

8

¿Es un cliente confiable?... Minería de datos tradicional y Fintech para el cálculo de la rentabilidad

Cristian Garcés Poveda¹, Alexandra González Eras ²,
1,2 Universidad Técnica Particular de Loja, 1 cagarces@utpl.edu.ec,
2 acgonzalez@utpl.edu.ec

Fecha de presentación: 08 de julio de 2019

Fecha de aceptación: 29 de julio de 2019

DOI: <https://doi.org/10.33386/593dp.2019.5-1.156>

RESUMEN

Este trabajo presenta una propuesta para la detección de clientes corporativos fiables en función de sus perfiles de usuario en entidades crediticias, según un modelo de clasificación automática basado en redes neuronales, el cual logra una alta precisión en relación con algoritmos de aprendizaje tradicionales. Para el desarrollo del modelo de minería usamos una adaptación de la metodología CRISP-DM que permite la creación de un modelo confiable que puede integrarse en plataformas de servicios en la nube. La implementación de este modelo ofrece una reducción de tiempo en el cálculo de la rentabilidad de los clientes, es de fácil aplicación en servicios corporativos y la visualización de los resultados para la consulta y la toma de decisión. De esta forma el modelo se constituye en una propuesta Fintech válida para las empresas corporativas y de finanzas.

Palabras clave: rentabilidad. CRISP-DM. modelos de clasificación automática. redes neuronales. fintech. minería de datos.

ABSTRACT

This paper presents a proposal for the detection of reliable corporate clients based on their user profiles in credit institutions, according to an automatic classification model based on neural networks, which achieves high accuracy in relation to traditional learning algorithms. For the development of the mining model, we use an adaptation of the CRISP-DM methodology that allows the creation of a reliable model that can be integrated into cloud service platforms. The implementation of this model offers reduced time to calculate customer profitability; it is easy to apply in corporate services and the visualization of the results for the consultation and decision-making. Thus, the model constitutes a valid Fintech proposal for corporate and finance companies.

Keywords: profitability. CRISP-DM. automatic classification models. neural networks. Fintech. data mining.

Introducción

Durante los últimos 30 años la globalización, el masivo uso de Internet, la aparición de nuevas tecnologías ha generado en las empresas constantes modificaciones e innovaciones. Precisamente el uso de las tecnologías financieras o Fintech ha logrado no solamente el incremento de sus beneficios sino la optimización de procesos, la gestión de grandes cantidades de datos para la toma de decisiones, la generación de nuevos servicios y productos; en definitiva, ha hecho que las empresas sean más competitivas en relación con aquellas que realizan sus actividades de la forma tradicional.

Este trabajo presenta una propuesta para la clasificación de clientes corporativos según su rentabilidad, para ello el modelo de minería de datos basado en redes neuronales se implementa, según una adaptación de la metodología CRISP-DM en el cual, en la fase de despliegue, ofrece una interfaz para la visualización interactiva de los resultados del modelo. La implementación de este modelo ofrece una reducción en el tiempo que conlleva el cálculo de la rentabilidad de los clientes, y al mismo tiempo se constituye en una herramienta fácil de integrar en servicios corporativos en la nube.

Antecedentes

“La tecnología financiera, a menudo reducida a Fintech, es la tecnología y la innovación que tiene como objetivo competir con los métodos financieros tradicionales en la prestación de servicios financieros” (Lin, 2015). Es una industria emergente que utiliza la tecnología para mejorar las actividades en finanzas (Schueffel, 2016); además se le considera como “una nueva industria financiera que aplica tecnología para mejorar las actividades financieras” (Scholten, 2017). FinTech son las nuevas aplicaciones,

procesos, productos o modelos de negocios en la industria de servicios financieros que, de la mano de las técnicas de aprendizaje proporcionadas por la inteligencia artificial, permiten la automatización de seguros, operaciones, servicios bancarios y gestión de riesgos (Aldridge, 2017). La novedosa Fintech representa un reto, tanto para las empresas como para los clientes, que requieren productos y servicios atractivos, rentables y, que cumplan con todas las normativas vigentes por las entidades financieras de control.

Debido al crecimiento y variedad de información disponible en las instituciones financieras, se hace necesario el análisis de datos en diferentes áreas (marketing, créditos, riesgos, etc.) (Tase, 2016). En relación con el análisis de la rentabilidad de los clientes de instituciones financieras existen trabajos que proponen modelos predictivos para optimizar campañas bancarias, por ejemplo, de tarjetas y líneas de crédito, por medio de una segmentación de cliente según sus niveles de riesgo, rentabilidad y morosidad (Venegas, 2016) (Pacco, 2015). De igual forma, mediante modelos de minería de datos se autoriza créditos según el historial crediticio del cliente y se establecen segmentos de acuerdo con las políticas de la institución (Medina, 2017). Otra área es la identificación de fraudes informáticos donde a través de un modelo de minería se detectan anomalías en la información (Nuñez, 2018).

Los trabajos de investigación expuestos identifican a CRISP-DM como la metodología escogida para este tipo de investigaciones, por facilitar el proceso de análisis y diseño de modelos; además el uso de técnicas de clasificaciones como Regresión Logística, Redes Neuronales, Arboles de Decisión, Redes Bayesianas (Venegas, 2016) permiten la obtención de resultados de alta precisión en la identificación de patrones

y predicción de hechos. Los aportes de estas investigaciones son de ayuda para las instituciones en especial para los departamentos comerciales, riesgos, finanzas, en donde los resultados que ofrecen los modelos de minería apoyan al diseño de estrategias, por ejemplo, incrementar ventas (cubrir segmentos que actualmente no se tiene posicionamiento), mejorar la cartera (identificación de mejores perfiles para la colocación de créditos) (Bonnes, 2017), campañas de servicios, productos según el perfil del cliente y su rentabilidad (de acuerdo a características y políticas de las instituciones) (Scholten, 2017).

La rentabilidad es considerada como el rubro que mide los resultados de una empresa en un período de tiempo determinado, que permite trazar estrategias para la toma de decisiones (Curto, 2015), en el control e identificación de oportunidades (Santiesteban, 2011), y las estrategias para incrementar la rentabilidad, según políticas institucionales (Scholten, 2017).

Consideremos el proceso clásico de obtención de la rentabilidad de clientes, que se presenta en la Figura 1, en donde el problema se centra en la dificultad de acceder a la información para realizar análisis oportunos y completos por parte de los ejecutivos, ya que por lo general los datos reposan en diferentes unidades y departamentos, no se encuentran estandarizados, su consolidación requiere de un tiempo adicional que retarda la generación de informes hacia directivos y, de esta forma, no se logra una visión integral de la situación financiera de un cliente corporativo.

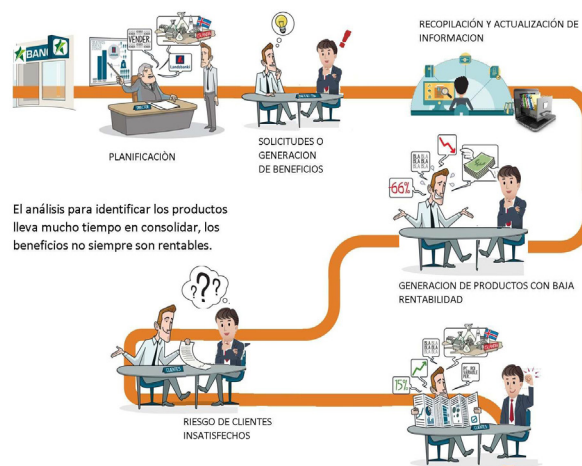


Figura 1. La problemática del cálculo tradicional de la rentabilidad.

Mediante un proceso de minería de datos apoyado en tecnologías financieras nuestra propuesta cumple con las expectativas de los ejecutivos comerciales, generando una alternativa de rentabilidad para la toma de decisiones, tomando en cuenta: (1) los productos y servicios necesarios para la generación de la rentabilidad de clientes, (2) clientes potenciales considerando la rentabilidad. (3) forma de despliegue de los resultados de rentabilidad en tiempo real. (4) tiempos de respuesta que optimicen recursos y mejoren los indicadores de eficiencia de los ejecutivos comerciales. (5) atención eficiente a clientes según su perfil rentable.

En los siguientes apartados se presenta la metodología empleada, experimentación y resultados, así como las conclusiones y recomendaciones del presente estudio.

Metodología

La Fig. 2 presenta la metodología de la propuesta, la cual se fundamenta en el modelo CRISP-DM, de amplio reconocimiento en el desarrollo de procesos de minería de datos (Bonnes, 2017). En este caso se ha hecho una selección de las actividades de cada fase según el propósito del modelo, las

6 fases que conforman la metodología son: Comprensión del negocio, donde se establece los objetivos del modelo de minería, en relación con los requisitos de la empresa; Comprensión de los datos, donde se define las características de los datasets, el tamaño de la muestra y empieza la recolección de los datos. En la Preparación de los datos se establece la consistencia de los datos y se toman decisiones sobre la inclusión o exclusión de estos; el Modelado es donde se seleccionan los algoritmos de aprendizaje automático idóneos para el modelo; en la Evaluación se analiza y depura el modelo en función de los resultados de la aplicación de los algoritmos. Finalmente, en el Despliegue se escogen los mecanismos adecuados para la visualización de los resultados hacia usuarios finales.



Figura 2. CRISP-DM adaptada al caso de estudio.

A continuación, se presenta cada una de estas fases alineada con nuestro caso de estudio, el cual corresponde a una entidad bancaria y el análisis de clientes corporativos, que generan importantes ingresos a este tipo de entidades financieras.

Comprensión del negocio

Se analiza la problemática en el

cálculo de la rentabilidad, se definen los subproblemas con su correspondiente causa y efecto y en función de esta relación se determinan los objetivos del modelo, los cuales se resumen en los siguientes: (1) Identificar y clasificar la rentabilidad de los clientes corporativos según su perfil de uso de productos y servicios, y (2) Visualizar la probabilidad de un cliente de aumentar o disminuir su rentabilidad en función de sus productos y comportamiento dentro de la institución financiera. Adicionalmente, en esta etapa se consideran las presunciones y restricciones con respecto al acceso y gestión de la información para el modelo de minería.

Comprensión de los datos

Para realizar la selección de datos se realiza un enfoque principal en el cliente corporativo, adicional se consultan todas las operaciones y productos financieros que puede tener dentro de la organización, y que serán utilizados para la generación de la rentabilidad dentro del modelo de minería de datos. De igual forma, se realiza la obtención estadística de la muestra según la cantidad de datos legalmente disponibles, quedando conformada por 311 casos. Otro aspecto de esta fase es la obtención de los indicadores y variables que forman parte del cálculo de la rentabilidad de los clientes. La Figura 3 presenta el conjunto de indicadores que, según los productos y servicios que ofrece la institución bancaria, conforman la rentabilidad de un cliente corporativo.



Figura 3. Comprensión de los datos Preparación de los datos

El proceso de preparación comienza con la revisión de la consistencia de los datos, así como con la selección o exclusión de aquellos registros que generan distorsión en el modelo de minería. La Figura 4 presenta el resultado del análisis de los datos, donde se observa que la mayor concentración de clientes se encuentra en las 3 primeras clases, es decir un 96% del total de clientes corporativos; además, de acuerdo a políticas de la institución aquellos clientes que presentan una rentabilidad menor a \$0.00 son una "perdida (ver la Figura 4 parte superior) en este caso, un total de 78 clientes y clientes rentables clasificados en (rentabilidad baja, rentabilidad media y rentabilidad alta).

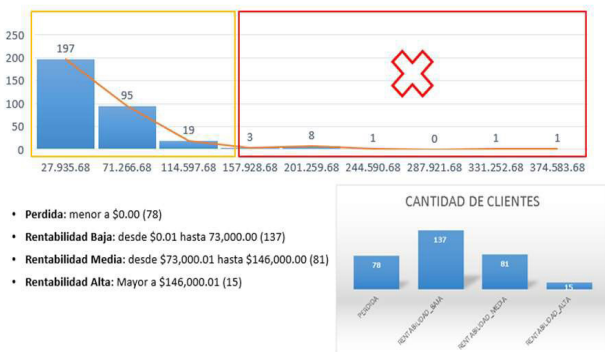


Figura 4. Preparación de los datos

Para optimizar el dataset se eliminan los ítems con rentabilidad cero y se realiza una nueva distribución de estos según los rangos de la rentabilidad definidos: pérdida, baja, media y alta, de tal forma que se logra un balance de los seleccionados según la cantidad de clientes que se ubican en cada rango (ver la Figura 4 parte inferior).

Modelado

El proceso de modelado comprende la selección de la técnica de aprendizaje automático para la clasificación de la rentabilidad, en función de una revisión de trabajos relacionados que se presenta en la Tabla 1, lo cual facilitó la selección del algoritmo de aprendizaje automático más

recomendable; en este caso, el modelo de redes neuronales por la precisión que alcanza en la clasificación de clientes según su perfil crediticio (Medina, 2017) (Pacco, 2015) (Gironés, 2017) (Martín del Peso, 2017).

Trabajo	Objetivo	Técnica utilizada	Cantidad de datos	Confiable
[6]	Computas bancarias por medio de segmentación de clientes según riesgo y rentabilidad.	Arboles de decisión Regresión Logística Redes Neuronales	Se detalla la posible construcción de un modelo de riesgo y rentabilidad	No se especifica
[8]	Pronóstico de riesgo de morosidad.	Redes Neuronales y Clusters	Se considera una muestra de 130 alumnos.	Precisión 88% Error 2.67
[7]	Modelo de créditos de consumo, identificando morosidad y Score.	Regresión Logística, Arboles de decisión y Redes Neuronales	3000 clientes que forman parte de la cartera.	Precisión 84.18% - Error 0.16
[9]	Identificar posibles fraudes tributario	Arboles de decisión	• 8191 para ejecución I • 113 para ejecución II • 629 para ejecución III	No se especifica
[13]	Determinación de compañías y análisis de consumos por tarjeta de crédito.	Redes Neuronales	No se especifica cantidad de datos	Precisión 89% Error 0.20
[14]	Problemas de predicción y clasificación financiera	Redes Neuronales	No se especifica cantidad de datos	Precisión entre el 85% y 90%

Tabla 1. Trabajos relacionados

El siguiente paso consiste en obtener la configuración óptima del modelo neuronal para nuestro caso de estudio. En la Figura 5 se observa las pruebas realizadas en el modelo neuronal, siendo la configuración de 3 capas la seleccionada para el modelo de minería de datos por los valores de precisión que ofrece al ser entrenada con el conjunto de datos de clientes corporativos.

Cantidad de capas	DETALLE			DETALLE		
	4 CAPAS			3 CAPAS		
Cantidad de nodos por capa	3,4,5,4	3,4,6,4	3,8,2,4	3,2,4	5,4,2	4,5,4
Correctly Classified Instances	42	42	42	83	83	83
Incorrectly Classified Instances	64	64	64	23	23	23
Kappa statistic	0	0	0	0.6745	0.6745	0.6745
Mean absolute error	0.3392	0.3392	0.3392	0.1747	0.1762	0.1749
Root mean squared error	0.4197	0.4197	0.4197	0.2949	0.2941	0.2949
Relative absolute error	99.6336%	99.6335%	99.6339%	51.3140%	51.7400%	51.3818%
Root relative squared error	100.5668%	100.5668%	100.5676%	70.6651%	70.4732%	70.6693%
Total Number of Instances	106	106	106	106	106	106

Cantidad de capas	DETALLE			DETALLE		
	2 CAPAS			1 CAPA		
Cantidad de nodos por capa	3,7	4,2	6,5	6	3	2
Correctly Classified Instances	83	83	83	83	83	83
Incorrectly Classified Instances	23	23	23	23	23	23
Kappa statistic	0.6745	0.6745	0.6745	0.6745	0.6745	0.6745
Mean absolute error	0.1745	0.174	0.1746	0.1734	0.173	0.196
Root mean squared error	0.2944	0.2945	0.2944	0.294	0.294	0.2943
Relative absolute error	51.2623%	51.1211%	51.2854%	50.9489%	51.0786%	49.6544%
Root relative squared error	70.5435%	70.5694%	70.5554%	70.4507%	70.4640%	70.5263%
Total Number of Instances	106	106	106	106	106	110

Figura 5. Modelado

Resultados y evaluación

La evaluación del modelo se realiza en función de los objetivos de los que debe

cumplir el modelo de minería, para ello se establecen dos casos: (1) Clasificación de productos y servicios con variables (si / no), valor mayor a cero rentable, menor a cero no rentable, y (2) la clasificación con relación a sus valores de rentabilidad. La Tabla 2 presenta el esquema de evaluación para los dos casos, el cual establece las condiciones en la que los resultados se consideran óptimos, como por ejemplo características de los datasets de prueba y validación en función de los variables de rentabilidad establecidas y la configuración del modelo neuronal.

Datos de prueba	Datos de validación	Cantidad de capas	Variables
66% = 205 clientes de entrenamiento	34% = 106 clientes para la evaluación de la red	una capa con 3 nodos. recomendado por Weka y validado por tener el menor valor de error	<ul style="list-style-type: none"> Cash Management SIT Adquierecia Rentabilidad

Tabla 2. Evaluación del Caso 1

Además, en esta fase se evalúan los resultados obtenidos en el momento del despliegue del modelo, tomando en consideración el tiempo que requiere el cálculo de la rentabilidad según el modelo de minería y la forma tradicional. Para la evaluación de tiempos de respuesta, se considera el tiempo del proceso actual y la simulación de la propuesta de minería de datos con los ejecutivos comerciales, estableciendo el promedio de las operaciones que realizan todos los ejecutivos de manera mensual. Este análisis se presenta en detalle en el apartado de Análisis de resultados.

Despliegue del modelo

El despliegue del modelo considera la estrategia mediante la cual los resultados del modelo serán presentados a los usuarios. En ese sentido se ha considerado el uso de la herramienta Power BI por su capacidad de integración con los programas de minería utilizados comúnmente en las instituciones financieras (Weka, RapidMiner), además

de que permite el aprovechamiento de los servicios de la Nube para la generación de servicios Fintech de visualización de datos en tiempo real (Metro, 2018) (Tangerine, 2018).

La visualización con Power BI, permite la interacción con diversas bases de datos, así como la actualización y visualización en diferentes dispositivos. Dentro de la plantilla de visualización se muestra todos los datos relevantes para el ejecutivo, como los valores de rentabilidad de los productos, la posición del cliente en fuentes externas, montos de depósitos, cartera, póliza, garantías, contingentes y la probabilidad que tiene el cliente de pertenecer a otra categoría.



Figura 6. Despliegue en Power BI

Experimentación

El proceso de experimentación se inicia con la selección de los datasets en función del tamaño de muestra y las consideraciones de la Tabla 1, con la configuración del modelo neuronal indicado en la Figura 5. A continuación se presenta el proceso para los dos casos definidos.

Caso 1

En el primer caso se considera el valor de rentabilidad en cada uno de los productos, permitiendo realizar una clasificación de cumplimiento de utilidad o perdida, en función a los resultados de ingresos y gastos en cada producto o según lo siguiente: (1) SI, considerando

que representan una utilidad, en donde la rentabilidad es mayor a \$0.01, (2) NO, considerando que representa una pérdida, en donde la rentabilidad es menor a \$0.00, y (3) NA, considerando que el cliente no dispone de ese producto. El modelo neuronal tomando en cuenta los cuatro rangos presentados en la Figura 4 obtiene un 78.30% de ítems correctamente clasificados, y un 21.69% que no son clasificados correctamente, con un error absoluto medio de 0.17.

La Figura 7 presenta los resultados, en donde se observa que los casos que no fueron clasificados correctamente se encuentran en el grupo de clientes con rentabilidad alta, que representan el 4.8% del total de clientes, los cuales se excluyen del dataset a efecto de analizar el modelo con la mayor concentración de datos. El resultado de esta nueva ejecución del modelo alcanza un 89% de ítems correctamente clasificados y un 10% de ítems no clasificados correctamente, con un error de 0.18, el cual se considera bajo en relación con el porcentaje de aciertos alcanzado por el modelo (vea la Figura 8), como se observa el modelo es muy cercano a los datos reales.

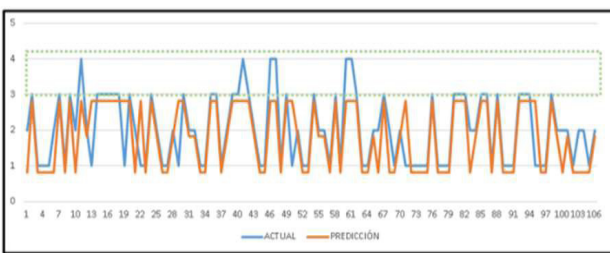


Figura 7. Comparación del modelo actual y la predicción para los 4 rangos definidos

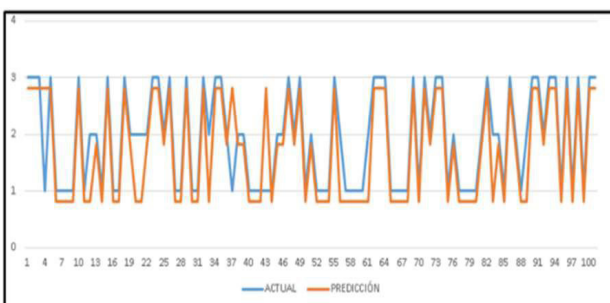


Figura 8. Comparación del modelo actual y la predicción para los 3 rangos definidos

Caso 2

En el segundo caso se considera el valor de rentabilidad en función de sus valores por cada producto con clasificación de los montos finales de rentabilidad. El modelo neuronal tomando en cuenta los cuatro rangos presentados en la Figura 4 obtiene un 87.73% de ítems correctamente clasificados, y un 12.26% que no son clasificados correctamente, con un error absoluto medio de 0,12.

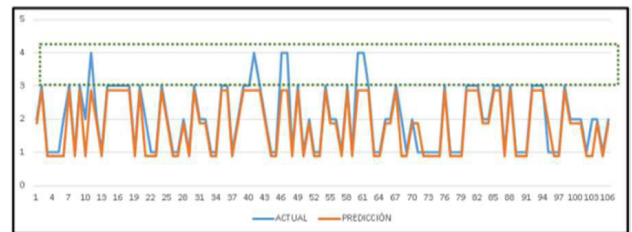


Figura 9. Comparación del modelo actual y la predicción para los 4 rangos definidos

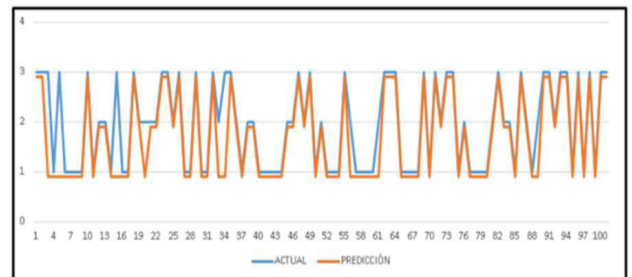


Figura 9. Comparación del modelo actual y la predicción para los 3 rangos definidos

La Figura 9 presenta los resultados, en donde se observa que los casos que no fueron clasificados correctamente se encuentran en el grupo de clientes con rentabilidad alta, que representan el 4.8% del total de clientes. De igual forma que en el caso 1 se excluyen del dataset este grupo de clientes y, el resultado de esta nueva ejecución del modelo alcanza un 91.08% de ítems correctamente clasificados y un 8.91% de ítems no clasificados correctamente, con un error de 0.095, el cual se considera bajo

en relación con el porcentaje de aciertos alcanzado por el modelo (vea la Figura 10). Se puede afirmar entonces que, en los dos casos tanto si consideramos el rango de alta rentabilidad como si no, los modelos arrojan resultados cercanos a los datos reales, con errores relativamente bajos.

Análisis y discusión

En este apartado se presentan los resultados obtenidos con nuestra propuesta desde dos perspectivas: en primer lugar, el modelo de minería de datos y en segundo lugar el despliegue del modelo en la institución financiera.

Resultados del modelo de minería de datos

Después de estudiar los casos propuestos, se considera los siguientes porcentajes promedios de éxito que se presentan en la Tabla 3, la cual recopila un resumen de todos los casos y la efectividad de cada uno de ellos, estos resultados permiten realizar una comparación y análisis.

Rentabilidad actual	Predicción	Caso 1			Caso 2		
		Cantidad de cliente	Aciertos general	Acierto por categoría	Cantidad de cliente	Aciertos general	Acierto por categoría
Pérdida	Perdida	53	17%	68%	66	21%	85%
	Rentabilidad_baja	24	8%	31%	12	4%	15%
	Rentabilidad_mediana	1	0%	1%	-		
	Total	78			78		
Rentabilidad_alta	Rentabilidad_mediana	15	5%	100%	15	5%	100%
	Total	15			15		
Rentabilidad_baja	Perdida	2	1%	1%	3	1%	2%
	Rentabilidad_baja	110	35%	80%	132	42%	96%
	Rentabilidad_mediana	25	8%	18%	2	1%	1%
	Total	137			137		
Rentabilidad_mediana	Rentabilidad_baja	-			2	1%	2%
	Rentabilidad_mediana	81	26%	100%	79	25%	98%
	Total	81			81		
Total general		311	78%		311	89%	

Tabla 3. Comparación del Caso 1 y 2

En el primer caso representa una efectividad mayor en la clasificación de rentabilidad media, clasificando el

100% de la categoría, para la categoría rentabilidad baja, clasifica el 80% de 137 clientes, y la categoría de perdida clasifica el 68% que representa 53 clientes. En el segundo caso la categoría rentabilidad media es clasificada un 98%, que representa 79 clientes, en la categoría rentabilidad baja, clasifica un 94% que representa 129 clientes, y en la categoría perdida clasifica el 85% que representa 66 clientes. También el caso dos representa una mejor clasificación con 93% de clientes, esto se debe a que el modelo en este caso clasifica mejor los clientes de rentabilidad baja pasando de 110 en el primer caso a 129 en el segundo se debe tomar en cuenta que esta categoría tiene una mayor concentración de datos comparado con las demás, por su parte la categoría pérdida pasa de una clasificación de 53 a 66 clientes en el segundo caso.

Los resultados muestran un alto nivel de confiabilidad, esto se debe principalmente al proceso de depuración y estandarización de los datos realizados en etapas previas, que permiten eliminar clientes y productos atípicos que causan distorsiones en el modelo de minería.

Resultados del despliegue del modelo

La Figura 11 presenta los resultados del despliegue del modelo de minería de datos, como se observa los tiempos de ejecución de los procesos realizados en los diferentes departamentos sufre una considerable disminución en relación con el proceso tradicional. En el presente caso de estudio, el volumen promedio de operaciones es de 92 operaciones a nivel nacional, con una concentración del 64% de la original (nuevos créditos), siendo el proceso de nuevos créditos el de mayor impacto, la variación con los otros procesos (Renovación 23%, Contingentes 9%, Refinanciamiento 3%) cumplen con un proceso similar con la diferencia en el tiempo de validación de documentación.

(Renovación 2:00:00, Contingentes 1:00:00, Refinanciamiento 2:00:00).

En el proceso actual (vea la Figura 11) se considera un promedio de 17 horas con 30 minutos en la aprobación de un nuevo crédito, y 16 horas con 30 minutos si el crédito es negado, este tiempo contempla desde la llegada del cliente a la institución financiera, la recepción de documentos, el análisis y validación de condiciones, y la notificación. Cabe indicar que el tiempo considerado es únicamente del proceso de ejecutivo comercial, debido a que el proceso se relaciona con otros procesos de otros departamentos, como control y evaluación quienes son los encargados de validar toda la documentación física del cliente y en casos de refinanciamiento los casos son expuestos a un comité de evaluación con el departamento de riesgos. Representado en la gráfica como (A).

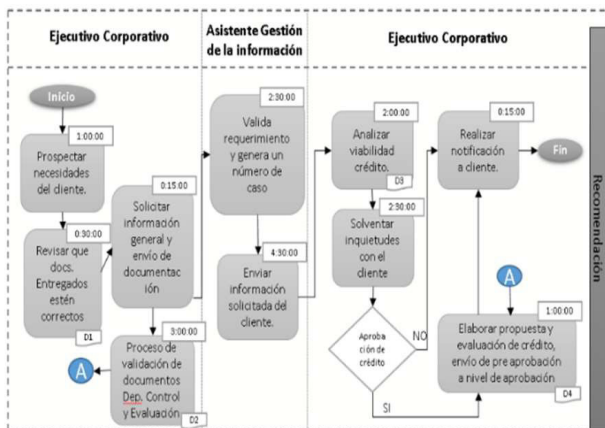


Figura 11. Flujo y tiempo del proceso de cálculo de rentabilidad actual.

Además, durante el proceso se genera una serie de documentos, como son: 1) D1.- Información Financiera, flujos de caja, Documentos Legales, Situper (Garante-Codeudores). 2) D2.- Validación de información. 3) D3.- Análisis y generación de rentabilidad y 4) D4.- Propuesta y evaluación de crédito. Como se puede observar en la Figura 10, la mayor concentración de tiempo la lleva la generación de un caso y envío

de información al ejecutivo comercial por parte de gestión de la información 7 horas, más 2 horas en el proceso de análisis y validación del crédito por parte del ejecutivo.

Con la propuesta generada en el proceso de minería de datos (vea la Figura. 12), se considera una reducción de tiempo a 9 horas en la aprobación y 8 horas en los créditos negados, lo que representa (47% de reducción de tiempo en aprobado y 50% de tiempo en créditos negados). Esto debido a la eliminación del proceso de envío de casos a otro departamento, así como la consolidación de la información. Otra reducción importante es el proceso de análisis y validación, en donde el ejecutivo pasa de un tiempo de 2 horas a solo 30 minutos con el modelo propuesto, esta reducción se debe principalmente a la disponibilidad de la información y el contar con información consolidada y analizada previamente.

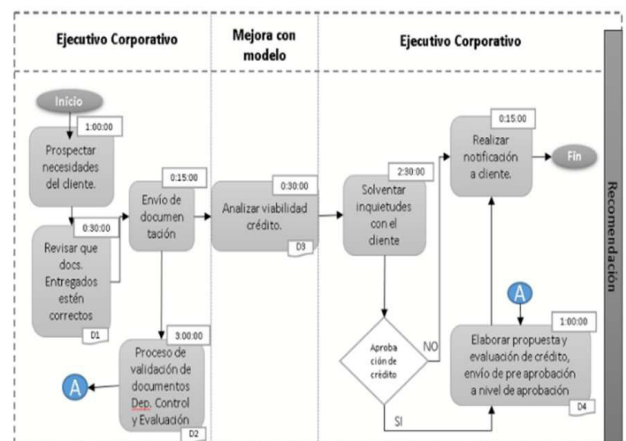


Figura. 12 Flujo y tiempo del proceso de cálculo con el modelo de minería

En la Figura 12 se observa la disminución del tiempo en el subproceso de verificación y generación de número de caso, así como el envío de la información solicitada, y el efecto positivo que genera el modelo de minería de datos, es importante considerar que los tiempos de recepción, así como de elaboración de propuesta no se ven afectados.

En consecuencia, el modelo de minería permite la optimización de los tiempos de respuesta del cálculo de la rentabilidad de los clientes, permitiendo el desarrollo de gestiones adicionales y así mejorar los indicadores de desempeño del personal de la institución. De igual forma, con el modelo de minería se evita la dependencia de otros departamentos para el proceso de análisis de los clientes, lo que contribuye a la generación de cálculos de rentabilidad en tiempo real y con ello disponer de información oportuna sobre los productos y servicios que podrían ser requeridos por los clientes.

Conclusiones y trabajos futuros

Este trabajo presenta un modelo de minería de datos orientado hacia la clasificación automática de clientes corporativos rentables, mediante un modelo neuronal que obtiene una precisión de Información relevante para la administración y gestión diaria mediante la aplicación Power BI. De esta forma el modelo de minería de datos de clientes corporativos permite a los ejecutivos tener, información relevante con fácil acceso para la administración y gestión diaria, con un 50% de reducción en el tiempo de procesamiento.

Además, el cálculo de rentabilidad proporciona una visión completa de la utilidad o pérdida que representa el cliente, incluyendo la probabilidad de presentar un cambio en la rentabilidad (aumentar o disminuir), permitiendo a los ejecutivos ofertar alternativas de productos. De igual forma, para todo proceso de actualización de datos que se requiera incorporar en el modelo, se deberá considerar un proceso previo de depuración y validación.

Con respecto a la distribución de la clasificación de rangos de utilidad puede ser modificada por la institución financiera en referencia a políticas internas que permitan generar grupo de clientes por niveles de rentabilidad. Así también el modelo

de cálculo de rentabilidad propuesto considera una rentabilidad por producto, con lo cual permite conocer que producto o servicio es relevante para la rentabilidad de los clientes.

El modelo de minería de datos permite el reconocimiento de productos y servicios que actualmente no se consideran dentro de la rentabilidad, mejorando la efectividad del cálculo. De esta manera, según los trabajos relacionados, se logra una variación de la rentabilidad comparada con el proceso actual con una confiabilidad del 89%.

Los trabajos futuros se encaminan hacia el desarrollo de modelos que consideren otros servicios y productos, además de la extensión del modelo hacia otros grupos de clientes. A nivel del despliegue del modelo, es necesaria la integración de los datos de los departamentos que intervienen en el cálculo de la rentabilidad en una infraestructura de información en la Nube que garantice eficiencia en la gestión de los datos.

Referencias bibliográficas

- Aldridge, I., & Krawciw, S. (2017). Real-time risk: What investors should know about FinTech, high-frequency trading, and flash crashes. John Wiley & Sons.
- Bonnes, K. (2017). Predicting mortgage demand using machine learning techniques (Master's thesis, University of Twente).
- Curto Díaz, J., & Braulio Gil, N. (2015). Customer analytics: mejorando la inteligencia del cliente mediante los datos. Customer analytics, 1-116.
- Gironés, J., Casas, J., & Minguillón, J. (2017). Minería de datos: modelos y algoritmos". Editorial UOC.
- Lin, T. C. (2015). Infinite financial

- intermediation. *Wake Forest L. Rev.*, 50, 643.
- Martin del Peso, M. (2017). Aplicaciones de las redes neuronales artificiales a problemas de predicción y clasificación financiera, Universidad Rey Juan Carlos, Madrid.
- Medina Rodríguez, M. D. P., & Ulfe Rentería, H. G. (2017). Modelo de Credit Scoring para Predecir el Otorgamiento de Crédito Personal en Una Cooperativa de Ahorro y Crédito.
- Microsoft Power BI customers. Metro Bank case study. Customers.microsoft.com. Retrieved from <https://customers.microsoft.com/en-US/story/building-a-bank-that-can-surprise-and-delight-with-pow>
- Microsoft power BI customers. Tangerine case study. Customers.microsoft.com. Retrieved from <https://customers.microsoft.com/Pages/CustomerStory.aspx?recid=14594>
- Nuñez, J. P., Savoy, F., Retamar, M. S., De Gracia, L., De Battista, A. C., & Ramos, L. M. M. (2018). Minería de datos para la detección de factores de influencia en el test Apgar.
- Pacco Palomino, R. (2015). Análisis predictivo basado en redes neuronales no supervisadas aplicando algoritmo de kmedias y crisp-dm para pronóstico de riesgo de morosidad de los alumnos en la Universidad Peruana Unión.
- Santiesteban, E., Fuentes, V., & Leyva, E. (2011). Análisis de la rentabilidad económica: tecnología propuesta para incrementar la eficiencia empresarial. Editorial Universitaria.
- Scholten, U. (2017). Banking-as-a-Service - what you need to know. *VentureSkies*.
- Schueffel, P. (2016). Taming the beast: a scientific definition of fintech. Available at SSRN 3097312.
- Tase, R. O. R., Cabrera, A. V., Naranjo, D. L. O., Diaz, A. A. O., & Blanco, I. F. (2016). Nuevo algoritmo multclasificador para flujos de datos con cambios de concepto/ensemble algorithm for data streams with concept drift. *Holos*, 32(2), 24.
- Venegas, L. (2016). Plan para enfocar las campañas Bancarias utilizando Datamining.