamos de capital, garantía desde 80% hasta 98%, de acuerdo con el tamaño de la organización, otra medida tomada por el gobierno fue el Programa de Apoyo Empresarial a los micro y pequeña empresa (PAE-MYPE).

A diciembre del 2020 el saldo de deuda de créditos de consumo e hipotecario en el sistema financiero ascendió a S/132 420 millones, de este monto el 56% corresponde a créditos de consumo y el 44% a créditos hipotecarios.

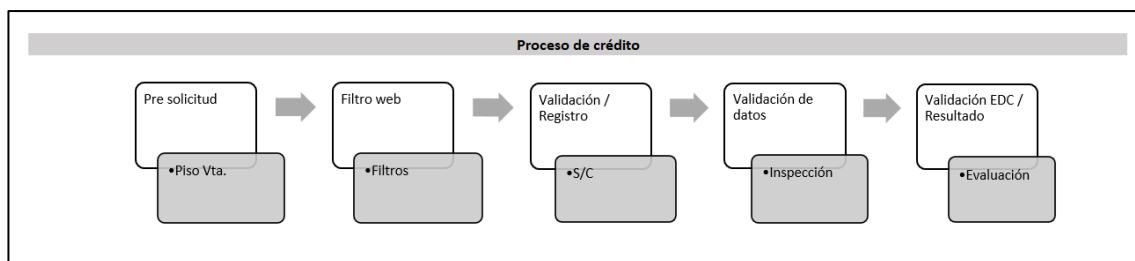
La morosidad es uno de los grandes problemas que afrontan las organizaciones financieras, los picos más altos de morosidad se dieron en el tercer y cuarto trimestre del 2020, para mayo del año 2021 se encontraba en S/11 628 millones y para mayo del 2022 la deuda de morosidad ascendía en S/10 629 millones según el informe de morosidad realizado por Equifax y Emprende UP. La Financiera no fue ajena este problema, durante la fase de levantamiento de información se pudo conocer el proceso de otorgamiento de créditos como se muestra en la Figura 1, además que su sistema de calificación es obsoleto, los créditos mal colocados, entre otros factores conllevan a tener un índice de morosidad de aproximadamente 7.4%. He ahí la importancia de desarrollar un modelo Credit Scoring (CS), [13] lo define como el conjunto de modelos

y técnicas que apoyan a las entidades financieras al momento de un otorgamiento de crédito. Hoy en día existen diferentes técnicas, pero para la presente investigación se analizaron diferentes técnicas: Regresión Logística, Redes Neuronales, arboles de decisión y reglas de clasificación.

Ante esta realidad problemática, se formuló la siguiente pregunta: ¿Cómo mejorar el proceso de otorgamiento de créditos, con la finalidad de tomar mejores decisiones dentro de la financiera de la ciudad de Chiclayo?

Debido a lo mencionado anteriormente se planteó realizar la presente investigación, planteando como objetivo general, implementar un modelo Credit Scoring, para mejorar la toma de decisiones en el proceso de otorgamiento de créditos en la Financiera Efectiva del grupo EFE. Para ello se plantearon objetivos específicos: Determinar la técnica credit scoring que reúna los requisitos de predictibilidad del sistema, para la toma de decisiones en la asignación de créditos, implementar escenarios de predicción aplicando la técnica de Credit Scoring seleccionada para predecir el nivel de riesgo en el otorgamiento de crédito, medir el nivel de precisión del sistema para determinar la confiabilidad de los resultados del proceso y por ultimo validar el grado de usabilidad del modelo credit scoring, por parte de los usuarios finales.

Ante la necesidad de una solución, se desarrolló la investigación tecnológica aplicada, que se justifica económicamente, porque ayudará con la reducción de tasas de morosidad de los clientes de la financiera Efectiva, a su vez se utilizará software libre lo cual no representará costos extras por obtención de licencias. A la vez en el aspecto científico se justifica porque se emplearán técnicas, herramientas, modelos y metodologías que servirán como base para futuros proyectos de investigación, además de asentar precedentes en la implementación de este tipo de soluciones. Además, desde el aspecto tecnológico debido a la necesidad de desarrollar un sistema basado en un modelo CS que ayude a predecir el riesgo en el otorgamiento de créditos en la financiera efectiva, para implementar dicho sistema se requiere de lenguajes de programación, además se utilizó la metodología CRISP-DM.



*Fig. 1. Proceso de crédito*

## II. Revisión de la literatura

### 2.1. Antecedentes

#### 2.1.1. Antecedentes internacionales

Lagua [4], desarrolló un Modelo CS que brinde soluciones a las organizaciones en la ciudad de Ambato – Ecuador, se utilizó el método Knowledge Discovery in Data (KDD), el cual permitió aplicar herramientas estadísticas como regresión logística. Finalmente concluyeron que el modelo cumple las condiciones para adaptarse a la organización, además cada cierto tiempo se tiene que monitorear el modelo con el objetivo de medir su eficiencia. Esta investigación fue considerada debido al enfoque y la utilización del algoritmo de Regresión Logística para darle solución a su problemática.

Aboobyda y Taring [5], en esta investigación se demuestra que la unión entre machine learning y la minería de datos permiten reducir el riesgo que genera el otorgamiento de créditos bancarios. En dicha investigación (Pag.11) utilizaron la herramienta weka con los siguientes algoritmos: Redes Bayesianas, Bayes Ingenuo y J48. Se obtuvieron los siguientes resultados, 73.8%, 77.4% y 78.3% respectivamente a los algoritmos mencionados anteriormente y como se aprecia el último modelo fue el mejor debido a que tiene una mayor exactitud. Esta investigación fue considerada debido a los tres algoritmos supervisados utilizados en la presente.

#### 2.1.2. Antecedentes nacionales

Alarcón y Mora [7], en la investigación el problema central es el alto nivel de morosidad de microfinancieras con respecto a los bancos del Perú, debido a ello se plantea como objetivo general “implementar un modelo predictivo de riesgo crediticio no revolvante, para las microfinancieras del Perú”. Se utilizó SPSS Modeler de minería de datos, además de la metodología Knowledge Discovery in

Data (KDD). Finalmente destacan que el algoritmo de redes Bayesianas tiene un alto porcentaje de probabilidad de pago de los clientes, dentro de las 5 iteraciones se logra un 83% en promedio. Esta investigación fue considerada debido a la comparación de las técnicas y la utilización de la metodología KDD, la cual nos permite conocer la forma en la que se maneja a diferencia de CRISP-DM.

Rodríguez y Miñano [6], menciona que para el crecimiento económico de las entidades financieras es importante la evaluación de préstamos crediticios. La presente investigación plantea como objetivo general “mediante un modelo de machine learning determinar la mejora en la evaluación de préstamos crediticios”. Finalmente concluyen que los préstamos crediticios se lograron mejorar, además de generar una mayor ganancia debido a los préstamos crediticios y por último disminuir el dinero perdido por préstamos crediticios. Esta investigación fue considerada debido a su forma de presentar los resultados finales del modelo y la utilización de la metodología CRISP-DM.

### **2.1.3. Antecedentes locales**

Medina y Ulfe [8], la cooperativa de ahorro y crédito afronta diversos problemas, siendo el principal problema la morosidad que cada año es variable. Dicha investigación plantea como objetivo general “construir un Modelo CS en la cartera de crédito personal”, se utilizó SAS Enterprise Miner 13.1 y para el modelamiento utilizaron técnicas de redes neuronales, arboles de clasificación y regresión logística. Finalmente determinaron que el último modelo mencionado fue el mejor que se basó en indicadores de predictibilidad y eficiencia. Esta investigación fue considerada debido a la realización de una comparativa entre las tres técnicas mencionadas y la utilización de la metodología CRISP-DM.

## **2.2. Bases teóricas científicas**

### **2.2.1. Credit scoring (CS)**

Thomas et al [13], lo define como el conjunto de modelos y técnicas que apoyan a las entidades financieras al momento de un otorgamiento de crédito. Además, Millán y Caicedo [14], define a CS como un modelo estadístico utilizando información histórica de los clientes permitiendo estimar la probabilidad de que

incumplan con los pagos y además de obtener un indicador donde se pueda distinguir entre un buen y mal pagador.

### **Tipos de CS**

Según Lawrence y Solomon [15], los tipos de CS son los que se mencionan a continuación:

- A. Applicant o de admisión:** Modelo aplicado cuando se requiere una nueva solicitud de crédito y a su vez también se utiliza para la aprobación de clientes.
- B. Behavior o modelos de comportamiento:** Modelo aplicado en los casos de seguimiento de los créditos, préstamos personales, créditos preaprobados, entre otros.
- C. Collection:** Modelo aplicado en la gestión de cobranzas en tramos de 30, 60 y 90 días de mora, a su vez se puede utilizar para los créditos castigados.
- D. Rating:** Modelo aplicado en la banca de negocios utilizando información financiera de los clientes, el uso de los sobregiros, garantías, entre otros. Es el más utilizado.

### **Técnicas aplicadas en los modelos CS:**

#### **A. Regresión logística**

Medina y Ulfe [8], pág.38, “es una técnica que se aplica cuando se relacionan las variables dependientes e independientes”. Además, Microsoft [16], menciona que la técnica utiliza una curva que lleva una forma de "S" la cual permite que los datos se puedan observar en grupos y además de ser una técnica eficaz.

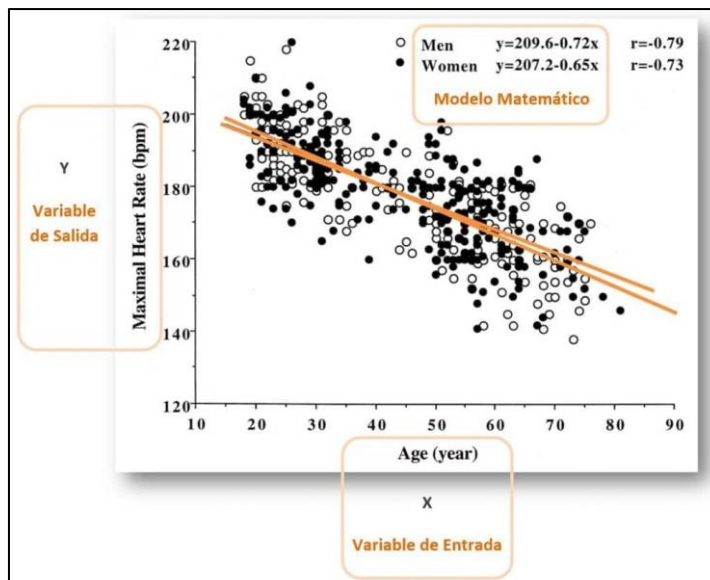


Fig. 2. Regresión Logística

## B. Árboles de clasificación

Gámez y García [17], son denominados arboles de decisión o arboles de identificación, dicha técnica se basa en atributos y que permite dividir en dos o más a una población. Fernández [18], “la técnica tiene forma de "árbol", donde "ramas" representan un conjunto de decisiones”.

Consta de un conjunto de atributos:

$$X = x_1, x_2, \dots, x_k$$

Donde:

$x_1, \dots, x_k$  = Entradas, estas entradas devuelven una respuesta Y.

Si tomamos valores discretos se denomina clasificación y cuando tomamos valores continuos se denomina regresión.

Estructura:

- **Nodo raíz:** Representa toda la población o muestra.
- **Nodo intermedio:** Denominado segmento intermedio, se pueden generar dos o más segmentos descendientes.
- **Nodo terminal:** Denominado segmento terminal, es el último de los segmentos el cual ya no se puede subdividir.

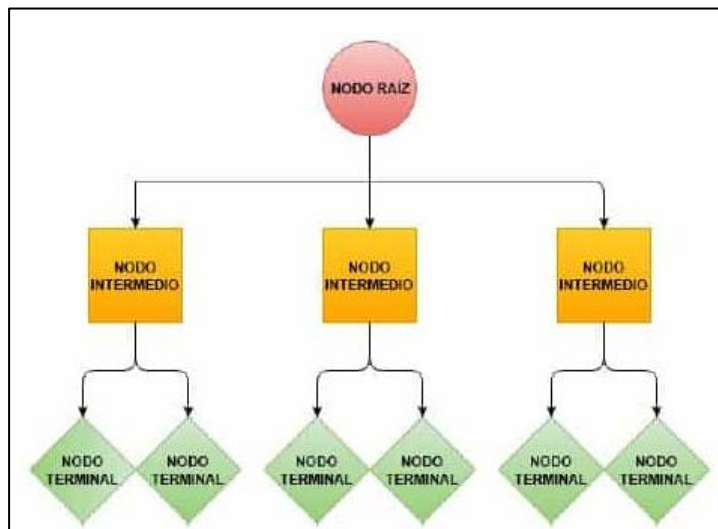


Fig. 3. Estructura de un árbol

### Algoritmos:

- **Arboles de regresión y clasificación (C&RT o CART):** Diseñado por L. Brieman, algoritmo binario, realiza particiones de los datos a su vez indica subconjuntos precisos e iguales.
- **CHAID:** Elaborado por Kass, algoritmo estadístico multidireccional, dicho algoritmo explora los datos y construye segmentos.
- **C4.5:** Diseñado por Ross Quinlan, extensión del algoritmo ID3, algoritmo que se utiliza para la clasificación de los datos.

### C. Redes neuronales

Marín [19], consta de un patrón de nodos y conexiones que replican las actividades del cerebro humano, son sistemas definidos por funciones ( $f$ ). Las entradas al sistema se denominan vector “ $X$ ” y las salidas denominadas vector “ $Y$ ”; la relación que existe entre ambas es denotada por la función  $Y = f(X, W)$ , donde “ $W$ ” representa los pesos de la red.

#### Componentes de una red neuronal:

- **Entradas:** Variables independientes “ $X_j$ ” ingresan a la neurona “ $j$ ” y se representa a continuación:

$$j = 1, 2, \dots, R$$

#### Donde:

“ $R$ ”  $\rightarrow$  N° de variables independientes.

- **Pesos:** Las entradas tienen un peso relativo “ $W_{ij}$ ” que determinan la importancia de cada una de ellas en la función.

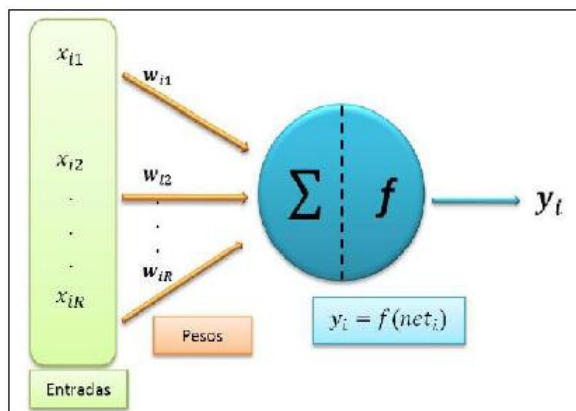


Fig. 4. Componentes de una red neuronal

#### Estructura de una red neuronal:

- **Función de entrada:** La función permite obtener el valor de la entrada neta “ $h_i$ ” a partir de las entradas y los pesos. Siendo la función suma ponderada la que más se utiliza, como se muestra a continuación:

$$h_i(t) = \sum_{j=1}^R W_{ij} * X_j$$

- **Función de activación:** La función es denotada como se muestra a continuación y se da entre la suma ponderada y los sesgos (*bias*).

$$a_i(t) = f_i(b_i, h_i(t))$$

- **Función de salida:** La función “ $Y_i(t)$ ” es equivalente al valor resultante de la función activación.

$$Y_i(t) = F_i(a_i(t)) = a_i(t)$$

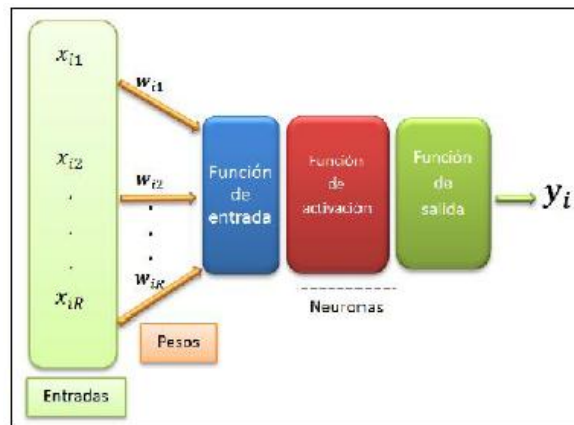


Fig. 5. Red con "n" entradas y una salida

**Modelos:** Según Serrano y Martin del Brío [20] existen diferentes modelos desarrollados en distintos estudios, entre ellos se encuentran los siguientes:

- Adaline
- Backpropagation
- Redes neuronales de aprendizaje competitivo.
- Redes de Hopfiel, entre otras.

#### D. Reglas de clasificación

Stefanowski y Wilk [33], las reglas de decisión son una de las formas de representación del conocimiento típicas para formalizar el conocimiento descubierto, además, las reglas constituyen bloques de conocimiento y los expertos del dominio pueden fácilmente analizarlas individualmente.

#### Estructura:

Se expresan de la forma:

*If P then Q*

#### Donde:

**P:** parte condicional formada por una conjunción de condiciones elementales ( $p_1, p_2, \dots, p_k$ ).

**Q:** parte de decisión que asigna un valor de decisión (clase) a un objeto que cumple la condición.

**Algoritmos:****➤ OneR (Una regla)**

[34], Es un algoritmo de clasificación simple pero preciso que genera una regla para cada predictor en los datos, luego selecciona la regla con el error total más pequeño como su "regla única".

**➤ ZeroR**

Vijayarani y Muthulakshmi [35], Este se basa en el objetivo e ignora todos los predictores. El clasificador ZeroR predice simplemente la categoría mayoritaria (clase).

**2.2.2. Minería de datos (MD)**

Gonzales [25], nos indica que después de identificar los patrones, se pueden hacer predicciones con nuevos datos que se incorporen al sistema.

A su vez Rivera [23], lo define como un mecanismo que a partir de los datos con lo que se cuenta se pueda crear un conocimiento con el objetivo de realizar un análisis exhaustivo.



*Fig. 6. Visión de MD*

## Tipos de datos:

Pérez [26], establece dos tipos de datos:

### A. Categóricos

Son los datos cualitativos, dichos datos se pueden asignar a una categoría correspondiente. Existen dos tipos:

#### ➤ Nominales

Datos con una asignación, por ejemplo: los colores, etc.

#### ➤ Ordinales

Datos ordinales, por ejemplo: grados académicos, etc.

### B. Magnitudes

Son los datos cuantitativos, es decir, que pueden medirse o contarse, existen dos tipos:

#### ➤ Discretos

Datos precisos, ejemplo: el n° de horas, la edad, etc.

#### ➤ Continuos

Datos definidos en un intervalo, por ejemplo: grados, el peso, etc.



Fig. 7. Tipos de datos

## Modelos de aprendizaje

### A. Aprendizajes supervisados

Alpaydin [27], se ingresa el conjunto de datos la cual se conoce cuál será el resultado. A su vez, Sánchez [28] menciona que se basa en una variable objetivo y variables que complementan y ayudan a la predicción de una variable salida.

## **B. Aprendizajes no supervisados**

Sánchez [28], aquellos que al conocerse los datos se identifican las pautas o estándares a seguir.

### **2.2.3. Términos financieros**

#### **Crédito**

Ross et al [9], Entendemos por crédito, a una inversión que realiza un cliente y que se relaciona con la venta de un producto o servicio, con el objetivo de generar e incrementar las ventas. [8] complementa que a través de una entidad financiera un cliente dispone de una cantidad de dinero pactada mediante determinadas condiciones.

#### **Riesgo**

Echemendía [11], define al riesgo como la posibilidad de incumplimientos o la probabilidad de tener un resultado no deseado.

#### **Crédito de riesgo**

Según Chorafas [12], menciona que un cliente a la fecha de vencimiento de su crédito no cumple con el contrato por diferentes causas.

#### **Riesgos en un préstamo crediticio**

Llorca [22], un préstamo crediticio, contiene una serie de beneficios, y a la vez generan riesgos que se deben tener en cuenta, algunos de ellos son los siguientes:

- Conocer el comportamiento del cliente
- Evitar tener varios créditos abiertos.
- Conocer el crédito obtenido para evitar situaciones inesperadas.

## **III. Materiales y métodos**

### **3.1. Tipo de investigación**

Se utilizará el tipo de investigación aplicada. Se desarrollará un modelo de CS para predecir el riesgo en el otorgamiento de créditos en la financiera Efectiva.

El nivel de la investigación es cuantitativa cuasiexperimental.

### 3.2. Diseño de investigación

Basándose en lo establecido por Hernández et al [22], se utilizó el diseño de contrastación preprueba/posprueba en un solo grupo, como se muestra a continuación.

$$O_1 \rightarrow X \rightarrow O_2$$

**Donde:**

$O_1$  = Medición del nivel de precisión en la predicción de riesgo en el otorgamiento de créditos antes de aplicar el modelo Credit Scoring.

$X$  = Desarrollo del modelo Credit Scoring

$O_2$  = Medición del nivel de precisión en la predicción de riesgo en el otorgamiento de créditos después de aplicar el modelo Credit Scoring.

### 3.3. Población y muestra

#### 1. Población

Basándose en Hernández [22], la población para la presente investigación será de tipo censal debido a que se tomarán todos los registros crediticios emitidos dentro de la financiera.

### 3.4. Variables

#### 1. Variable independiente

Modelo de Credit Scoring.

#### 2. Variable dependiente

Predicción de riesgo en el otorgamiento de créditos.

### 3.5. Métodos de investigación

Los métodos que se utilizaran son los siguientes.

TABLA I. Métodos de investigación

Método	Descripción
Analítico	Se realiza un análisis completo de la organización y conocer los problemas que presenta la financiera de la ciudad de Chiclayo.
Deductivo	Desarrollar la estrategia de solución frente a los problemas en el proceso de otorgamiento de créditos.
Implementación	Se pone en marcha la estrategia de solución para apoyar en el proceso de otorgamiento de créditos.

### 3.6. Técnicas e instrumentos de recolección de datos

Los métodos e instrumentos que se utilizaran son los siguientes.

TABLA II. Técnicas e instrumentos

TÉCNICAS	INSTRUMENTOS	ELEMENTOS DE LA POBLACIÓN	PROPÓSITO
Revisión de la literatura	Fichas bibliográficas	Artículos científicos y tesis	Conocer el estado del arte como sustento del proyecto.
Fuentes internas	Reportes, BD (Excel)	Clientes	Comportamiento de clientes para la futura predicción.
Entrevista	Guía de entrevista (Anexo 06)	Gerente de la financiera Efectiva y Analista de créditos	Conocer el giro del negocio de la Financiera.

### 3.7. Procedimientos

#### 3.7.1. Metodología de desarrollo

A continuación, podemos observar el proceso que se lleva a cabo en la metodología CRISP-DM con sus respectivas actividades, además las fases están relacionadas de forma bidireccional lo que permite retroceder de una fase a otra.

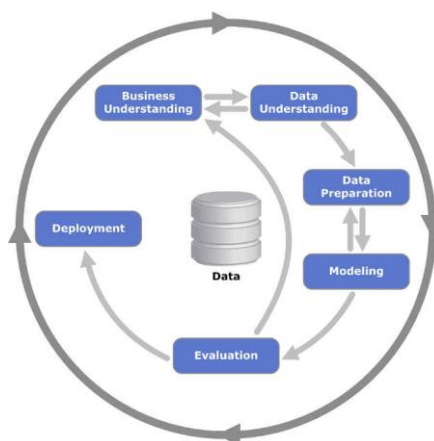


Fig. 8. Ciclo de vida – CRISP-DM

## A. Fase #1: Comprensión del negocio

Actividades:

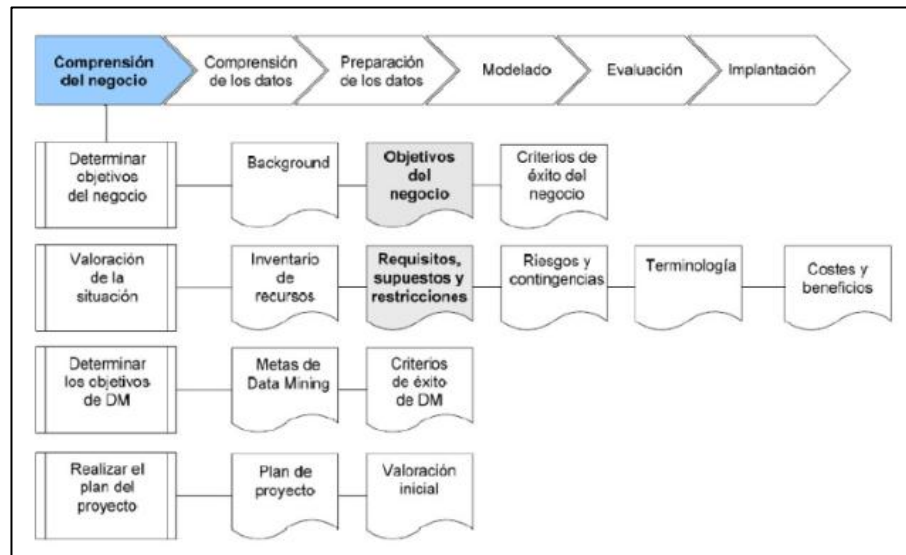


Fig. 9. Fase 1: Comprensión del negocio

## B. Fase #2: Comprensión de los datos

Actividades:

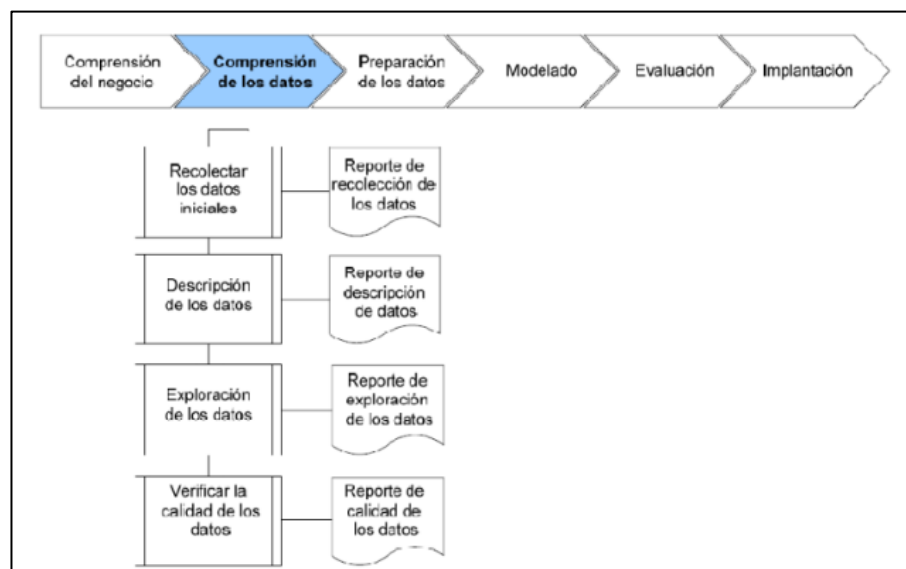


Fig. 10. Fase 2: Comprensión de los datos

### C. Fase #3: Preparación de los datos

Actividades:

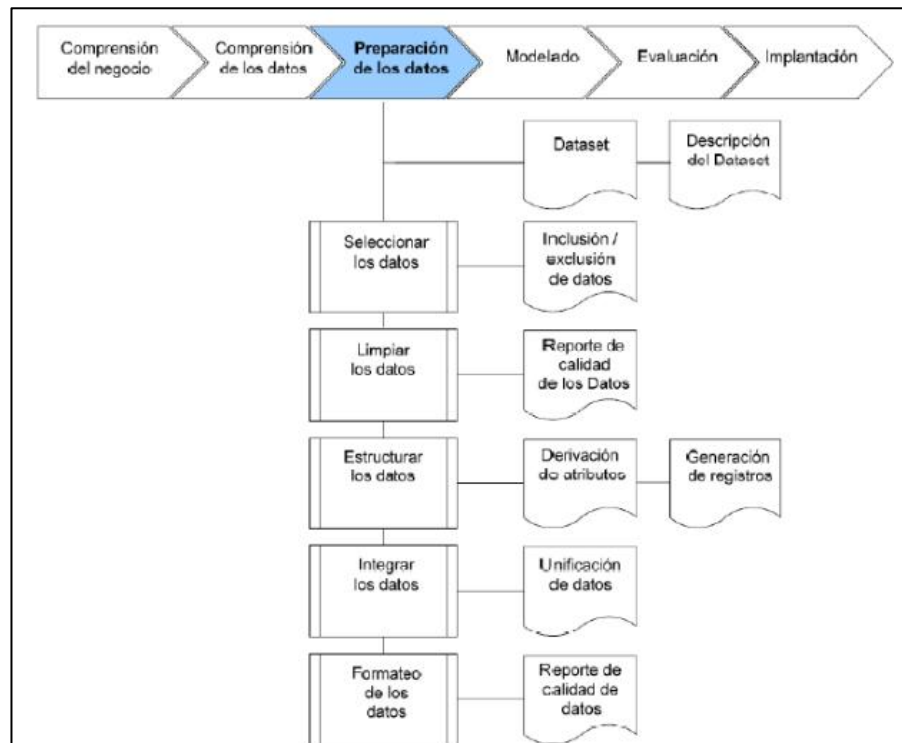


Fig. 11. Fase 3: Preparación de los datos

### D. Fase #4: Modelamiento

Actividades:

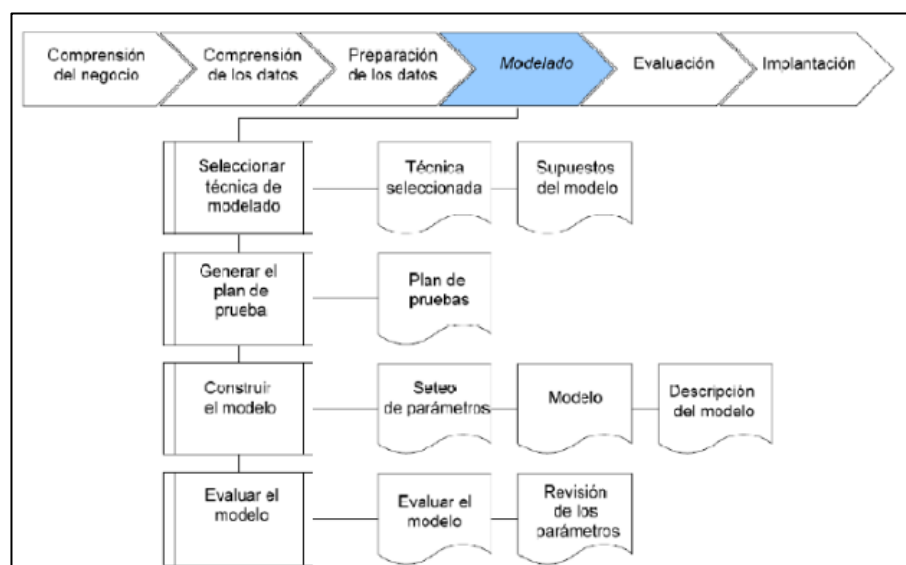
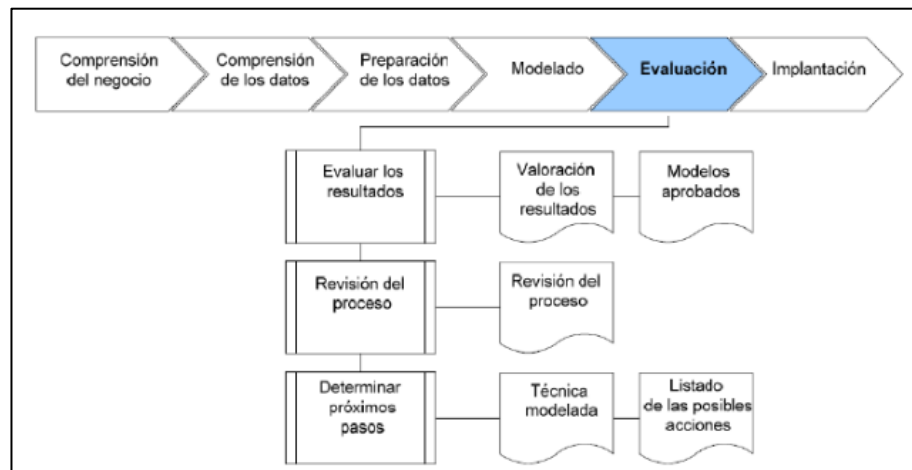


Fig. 12. Fase 4: Modelado

## E. Fase #5: Evaluación

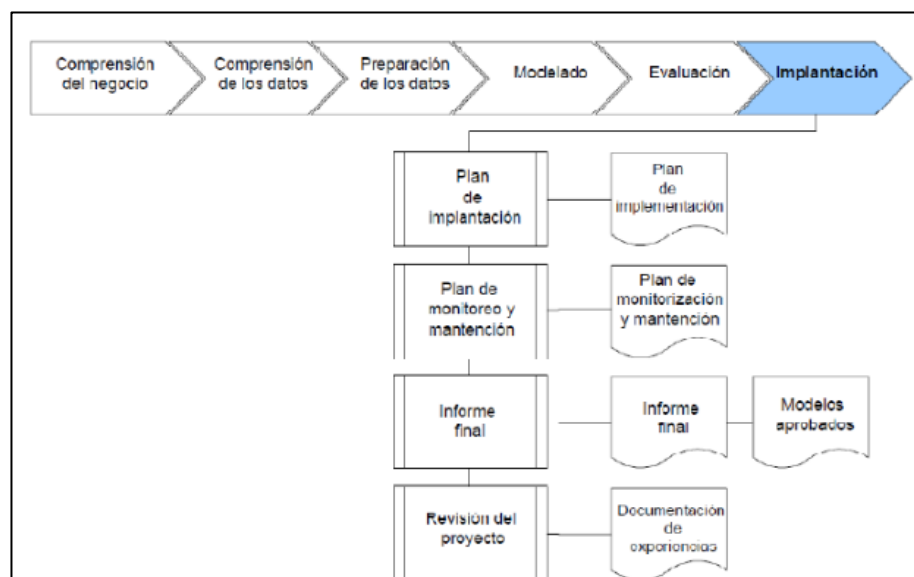
Actividades:



*Fig. 13. Fase 5: Evaluación*

## F. Fase #6: Despliegue

Actividades:

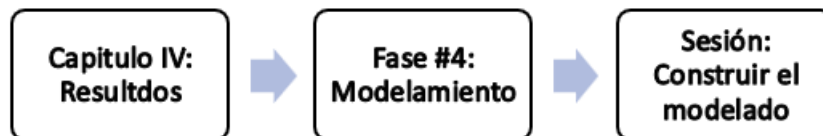


*Fig. 14. Fase 6: Implantación*

### 3.7.2. Producto acreditable

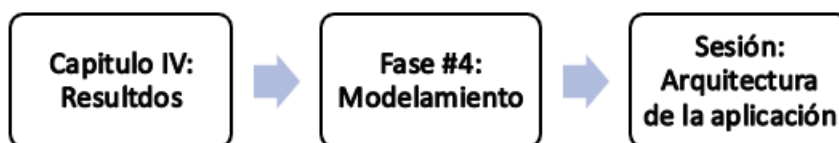
#### A. Interfaces

Las interfaces del sistema propuesto fueron diseñadas utilizando los lenguajes HTML, Javascript, CSS, PHP y el framework Bootstrap, las mismas que se presentan en:



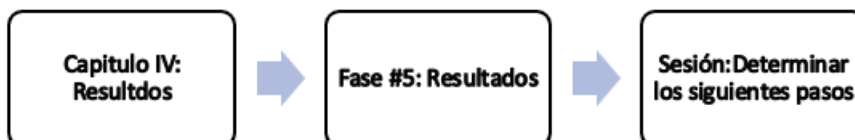
#### B. Arquitectura

Se ha creado una arquitectura ideal para la operación del sistema, que se detalla en:



#### C. Infraestructura tecnológica

En base a la arquitectura descrita en el punto anterior, se describe los componentes y sus respectivas características:



### 3.7.3. Manual de usuario

Se realizó, para ayudar a los usuarios en la utilización adecuada del sistema. Se explica detalladamente el proceso de registro de información de los clientes, el procesamiento y la obtención de resultados (Anexo N° 03).

### 3.8. Matriz de consistencia

*TABLA III. Matriz de consistencia*

<u>FORMULACIÓN DEL PROBLEMA</u>		<u>MÉTODOLÓGÍA DE INVESTIGACIÓN</u>		
¿Cómo mejoraría la implementación de un sistema en la predicción de riesgo en el otorgamiento de créditos en la Financiera Efectiva?		<u>TIPO DE INVESTIGACIÓN</u>		
		Investigación aplicada		
<u>OBJETIVO GENERAL</u>	<u>MÉTODO</u>	<u>DESCRIPCIÓN</u>		
Implementar un modelo Credit Scoring, para mejorar la toma de decisiones en el proceso de otorgamiento de créditos en la Financiera Efectiva del grupo EFE	Analítico	Se realiza un análisis completo de la organización y conocer los problemas que presenta la financiera de la ciudad de Chiclayo		
	Deductivo	Desarrollar la estrategia de solución frente a los problemas en el proceso de otorgamiento de créditos.		
	Implementación	Se pone en marcha la estrategia de solución para apoyar en el proceso de otorgamiento de créditos.		
	<u>TÉCNICAS</u>	<u>INSTRUMENTOS</u>	<u>ELEMENTOS DE LA POBLACIÓN</u>	<u>PROPÓSITO</u>
	Revisión de la literatura	Fichas bibliográficas	Artículos científicos y tesis	Conocer el estado del arte como sustento del proyecto.
	Fuentes internas	Reportes, BD (Excel)	Clientes	Comportamiento de clientes para la futura predicción.
	Entrevista	Guía de entrevista (Anexo 06)	Gerente de la financiera Efectiva y Analista de créditos	Conocer el giro del negocio de la Financiera.
<u>OBJETIVOS ESPECÍFICOS</u>	<u>DESCRIPCIÓN DEL LOGRO DE LOS OBJETIVOS ESPECÍFICOS</u>		<u>INDICADORES</u>	
Determinar la técnica de credit scoring que reúna los requisitos de predictibilidad del sistema, para la toma de decisiones en la asignación de créditos.	Se pretende lograr determinar la técnica que ayude al modelo predictivo evaluándolas para lograr un nivel de eficiencia del modelo.		Métricas de evaluación Dimensiones por evaluar.	

<p>Implementar escenarios de predicción aplicando la técnica de Credit Scoring seleccionada para predecir el nivel de riesgo en el otorgamiento de crédito.</p>	<p>Se pretende implementar diferentes escenarios de predicción para predecir el nivel de riesgo en el otorgamiento de crédito.</p>	<p># Escenarios de predicción Construcción del modelo CS</p>
<p>Medir el nivel de precisión del sistema para determinar la confiabilidad de los resultados del proceso.</p>	<p>Se obtendrá un grado de precisión calculada a través de la comparación de los resultados del algoritmo con respecto a la realidad del contexto.</p>	<p>Porcentaje de nivel precisión = aciertos / totales</p>
<p>Validar el grado de usabilidad del modelo credit scoring por parte de los usuarios finales.</p>	<p>Se obtendrá un grado de usabilidad calculada a través del System Usability Scale (SUS), para conocer el puntaje de eficiencia y usabilidad del sistema.</p>	<p>Puntaje obtenido a través de System Usability Scale (Puntaje de usabilidad y eficiencia).</p>

---

### 3.9. Consideraciones éticas

A continuación, listamos las consideraciones tomadas en cuenta con el objetivo de la protección, bienestar y seguridad de los datos (protección) de los participantes, en este caso una empresa financiera ubicada en la ciudad de Chiclayo, teniendo en cuenta las siguientes consideraciones:

- La información obtenida de las entrevistas, análisis documental y observación es de carácter privado y pertenece a la empresa, siendo su tratamiento únicamente para fines de esta investigación.
- La información registrada y obtenida en la aplicación de la presente investigación es de carácter privado y pertenece a la empresa, siendo su tratamiento confidencial.

## IV. Resultados y discusión

### 4.1. En base a la metodología utilizada

La metodología que se desarrolla en el presente proyecto es CRISP-DM. [12], ayuda a orientar proyectos basados en minería de datos, además de ser una metodología flexible y fácil de personalizar.

#### 4.1.1. Fase 1: Comprensión del negocio

##### 1. Determinar los objetivos del negocio

###### A. Contexto

En este apartado podremos conocer en qué estado se encuentra la organización (Financiera Efectiva), se realizó una entrevista (Anexo 06), al gerente para conocer el Core del negocio y al analista de créditos se pudo conocer que actualmente el proceso de crédito es obsoleto (Ver figura 1), además de una demora en el proceso de otorgamiento de crédito y que por último se tiene un 7.4% de índice de morosidad.

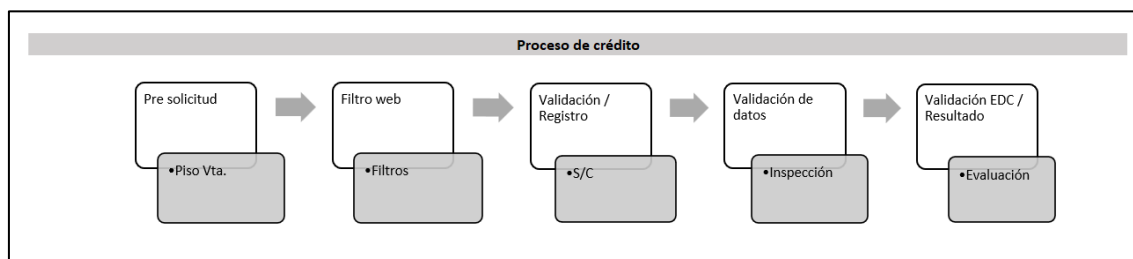


Fig. 15. Proceso de otorgamiento de créditos

## B. Objetivos del negocio

La organización en estudio ha brindado los siguientes objetivos que ha definido en su PLAN ESTRATEGICO:

- 1) Brindar un servicio de calidad
- 2) Mejorar el proceso de evaluación de créditos, con el objetivo de minimizar el índice de morosidad (actualmente 7.4%).
- 3) Maximizar la rentabilidad.
- 4) Minimizar el riesgo de incumplimiento de pago.

## 2. Evaluar la situación problemática del negocio

### A. Inventario de recursos

#### Recursos humanos

*TABLA IV. Recursos humanos*

<b>RECURSOS HUMANOS</b>	
Gerente General	C. H.
Gerente de Cobranzas	M. P.
Gerente de Riesgos.	J. B.
Gerente Legal	E. C.
Analista de Créditos	M. Q.
Asesores de Créditos	<ul style="list-style-type: none"> <li>✓ M. S.</li> <li>✓ V. S.</li> <li>✓ D. D.</li> <li>✓ M. Á.</li> <li>✓ L. V.</li> <li>✓ E. G.</li> <li>✓ M. C.</li> <li>✓ L. U</li> </ul>

**Fuente de datos***TABLA V. Fuente de datos*

<b>FUENTE DE DATOS</b>	
Base de Datos	Data de la Financiera Efectiva.
Informes de las diferentes áreas	Información que brindan las áreas de la empresa.
Normas	Normas técnicas (precisar).
Recurso Humano	Analista y asesores de créditos

**B. Requisitos y restricciones****Requisitos para solicitar un crédito***TABLA VI. Requisitos para un crédito*

<b>REQUISITOS</b>	
1	Copia de DNI por ambos lados.
2	Copia de recibo de luz o agua.
3	Si es formal, boletas de pago.
4	Si es comerciante, boletas o facturas del giro del negocio.

**Restricciones de los involucrados***TABLA VII. Restricciones de los involucrados*

<b>RESTRICCIONES DE LOS INVOLUCRADOS</b>	
1	Falta de liquidez.
2	Falta de capacidad para prestar o emitir acciones.
3	Falta de recursos humanos en el otorgamiento de créditos.
4	Demora en las solicitudes de crédito.
5	Un cliente no puede ser mayor de los 70 años.
6	Nivel de endeudamiento del cliente (No mayor al 35% de sus ingresos).

### C. Terminología financiera

*TABLA VIII. Terminología financiera*

<b>TERMINOLOGÍA FINANCIERA</b>	
Activo	Elemento que tiene valor patrimonial.
Amortización	Reducción gradual de una deuda a través de pagos periódicos.
Capital	Es el monto del préstamo o crédito.
Fecha de desembolso	Fecha en la que el cliente recibe el crédito.
ITF	Impuesto a las transacciones financieras.
Monto de la cuota	Monto que se genera para la cancelación del préstamo.
Nº de cuotas	Número de cuotas pactadas para hacer efectivo la cancelación del préstamo.
Score	Utilizado por las entidades financieras para aprobar o denegar créditos.
Seguro de Desgravamen	Es la póliza de seguro vinculada al préstamo.
Tasa de interés	Es la valoración del precio del dinero, es decir, el reconocimiento por cada 100 unidades prestadas en un periodo tiempo.
Tasa Efectiva Anual (TEA)	Es la tasa de interés expresada anualmente.
Tasa de Costo Efectiva Anual (TCEA)	Es la tasa de interés que expresa el costo real del crédito, aquí se incluyen los intereses, comisiones y gastos.
Tasa Efectiva Mensual (TEM)	Es la tasa de interés expresada mensualmente.
ROI	Retorno de la inversión, utilizada para medir el resultado económico generado por las inversiones realizadas.

### 3. Determinar el objetivo que cumplirá la metodología

Construir el modelo CS con la técnica seleccionada que permita predecir el otorgamiento de crédito a clientes clasificados, según escenarios de predicción.

Los escenarios se definen de acuerdo con las siguientes variables.

- Código Cliente (dni)
- Datos de identificación del Cliente (Apellidos y nombres)
- Monto Solicitado (con tipo de moneda)
- Situación laboral (independiente / dependiente)
- Formalidad laboral (informal / formal)
- Casa Propia (Con / sin)
- Ingresos del cliente
- Historial crediticio

### 4. Desarrollar el plan del proyecto

#### A. Equipo de trabajo

*TABLA IX. Equipo de trabajo*

N°	ROL	MIEMBRO
1	Estuante	Carlos Fabricio Olivos Seclén
2	Asesor	María Ysabel Aranguri García
3	Analista de crédito	Mario Q.
4	Gerente General	Cesar H.

#### B. Requerimientos

Se desarrollaron en base al estándar IEEE Std. 830 [31], la cual se basa en un formato para llevar a cabo la Especificación de Requisitos de Software (ERS).

**Requerimientos funcionales (R. F.)***TABLA X. R.F. 01*

<b>Identificación del requerimiento</b>	<b>del</b>	RF01
<b>Nombre del requerimiento</b>	<b>del</b>	Autenticación de usuarios
<b>Descripción del requerimiento</b>	<b>del</b>	Los usuarios deben identificarse para acceder al sistema (Administrador / asesor de crédito).
<b>Requerimiento funcional</b>	<b>no</b>	<ul style="list-style-type: none"> <li>✓ RNF01</li> <li>✓ RNF02</li> <li>✓ RNF05</li> <li>✓ RNF06</li> </ul>

*TABLA XI. R.F. 02*

<b>Identificación del requerimiento</b>	<b>del</b>	RF02
<b>Nombre del requerimiento</b>	<b>del</b>	Registro de usuarios
<b>Descripción del requerimiento</b>	<b>del</b>	El usuario administrador tendrá la opción de registrar usuarios con los permisos correspondientes para acceder al sistema.
<b>Requerimiento no funcional</b>	<b>no</b>	<ul style="list-style-type: none"> <li>✓ RNF01</li> <li>✓ RNF02</li> <li>✓ RNF05</li> <li>✓ RNF06</li> </ul>

*TABLA XII. R.F. 03*

<b>Identificación del requerimiento</b>	<b>del</b>	RF03
<b>Nombre del requerimiento</b>	<b>del</b>	Modificar usuarios
<b>Descripción del requerimiento</b>	<b>del</b>	El sistema ofrecerá modificar un usuario y otorgarles o denegarles algún permiso.
<b>Requerimiento funcional</b>	<b>no</b>	<ul style="list-style-type: none"> <li>✓ RNF01</li> <li>✓ RNF02</li> <li>✓ RNF05</li> <li>✓ RNF06</li> </ul>

TABLA XIII. R.F. 04

<b>Identificación del requerimiento</b>	<b>del</b>	RF04
<b>Nombre del requerimiento</b>	<b>del</b>	Información del sistema
<b>Descripción del requerimiento</b>	<b>del</b>	El sistema ofrecerá a los usuarios información de los diferentes módulos de acuerdo con el permiso que tengan.
<b>Requerimiento funcional</b>	<b>no</b>	<ul style="list-style-type: none"> <li>✓ RNF01</li> <li>✓ RNF02</li> <li>✓ RNF05</li> <li>✓ RNF06</li> </ul>

TABLA XIV. R.F. 05

<b>Identificación del requerimiento</b>	<b>del</b>	RF05
<b>Nombre del requerimiento</b>	<b>del</b>	Registro de clientes
<b>Descripción del requerimiento</b>	<b>del</b>	El sistema ofrecerá registrar un nuevo cliente.
<b>Requerimiento funcional</b>	<b>no</b>	<ul style="list-style-type: none"> <li>✓ RNF01</li> <li>✓ RNF02</li> <li>✓ RNF05</li> <li>✓ RNF06</li> </ul>

TABLA XV. R.F. 06

<b>Identificación del requerimiento</b>	<b>del</b>	RF06
<b>Nombre del requerimiento</b>	<b>del</b>	Modificar clientes
<b>Descripción del requerimiento</b>	<b>del</b>	El sistema ofrecerá modificar un cliente.
<b>Requerimiento funcional</b>	<b>no</b>	<ul style="list-style-type: none"> <li>✓ RNF01</li> <li>✓ RNF02</li> <li>✓ RNF05</li> <li>✓ RNF06</li> </ul>

TABLA XVI. R.F. 07

<b>Identificación del requerimiento</b>	<b>del</b>	RF07
<b>Nombre del requerimiento</b>	<b>del</b>	Exportar a Excel y PDF.
<b>Descripción del requerimiento</b>	<b>del</b>	El sistema ofrecerá exportar la tabla de datos del cliente a Excel y PDF.
<b>Requerimiento funcional</b>	<b>no</b>	<ul style="list-style-type: none"> <li>✓ RNF01</li> <li>✓ RNF02</li> <li>✓ RNF05</li> <li>✓ RNF06</li> </ul>

TABLA XVII. R.F. 08

<b>Identificación del requerimiento</b>	<b>del</b>	RF08
<b>Nombre del requerimiento</b>	<b>del</b>	Predecir el riesgo en el otorgamiento de créditos
<b>Descripción del requerimiento</b>	<b>del</b>	El sistema ofrecerá predecir el nivel de riesgo de un cliente para acceder a un crédito.
<b>Requerimiento funcional</b>	<b>no</b>	<ul style="list-style-type: none"> <li>✓ RNF01</li> <li>✓ RNF03</li> <li>✓ RNF04</li> <li>✓ RNF05</li> <li>✓ RNF06</li> </ul>

TABLA XVIII. R.F. 09

<b>Identificación del requerimiento</b>	<b>del</b>	RF09
<b>Nombre del requerimiento</b>	<b>del</b>	Resultado de predicción
<b>Descripción del requerimiento</b>	<b>del</b>	El sistema ofrecerá al usuario una respuesta a la predicción mediante una semaforización:
<b>Requerimiento funcional</b>	<b>no</b>	<ul style="list-style-type: none"> <li>✓ Nivel alto (Color Rojo)</li> <li>✓ Nivel medio (Color Amarillo)</li> <li>✓ Nivel bajo (Color Verde)</li> <li>✓ RNF05</li> <li>✓ RNF06</li> </ul>

TABLA XIX. R.F. 10

<b>Identificación del requerimiento</b>	<b>del</b>	RF10
<b>Nombre del requerimiento</b>	<b>del</b>	Información de créditos
<b>Descripción del requerimiento</b>	<b>del</b>	El sistema ofrecerá al usuario consultar sobre los créditos otorgados y denegados.
<b>Requerimiento funcional</b>	<b>no</b>	<ul style="list-style-type: none"> <li>✓ RNF05</li> <li>✓ RNF06</li> </ul>

TABLA XX. R.F. 11

<b>Identificación del requerimiento</b>	<b>del</b>	RF11
<b>Nombre del requerimiento</b>	<b>del</b>	Importar data en formato (CSV)
<b>Descripción del requerimiento</b>	<b>del</b>	El sistema ofrecerá al usuario importar data de forma masiva para su predicción.
<b>Requerimiento funcional</b>	<b>no</b>	<ul style="list-style-type: none"> <li>✓ RNF01</li> <li>✓ RNF05</li> <li>✓ RNF06</li> </ul>

### Requerimientos no funcionales (R.N. F.)

TABLA XXI. R.N.F. 01

<b>Identificación del requerimiento</b>	<b>del</b>	RNF01
<b>Nombre del requerimiento</b>	<b>del</b>	Interfaz del sistema
<b>Descripción del requerimiento</b>	<b>del</b>	El sistema presentará interfaces de usuario sencillas y amigables.
<b>Prioridad del requerimiento: Alta</b>		

TABLA XXII. R.N.F.02

<b>Identificación requerimiento</b>	<b>del</b>	RNF02
<b>Nombre requerimiento</b>	<b>del</b>	Nivel de usuario
<b>Descripción requerimiento</b>	<b>del</b>	Garantiza el acceso al sistema de acuerdo con el nivel de usuario.
<b>Prioridad del requerimiento: Alta</b>		

TABLA XXIII. R.N.F.03

<b>Identificación requerimiento</b>	<b>del</b>	RNF03
<b>Nombre requerimiento</b>	<b>del</b>	Evaluación de créditos rápida.
<b>Descripción requerimiento</b>	<b>del</b>	El sistema ofrecerá al usuario determinar en menor tiempo la evaluación del crédito.
<b>Prioridad del requerimiento: Alta</b>		

TABLA XXIV. R.N.F.04

<b>Identificación requerimiento</b>	<b>del</b>	RNF04
<b>Nombre requerimiento</b>	<b>del</b>	Tiempos de respuesta.
<b>Descripción requerimiento</b>	<b>del</b>	El sistema ofrecerá navegar en un tiempo menor con menor cantidad de “clics”.
<b>Prioridad del requerimiento: Alta</b>		

TABLA XXV. R.N.F.05

<b>Identificación requerimiento Nombre requerimiento</b>	<b>del del</b>	RNF05 Confiabilidad del sistema
<b>Descripción requerimiento</b>	<b>del</b>	El sistema tendrá que estar en funcionamiento las 24 horas del día / 7 días a la semana, debido a que es una página web.
<b>Prioridad del requerimiento: Alta</b>		

TABLA XXVI. R.N.F.06

<b>Identificación requerimiento Nombre requerimiento</b>	<b>del del</b>	RNF06 Seguridad en información
<b>Descripción requerimiento</b>	<b>del</b>	Garantizar la seguridad del sistema con respecto a la información y datos (Documentos, datos personales, contraseñas).
<b>Prioridad del requerimiento: Alta</b>		

### C. Fases de desarrollo

*TABLA XXVII. Fase 1*

<b>Fase 1</b>	
<b>Nombre:</b>	Comprensión del negocio
<b>Descripción:</b>	<p>En esta fase se evalúa la situación actual del negocio a su vez se conoce los objetivos planteados del negocio:</p> <p>Actividades:</p> <ul style="list-style-type: none"> <li>✓ Determinar los objetivos del negocio.</li> <li>✓ Evaluar la situación actual del negocio.</li> <li>✓ Determinar el objetivo que cumplirá la metodología CRISP-DM.</li> <li>✓ Desarrollar el plan del proyecto.</li> </ul>

*TABLA XXVIII. Fase 2*

<b>Fase 2</b>	
<b>Nombre:</b>	Comprensión de los datos
<b>Descripción:</b>	<p>En esta fase se accede a los datos con los que se cuenta de parte de la Financiera Efectiva y exploramos los datos.</p> <p>Actividades:</p> <ul style="list-style-type: none"> <li>✓ Recopilación de datos iniciales.</li> <li>✓ Descripción de los datos.</li> <li>✓ Base de datos (modelo físico).</li> <li>✓ Verificar la calidad de los datos.</li> </ul>

TABLA XXIX. Fase 3

<b>Fase 3</b>	
<b>Nombre:</b>	Preparación de los datos
<b>Descripción:</b>	<p>En esta fase se procede a seleccionar, limpiar e integrar los datos.</p> <p>Actividades:</p> <ul style="list-style-type: none"> <li>✓ Seleccionar los datos.</li> <li>✓ Limpiar los datos.</li> <li>✓ Exclusión de variables</li> <li>✓ Construir los datos.</li> <li>✓ Integrar los datos</li> <li>✓ Formateo los datos</li> </ul>

TABLA XXX. Fase 4

<b>Fase 4</b>	
<b>Nombre:</b>	Modelamiento
<b>Descripción:</b>	<p>En esta fase se determina que técnica se va a implementar en el modelo CS.</p> <p>Actividades:</p> <ul style="list-style-type: none"> <li>✓ Seleccionar la técnica de modelamiento.</li> <li>✓ Plan de pruebas del modelado.</li> <li>✓ Arquitectura de la aplicación.</li> <li>✓ Construir el modelado.</li> <li>✓ Evaluar el modelado.</li> </ul>

TABLA XXXI. Fase 5

<b>Fase 5</b>	
<b>Nombre:</b>	Evaluación
<b>Descripción:</b>	<p>En esta fase se evalúa los resultados obtenidos.</p> <p>Actividades:</p> <ul style="list-style-type: none"> <li>✓ Evaluar los resultados.</li> <li>✓ Proceso de revisión.</li> <li>✓ Determinar los siguientes pasos.</li> </ul>

TABLA XXXII. Fase 6

<b>Fase 6</b>	
<b>Nombre:</b>	Despliegue
<b>Descripción:</b>	<p>En esta fase se procede a implantar el modelo en el entorno de desarrollo y/o producción.</p> <p>Actividades:</p> <ul style="list-style-type: none"> <li>✓ Planear la implantación.</li> <li>✓ Producir el informe final.</li> </ul>

#### D. Planificación inicial

Se tomará en cuenta el nivel de riesgo de las fases mediante la siguiente formula:

$$NR = C * A$$

Donde:

NR: Nivel de riesgo

C: Consecuencia

A: Amenaza

A su vez, se señala una semaforización: Nivel Bajo (Color Verde), Nivel Medio (Color Amarillo), Nivel Alto (Color Rojo).

TABLA XXXIII. Nivel de riesgo de las fases

Fase	Nombre	Valor de la consecuencia	Probabilidad de amenaza	Nivel del Riesgo (C*A)
1	Comprensión del negocio	2	1	Bajo (2)
2	Comprensión de los datos	2	1	Bajo (2)
3	Preparación de los datos	3	2	Medio (6)
4	Modelamiento	3	5	Alto (15)
5	Evaluación	2	4	Medio (8)
6	Despliegue	2	1	Bajo (2)

#### 4.1.2. Fase 2: Comprensión de los datos

##### 1. Recopilación de datos iniciales

Los datos fueron recogidos de la base de datos de financiera Efectiva, de la cual se tiene entre los años 2019, 2020 y 2021.

A su vez, se realizó un estudio del negocio en base a una entrevista con el analista de créditos de la financiera (Anexo 06) para comprender la naturaleza de los datos obtenidos.

##### 2. Descripción de los datos

En el presente proyecto se tomaron los datos de clientes con un total de 11936 evaluaciones crediticias. Se elaboro un diccionario de datos que se van a utilizar como se muestra a continuación:

TABLA XXXIV. Diccionario tabla 'usuario'

Nombre de tabla: usuario					
Campos	Descripción	Tipo de dato	Tipo de Restricción	Descripción de Restricción	Restricción de Nulidad
idusuario	Identificador de la tabla USUARIO	Integer (11)	Primary key	Tabla: usuario	Not Null
nombre	Nombre del usuario	Varchar (50)	-	-	Not Null
direccion	Dirección del usuario	Varchar (100)	-	-	Null
telefono	Teléfono del usuario	Varchar (15)	-	-	Null
login	Nombre de usuario con el cual inicia sesión	Varchar (20)	-	-	Not Null
clave	Clave del usuario con el cual inicia sesión	Varchar (64)	-	-	Not Null
Imagen	Foto del usuario	Varchar	-	-	Null
estado	Estado (Activo/Inactivo) del usuario	Boolean	check	1: Activo 2: Inactivo	Null

TABLA XXXV. Diccionario tabla 'permisos'

Nombre de tabla: permisos					
Datos	Descripción	Tipo de dato	Tipo de Restricción	Descripción de Restricción	Restricción de Nulidad
Idpermiso	Identificador de la tabla PERMISO	Integer (11)	Primary key	Tabla: permisos	Not Null
Permiso	Descripción del permiso al sistema	Varchar (50)	-	-	Not Null

TABLA XXXVI. Diccionario tabla 'usuariopermiso'

Nombre de tabla: usuariopermiso					
Datos	Descripción	Tipo de dato	Tipo de Restricción	Descripción de Restricción	Restricción de Nulidad
idusuariopermiso	Identificador de la tabla USUARIO-PERMISO	Integer (11)	Primary key	Tabla: usuario-permiso	Not Null
Idusuario	Nombre del usuario	Integer (11)	-	-	Null
Idpermiso	Dirección del usuario	Integer (11)	-	-	Null

TABLA XXXVII. Diccionario tabla 'clientes'

TABLA XXXVIII. Diccionario tabla 'prestamos'

Nombre de tabla: clientes					
Datos	Descripción	Tipo de dato	Tipo de Restricción	Descripción de Restricción	Restricción de Nulidad
idcliente	Identificador de la tabla Clientes	Integer (11)	Primary key	Tabla: Cliente	Not Null
nombre	Nombre del cliente	Varchar (50)	-	-	Not Null
cedula	Cedula del cliente	Varchar (50)	-	-	Not Null
direccion	Dirección del cliente	Varchar (100)	-	-	Null
telefono	Teléfono del cliente	Varchar (15)	-	-	Null
situación_laboral	Situación laboral del cliente	Varchar (150)	-	1: Dependiente 2: Independiente	Not Null
formalidad	Formalidad del cliente	Varchar (150)	-	1: Formal 2: Informal	Not Null
tipo_casa	Tipo de casa del cliente	Varchar (150)	-	1: Propia 2: Sin casa	Not Null
ingresos	Ingresos que percibe el cliente	Integer (11)	-	-	Not Null
estado	Estado del cliente	Boolean	check	1: Validado 2: No validado	Not Null
Nombre de tabla: prestamos					

Datos	Descripción	Tipo de dato	Tipo de Restricción	Descripción de Restricción	Restricción de Nulidad
Idprestamo	Identificador de la tabla PRESTAMOS	Integer (11)	Primary key	Tabla: prestamos	Not Null
Idcliente	Identificador de la tabla CLIENTE	Integer (11)	Foreign Key	Tabla: cliente	Not Null
Idusuario	Identificador de la tabla USUARIO	Integer (11)	Foreign Key	Tabla: usuario	Not Null
Fprestamo	Fecha que se realiza el otorgamiento de crédito	datetime	-	-	Not Null
Monto	Monto que solicita el cliente	Decimal (16,2)	-	-	Not Null
Interés	Interés con el cual se le otorga el crédito	Decimal (16,2)	-	-	Not Null
Formapago	La forma de pago que va a realizar el cliente	Varchar (20)	-	-	Null
Fpago	La fecha de pago del crédito	datetime	-	-	Null
Plazo	Plazo del crédito	Varchar (20)	-	-	Null
Fplazo	La fecha de plazo del crédito	datetime	-	-	Null
Estado	Estado del préstamo	boolean	check	1: Activo 2: Inactivo	Null

TABLA XXXIX. Diccionario tabla 'pagos'

Nombre de tabla: pagos					
Datos	Descripción	Tipo de dato	Tipo de Restricción	Descripción de Restricción	Restricción de Nulidad
Idpago	Identificador de la tabla PAGO	Integer (11)	Primary key	Tabla: pago	Not Null
Idprestamo	Identificador de la tabla PERMISOS	Integer (11)	Foreign Key	Tabla: prestamos	Not Null
Usuario	Usuario que registra el pago	Varchar (50)	-	-	Not Null
Fecha	Fecha que se realiza el pago	datetime	-	-	Not Null
Cuota	Monto que con el cual se realiza el pago	Decimal (16,2)	-	-	Not Null

### 3. BD transaccional (Script)

Se creó la BD en MySQL.

Creación tabla 'usuario':

```
CREATE TABLE `usuarios` (
  `idusuario` int(11) NOT NULL AUTO_INCREMENT,
  `nombre` varchar(50) NOT NULL,
  `direccion` varchar(100) DEFAULT NULL,
  `telefono` varchar(15) DEFAULT NULL,
  `login` varchar(20) NOT NULL,
  `clave` varchar(64) NOT NULL,
  `imagen` varchar(50) DEFAULT NULL,
  `estado` tinyint(4) DEFAULT '1',
  PRIMARY KEY (`idusuario`),
  UNIQUE KEY `idusuario_UNIQUE` (`idusuario`)
) ENGINE=InnoDB AUTO_INCREMENT=6 DEFAULT CHARSET=utf8
```

Fig. 16. Tabla 'usuario'

Creación tabla 'permisos':

```
CREATE TABLE `permisos` (
  `idpermiso` int(11) NOT NULL AUTO_INCREMENT,
  `permiso` varchar(50) NOT NULL,
  PRIMARY KEY (`idpermiso`)
) ENGINE=InnoDB AUTO_INCREMENT=8 DEFAULT CHARSET=utf8
```

Fig. 17. Tabla 'permisos'

Creación tabla 'usuariopermiso':

```
CREATE TABLE `usuariopermiso` (
  `idusuariopermiso` int(11) NOT NULL AUTO_INCREMENT,
  `idusuario` int(11) DEFAULT NULL,
  `idpermiso` int(11) DEFAULT NULL,
  PRIMARY KEY (`idusuariopermiso`),
  UNIQUE KEY `idusuariopermiso_UNIQUE` (`idusuariopermiso`),
  KEY `fkPermiso_idx` (`idpermiso`),
  KEY `fkUsuario_idx` (`idusuario`),
  CONSTRAINT `fkPermiso` FOREIGN KEY (`idpermiso`) REFERENCES `permisos` (`idpermiso`) ON DELETE NO ACTION ON UPDATE NO ACTION,
  CONSTRAINT `fkUsuario` FOREIGN KEY (`idusuario`) REFERENCES `usuarios` (`idusuario`) ON DELETE NO ACTION ON UPDATE NO ACTION
) ENGINE=InnoDB AUTO_INCREMENT=41 DEFAULT CHARSET=utf8
```

Fig. 18. Tabla 'usuariopermiso'

Creación tabla 'clientes':

```
CREATE TABLE `clientes` (
  `idcliente` int(11) NOT NULL AUTO_INCREMENT,
  `nombre` varchar(50) NOT NULL,
  `cedula` varchar(50) NOT NULL,
  `direccion` varchar(100) DEFAULT NULL,
  `telefono` varchar(15) DEFAULT NULL,
  `situacion_laboral` varchar(150) CHARACTER SET utf8 COLLATE utf8_spanish_ci NOT NULL DEFAULT '',
  `formalidad` varchar(150) CHARACTER SET utf8 COLLATE utf8_spanish_ci NOT NULL DEFAULT '',
  `tipo_casa` varchar(150) CHARACTER SET utf8 COLLATE utf8_spanish_ci NOT NULL DEFAULT '',
  `nivel_riesgo` varchar(150) CHARACTER SET utf8 COLLATE utf8_spanish_ci NOT NULL DEFAULT '',
  `estado` tinyint(4) DEFAULT '1',
  PRIMARY KEY (`idcliente`),
  UNIQUE KEY `idclientes_UNIQUE` (`idcliente`)
) ENGINE=InnoDB AUTO_INCREMENT=182 DEFAULT CHARSET=utf8
```

Fig. 19. Tabla 'clientes'

### Creación tabla 'prestamos'

```
CREATE TABLE `prestamos` (
  `idprestamo` int(11) NOT NULL AUTO_INCREMENT,
  `idcliente` int(11) NOT NULL,
  `usuario` int(11) NOT NULL,
  `prestamo` datetime NOT NULL,
  `monto` decimal(16,2) NOT NULL,
  `interes` decimal(16,2) NOT NULL,
  `saldo` decimal(16,2) NOT NULL,
  `formapago` varchar(20) DEFAULT NULL,
  `fpago` datetime DEFAULT NULL,
  `plazo` varchar(20) DEFAULT NULL,
  `fplazo` datetime DEFAULT NULL,
  `estado` tinyint(4) DEFAULT '1',
  PRIMARY KEY (`idprestamo`),
  KEY `fkClientes_idx` (`idcliente`),
  KEY `fkUsuarios_idx` (`usuario`),
  CONSTRAINT `fkClientes` FOREIGN KEY (`idcliente`) REFERENCES `clientes` (`idcliente`) ON DELETE NO ACTION ON UPDATE NO ACTION,
  CONSTRAINT `fkUsuarios` FOREIGN KEY (`usuario`) REFERENCES `usuarios` (`idusuario`) ON DELETE NO ACTION ON UPDATE NO ACTION
) ENGINE=InnoDB AUTO_INCREMENT=12 DEFAULT CHARSET=utf8
```

Fig. 20. Tabla 'prestamos'

### Creación tabla 'pagos'

```
CREATE TABLE `pagos` (
  `idpago` int(11) NOT NULL AUTO_INCREMENT,
  `idprestamo` int(11) NOT NULL,
  `usuario` varchar(50) NOT NULL,
  `fecha` datetime NOT NULL,
  `cuota` decimal(16,2) NOT NULL,
  PRIMARY KEY (`idpago`),
  KEY `fkPrestamos_idx` (`idprestamo`),
  CONSTRAINT `fkPrestamos` FOREIGN KEY (`idprestamo`) REFERENCES `prestamos` (`idprestamo`) ON DELETE NO ACTION ON UPDATE NO ACTION
) ENGINE=InnoDB AUTO_INCREMENT=7 DEFAULT CHARSET=utf8
```

Fig. 21. Tabla 'pagos'

## 4. BD modelo físico transaccional

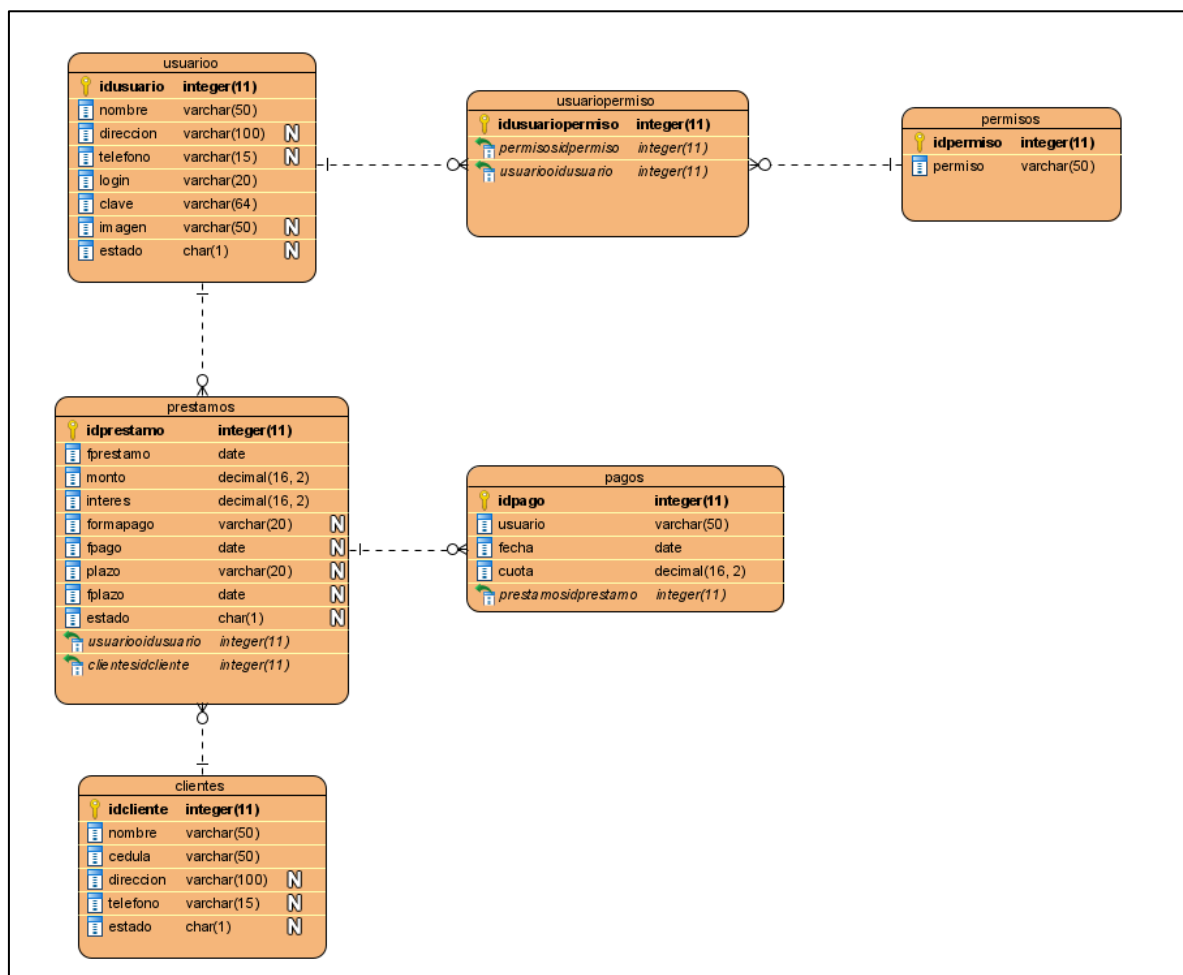


Fig. 22. Modelo físico transaccional

## 5. Verificar la calidad de los datos

A continuación, se puede observar los puntos a considerar para verificar la calidad de los datos.

- 1) Los campos de las tablas no deben tener valores nulos.
- 2) Los campos de las tablas que guarda el “monto” no deben tener un valor “cero”.
- 3) No debe ver duplicidad en la solicitud de créditos.

### 4.1.3. Fase 3: Preparación de los datos

#### 1. Seleccionar los datos

De acuerdo con el diccionario de datos, se procede a seleccionar los datos requeridos para ser usados en el algoritmo de predicción.

TABLA XL. Selección de datos

Nº TABLA	DESCRIPCIÓN
Tabla XXXIV	Tabla usuario
Tabla XXXV	Tabla permiso
Tabla XXXVI	Tabla usuariopermiso
Tabla XXXVII	Tabla clientes
Tabla XXXVIII	Tabla prestamos
Tabla XXXIX	Tabla pagos

#### 2. Limpiar los datos

Para este proceso se tomó en cuenta 3 puntos que se mencionaron en el apartado “5” de la fase #2: Verificación de calidad de los datos.

##### A. Los campos de las tablas no deben tener valores nulos:

Se observó que no existen campos nulos en la base de datos.

Cod.Client	Direccion	Fecha Vencimiento	Valor Cuota	Importe Monto	Nro Cuota	Campaña	Prod.Principal	Fec.Desembols
45995839	CALLE CAJAMARCA 1100 SEC PUEBLO LIBRE	15/09/2021	215.01	1290.06	6	6309 - ECOMMERCE_ELECTRO	ECOMMERCE_ELECTRO	15/05/2021
80375694	MORERIA ALTA	3/09/2021	204.3	2451.6	12	6505 - ELECTRO PT	ELECTRO PT	21/05/2021
70040833	LASALMENDRA	15/09/2021	129.31	1551.72	12	6505 - ELECTRO PT	ELECTRO PT	10/06/2021
46024487	SECTOR JOSE OLAYA SN C.P. JOSE OLAYA	15/09/2021	210.66	2527.92	12	6457 - CAMP_ECOMMERCE_CUPONES	CAMP_ECOMMERCE_CUPONES	12/06/2021
33826382	CAJARIUO Q-C RUTA 09 REF	3/09/2021	323.08	3876.96	12	6309 - ECOMMERCE_ELECTRO	ECOMMERCE_ELECTRO	29/08/2021
47733663	CALLE SAN ANTONIO N-215	15/09/2021	215.07	2181.05	15	6240 - ECOMMERCE_MITO	ECOMMERCE_MITO	11/08/2021
16484661	CA CAPITAN QUIMON N3300	3/10/2021	278.78	2787.8	10	6309 - ECOMMERCE_ELECTRO	ECOMMERCE_ELECTRO	20/08/2021
46687894	CA. FRANCISCO CABRERA 1183 CH	15/09/2021	194.88	4676.4	24	6309 - ECOMMERCE_ELECTRO	ECOMMERCE_ELECTRO	17/07/2021
33591617	AV MARIANO MELGAR 346	3/09/2021	251.23	2009.84	8	6309 - ECOMMERCE_ELECTRO	ECOMMERCE_ELECTRO	31/07/2021
42847612	SAN JUAN / LAS MAGNOLIAS S/N	15/09/2021	289.05	5202.9	18	6309 - ECOMMERCE_ELECTRO	ECOMMERCE_ELECTRO	13/08/2021
43969556	MIRAFLORES-JAEN	5/09/2020	71.17	1708.08	24	951 - CREDIPREFERENTE	PCHA ID1 LHVANA B/ALLMINIO IM	9/10/2018
77130464	CA PEDRO RUIZ C-6-MAGLALAN-JAEN	5/10/2020	413.09	6762.19	17	1873 - CREDIEFECTIVO CONECTA SELVA	CREDIEFECTIVO CONECTA SELVA	4/03/2019
43366613	LLEGAR AL RECREO EL DORADO	15/10/2020	141.31	1823.92	13	1873 - CREDIEFECTIVO CONECTA SELVA	CREDIEFECTIVO CONECTA SELVA	14/03/2019
27716463	CA. SAN JOSE NPO. 108 PUEBLO NUEVO	10/10/2020	247.26	3214.42	13	1935 - BNK MEJORADA	43UK6200 UHD 43" LG SMART TV A	7/05/2019
2847264	MOLINO NUEVA VALLA	10/10/2020	245.48	4664.09	19	5648 - PAGO_DIFERIDO_80D	EQ MHC-V81D BT/ONEBOX/INFC/HDMI	16/05/2019
27665661	VILLANUEVA PINILLOS 1300 P UEBLO LIBRE	3/07/2021	133.05	2614	20	1344 - CAMPAÑA 1344	TV UN-49NU7300G 49" UHD SMART	26/05/2019
46483967	FRENTE A LA PANADERIA	3/02/2021	253.79	4993.45	19	1344 - CAMPAÑA 1344	TV 50UB285 50" UHD DIGITAL DMA	28/05/2019

Fig. 23. Limpieza de datos nulos

**B. Los campos de la tabla que guarda el monto no deben tener un valor cero.**

Se puede observar que no existe un importe con valor “0.00” en la columna MONTO.

Direccion	Telefono	Fecha Vencimiento	Valor Cuota	Importe Monto
CALLE CAJALIBRE				1290.06
CALLE EMILIA MORERIA				2451.6
CA. LAS ALAS				1551.72
LASALMEN				
CA. NICOLA				2527.00

Fig. 24. Limpieza de datos con monto '0.00'

**C. No debe haber duplicidad en la solicitud de créditos.**

Existen datos duplicados pero diferentes solicitudes de créditos, lo cual nos genera un conflicto con el algoritmo de predicción.

Cod.Cliente	Nombre	Direccion	Telefono	Fecha Vencimiento
43969556	HECTOR G...			021
77130464	LUIS EDER...			021
43366613	MAURICIO...			020
27716463	JOSE BAL...			020

Fig. 25. Limpieza de datos duplicados - Ej.1

Cod.Cliente	Nombre	Direccion	Telefono	Fecha Vencimiento
43660858	NANC...			
27716158	LUIS...			
47421994	JOSE...			
80544842	EDW...			
41731624	CELS...			
27715938	MAR...			
73586811	IVAN...			
48584103	LUCE...			
42632530	TERE...			
80643184	LEON...			
43879773	CINTI...			
27702059	TOMA...			
47113555	JAVIE...			

Fig. 26. Limpieza de datos duplicados - Ej.2

### 3. Exclusión de variables

Después de analizar la base de datos original, se observó y se logró determinar que existen variables que no se tomarían es por lo que se excluyeron. Siendo la tabla XLI la que nos permitirá conocer cuáles son las variables para considerar para la predicción.

*TABLA XLI. Variables BD original*

<b>Datos</b>	<b>Descripción</b>
codCliente	Identificador del cliente
Nombre	Nombre del cliente
Dirección	Dirección del cliente
Teléfono	Teléfono del cliente
fechaVencimiento	Fecha de vencimiento de la cuota
valorCuota	Valor de la cuota
Monto	Monto solicitado
nroCuota	Numero de cuotas
Campaña	Campaña
prodPrincipal	Producto solicitado por el cliente
fecDesembolso	Fecha de desembolso de la cuota
tipoEvaluacion	Tipo de evaluación
situacionLaboral	Situación de laboral (Dependiente / Independiente)
formalidadLaboral	Formalidad laboral (Formal / Informal)
casaPropia	Con / Sin Casa propia
Departamento	Departamento del cliente
Provincia	Provincia del cliente

TABLA XLII. Variables BD prestamos\_modelo

Datos	Descripción
codCliente	Identificador del cliente
Nombre	Nombre del cliente
Monto	Monto solicitado
situacionLaboral	Situación de laboral (Dependiente / Independiente)
formalidadLaboral	Formalidad laboral (Formal / Informal)
casaPropia	Con / Sin Casa propia
Ingresos	Ingresos del cliente

#### 4. Construir los datos

En este proceso se va a construir una nueva columna en la base de datos. Columna **Historial Crediticio**, nos permitirá conocer el comportamiento con la financiera de un cliente, así poder evaluar el otorgamiento de créditos, a su vez se construye la columna **Nivel de Riesgo** donde se almacenará el resultado de la predicción.

#### 5. Integrar los datos

Se realizó la limpieza de los datos y se generó una nueva base de datos, en la figura N° 27 se puede observar que la primera tabla contiene la data original que nos otorgó la organización, en la segunda tabla se tiene la data que se usará en el modelado.

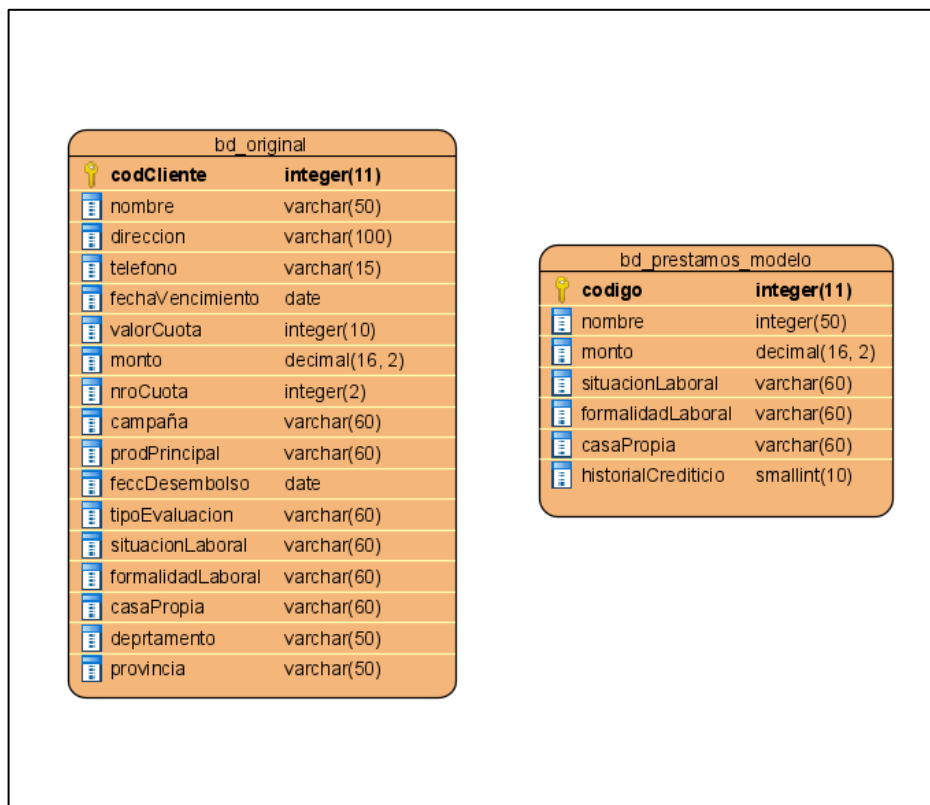


Fig. 27. BD transaccional y BD de predicción

## 6. Formateo de datos

El campo que hace referente a situación laboral, formalidad laboral y casa propia, ha sido modificado con los valores ya que permite tomarlos de mejor manera en el procesamiento de datos.

TABLA XLIII. Formateo columna 'situacion\_laboral'

Columna situación laboral	
Valores	Nuevos valores
Dependiente	0
Independiente	1

TABLA XLIV. Formateo columna 'formalidad\_laboral'

Columna formalidad laboral	
Valores	Nuevos valores
Formal	0
Informal	1

TABLA XLV. Formateo columna 'casa\_propia'

Columna casa propia	
Valores	Nuevos valores
Con casa propia	0
Sin casa propia	1

#### 4.1.4. Fase 4: Modelamiento

##### 1. Seleccionar la técnica de modelamiento

En este punto se tomó en cuenta diferentes aspectos para la selección de la técnica:

- A. El tipo de dato de cada variable.
- B. El objetivo que cumplirá la metodología.
- C. Los requerimientos tanto funcionales como no funcionales mencionados previamente.

Contando con la información presentada previamente y tomando en cuenta los aspectos mencionados anteriormente podemos observar que posee una variable objetivo (target) que nos indicará el nivel de riesgo de un cliente y un conjunto de variables sobre información de los clientes.

Por ello se determinó que el modelo a seguir debe estar basado en modelos de aprendizaje supervisados, ya que éstos pueden trabajar en base a una variable objetivo.

Debido a lo expuesto por [6], [23], [30] se determinó que los métodos de clasificación están más alineados con la presente investigación además que podemos cumplir los objetivos trazados por la metodología, debido a que son los más utilizados, a su vez se tiene un alto grado de precisión y por último dichos métodos se genera dos conjuntos de datos (conjunto de entrenamiento y conjunto de prueba).

Tomando en cuenta lo mencionado anteriormente y haberse analizado los precedentes [5], [6], [8], [36] y [37] se consideró pertinente utilizar los siguientes algoritmos para su evaluación debido a que son los más empleados por los autores.

*TABLA XLVI. Comparación de algoritmos empleados*

Algoritmo	Dificultad	Nivel	Tipo de aprendizaje
J48	Media	Medio	Supervisado
Naive Bayes	Media	Medio	Supervisado
OneR	Fácil	Medio	Supervisado

## 2. Plan de pruebas del modelo

En este punto se detalla el cómo se midió inicialmente el rendimiento de los modelos implementados con el objetivo de obtener mejores resultados.

A su vez se determinó dividir los datos en dos conjuntos, ver la tabla XLVI, el primero conjunto nos servirá para el entrenamiento mientras el segundo conjunto nos servirá para la prueba de la predicción.

*TABLA XLVII. Plan de pruebas*

	PORCENTAJE
<b>DATA DE ENTRENAMIENTO</b>	75%
<b>DATA DE PRUEBAS</b>	25%

Para evaluar el modelo se utilizó la matriz de confusión, la cual permite evaluar el desempeño del modelo empleado en el aprendizaje supervisado. Esta matriz nos permite visualizar los tipos de aciertos y errores en el proceso de aprendizaje de los datos.

*TABLA XLVIII. Matriz de confusión*

		Predicción	
		Positivos	Negativos
Observación	Positivos	Verdaderos Positivos (TP)	Falsos Negativos (FN)
	Negativos	Falsos Positivos (FP)	Verdaderos Negativos (TN)

### 3. Arquitectura de la aplicación

Se tomará en cuenta los siguientes componentes:

#### A. Componente Web

En el presente componente contiene la página web donde se va a poder ingresar los datos que se usaran para clasificar un nuevo crédito e indicar el nivel de riesgo para poder así determinar si se acepta o rechaza un crédito. Se utilizo el lenguaje JavaScript y HTML para la elaboración de la página web.

Se mostrará a continuación las principales funciones desarrolladas para la creación del sistema.

Función que permite la conexión entre la BD y el sistema.

```
4 $conexion = new mysqli(DB_HOST,DB_USERNAME,DB_PASSWORD,DB_NAME);
```

*Fig. 28. Conexión a la BD*

Función que se ejecuta automáticamente al iniciar el sistema.

```
3 //Funcion que se ejecuta al inicio
4
5 function init() {
6     mostrarform(false);
7     listar();
8
9     $("#formulario").on("submit", function (e) {
10        guardaryeditar(e);
11    });
12 }
```

*Fig. 29. Función inicial*

Parámetros de entrada que se toman en cuenta para el registro de un cliente.

```
4 $cliente = new Clientes();
5
6 $idcliente = isset($_POST["idcliente"]) ? limpiarCadena($_POST["idcliente"]) : "";
7 $cedula = isset($_POST["cedula"]) ? limpiarCadena($_POST["cedula"]) : "";
8 $nombre = isset($_POST["nombre"]) ? limpiarCadena($_POST["nombre"]) : "";
9 $direccion = isset($_POST["direccion"]) ? limpiarCadena($_POST["direccion"]) : "";
10 $telefono = isset($_POST["telefono"]) ? limpiarCadena($_POST["telefono"]) : "";
11 $ingresos = isset($_POST["ingresos"]) ? limpiarCadena($_POST["ingresos"]) : "";
12 $situacionLaboral = isset($_POST["situacionLaboral"]) ? limpiarCadena($_POST["situacionLaboral"]) : "";
13 $formalidad = isset($_POST["formalidad"]) ? limpiarCadena($_POST["formalidad"]) : "";
14 $tipoCasa = isset($_POST["tipoCasa"]) ? limpiarCadena($_POST["tipoCasa"]) : "";
15 $validacion = isset($_POST["procesoValidacion"]) ? limpiarCadena($_POST["procesoValidacion"]) : "";
16 $nivel = isset($_POST["nivel"]) ? limpiarCadena($_POST["nivel"]) : "";
```

*Fig. 30. Parámetros de entrada de un nuevo cliente*

Parámetros de entrada que se toman en cuenta para el registro de un préstamo.

```

5  $prestamo = new Prestamo();
6
7  $idprestamo = isset($_POST["idprestamo"]) ? limpiarCadena($_POST["idprestamo"]) : "";
8  $idcliente = isset($_POST["idcliente"]) ? limpiarCadena($_POST["idcliente"]) : "";
9  $usuario = isset($_POST["usuario"]) ? limpiarCadena($_POST["usuario"]) : "";
10 $fprestamo = isset($_POST["fprestamo"]) ? limpiarCadena($_POST["fprestamo"]) : "";
11 $monto = isset($_POST["monto"]) ? limpiarCadena($_POST["monto"]) : "";
12 $interes = isset($_POST["interes"]) ? limpiarCadena($_POST["interes"]) : "";
13 $saldo = isset($_POST["saldo"]) ? limpiarCadena($_POST["saldo"]) : "";
14 $formapago = isset($_POST["formapago"]) ? limpiarCadena($_POST["formapago"]) : "";
15 $fechapago = isset($_POST["fechapago"]) ? limpiarCadena($_POST["fechapago"]) : "";
16 $plazo = isset($_POST["plazo"]) ? limpiarCadena($_POST["plazo"]) : "";
17 $fplazo = isset($_POST["fplazo"]) ? limpiarCadena($_POST["fplazo"]) : "";

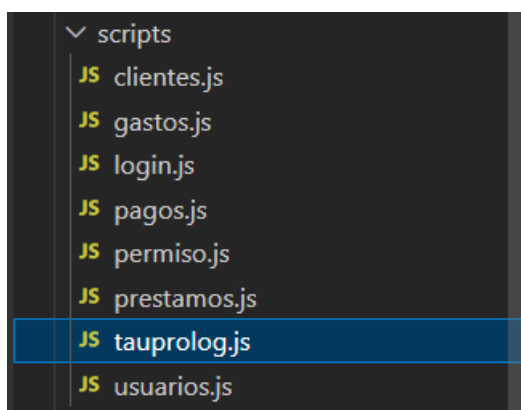
```

*Fig. 31. Parámetros de entrada de un préstamo*

## B. Componente de predicción

En este punto se considera los componentes para el desarrollo del sistema web. Los métodos permiten calcular el nivel de exactitud de la clasificación para el otorgamiento de crédito en la Financiera. Se ha utilizado la librería “Tau Prolog”.

Tau Prolog, es un intérprete de Prolog implementado en lenguaje JavaScript, está dirigida por ISO Prolog, a su vez fue diseñado para promover la aplicabilidad y portabilidad del texto y los datos de Prolog entre varios sistemas de procesamiento de datos.



*Fig. 32. Librería Tau Prolog*

En la figura 33, se puede observar una parte de la librería Tau Prolog.

```

149 // Virtual file system for browser
150 tau_file_system = {
151 // Current files
152 files: new TauDirectory("/", "/", null),
153 // Open file
154 open: function(path, type, mode) {
155     var dirs = path.replace(/\/$/, "").split("/");
156     var dir = tau_file_system.files;
157     var name = dirs[dirs.length-1];
158     for(var i = 1; i < dirs.length-1; i++) {
159         dir = dir.lookup(dirs[i]);
160         if(!pl.type.is_directory(dir))
161             return null;
162     }
163     var file = dir.lookup(name);
164     if(file === null) {
165         if(mode === "read")
166             return null;
167         file = new TauFile(name, type, dir);
168         dir.push(name, file);
169     } else if(!pl.type.is_file(file)) {
170         return null;
171     }
172     if(mode === "write")
173         file.text = "";
174     return file;
175 },

```

Fig. 33. Tau Prolog

### C. Componente Inteligencia Artificial (IA)

El presente componente es el más importante pues en este componente se va a implementar la parte lógica del algoritmo de predicción, además de consumir la data que servirá para el entrenamiento del algoritmo.

En la figura 34, línea 200 llamamos a la librería que se ha importado y se crean las reglas para la predicción.

```

200 var session = pl.create();
201 //VARIABLES
202 //var nivel = -1;
203 const reglas = `
204     %Hechos
205     %calcular_nivel
206     calcular_nivel(dependiente,formal,propia,0).
207     calcular_nivel(independiente,formal,propia,0).
208     calcular_nivel(dependiente,formal,sincasa,0).
209     calcular_nivel(independiente,formal,sincasa,1).
210     calcular_nivel(independiente,informal,sincasa,2).
211     calcular_nivel(independiente,informal,precaria,2).
212
213     %Reglas
214     calcular_nivel_final(A,B,C,X):- calcular_nivel(A,B,C,X).
215
216 `;

```

Fig. 34. Reglas de predicción

La función `caclular_nivel` permitirá obtener el resultado de la predicción, ingresándole los parámetros correspondientes.

```

219 const calcular_nivel = ([situacionLaboral, formalidad, ingresos, tipoCasa]) => {
220   return new Promise((resolve, reject) => {
221     session.consult(reglas, {
222       success: function () {
223         /* Program loaded correctly */
224         //console.log("PL Succes!!");
225         session.query(
226           "calcular_nivel_final(" +
227             situacionLaboral +
228             "," +
229             formalidad +
230             "," +
231             ingresos +
232             "," +
233             tipoCasa +
234             ", X ).",

```

Fig. 35. Func. `Calcular_nivel`

En la figura 36, se puede observar los componentes que se utilizaron para la elaboración del sistema de predicción.

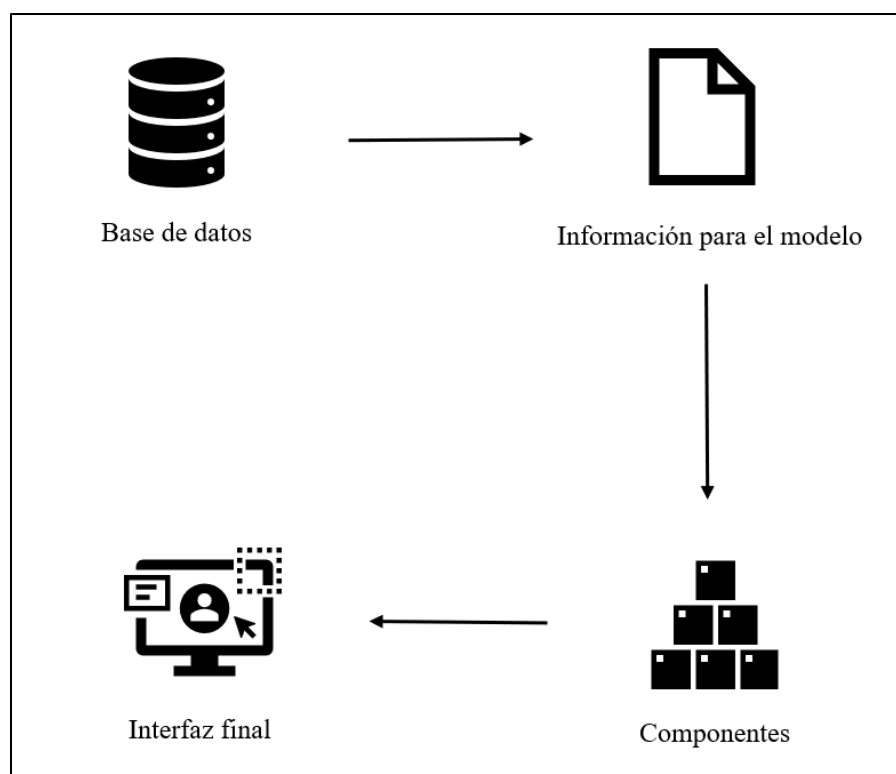


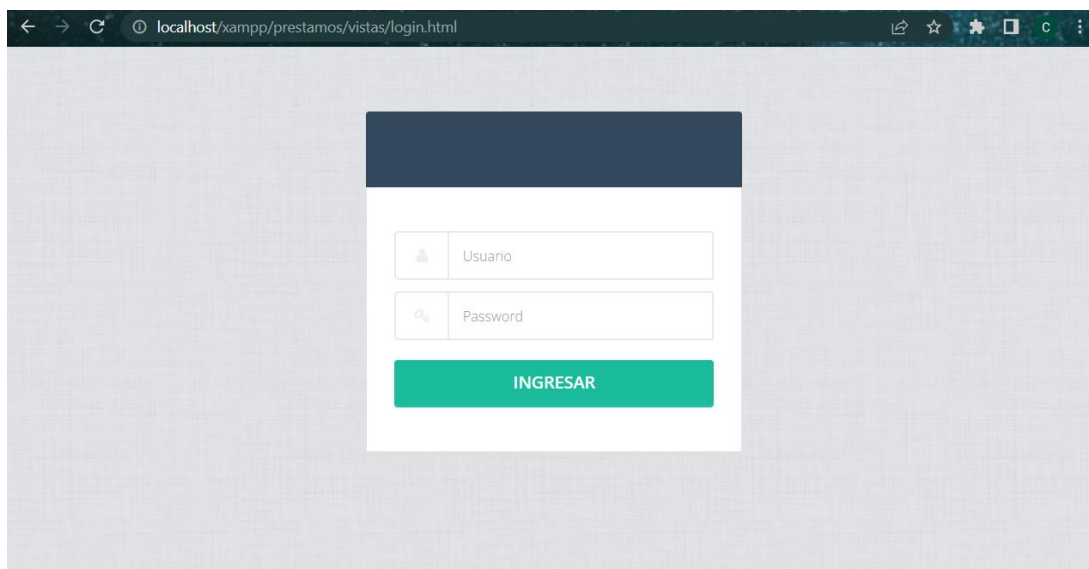
Fig. 36. Componentes

#### 4. Construir el modelo

En este apartado se procede a construir el modelo, para después proceder a su evaluación.

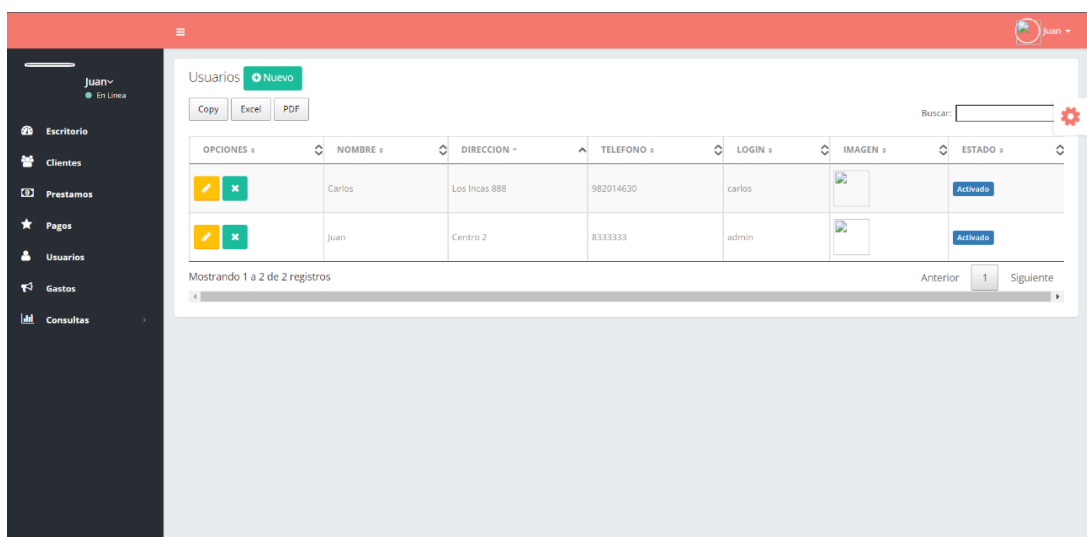
##### Interfaz de usuario:

- **Interfaz Login**, el usuario (Administrador / Asesor de créditos) podrá ingresar al sistema mediante su usuario y clave.



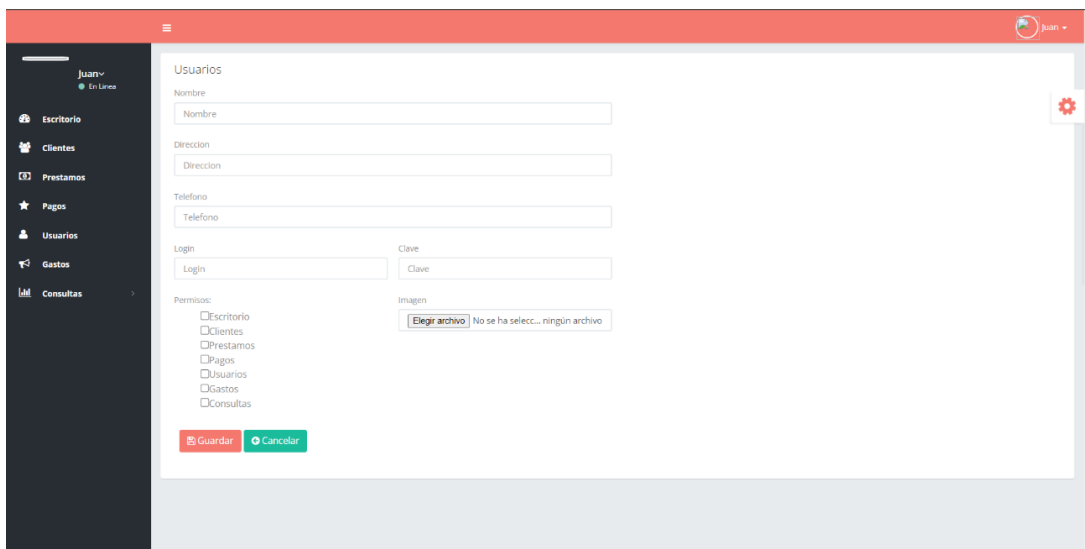
*Fig. 37. Interfaz Login*

- **Interfaz Gestionar Usuarios**, está permitirá visualizar la lista de usuarios, a su vez registrar un nuevo usuario la cual solo la puede realizar el usuario Administrador.



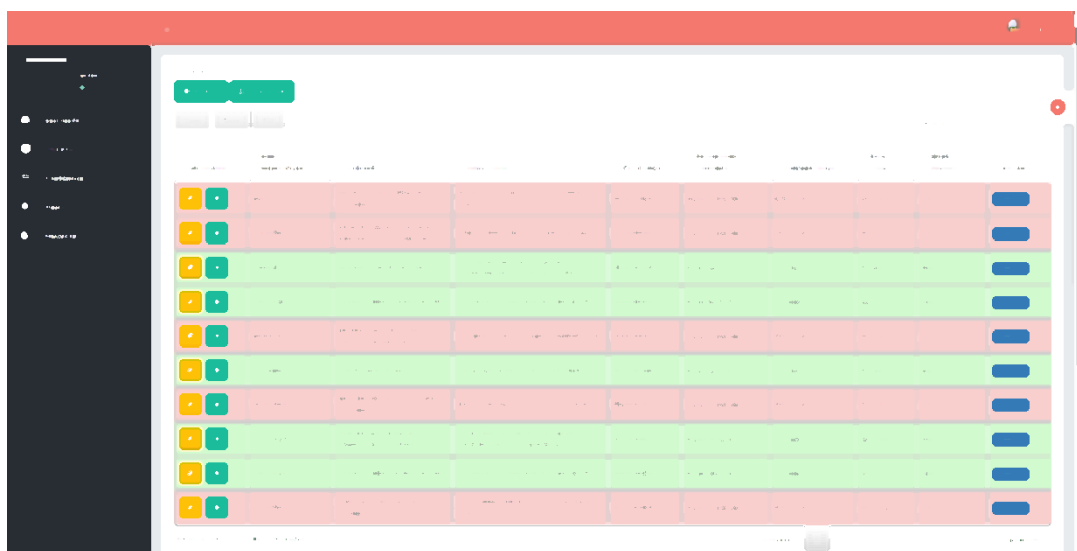
*Fig. 38. Interfaz gestión de usuarios*

- **Interfaz Registro de Usuario**, el usuario Administrador podrá registrar un nuevo usuario (Administrador o asesor de créditos) y otorgarle los permisos correspondientes dentro del sistema.



*Fig. 39. Interfaz de registro de usuario*

- **Interfaz Gestionar Clientes**, permite visualizar la lista de clientes con la que cuenta la financiera, a su vez permite al usuario registrar, analizar y predecir el riesgo en el otorgamiento de créditos.



*Fig. 40. Interfaz gestionar clientes*

- **Interfaz registro de clientes**, permite registrar un nuevo cliente a su vez predecir el otorgamiento de crédito.

The screenshot shows a web application interface for registering a new client. The page title is 'Clientes' and there is an 'Importar' button. The form contains the following fields:

- Documento Identidad: Cedula
- Nombre: Nombre
- Direccion: Direccion
- Telefono: Telefono
- Sueldo: Sueldo
- Situacion laboral: Dependiente
- Formalidad: Formal
- Tipo de casa: Propia
- Proceso de Validación: Validado

At the bottom of the form, there are two buttons: 'Guardar' (Save) and 'Cancelar' (Cancel).

Fig. 41. Interfaz registro de clientes

- **Interfaz de Gestionar Préstamos**, permite visualizar una lista de los préstamos otorgados, a su vez, permite registrar un nuevo préstamo.

The screenshot shows a web application interface for managing loans. The page title is 'Préstamos' and there is a 'Nuevo' button. The interface includes a search bar and a table with the following columns:

- OPCIONES
- CLIENTES
- USUARIOS
- FECHA
- MONTO
- INTERES
- SALDO
- PAGOS
- FECHAS
- PLAZO
- FECHA
- ESTADO

The table contains three rows of data:

OPCIONES	CLIENTES	USUARIOS	FECHA	MONTO	INTERES	SALDO	PAGOS	FECHAS	PLAZO	FECHA	ESTADO
[Icon]	Mario Casas	Juan	2019-04-30	10000.00	20.00	12000.00	Diario	2019-04-01	Mes	2019-05-01	[Icon]
[Icon]	Carlos O	Juan	2022-11-23	5000.00	0.00	5650.00	Mensual	2022-11-30	Mes	2022-12-30	[Icon]
[Icon]	Juan Camilo	Juan	2022-11-23	1000.00	0.00	1130.00	Quincenal	2022-11-01	Mes	2022-12-01	[Icon]

At the bottom of the table, there is a pagination control showing 'Mostrando 1 a 3 de 3 registros' and navigation buttons for 'Anterior', '1', and 'Siguiete'.

Fig. 42. Interfaz gestionar prestamos

- **Interfaz Registro de préstamo**, permite registrar un nuevo préstamo a un cliente después de la predicción del riesgo en el otorgamiento de crédito.

Fig. 43. Interfaz registro de prestamos

- **Interfaz resultado de la predicción**, la interfaz muestra el resultado de la predicción, mediante una semaforización: Nivel bajo (color verde), Nivel medio (color amarillo), Nivel Alto (color rojo).

OPCIONES	CEDULA	CLIENTE	DIRECCIÓN	TELÉFONO	SITUACIÓN LABORAL	FORMALIDAD	TIPO CASA	NIVEL RIESGO	ESTADO
[Icon]	10223878	Mario Casas	Los Lago	714545	Dependiente	Formal	Propia	Bajo	Activado
[Icon]	281-001085-00126	Juan Camilo	Torres	333323	Dependiente	Formal	Propia	Bajo	Activado
[Icon]	12345678	Prueba Medio	prueba 123	973456785	Independiente	Informal	Sin casa	Medio	Activado
[Icon]	12345655	Prueba Alto	prueba 123	973456785	Independiente	Informal	Precaria	Alto	Activado
[Icon]	12345666	Prueba Bajo	prueba 123	973456785	Dependiente	Formal	Propia	Bajo	Activado
[Icon]	24242424	Prueba 21	prueba 123	973456785	Independiente	Informal	Precaria	Alto	Activado
[Icon]	12345655	Pablo Guerrero	prueba 123	973456785	Independiente	Informal	Sin casa	Alto	Activado
[Icon]	12345678	Pablo Guerrero	prueba 123	973456785	Dependiente	Formal	Propia	Bajo	Activado
[Icon]	12345666	Prueba Cat	prueba 123	973456785	Independiente	Informal	Precaria	Alto	Activado
[Icon]	73201845	Juan Carlos	Av. Gh 234	9870123456	Dependiente	Formal	Sin casa	Bajo	Activado

Fig. 44. Interfaz resultado de predicción

- **Interfaz recomendación**, la interfaz mostrará una recomendación cuando se solicite un crédito en caso de ser un cliente en Nivel alto.

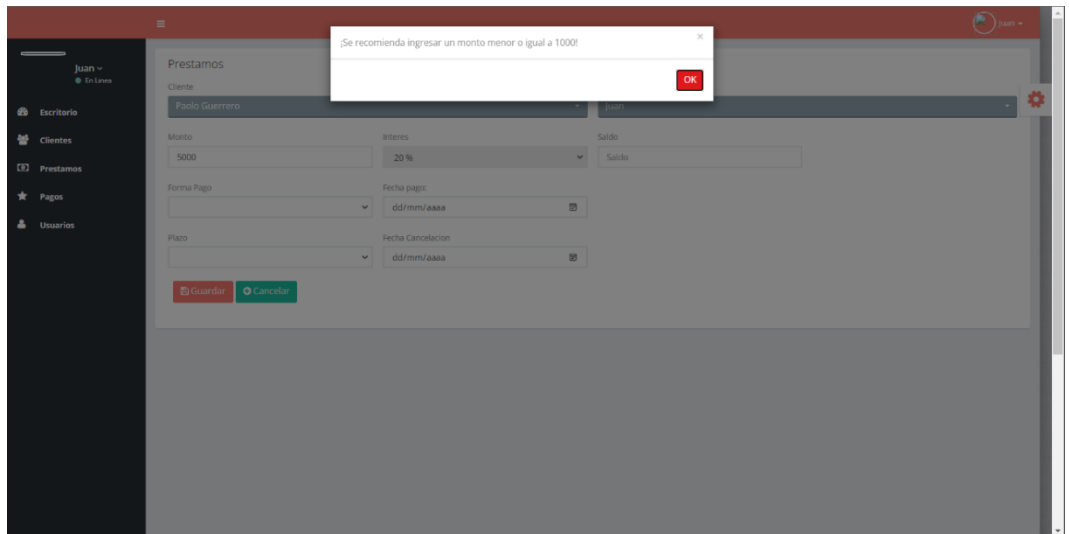


Fig. 45. Interfaz de recomendación

## 5. Evaluación del modelo

Las pruebas realizadas al modelo que se han desarrollado en el presente proyecto fueron hechas en la herramienta Weka, esta herramienta brindara la información de los créditos clasificados para el conjunto de pruebas (25%).

Lo expuesto por Géron [32], se consideró evaluar el modelo utilizando las siguientes métricas de evaluación:

*TABLA XLIX. Métricas de evaluación*

Métricas	Descripción	Formula
Precisión	Representa el # de verdaderos positivos	$\frac{TP}{TP + FP} * 100$ <p><b>Donde:</b>  <b>TP:</b> Verdaderos positivos  <b>FP:</b> Falsos positivos</p>
Recall (Exhaustividad)	Nos informa sobre la cantidad que el modelo es capaz de identificar	$\frac{TP}{TP + FN} * 100$ <p><b>Donde:</b>  <b>TP:</b> Verdaderos positivos  <b>FN:</b> Falsos negativos</p>
AUC (Área bajo la curva)	Nos indica la relación entre la tasa de verdaderos positivos y la tasa de falsos positivos	$1 - \text{especificidad} = \frac{FP}{TN + FP}$ <p><b>FP:</b> Falsos positivos  <b>TN:</b> Verdaderos negativos  <b>FP:</b> Falsos positivos</p>

A continuación, se muestra los pasos que se siguió utilizando la herramienta Weka.

1) Seleccionar el archivo a utilizar.

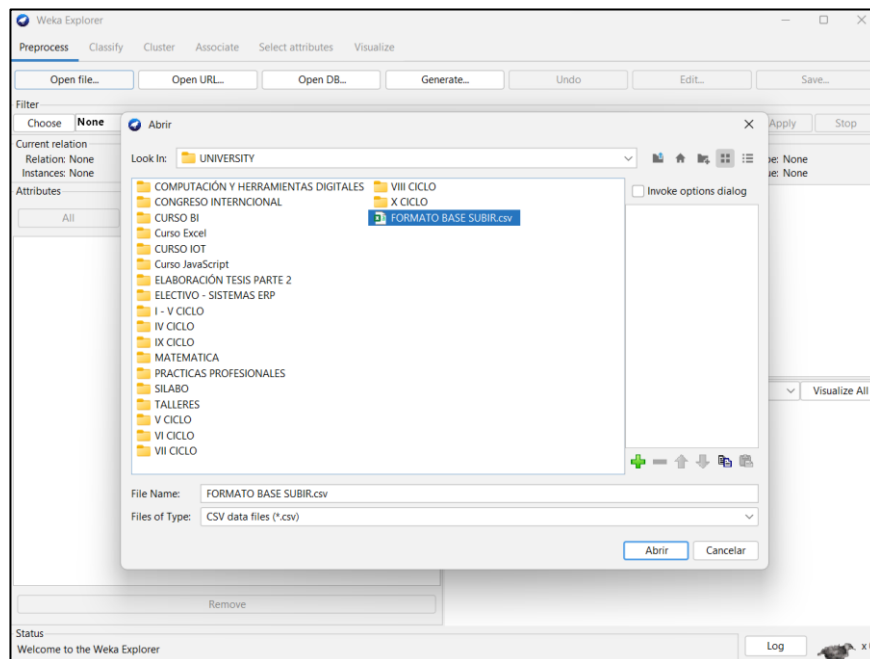


Fig. 46. Selección de archivo -WEKA

2) Indicamos el modelo a utilizar.

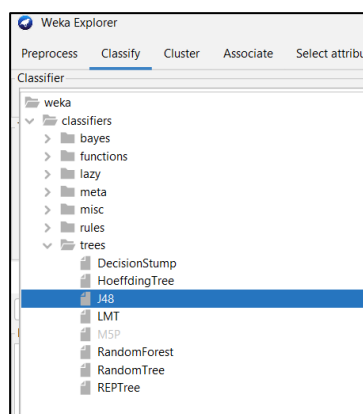


Fig. 47. Selección de algoritmo

3) Indicamos el porcentaje de entrenamiento.

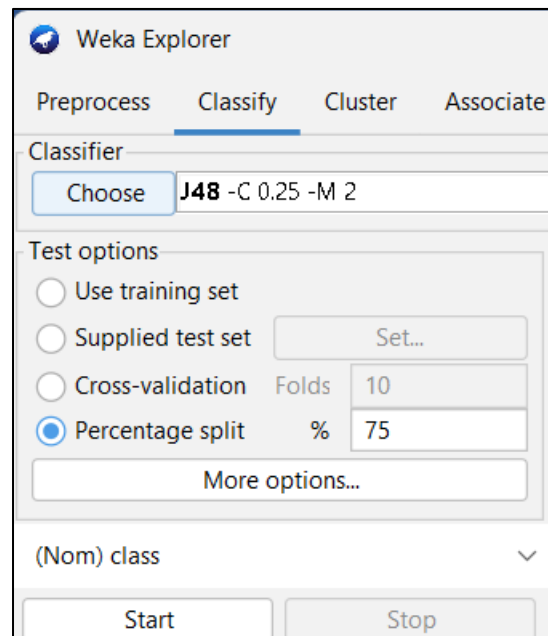


Fig. 48. Indicamos el porcentaje

A continuación, se elaboraron las pruebas con los algoritmos supervisados a utilizar, se tomó en cuenta el 25% de la data.

➤ J48

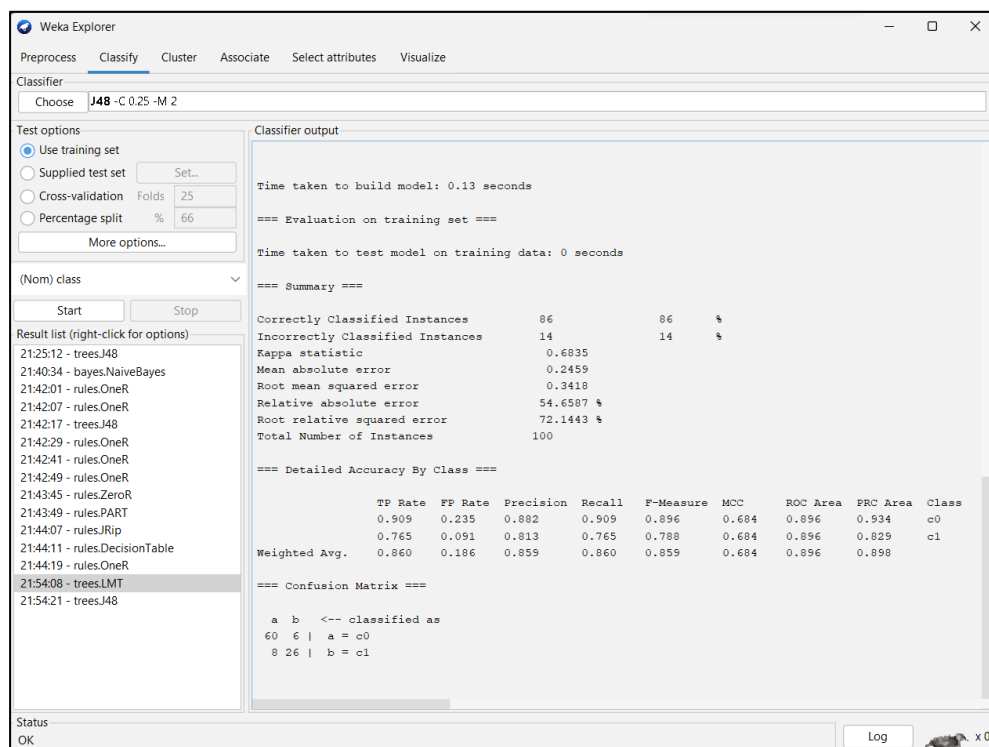


Fig. 49. Prueba del algoritmo J48

## ➤ Naive Bayes

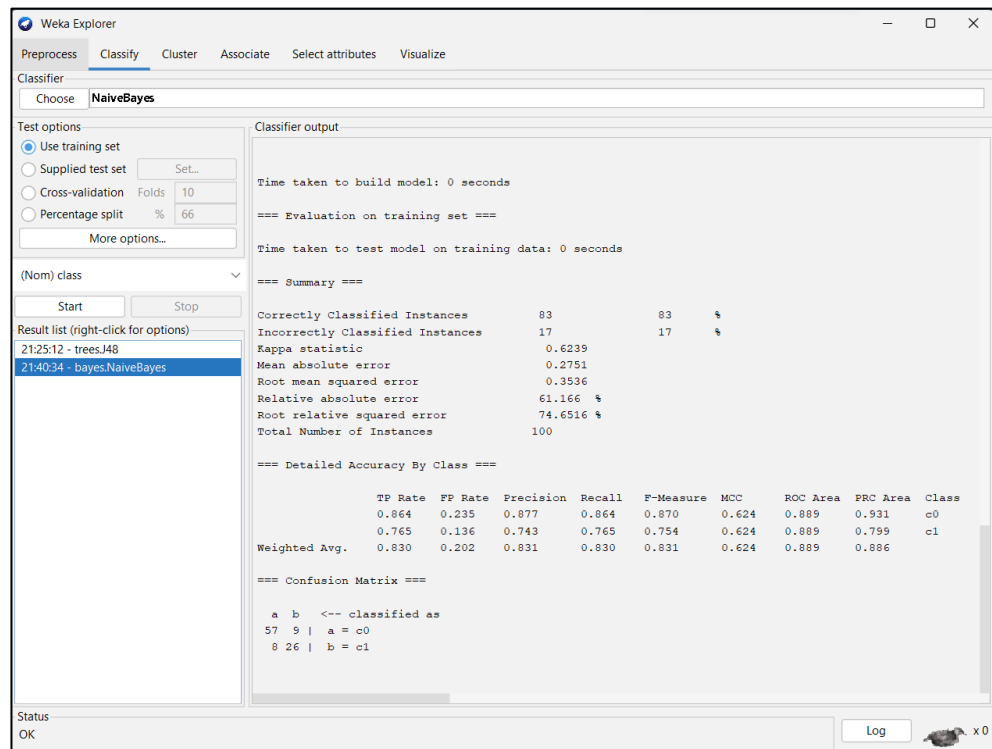


Fig. 50. Prueba del algoritmo Naive Bayes

## ➤ OneR

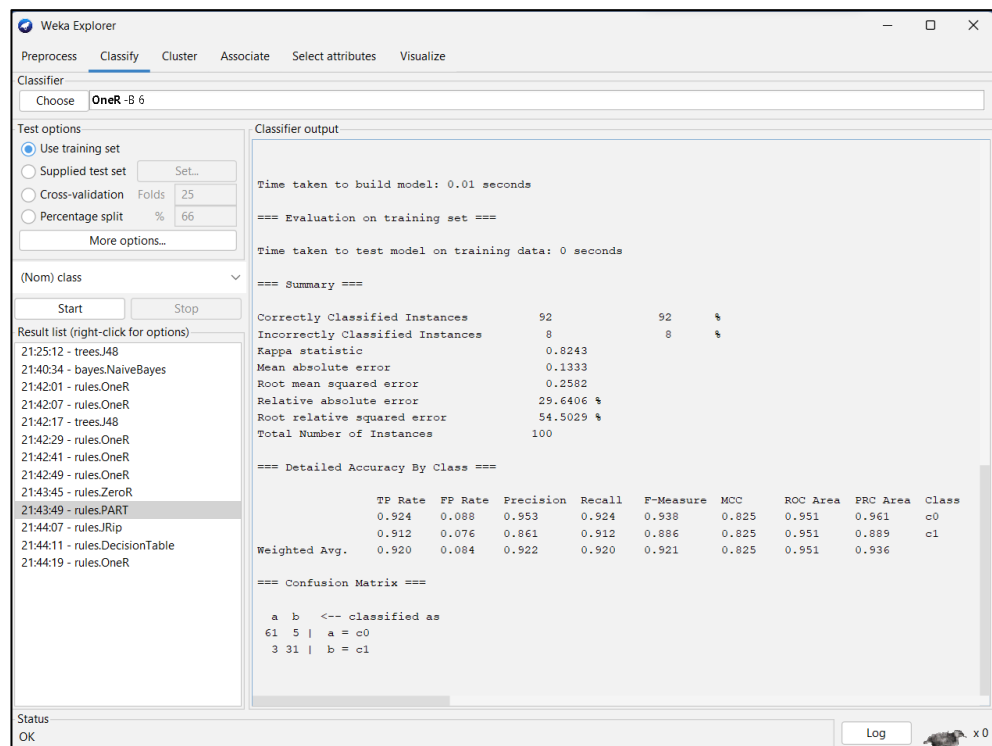


Fig. 51. Prueba del algoritmo OneR

Como se muestra a continuación el porcentaje de precisión de los tres algoritmos utilizados para la prueba.

*TABLA L. Comparativa de precisión de los tres algoritmos*

<b>Algoritmo</b>	<b>Precisión</b>
J48	86%
Naive Bayes	83%
OneR	92%

Después de aplicar los algoritmos de clasificación que son J48, Naive Bayes y OneR. Se realizó la comparativa de los resultados obtenidos en la tabla L, en la que se puede observar que el mejor algoritmo para la clasificación de créditos es el algoritmo OneR, debido a su alto nivel de precisión 92% del 25% de los datos. Las pruebas se realizaron en diferentes ocasiones para mejorar la precisión del algoritmo.

#### **4.1.5. Fase 5: Evaluación**

##### **1. Evaluar los resultados**

Según lo presentado en los puntos previos y visualizando los resultados de los modelos supervisados en base a los algoritmos empleados, se pudo concluir que el algoritmo que mejor se adapta a la realidad que se presenta en el presente proyecto fue el algoritmo OneR, en comparación a los otros algoritmos empleados, se obtuvo un Recall de 92%, un AUC de 94% y una precisión del 92%.

<b>Algoritmo</b>	<b>Precisión</b>	<b>AUC</b>	<b>Recall</b>
J48	86%	90%	86%
Naive Bayes	83%	89%	83%
OneR	92%	94%	92%

En la tabla LI, se puede visualizar la matriz de confusión resultante al evaluar este algoritmo con su respectiva conclusión.

*TABLA LI. Matriz de confusión del algoritmo*

		Valor actual del crédito	
		Yes	No
Valor de predicción	Yes	61	5
	No	3	31

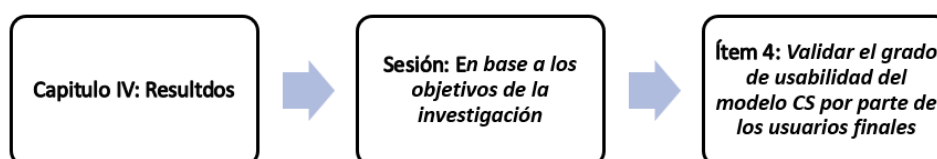
Como se puede apreciar en la matriz de confusión del algoritmo OneR, el modelo clasificó de 100 solicitudes de créditos, clasificó correctamente 61 solicitudes de créditos correctamente (representan un 92%), 5 solicitudes fueron clasificados erróneamente (representan un 8%). A su vez, 31 solicitudes cumplieron con el crédito aprobado (representan un 90.3%) y por otro lado 3 no cumplieron con el crédito otorgado (representan un 9.7%).

## 2. Ejecución de pruebas del sistema

En este punto se procedió a probar la correcta funcionalidad del sistema, ejecutando las pruebas de caja blanca donde se conocerá a detalle el funcionamiento del modelo CS y por otro lado las pruebas de caja negra donde se describe la función o método de la parte interna del modelo CS (ver Anexo 04 y 05):

## 3. Evaluar el grado de usabilidad del sistema

Se utilizó la herramienta de System Usability Scale (SUS), creada por John Brooke (1986), los resultados después de aplicar la herramienta SUS, se mostrarán en el siguiente apartado:



### Cómo administrar la escala de usabilidad del sistema:

La Escala de Usabilidad del Sistema es una Escala Likert que incluye 10 preguntas (ver Anexo 02) que responderán los usuarios de tu sitio web.

El puntaje de usabilidad utilizando SUS, se obtendrá de la siguiente manera:

1. Sumar las preguntas impares y restarle 5 puntos, se obtendrá el valor “X”:

$$(P1 + P3 + P5 + P7 + P9) - 5 = X$$

2. A 25 réstele la suma de las preguntas pares, se obtendrá el valor “Y”.

$$25 - (P2 + P4 + P6 + P8 + P10) = Y$$

3. Después sumar “X + Y”.

$$X + Y = Z$$

4. Finalmente, multiplicar dicho resultado por 2.5.

$$Z * 2.5 = N$$

La puntuación aceptable va desde los 68 puntos hasta los 100 puntos, el promedio de la escala de usabilidad del sistema es de 68 (Suficiencia), como se aprecia en la siguiente figura:

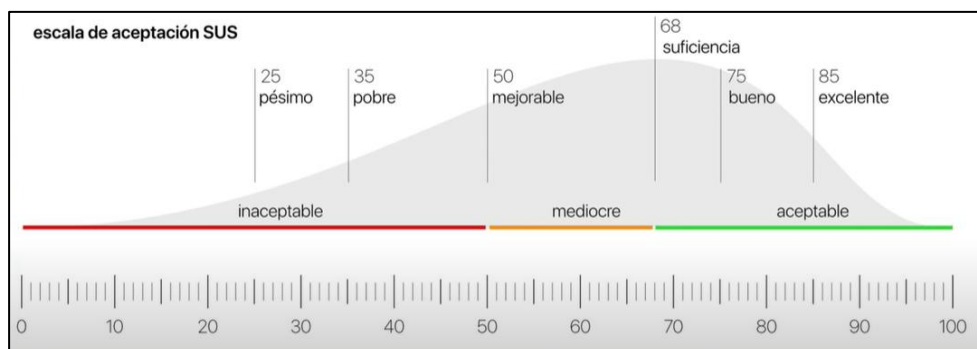


Fig. 52. Escala de aceptación de SUS

#### 4. Determinar los siguientes pasos

Después de determinar la técnica seleccionada, desarrollar el sistema y evaluar los resultados, se tendrá que tomar en cuenta lo siguiente con respecto a Hardware y Software:

### A. Hardware

- **Ordenador:** Se requiere un ordenador con suficiente capacidad para desarrollar y ejecutar el problema.
- **Servidor web HTTP:** Se empleo el servidor web local.
- **Conexión:** Con respecto a la conectividad puede ser cableada o por Wifi.

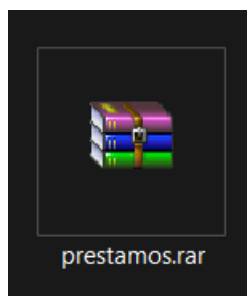
### B. Software:

- **Sistema operativo (SO):** El sistema es adaptable a cualquier SO.
- **Lenguajes de programación:** Se empleo lenguaje de programación JavaScript y el lenguaje HTML.
- **Herramientas:** Se utilizaron herramientas como Visual Studio Code para el desarrollo web del modelo, Visual Studio Paradigm para la creación del modelo E-R, Xamp para el desarrollo de la BD e iniciar de forma local el sistema, Microsoft Excel para la creación del formato (CSV) que se utilizara para la carga masiva, Weka y Postamn para comprobar el funcionamiento del sistema.

## 4.1.6. Fase 6: Despliegue

### 1. Desplegar el plan

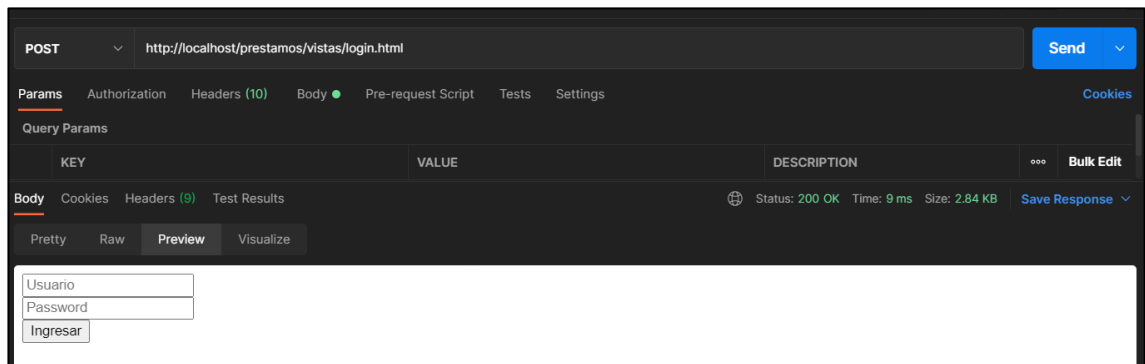
En primer lugar, se procedió a generar un archivo “rar” que se le denomino “prestamos.rar”, este archivo contiene el sistema integrado y la base de datos del sistema. A su vez, se tomará en cuenta las características mencionadas en el *ítem 4. Fase #5: Resultados, sesión Determinar los siguientes pasos, en el Capítulo IV. Resultados*, para la implantación del sistema.



*Fig. 53. Archivo .rar*

## 2. Monitorear y mantener

En este punto se utilizó el método GET y POST, utilizando la herramienta “**Postman**”, se procede a realizar peticiones con POST, con el objetivo de comprobar que funcione correctamente, como se muestra en la figura 47, el sistema responde correctamente.



*Fig. 54. Petición POST.*

## 3. Desarrollar el reporte final

Para el desarrollo de la interfaz final se empleó Bootstrap y el lenguaje de programación JavaScript, así como el lenguaje HTML, lenguaje de diseño CSS, librerías como fullcalendar.css para mostrar el calendario de forma visual, dropzone.css esta librería proporciona la carga de archivos con una vista previa, jquery-ui.css facilita la creación de las interfases de usuario, entre otras librerías.

Para el desarrollo de la interfaz final, se tomó en cuenta que la interfaz final sea una simple, amigable e interactiva con el usuario final.

## 4. Revisión del proyecto

Después de realizar las respectivas comparaciones de las técnicas CS y haber seleccionado la que mejor se adapte a la organización se obtuvo un grado de precisión del 86% de efectividad en la predicción de riesgo en el otorgamiento de créditos. A su vez en el presente proyecto se desarrolló todas las fases de la metodología CRISP-DM, con el objetivo de que los asesores comerciales puedan predecir a través de una interfaz web, está interfaz permite cargar de forma individual y data masiva a través de un formato “.CSV” y comprobar el nivel de riesgo de un cliente para el otorgamiento de créditos.

La precisión que brinda este sistema para determinar el nivel de riesgo de un cliente permitirá a la financiera reducir la tasa de morosidad, reducir el tiempo en proceso de otorgamiento de créditos y observar el nivel de riesgo que se encuentra el cliente y así poder tomar decisiones oportunas.

Y para finalizar, se realizaron pruebas de caja blanca y caja negra al producto final las cuales pueden apreciarse en los Anexos 04 y 05

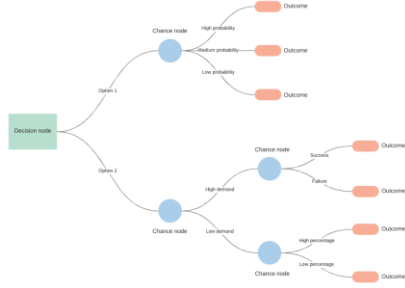
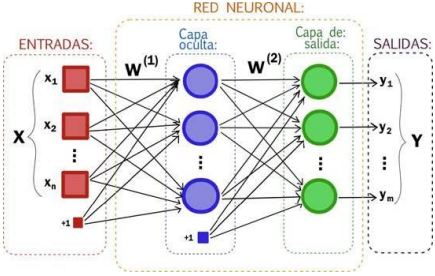
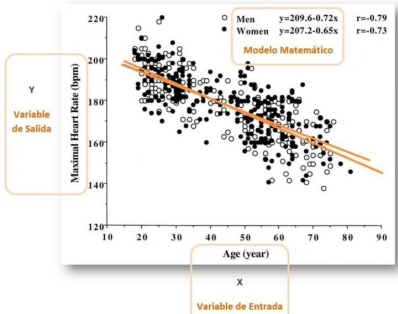
## **4.2. En base a los objetivos de la investigación**

### **4.2.1. Determinar la técnica Credit Scoring que reúna los requisitos de predictibilidad del sistema, para la toma de decisiones en la asignación de créditos.**

Para el cumplimiento de dicho objetivo se definieron variables recolectadas a través del conocimiento de expertos del tema e investigaciones por otros autores. A su vez se tomará en cuenta lo mencionado en el ítem **4.1.4. Fase #4: Modelamiento, sección Seleccionar la técnica de modelamiento, en el Capítulo IV. Resultados.**

En la tabla LII, se puede observar una comparativa entre las tres técnicas utilizadas en la implementación de minería de datos, después de analizar la información, se logró determinar que los métodos de clasificación son los que mejor se adapta debido a su facilidad de uso y su alto nivel en la precisión para este tipo de casos donde tenemos una variable objetivo (modelos supervisados). Por último, en la tabla LIII, se definieron las dimensiones a tomarse en cuenta después de la selección de la técnica y su respectivo algoritmo.

TABLA LII. Comparativa de técnicas CS

Técnicas	Algoritmos	Precisión	Estructura
Arboles de clasificación	J48	[8], concluyen que el algoritmo J48 logró un 78.3% de precisión.	
	Naive Bayes	[9], destaca que el algoritmo de Redes Bayesianas tiene un alto nivel de porcentaje de predictibilidad con un 83%.	
	OneR	[6], este modelo logró clasificar de forma correcta con un 75% los solicitantes de préstamos.	
Redes neuronales	Adaline	[7], determinando que el modelo redes neuronales con un 78% fue el mejor para la clasificación de agricultores.	
	Backpropagation Redes neuronales de aprendizaje competitivo	[10], concluyeron que el modelo cumple las condiciones para adaptarse a la organización, con un 84% de precisión.	
Regresión Logistica	Reg. Log. Bucket	[11], el modelo de regresión logística fue el mejor que se basó en indicadores de predictibilidad y eficiencia, con 86%.	
	Reg. Log. Quartil		
	Reg. Log. Optimal		

La tabla se elaboró con el objetivo de elaborar una comparativa entre las diferentes técnicas analizadas en esta investigación y su nivel de precisión a la que llegaron en cada una de sus investigaciones, a su vez, se puede observar los algoritmos que se utilizan en cada una de las técnicas y su respectiva estructura. Se puede observar que la técnica de clasificación es la que mayor porcentaje de precisión se obtuvo en las respectivas investigaciones, la cual nos permite asegurar que las técnicas de clasificación son las que mejor se adaptan para este tipo de investigaciones.

*TABLA LIII. Indicadores del Obj. Esp. 01*

<b>Variable independiente</b>	<b>Dimensión</b>	<b>Indicador</b>	<b>Método</b>
	Sensibilidad	Porcentaje de créditos clasificados como aprobados.	$\frac{VP}{VP + FN} * 100$ <b>Donde:</b> <b>VP:</b> Verdaderos positivos <b>FN:</b> Falsos negativos
Modelo de Credit Scoring	Especificidad	Porcentaje de créditos clasificados como rechazados.	$\frac{VN}{VN + FP} * 100$ <b>Donde:</b> <b>VN:</b> Verdaderos negativos <b>FP:</b> Falsos positivos
	Eficacia	Porcentaje de créditos clasificados como correctos.	$\frac{VP + VN}{TI} * 100$ <b>Donde:</b> <b>VP:</b> Verdaderos positivos <b>VN:</b> Verdaderos negativos <b>TI:</b> Total de instancias

Después, de analizar comparativamente las técnicas, se tomará en cuenta las tres dimensiones que se pueden observar en la tabla LIII, con el objetivo de medir el modelo empleando cada una de dichas dimensiones.

**4.2.2. Implementar escenarios de predicción aplicando la técnica de Credit Scoring seleccionada para predecir el nivel de riesgo en el otorgamiento de crédito.**

Durante el análisis realizado, se desarrollaron los escenarios que se tomaran cuenta para la predicción de créditos personales tomando en cuenta los siguientes parámetros de entrada (situación laboral, formalidad, tipo de casa, entre otras), a su vez se tomará en cuenta el porcentaje de interés de acuerdo con el nivel (bajo, medio o alto), en el que se encuentra un cliente para otorgarle un nuevo crédito. Tomando en cuenta todo lo indicado el sistema nos arrojará el nivel de riesgo identificándolos mediante una semaforización con su respectivo color como se muestra a continuación en la tabla XLIX.

*TABLA LIV. Escenarios de predicción*

Nivel de riesgo	Parámetros de entrada						Interés	Identificación (color)
	Situación laboral	Valor	Formalidad	Valor	Tipo casa	Valor		
<b>Bajo</b>	Dependiente	0	Formal	0	Propia	0	- 2 < X < 4 solicitudes pagadas = 12% de interés.	<b>Verde (0)</b>
	Independiente	1	Formal	0	Propia	0		
<b>Medio</b>	Dependiente	0	Formal	0	Sin casa	1	- 4 < x solicitudes pagadas = 10% de interés.	<b>Amarillo (1)</b>
	Independiente	1	Formal	0	Sin casa	1		
<b>Alto</b>	Independiente	1	Informal	1	Sin casa	1	- Cliente deudor = 20% de interés → Se recomienda un préstamo menor o igual 1000.	<b>Rojo (2)</b>
	Independiente	1	Informal	1	Precaria	1		

En la siguiente figura se puede observar las reglas que fueron creadas para la predicción y el resultado que nos arrojará el algoritmo de predicción.

```

200 var session = pl.create();
201 //VARIABLES
202 //var nivel = -1;
203 const reglas = `
204     %Hechos
205     %calcular_nivel
206     calcular_nivel(dependiente,formal,propia,0).
207     calcular_nivel(independiente,formal,propia,0).
208     calcular_nivel(dependiente,formal,sincasa,0).
209     calcular_nivel(independiente,formal,sincasa,1).
210     calcular_nivel(independiente,informal,sincasa,2).
211     calcular_nivel(independiente,informal,precaria,2).
212
213     %Reglas
214     calcular_nivel_final(A,B,C,X):- calcular_nivel(A,B,C,X).
215
216 `;

```

*Fig. 55. Reglas de predicción*

En la siguiente figura se muestra la función del cálculo del nivel de acuerdo con el comportamiento del cliente.

```

217 const calcular_nivel = (situacionLaboral, formalidad, tipoCasa) => {
218     return new Promise((resolve, reject) => {
219         session.consult(reglas, {
220             success: function () {
221                 /* Program loaded correctly */
222                 //console.log("PL Succes!!");
223                 session.query(
224                     "calcular_nivel_final(" +
225                     situacionLaboral +
226                     "," +
227                     formalidad +
228                     "," +
229                     tipoCasa +
230                     ", X ).",

```

*Fig. 56. Función del cálculo del nivel*

Se implementaron tres escenarios de predicción, tomando en cuenta la información necesaria brindada por la organización, los escenarios nos permitirán observar en qué nivel de riesgo se encuentra un cliente con el objetivo de poder tomar mejores decisiones dentro de la organización con el objetivo de disminuir el porcentaje de morosidad.

#### 4.2.3. Medir el nivel de precisión del sistema para determinar la confiabilidad de los resultados del proceso.

Se utilizaron las siguientes métricas de evaluación, para demostrar la confiabilidad de los resultados del proceso. A continuación, se mostrará la matriz de confusión, la que permitirá realizar los cálculos de las métricas mencionadas en la *tabla LIII*.

*TABLA 55. Matriz de confusión del modelo*

		Valor actual del crédito	
		Yes	No
Valor de predicción	Yes	1419	340
	No	225	1000

En primer lugar, se calculará el nivel de sensibilidad del sistema, como se muestra a continuación:

*TABLA LVI. Resultado de sensibilidad del modelo*

Métricas	Formula	Cálculo
	$\frac{TP}{TP + FP} * 100$	
Sensibilidad	Donde: TP: Verdaderos positivos FP: Falsos positivos	$\frac{1419}{1419 + 225} * 100 = 86\%$

Después, se procede a realizar el cálculo de la especificidad del sistema, como se muestra a continuación:

*TABLA LVII. Resultado de especificidad del modelo*

Métricas	Formula	Cálculo
	$\frac{VN}{VN + FP} * 100$	
Especificidad	Donde: VN: Verdaderos negativos FP: Falsos positivos	$\frac{1000}{1000 + 340} * 100 = 75\%$

Po último, se procede a realizar el cálculo de la eficacia del sistema, como se muestra a continuación:

*TABLA LVIII. Resultado de eficacia del modelo*

Métricas	Formula	Cálculo
	$\frac{VP + VN}{TI} * 100$	
Eficacia	<p><b>Donde:</b></p> <p><b>VP:</b> Verdaderos positivos</p> <p><b>VN:</b> Verdaderos negativos</p> <p><b>TI:</b> Total de instancias</p>	$\frac{1419 + 1000}{1419 + 340 + 225 + 100} * 100$ <p>= 81%</p>

Después de realizar los respectivos cálculos, se puede observar que el modelo tiene un 86% de sensibilidad que indica el porcentaje de créditos clasificados como aprobados, 75% especificidad que indica el porcentaje de créditos clasificados como rechazados y 81% de eficacia que indica el porcentaje de créditos clasificados como correctos.

#### 4.2.4. Validar el grado de usabilidad del modelo Credit Scoring por parte de los usuarios finales

Para determinar el cumplimiento de este objetivo, se realizó un cuestionario de validación a los usuarios finales que utilizarán el sistema, teniendo en cuenta parámetros de definidos por **System Usability Scale (SUS)**, donde se pretende evaluar la usabilidad y satisfacción. (Anexo 02)

Para analizar los resultados del cuestionario, se estableció la siguiente ponderación:

*TABLA LIX. Ponderación de las respuestas (SUS)*

ESCALA	PONDERACIÓN
Muy en desacuerdo	1
En desacuerdo	2
No estoy seguro	3
De acuerdo	4
Muy de acuerdo	5

Esta encuesta fue realizada a un total de 14 usuarios finales (Gerente general, analista de créditos y 12 trabajadores de la empresa) involucrados en el proceso de otorgamiento de créditos durante el último año.

La puntuación aceptable va desde los 68 puntos hasta los 100 puntos, el promedio de la escala de usabilidad del sistema es de 68 (Suficiencia),

Como se presenta en la tabla LI, la puntuación de validación final de los usuarios con respecto al sistema es de 85.96.

TABLA LX. Resultados de la herramienta SUS

ENCUESTADO	PREGUNTAS										P. PARES	P. IMPARES	SUB TOTAL	PUNTAJE TOTAL
	Usabilidad y Satisfacción													
	p1	p2	p3	p4	p5	p6	p7	p8	p9	p10				
N°01	4	2	5	2	4	3	4	1	4	1	16	16	32	80
N°02	5	2	4	2	4	3	5	1	4	1	17	16	33	82.5
N°03	5	2	5	2	5	3	5	2	5	1	20	15	35	87.5
N°04	5	2	4	2	5	3	4	1	4	1	17	16	33	82.5
N°05	5	2	5	1	5	3	5	2	5	1	20	16	36	90
N°06	5	2	5	2	5	3	5	1	5	1	20	16	36	90
N°07	5	2	4	1	5	3	4	1	5	1	18	17	35	87.5
N°08	5	2	5	2	5	3	5	1	5	1	20	16	36	90
N°09	4	2	4	1	4	3	4	1	4	1	15	17	32	80
N°10	5	2	5	1	4	3	5	2	5	1	19	16	35	87.5
N°11	5	2	5	2	5	3	4	2	4	1	18	15	33	82.5
N°12	5	2	5	2	5	3	5	1	5	1	20	16	36	90
N°13	5	2	4	2	5	3	5	1	5	1	19	16	35	87.5
N°14	4	2	5	2	4	3	4	1	4	1	16	16	32	80
<b>TOTAL</b>	<b>63</b>	<b>26</b>	<b>60</b>	<b>22</b>	<b>61</b>	<b>39</b>	<b>60</b>	<b>17</b>	<b>60</b>	<b>13</b>				
<b>%TOTAL</b>	90%	37.1%	85.7%	31.4%	87.1%	55.7%	85.7%	24.3%	85.7%	18.6%	<b>18.38</b>	<b>16</b>	<b>34.38</b>	<b>85.96</b>
<b>TOTAL</b> <b>CARACTERÍSTICA</b>	<b>POR</b>										<b>85.96</b>			

### **4.3. Impactos esperados**

Los impactos esperados son:

#### **4.3.1. Impactos económicos**

El modelo ayudará a la organización a identificar a los clientes con riesgo de morosidad para así planificar estrategias que le permitan reducir la pérdida de dinero por el otorgamiento de créditos, además de aumentar sus ganancias.

#### **4.3.2. Impactos sociales**

El modelo permitirá la identificación de las causas que influyen en la morosidad de los clientes, así como la insatisfacción de estos con el objetivo de desarrollar estrategias y así solucionar oportunamente la calidad del servicio y trato del cliente.

#### **4.3.3. Impactos en tecnología**

El modelo servirá de precedente para futuras investigaciones puedan aplicar estrategias de minería de datos para predecir el riesgo de morosidad de los clientes.

#### **4.3.4. Impactos ambientales**

El modelo no requiere implementación tecnológica por lo que la amenaza medioambiental será muy baja.

**Discusión:**

En la presente investigación pretende analizar comparativamente las diferentes técnicas CS con el fin de determinar la que mejor se adapte a la organización, a su vez, se realizó una comparación entre los algoritmos supervisados que mayormente usan los autores y realizarse las respectivas pruebas, obteniendo un 86% de precisión en el modelo basado en el algoritmo OneR, esto demostró que dicho algoritmo es el que mejor se adapte a la organización. El algoritmo OneR fue empleado por [36], [37], en sus respectivos contextos, demostrando la facilidad de uso y su efectividad. Después de analizar los algoritmos utilizados en esta investigación se puede afirmar que el algoritmo OneR es factible utilizar en contextos de clasificación.

Con respecto al desarrollo del modelo CS mediante la técnica seleccionada se desarrolló el módulo de predicción, teniendo en cuenta los parámetros de entradas (situación laboral, formalidad laboral, con/sin casa propia, entre otras) que definen el comportamiento del cliente que a su vez nos permitirá conocer el nivel de riesgo en el otorgamiento del crédito. Siguiendo las fases de la metodología CRISP-DM que está compuesta por seis fases y en cada una de ellas podemos encontrar sus respectivas actividades. [6], [8] emplearon dicha metodología debido que les permite tener una guía de referencia en el desarrollo de proyectos de MD.

La presente investigación se centró en el desarrollo de modelos predictivos de aprendizaje supervisado, como lo son los algoritmos de clasificación debido a que éstos pueden trabajar en base a una variable objetivo, la cual permitirá conocer el nivel de riesgo del cliente.

Finalmente, se puede afirmar que la metodología CRISP-DM sirve como guía para el desarrollo y evaluación del modelo CS, así mismo los pasos a seguir para el despliegue de mismo.

A su vez, con respecto a los escenarios de predicción como apoyo a la toma de decisiones en el proceso de otorgamiento de créditos, se desarrolló una interfaz web, utilizando el lenguaje JavaScript, HTML y con respecto a la gestión de la base de datos, MySQL. Se logró crear una interfaz interactiva para el usuario que permita visualizar el nivel de riesgo en el que se encuentra un cliente. Con respecto a los escenarios de predicción como apoyo a la toma de decisiones en el proceso de otorgamiento de créditos.

Teniendo en cuentas las investigaciones [4], [6], [7] y [8], se desarrolló la interfaz web local que permite la conexión entre el usuario y el modelo predictivo, a su vez permitirá registrar la información en base a la información brindada por el cliente, después de ello se podrá visualizar la predicción mostrando el nivel de riesgo para su posible otorgamiento de

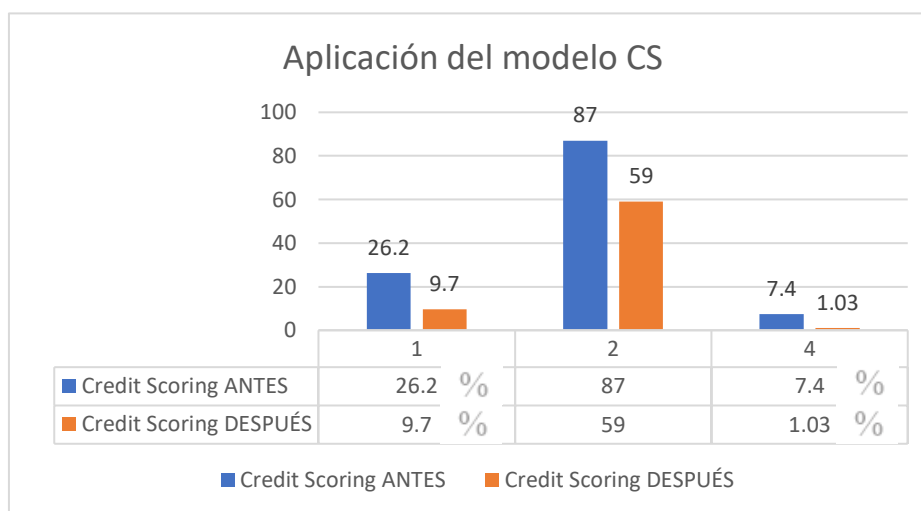
crédito, por último, dicha interfaz permitirá al usuario final visualizar reportes en formatos PDF y Excel. Todo ello permitirá mejorar la situación actual en la que encuentra la organización tomando en cuenta la entrevista realizada con el analista de créditos y el gerente (*Ver Anexo 06*), acelerando el proceso en el otorgamiento de los créditos, disminuyendo el porcentaje de morosidad debido a los créditos mal colocados, además, se realizaron pruebas de caja blanca y caja negra para el correcto funcionamiento de ejecución y la correcta funcionalidad de la interfaz final. Con respecto al diseño de la investigación como se puede mostrar en la tabla LXI, se logró determinar que después de aplicar el modelo CS la financiera puede tomar mejores decisiones, con una mejora en los créditos la morosidad se vería reducida en un 6.1% y además de poder disminuir el tiempo en el proceso de otorgamiento de créditos

Por último, para obtener la puntuación de la usabilidad y satisfacción del sistema se ha utilizado el cuestionario desarrollado por SUS, como resultado se obtuvo 85.96 puntos en promedio a los 14 participantes, con ello se puede afirmar que el sistema es útil, funcional y satisface las necesidades de la financiera.

Tomando en cuenta todo lo mencionado anteriormente, se realizó el documento de aprobación del producto acreditable por parte del gerente general de la financiera (*Ver Anexo 07*), se afirma que se logrará reducir el tiempo en el proceso de otorgamiento de créditos personales y determinar el nivel de riesgo en el que se encuentra un cliente para su posible otorgamiento de crédito, encontrarán en el modelo una herramienta fácil de utilizar y precisa para determinar el nivel de riesgo de los clientes.

*TABLA LXI. Comparativa antes/después de aplicar el modelo CS*

PASO	DESCRIPCIÓN	Credit Scoring	
		ANTES	DESPUÉS
1	Créditos mal colocados (%)	26.2 %	9.7 %
2	Demora en el proceso (h)	87	59
4	Índice de morosidad (%)	7.4 %	1.03 %



*Fig. 57. Comparativa antes/después de aplicar el modelo CS*

## V. Conclusiones

Después de aplicar el modelo CS, para predecir el riesgo en el otorgamiento de créditos en la financiera, se llegaron a las siguientes conclusiones:

1. Después de analizar comparativamente los tres algoritmos de clasificación empleados en esta investigación, se logró identificar que algoritmo que mejor se adapta a la organización fue el algoritmo OneR con un nivel de precisión del 86%. A su vez, se obtuvo un total de 81% de eficacia que representa la proporción de verdaderos positivos que el modelo pudo predecir correctamente, es decir, el porcentaje de predicciones realizadas correctamente.
2. Se logró construir en base al algoritmo seleccionado el modelo de predicción considerando las variables que logran definir el comportamiento del cliente empleando la metodología CRISP-DM como guía de desarrollo ejecutando sus 6 fases, manteniendo el nivel de precisión obtenido por el sistema.
3. Se logró identificar los tres escenarios de predicción en base a las variables del modelo, todo el modelo predictivo se implementó en una interfaz web local construida utilizando los lenguajes de programación JavaScript, HTML y el entorno de base de datos MySQL. La interfaz permite generar la predicción que determina el nivel de riesgo en el proceso de otorgamiento de créditos, a su vez permite generar reportes en formatos PDF y EXCEL. Además, el usuario final podrá identificar a un cliente de forma más rápida con la semaforización asignada a cada nivel: Nivel bajo (*Verde*), Nivel medio (*Amarillo*) y Nivel alto (*Rojo*).

4. Finalmente, se determinó el puntaje de usabilidad del modelo utilizando la herramienta SUS. Con respecto a ello se obtuvo 85.96 puntos lo que nos permite determinar que el modelo supera la puntuación aceptable establecida por la herramienta SUS.

## **VI. Recomendaciones**

1. Se recomienda utilizar algoritmos empleando la técnica de Redes Neuronales para mejores resultados debido a la capacidad de aprender progresivamente sobre la Base de Datos con la que se cuenta.
2. Se recomienda mejorar el modelo planteado utilizando diferentes variables que puedan ajustar los resultados de este para obtener un mayor grado de precisión y exactitud para predecir el otorgamiento de riesgo.
3. A fin de lograr una herramienta dinámica y aceptable, el modelo Credit Scoring construido se deberá calibrar al menos cada seis meses para que se mantenga actualizado.
4. Finalmente, si bien es cierto la presente investigación consideró utilizar el algoritmo de clasificación OneR, existen una variedad de algoritmos que pueden servir para futuras investigaciones y mejorar el nivel de precisión del sistema.

## VII. Referencias

- [1] Banco Mundial, América Latina y el Caribe: Panorama general, oct. 2022, [En línea]. Disponible: <https://www.bancomundial.org/es/region/lac/overview>, [Accedido 24-abr-2023]
- [2] Superintendencia de Banca y Seguros (SBS), 2022. [En línea], Disponible: [https://www.sbs.gob.pe/app/stats\\_net/stats/EstadisticaBoletinEstadistico.aspx?p=1#](https://www.sbs.gob.pe/app/stats_net/stats/EstadisticaBoletinEstadistico.aspx?p=1#), [Accedido 24-abr-2023]
- [3] Ministerio de Economía y Finanzas (MEF), 2022. [En línea]. Disponible: <https://www.gob.pe/mef>, [Accedido 24-abr-2023]
- [4] V. M. Laguna, “Diseño de un modelo de credit scoring en la gestión de riesgo crediticio en la cartera de microempresa de las cooperativas de ahorro y crédito”, Ecuador, 2015.
- [5] J. Aboobyda y M. Taring, “*Developing prediction model of loan risk in banks using data mining*,” vol. 3, Khartoum, Sudan, 2016.
- [6] J. J. Rodríguez y M. Miñano, “Desarrollo de una aplicación informática basada en un modelo de machine learning para mejorar la evaluación de préstamos crediticios”, Universidad Privada del Norte, Trujillo, 2017.
- [7] E. M. Alarcón y B. J. Mora, “Modelo para la evaluación del riesgo crediticio para los clientes de las microfinancieras del Perú”, Universidad Peruana de Ciencias Aplicadas, Lima, 2020.
- [8] M. P. Medina y H. G. Ulfe, “Modelo de credit scoring para predecir el otorgamiento de crédito personal en una cooperativa de ahorro y crédito”, Lambayeque - Perú, 2015.
- [9] S. A. Ross, R. W. Westerfield y J. F. Jaffe, *Finanzas corporativas*, 9<sup>o</sup> edición, México, McGraw Hill, 2012.
- [10] M. Medina y H. Ulfe, “Modelo de credit scoring para predecir el otorgamiento de crédito personal en una cooperativa de ahorro y crédito”, Dep. Estadística, Universidad Nacional Pedro Ruiz Gallo, Lambayeque, 2015.
- [11] B. Echemendía, “Definiciones acerca de riesgo y sus aplicaciones”, vol. 49, n<sup>o</sup> 3, pp. 470-481, sep.-dic. 2011.
- [12] D. Chorafas, *managing credit risk, analyzing rating and pricing the probability of default*, Euromoney Institutional Investor, Londres, 2000.
- [13] L. Thomas, B. Edelman y J. Crook, “Credit scoring and its applications,” SIAM modeling and computation Philadelphia, 2002.

- [14] J. C. Millán y E. Caicedo, “Modelos para otorgamiento y seguimiento en la gestión de riesgo de crédito”, *Revista de métodos cuantitativos para la economía y la empresa*, pp. 23-41, jul. 2017.
- [15] D. Lawrence y A. Solomon, “Managing a consumer lending business,” 1° edición, New York, 2002.
- [16] Microsoft 2016, Como elegir algoritmos para aprendizaje automático de Microsoft Azure, [En línea]. Disponible en: <https://docs.microsoft.com/es-es/azure/machine-learning/how-to-select-algorithms>, [Accedido: 24-abr-2023].
- [17] M. Gámez y N. García, “Rating de pequeñas y medianas empresas mediante Árboles de clasificación”, España, 2000.
- [18] G. Fernández, “*Data mining using SAS applications*,” 1° ed., Nueva York, Chappman & Hall, 2002.
- [19] J. Marín, “*Introducción a las redes neuronales aplicadas*”, vol. 3, n° 2, jul. 2012.
- [20] C. Serrano y B. Martin del Brío, “Fundamentos de las redes neuronales artificiales: Hardware y Software”, vol. 1, n° 1, pp. 103-125, ene-jun 1995.
- [21] R. Hernández, C. Fernández y M. Baptista, *Metodología de la Investigación*, 6ª ed. México D.F., México: McGRAW-HILL/Interamericana Editores, S.A. de CV., 2014.
- [22] E. Llorca, “7 riesgos de los créditos personales y como superarlos”, 2020.
- [23] I. Wilford Rivera, «MINERÍA DE DATOS: HERRAMIENTA DE APOYO EN LA SELECCIÓN DE EQUIPOS DE PROYECTOS INFORMÁTICOS,» *Industrial*, vol. 27, N° 3, 2006
- [24] M. I. Mejía Rocha y M. Colín Salgado, «GESTIÓN DEL CONOCIMIENTO Y SU IMPORTANCIA EN LAS ORGANIZACIONES,» *Revista TRILOGÍA*, N° 9, pp. 25-35, 2013
- [25] A. Gonzales, “¿Qué es Machine Learning?”, 2020, [En línea], Disponible: ¿Qué es Machine Learning? – Cleverdata, [Accedido 24-abr-2023].
- [26] M. Pérez, *Minería de Datos a través de ejemplos*, Ciudad de México: Editorial Alfaomega, 2015.
- [27] E. Alpaydin, “*Introduction to Machine Learning*,” 2da edición, Cambridge, Estados Unidos: Massachusetts Institute of Technology, 2010.
- [28] J. L. Sánchez Ramírez, «*Fundamentos de la Minería de Datos*,» Chalco de Díaz Covarrubias, 2018.

- [29] IBM Corporation, Manual CRISP-DM de IBM SPSS Modeler, 2012 [En línea]. Disponible en: <https://www.ibm.com/docs/es/spss-modeler/SaaS?topic=overview-crisp-dm-in-spss-modeler> [Accedido: 24-abr-2023].
- [30] D. A. Pinto Galindo, «Diseño de un Modelo Predictivo de Fuga de Clientes Utilizando Algoritmos Machine Learning,» Universidad ECCI, Bogotá, 2020.
- [31] IEEE Std.830 - 1998, “Especificación de Requisitos según el estándar de IEEE830”, oct. 2008.
- [32] A. Géron, Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow, 2nd Edition, O'Reilly Media, Inc., 2019.
- [33] J. Stefanowski, and S. Wilk, Extending Rule-Based Classifiers to Improve Recognition of Imbalanced Classes. *Advances in Data Management SCI*, 223, pp. 131-154, 2009.
- [34] Data Science, Predicting the Future, Disponible en: <https://www.saedsayad.com/oner.htm> [Accedido: 24-abr-2023].
- [35] S. Vijayarani y M. Muthulakshmi, Comparative analysis of bayes and lazy classification algorithms, *International Journal of Advanced Research in Computer and Communication Engineering*, pp. 3118-3124, 2013.
- [36] Y. Cardoso y L. Arza, “OneR Algorithm. Its application on Clinical Trials”, *Revista Cubana de Ciencias Informáticas*, vol. 11, Cuba, 2017.
- [37] K. Eckert y R. Suénaga, “Análisis de Deserción-Permanencia de Estudiantes Universitarios Utilizando Técnica de Clasificación en Minería de Datos”, vol. 11, Argentina, 2015.