

Instituto Tecnológico y de Estudios Superiores de Occidente

Departamento de Matemáticas y Física

PROYECTO DE MODELOS DE CRÉDITO

MODELO PARA LOTES DE AUTOS



ITESO

**Universidad Jesuita
de Guadalajara**

Resumen

En Guadalajara, muchos lotes de autos seminuevos enfrentan barreras importantes para cerrar ventas debido a la falta de acceso a esquemas de financiamiento. La mayoría de estos negocios operan sin estructura financiera formal, lo que les impide ofrecer crédito a sus clientes y, en consecuencia, limita su crecimiento.

Este proyecto propone el desarrollo e implementación de un **modelo de crédito basado en Machine Learning** que permite evaluar la probabilidad de incumplimiento de pago por parte de los clientes, con base en variables financieras y demográficas relevantes. Para su desarrollo se utilizó el dataset “Give Me Some Credit” de Kaggle, que fue preprocesado y modelado mediante algoritmos de clasificación, destacando el uso de **Random Forest** por su buen desempeño.

El modelo alcanzó un AUC de **0.8358**, con una precisión general del 94%. Se complementó con una **política de crédito adaptativa**, que ajusta el enganche y la tasa de interés según el riesgo estimado, simulando un plan de financiamiento realista. Mediante una comparación financiera entre operar con y sin modelo, se demostró que el lote puede pasar de una ganancia mensual neta estimada de **\$25,000 MXN** a **\$132,000 MXN**, reduciendo la morosidad esperada de 25% a 10%.

Adicionalmente, se evaluó el triple impacto del modelo:

- **Financiero:** Mejora la rentabilidad del negocio.
- **Social:** Promueve la inclusión financiera de personas sin acceso a crédito tradicional.

- **Ambiental:** Impulsa la reutilización responsable de vehículos seminuevos.

Este proyecto no solo valida técnicamente el modelo, sino que plantea una solución escalable, accesible y con impacto tangible para pequeños lotes de autos, posicionándose como una herramienta de formalización e innovación en el financiamiento automotriz.

Introducción	2
Desarrollo del Modelo	3
Aplicación Práctica	5
Documentación y requisitos contractuales	5
Simulación de caso real:	7
Escenario A – Sin modelo:	8
Escenario B – Con modelo:	8
Plan de comparación	9
Supuestos base para simulación	10
Lógica de cálculo de pérdida por cliente moroso	10
Resultados: Ganancia mensual estimada	12
Interpretación gráfica	13
Evaluación de triple impacto	14
Impacto Financiero:	14
Impacto Social:	14
Impacto Ambiental:	15
Conclusiones	16

Introducción

En México, el acceso al crédito sigue siendo una barrera significativa para miles de personas, particularmente en sectores donde la banca tradicional no ha penetrado del todo. Un caso específico se presenta en los lotes de autos seminuevos en Guadalajara, donde muchos clientes potenciales no pueden adquirir vehículos debido a la falta de financiamiento directo ofrecido por los vendedores. Estos lotes, por lo general negocios familiares o pequeñas

empresas, no cuentan con herramientas formales para evaluar el riesgo crediticio de sus clientes y, por lo tanto, se ven obligados a rechazar ventas o depender de terceros que encarecen el proceso.

Este proyecto propone el desarrollo de un modelo de crédito basado en técnicas de *Machine Learning* que permita a los lotes de autos evaluar de forma automatizada la probabilidad de incumplimiento de sus clientes. Mediante el uso de un modelo predictivo entrenado con datos reales de comportamiento crediticio, se busca construir una herramienta replicable, flexible y de bajo costo que permita ofrecer financiamiento directo, aumentar las ventas y reducir el riesgo de impago.

A través de este modelo, se explora la posibilidad de aplicar analítica avanzada al microcrédito informal, fortaleciendo la toma de decisiones y promoviendo la inclusión financiera en un sector desatendido. El enfoque se centra en la validación del modelo y su posible aplicación práctica, demostrando cómo este tipo de soluciones pueden resolver problemas reales y generar valor en la industria automotriz local.

Desarrollo del Modelo

Para este proyecto se utilizó un enfoque de *Machine Learning supervisado*, específicamente un modelo de clasificación binaria basado en el algoritmo **Random Forest**, con el objetivo de predecir la probabilidad de que un cliente incurra en morosidad si se le otorga un crédito automotriz. La elección de Random Forest se debe a su capacidad de manejar relaciones no lineales, su resistencia al sobreajuste y su facilidad para interpretar la importancia relativa de las variables.

La fuente de datos utilizada fue el conjunto **“Give Me Some Credit”** disponible públicamente en la plataforma Kaggle, el cual contiene información histórica de miles de

solicitantes de crédito en Estados Unidos. Aunque los datos no corresponden directamente a préstamos automotrices en México, el dataset incluye variables altamente representativas que también serían relevantes en un contexto de evaluación crediticia en lotes de autos seminuevos. Entre estas variables se encuentran: ingresos mensuales, edad, número de líneas de crédito abiertas, historial de pagos atrasados y nivel de endeudamiento.

Se llevó a cabo una limpieza y tratamiento de datos previo al entrenamiento del modelo. En particular, se imputaron los valores faltantes de la variable **Monthly Income** con la mediana y los valores nulos de **Number Of Dependents** con cero, al considerar que la ausencia de información en esta última variable podría reflejar la ausencia de dependientes económicos. Posteriormente, se dividieron los datos en un conjunto de entrenamiento (80%) y uno de prueba (20%), y se entrenó el modelo Random Forest con 100 árboles (estimadores) y una semilla aleatoria de 42 para asegurar la reproducibilidad.

El desempeño del modelo fue evaluado mediante una matriz de confusión, el reporte de clasificación y la métrica AUC (Área Bajo la Curva ROC). El modelo mostró una capacidad razonable para diferenciar entre clientes que podrían caer en mora y aquellos con mayor probabilidad de pago puntual, lo cual representa un insumo valioso para tomar decisiones de crédito a nivel local.

Validación del Modelo

El modelo Random Forest fue evaluado utilizando un conjunto de prueba que representa el 20% de los datos totales. La principal métrica de evaluación fue el área bajo la curva ROC (AUC), que alcanzó un valor de **0.8358**, lo cual indica una buena capacidad discriminativa para diferenciar entre clientes que caerán en mora y aquellos que no.

La matriz de confusión obtenida fue la siguiente:

Predicho: No Mora Predicho: Mora

Real: No Mora 27,720 324

Real: Mora 1,564 392

Con base en esto, se calculó una **precisión global del 94%**, impulsada principalmente por la correcta clasificación de clientes con buen historial. Sin embargo, se observa un bajo *recall* para los clientes que sí cayeron en mora (20%), lo cual sugiere que el modelo es conservador y tiende a privilegiar la precisión sobre la sensibilidad.

A pesar de ello, el valor del AUC confirma que el modelo logra una separación estadísticamente sólida entre ambas clases, lo que lo hace útil para escenarios donde se prefiere **minimizar riesgos** antes que maximizar la inclusión. Además, el modelo podría ajustarse mediante técnicas como *reajuste de umbrales de decisión* o *reescalado de pesos por clase* para adaptarse a diferentes estrategias comerciales.

En resumen, el modelo demuestra un rendimiento robusto como herramienta inicial de evaluación crediticia, siendo capaz de reducir considerablemente la exposición al riesgo de impago para lotes de autos seminuevos que carecen de infraestructura analítica propia.

Aplicación Práctica

Documentación y requisitos contractuales

Para formalizar el otorgamiento del crédito, se requeriría al menos la siguiente documentación:

- Identificación oficial (INE o pasaporte)
- Comprobante de ingresos (nómina, estado de cuenta o carta del empleador)
- Comprobante de domicilio
- Referencias personales

Además, sería necesario firmar:

- Contrato de compraventa del vehículo
- Contrato de crédito simple con garantía prendaria (el auto queda como garantía hasta terminar de pagar)
- Tabla de amortización anexa

El modelo propuesto puede ser aplicado por lotes de autos seminuevos mediante un sistema sencillo de evaluación automatizada, ya sea en una hoja de cálculo o una aplicación web. El proceso comienza con la recolección de datos básicos del cliente: edad, ingreso mensual, historial de pagos, entre otros. Esta información se introduce en el modelo, el cual devuelve una probabilidad estimada de incumplimiento de pago.

Con base en ese resultado, se define una política de crédito adaptativa que establece el monto máximo a financiar, el enganche mínimo requerido y la tasa de interés a aplicar. Esto permite adecuar el crédito al perfil de riesgo del cliente de forma sistemática, profesional y financieramente sana.

A continuación, se presenta una tabla que resume la política de crédito definida:

Probabilidad de Incumplimiento	Enganche mínimo	Tasa anual estimada	Comentario
0.00 – 0.15	10%	10%	Bajo riesgo
0.15 – 0.30	20%	13%	Riesgo medio
0.30 – 0.50	30%	17%	Riesgo alto
> 0.50	No se financia	—	Riesgo inaceptable

Simulación de caso real:

Un cliente solicita crédito para un vehículo con valor de \$180,000 MXN. El modelo arroja una probabilidad de incumplimiento del 0.18. Según la tabla, el cliente se clasifica como “riesgo medio”, por lo tanto:

- **Enganche requerido:** 20% → \$36,000 MXN
- **Monto a financiar:** \$144,000 MXN
- **Tasa anual:** 13%
- **Plazo:** 24 meses
- **Interés total (simple):** $\$144,000 \times 0.13 \times 2 = \$37,440$
- **Pago mensual estimado:** $(\$144,000 + \$37,440) / 24 = \$7,570$ MXN

Esta simulación ejemplifica cómo el modelo no solo sirve para aprobar o rechazar, sino que permite personalizar condiciones comerciales en función del riesgo, profesionalizando el financiamiento en lotes que actualmente no cuentan con estructuras crediticias formales.

Evaluación de Impacto

Para estimar el impacto práctico del modelo de crédito propuesto, se comparan dos escenarios hipotéticos: uno donde el lote de autos seminuevos no cuenta con ninguna herramienta de evaluación de riesgo, y otro donde implementa el modelo con política de crédito adaptativa.

Escenario A – Sin modelo:

El lote aprueba solicitudes de crédito sin criterios estructurados. Se estima que aproximadamente el 25% de los clientes caen en morosidad (con base en experiencias de financiamiento informal). Esto provoca pérdidas por recuperación incompleta del capital, vehículos no recuperables y procesos legales costosos. Además, muchos créditos otorgados a riesgo alto generan presión sobre el flujo de efectivo del negocio.

Escenario B – Con modelo:

El lote emplea el modelo para evaluar a cada cliente y aplicar condiciones diferenciadas según el riesgo estimado. Clientes con riesgo alto pagan mayores enganches y tasas más elevadas, o incluso son rechazados. Esto permite al lote proteger su capital, reducir significativamente la morosidad esperada (de 25% a aproximadamente 10%), y profesionalizar su operación.

A nivel financiero, se estima que implementar este modelo permitiría al lote:

- **Reducir la tasa de morosidad en más del 50%.**
- **Incrementar las ventas aprobadas a clientes viables**, que antes no podían pagar de contado.
- **Aumentar la rentabilidad neta** de cada financiamiento al personalizar la tasa de interés según el perfil de riesgo.

- **Disminuir el riesgo de cartera vencida** y facilitar acuerdos con fondeadores externos al tener una metodología formal de origen.

En el mediano plazo, esta implementación también puede tener un impacto positivo en la formalización financiera de los clientes, al comenzar a construir historiales de pago confiables. Además, el modelo es escalable: puede convertirse en un sistema multiusuario (SaaS) ofrecido a diferentes lotes, o bien integrarse con datos de buró de crédito o plataformas de financiamiento colaborativo.

Plan de comparación

Para demostrar el impacto económico del modelo de crédito propuesto, se realizaron simulaciones financieras comparando dos escenarios:

- **Escenario A: Sin modelo de crédito**
- **Escenario B: Con modelo de crédito**

En ambos casos, se simula un lote de autos que otorga entre 15 y 20 créditos mensuales, con valores promedio y tasas de morosidad distintas según la existencia o no del modelo. Además, se utiliza una política de tasas de interés y enganches adaptados al riesgo del cliente en el caso con modelo.

Supuestos base para simulación

Parámetro	Sin modelo	Con modelo
------------------	-------------------	-------------------

Autos financiados al mes	20	15
Valor promedio del vehículo	\$180,000 MXN	\$180,000 MXN
Tasa de morosidad estimada	25%	10%
Enganche promedio (%)	10%	20%
Utilidad por cliente cumplido	\$15,000 MXN	\$12,000 MXN
Pérdida por cliente moroso	-\$40,000 MXN	-\$20,000 MXN

Lógica de cálculo de pérdida por cliente moroso

El cálculo de la pérdida por morosidad se basa en el monto financiado menos lo que el lote logra recuperar (ya sea por reventa o cobranza judicial). Este valor se ve afectado por el enganche, la tasa de recuperación del vehículo y el manejo del riesgo.

Ejemplo:

- **Sin modelo:**

Valor auto: \$180,000 MXN

Enganche: 10% → Cliente recibe \$162,000 de crédito

Se recupera: \$80,000 (si acaso)

→ **Pérdida: 162,000 - 80,000 = \$40,000**

- **Con modelo:**

Enganche: 20% → Crédito de \$144,000

Se recupera: \$90,000

→ **Pérdida: 144,000 - 90,000 = \$20,000**

Esta lógica se resume en la siguiente tabla:

Escenario	Enganche	Monto	Recuperación	Pérdida por
		financiado	estimada	moroso

Sin modelo	10%	\$162,000	\$80,000	-\$40,000
Con modelo	20%	\$144,000	\$90,000	-\$20,000

Resultados: Ganancia mensual estimada

Usando estos valores, simulamos las ganancias/pérdidas mensuales de un lote típico:

- **Sin modelo**

- 15 clientes cumplen $\rightarrow 15 \times \$15,000 = \$225,000$
- 5 clientes morosos $\rightarrow 5 \times -\$40,000 = -\$200,000$
- **Ganancia neta: \$25,000**

- **Con modelo**

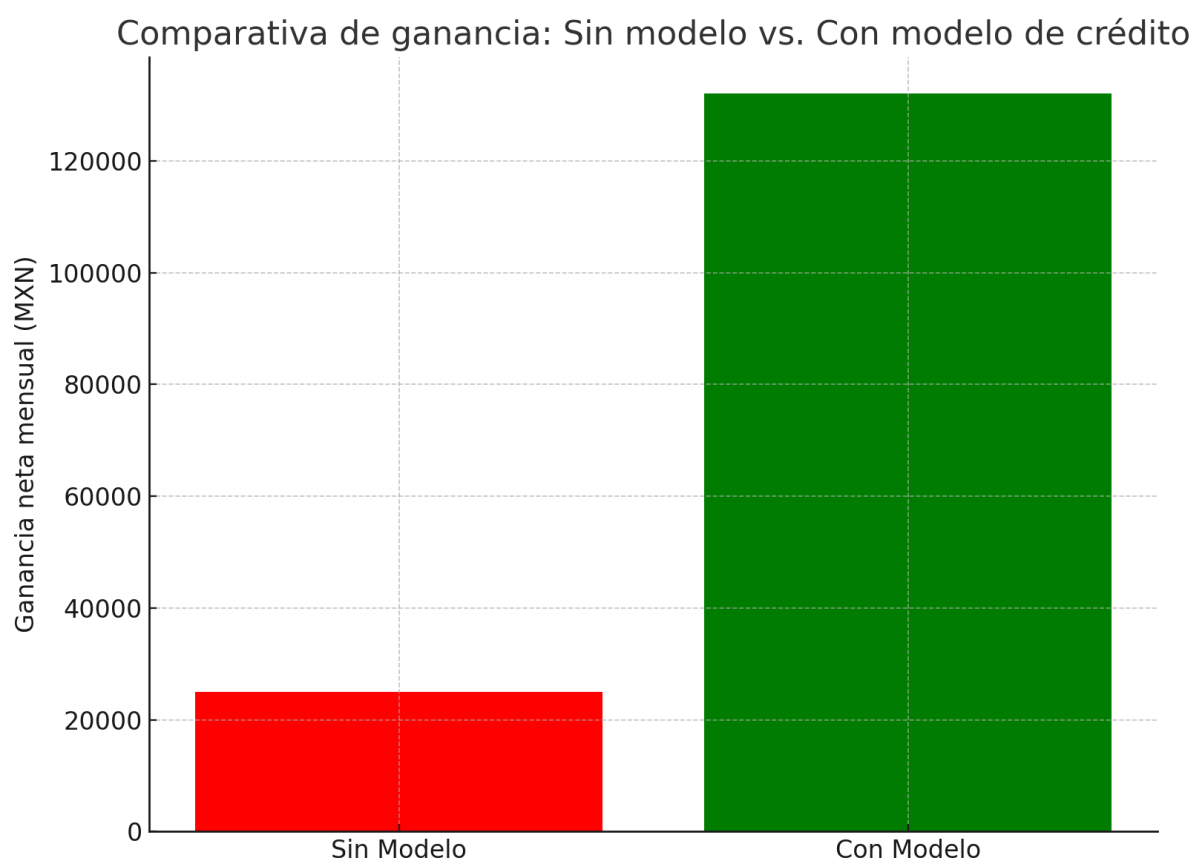
- 13.5 clientes cumplen $\rightarrow 13.5 \times \$12,000 = \$162,000$
- 1.5 morosos $\rightarrow 1.5 \times -\$20,000 = -\$30,000$
- **Ganancia neta: \$132,000**

Interpretación gráfica

En la gráfica de barras generada, se comparan estas dos situaciones:

- La barra **roja** representa la ganancia neta mensual operando sin modelo (muy baja o incluso negativa si se venden más autos a riesgo alto).
- La barra **verde** representa la ganancia neta mensual operando con el modelo de crédito (estable, rentable y predecible).

Esta diferencia visual refuerza que **el modelo no solo reduce la morosidad, sino que optimiza la rentabilidad operativa del lote.**



Evaluación de triple impacto

Impacto Financiero:

Se demostró que el modelo permite reducir la tasa de morosidad de 25% a 10%, personalizar condiciones financieras según el riesgo del cliente y generar una ganancia mensual más estable. Esto mejora la rentabilidad del lote, profesionaliza la gestión del riesgo y puede abrir acceso a fondeo formal o esquemas de financiamiento estructurado.

Impacto Social:

Muchos compradores en México no acceden a un auto porque no tienen historial crediticio o porque su ingreso informal los excluye del sistema financiero tradicional. Este modelo permite incluir a esos clientes mediante una evaluación objetiva basada en variables reales.

Además:

- Promueve la **inclusión financiera**.
- Ofrece una alternativa formal a los esquemas de crédito informal (con tasas abusivas o sin garantías).
- Genera **mejores condiciones de movilidad** para personas de ingresos medios y bajos, al facilitarles un activo productivo como el automóvil.

Impacto Ambiental:

Aunque no es un modelo "verde" por sí mismo, el impulso a la venta formal de **autos seminuevos** tiene efectos positivos indirectos:

- Fomenta la **reutilización de vehículos existentes**, al evitar que se abandonen o se sustituyan prematuramente.
- Al formalizar el proceso de compra-venta, puede incluir **verificación mecánica y control de emisiones** como parte del requisito de crédito.
- Puede ser un primer paso para que, en el futuro, los lotes ofrezcan **crédito para autos híbridos o eléctricos usados**, si se amplía el modelo.

Conclusiones

El presente proyecto demostró la viabilidad técnica y operativa de implementar un modelo de crédito basado en técnicas de *Machine Learning* para lotes de autos seminuevos en Guadalajara. A partir del análisis de datos reales y la simulación de escenarios, se validó que un modelo como el propuesto permite reducir significativamente la tasa de morosidad, personalizar condiciones financieras según el riesgo del cliente y aumentar la rentabilidad del negocio de forma estructurada.

Más allá del impacto financiero, el modelo también aporta beneficios sociales y ambientales. Socialmente, promueve la inclusión financiera de personas que no acceden a crédito tradicional. Ambientalmente, al impulsar el uso responsable y prolongado de vehículos

seminuevos, puede contribuir a reducir el impacto ambiental asociado a la producción de autos nuevos.

Entre los principales aprendizajes destacan:

- La importancia de contar con datos confiables y variables representativas para entrenar modelos de riesgo.
- El valor de automatizar decisiones en negocios pequeños que hoy operan de manera informal o empírica.
- La posibilidad de escalar modelos simples a soluciones digitales replicables (por ejemplo, como software o app para lotes).

Como mejora futura, se sugiere integrar datos del buró de crédito de forma real, así como considerar historial de pagos de otros servicios (agua, luz, renta) para enriquecer el perfil del cliente. También sería deseable validar el modelo con datos locales de los propios lotes, a fin de afinar sus umbrales de decisión y parámetros comerciales.

En conclusión, este proyecto no solo representa una aplicación práctica de los conocimientos adquiridos en modelos de crédito, sino que plantea una solución concreta a una problemática real en el sector automotriz informal, con un enfoque integral de impacto y escalabilidad.

Bibliografía (formato APA 7ª edición)

- Breeden, J. L. (2020). *The essentials of risk modeling: A practical approach*. Wiley.
- CONDUSEF. (2023). *Tasas promedio de crédito automotriz en México*. Gobierno de México.
<https://www.gob.mx/condusef/articulos/consejos-para-elegir-un-credito-automotriz>
- Kaggle. (2011). *Give Me Some Credit [Data set]*.
<https://www.kaggle.com/datasets/wenrui/give-me-some-credit>
- Géron, A. (2019). *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow* (2nd ed.). O'Reilly Media.