Angie Tatiana Aparicio Ochoa – Project Manager Harold Muñoz – Machine Learning Engineer Jhon Brandon Idarraga Cardenas – Data Engineer Jhonn Sebastian Robles Robles – Data Analyst

uNIVERSIDAD AUTONOMA DE OCCIDENTE  Cll 25 # 115-85 Km 2 Vía Cali - Jamundí

Software Modelador de Impagos (SMI)

Contenido

[Software Modelador de Impagos (SMI) 2](#_Toc175166995)

[Alineación Estratégica 2](#_Toc175166996)

[Propósito 2](#_Toc175166997)

[Alcance 2](#_Toc175166998)

[Producto 2](#_Toc175166999)

[Criterios de Éxito 2](#_Toc175167000)

[Planificación 2](#_Toc175167001)

[Hitos y Actividades 2](#_Toc175167002)

[Equipo 3](#_Toc175167003)

[Interesados 3](#_Toc175167004)

[Usuarios 3](#_Toc175167005)

[Recursos y Presupuesto 4](#_Toc175167006)

[Restricciones 4](#_Toc175167007)

[Riesgos 5](#_Toc175167008)

[Aprobación 5](#_Toc175167009)

# Software Modelador de Impagos (SMI)

## Alineación Estratégica

### Propósito

El proyecto tiene como propósito desarrollar un modelo predictivo de ML (Machine Learning) para identificar clientes con alta probabilidad de morosidad en el pago de cuotas mensuales para un banco en Taiwán; para ayudar a la entidad financiera a estar mas preparada ante una inminente morosidad por parte de cualquier cliente.

### Alcance

El alcance del proyecto es desarrollar un modelo predictivo de ML (Machine Learning) para identificar solo a clientes con alta probabilidad de morosidad en el pago de cuotas mensuales de sus tarjetas de crédito para un banco en Taiwán.

### Producto

Software Modelador de Impagos (SMI)

### Criterios de Éxito

|  |  |
| --- | --- |
| **Criterio** | **Indicador** |
| Precisión del modelo | Matriz de confusión:  Esta herramienta permite visualizar el desempeño del algoritmo, y también facilita ver que tipos de errores y aciertos tiene el modelo cuando está en el proceso de aprendizaje con los datos. |
| Reducción de la tasa de impagos | Indicador propio del banco. |
| Satisfacción de las partes interesadas (stakeholders) | Indicadores propios de cada departamento del banco mencionado como “stakeholders” en este documento. |

## Planificación

### Hitos y Actividades

|  |  |  |  |
| --- | --- | --- | --- |
| **Hito** | **Marco de Tiempo** | **Actividades** | **Entregables** |
| 1. Análisis exploratorio de los datos (EDA) | 6 días calendario | * 1. Cargue y visualización del dataset.   2. Explorar variables.   3. Identificar y manejar valores nulos o faltantes.   4. Detectar outliers.   5. Generar gráficos y visualizaciones. | Dataset con los datos completamente comprendidos y estadísticas del mismo junto con graficas. |
| 1. Selección y preparación de los datos | 6 días calendario | * 1. Realizar limpieza de datos.   2. Seleccionar características relevantes.   3. Codificar variables categóricas.   4. Normalizar o escalar variables numéricas.   5. Dividir el conjunto de entrenamiento y prueba. | Dataset con transformaciones necesarias de los datos y variables para su división en dos conjuntos (de entrenamiento y prueba para el modelo). |
| 1. Desarrollo del modelo ML | 13 días calendario | * 1. Seleccionar algoritmo de machine learning.   2. Entrenar modelo con datos de entrenamiento.   3. Evaluar rendimiento del modelo.   4. Optimizar hiperparámetros.   5. Aplicar técnicas de validación.   6. Documentar. | El mejor modelo de predicción posible de acuerdo con los datos de entrenamiento y prueba. |
| 1. Despliegue del modelo ML | 6 días calendario | * 1. Desplegar modelo final en AWS.   2. Configurar modelo para consulta desde aplicaciones externas.   3. Documentar proceso de despliegue. | El modelo de ML (Machine Learning) desplegado en AWS. |

### Equipo

|  |  |
| --- | --- |
| **Roles** | **Responsabilidades** |
| Angie Tatiana Aparicio Ochoa – Project Manager | * Gestionar el proyecto. * Coordinar las tareas de cada integrante del equipo. * Dirigir cada etapa del proyecto. |
| Harold Muñoz – Machine Learning Engineer | * Diseñar, construir y desplegar el modelo de Machine Learning que mejor se adapte a los datos en los diferentes ambientes. |
| Jhon Brandon Idarraga Cardenas – Data Engineer | * Asegurarse de que los datos estén disponibles y limpios para su análisis, correcta consulta y descarga para entrenar el modelo. |
| Jhonn Sebastian Robles Robles – Data Analyst | * Analizar información valiosa del dataset. * Identificar patrones y tendencias de los datos. |

### Interesados

|  |  |  |  |
| --- | --- | --- | --- |
| **Grupo de Interesados** | **Descripción del Interés** | **Nivel de Poder** | **Nivel de Influencia** |
| Departamento de Riesgos del banco | Este departamento le interesa la implementación del proyecto ya que reduciría la tasa de riesgos del banco de los tipos de clientes que pagan tarjetas de crédito. | Medio | Alto |
| Departamento de IT/Data Science | Tener una herramienta de Machine Learning que ayude no solo al departamento a tener más tecnología sino al banco para mejorar en este ámbito. | Medio | Medio |
| Gerencia del banco | Mejorar en la tasa de riesgos de los clientes que pagan tarjetas de crédito. | Alto | Alto |
| Departamento Legal y de Cumplimiento | Tener un mejor margen de cumplimiento por parte de los clientes que pagan tarjetas de crédito. | Medio | Alto |
| Clientes corporativos y accionistas | No tener pérdidas significativas por desfinanciamiento por incumplimientos de los clientes que pagan tarjetas de crédito en el banco. | Alto | Alto |

### Usuarios

|  |  |  |
| --- | --- | --- |
| **Grupo de Usuarios** | **Descripción** | **Beneficio** |
| Departamento de cobranzas | Personas que trabajan dentro del banco cuyo objetivo principal es cerrar el ciclo de ventas al asegurar que los pagos se realicen en los plazos y términos establecidos. | Anticiparse mediante el uso de la herramienta de Machine Learning a los posibles incumplimientos de los clientes que tengan tarjetas de crédito para realizar planes de mejora con respecto a la situación. |
| Gerencia y directivos del banco | Personas que administran todo el banco en general. | Disminuir la tasa de incumplimientos de los clientes de tarjetas de crédito para así también bajar el desfinanciamiento en ese sentido. |
| Auditores y equipos de cumplimiento | Son personas que dan recomendaciones y planes de mejora ya sea para un área en especifico o todas las áreas en este caso del banco tomando en cuenta los mismos procesos que realiza el banco en general. | Tener información mas clara de los clientes que probablemente incumplan con su tarjeta de crédito gracias al modelo de Machine Learning para así saber dar recomendaciones basadas en estos datos importantes. |

### Recursos y Presupuesto

|  |  |  |
| --- | --- | --- |
| **Tipo** | **Descripción** | **Estimación** |
| Equipo (Personas) | Hace referencia al tiempo invertido por las personas que intervienen dentro del proyecto. | * Project Manager   62 horas   * Machine Learning Engineer   155 horas   * Data Engineer   124 horas   * Data Analyst   93 horas |
| Dataset de clientes | Datos proporcionados por el banco de los clientes que tienen tarjeta de crédito. | No aplica |
| Python y sus librerías | Lenguaje de programación en donde se desarrollará el modelo de Machine Learning. | No aplica |
| Herramientas en la nube: Google colaboratory, AWS SageMaker, Sage Maker Studio, AWS Autocopilot | Herramientas que se encuentran en la nube y en donde se desplegara el modelo de Machine Learning para su consumo. | Se utilizarán las versiones gratuitas. |

### Restricciones

|  |  |
| --- | --- |
| **Tipo de Restricción** | **Descripción** |
| Disponibilidad y calidad de los datos | Que los datos proporcionados por parte del banco no estén en optimas condiciones como por ejemplo: datos erróneos, datos desactualizados, datos incompletos, datos mal clasificados; y que no siempre estén disponibles por parte del banco. |
| Regulaciones de privacidad | Que las leyes que cumple el banco impidan la transmisión, divulgación y tratamiento de los datos que sean fundamentales para el desarrollo del modelo de Machine Learning. |
| Recursos computacionales | Que el banco no tenga un recurso computacional potente para desplegar este modelo y consultarlo en la nube, que no tenga una ciberseguridad sólida. |

### Riesgos

|  |  |  |  |
| --- | --- | --- | --- |
| Descripción | Probabilidad | Impacto | Tratamiento |
| Sesgo en el modelo:  Que calcule mal la probabilidad de que un cliente incumpla sus obligaciones con su tarjeta de crédito. | Muy probable | Alto | Entrenar bien el modelo con los datos de entrenamiento y prueba. |
| Cambios en los patrones de comportamiento del cliente:  Que afecte la capacidad de cálculo de probabilidad del modelo. | improbable | Alto | Evaluar y actualizar el dataset con el que se entrenará y pondrá a prueba el modelo con los nuevos comportamientos de los clientes |
| Errores de predicción:  Obtener información errónea por parte del modelo y por lo tanto en las áreas que utilizan el modelo dentro del banco. | Probable | Alto | Revisando el modelo a profundidad y resolviendo el problema. |

### Aprobación

|  |  |  |
| --- | --- | --- |
| **Nombres** | **Rol** | **Firma** |
|  |  |  |
|  |  |  |
|  |  |  |