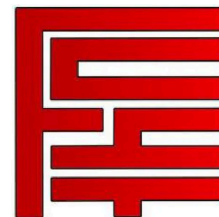


UNIVERSIDAD MAYOR DE SAN SIMÓN
FACULTAD DE CIENCIAS Y TECNOLOGÍA



CARTERA DE CLIENTES EN MICROFINANCIERAS

Documentación Técnica

Docente: Henry Frank Villarroel Tapia

Universitario(a): Coca Quiroz Ever

Carrera: Ing. de Sistemas

Asignatura: Simulación de Sistemas

Cochabamba – Bolivia

ÍNDICE

1. Introducción.....	3
2. Objetivo.....	4
3. Arquitectura del Proyecto.....	4
4. Librerías Necesarias.....	4
Descripción de las librerías:.....	5
5. Implementación Técnica.....	5
5.1 Generador de Datos (generadorDatos.py).....	5
5.2 Análisis y Modelo (analisisDatos.py).....	5
5.3 Evaluación y Resultados.....	5
6. Conclusiones.....	6
7. Presentación y Entrega.....	6

1. Introducción

Este documento detalla el desarrollo de un modelo de simulación para evaluar la probabilidad de deserción de clientes en microfinancieras. La herramienta permite a las entidades financieras optimizar sus estrategias de retención mediante el uso de técnicas de Machine Learning, lo que facilita la toma de decisiones basadas en datos.

El análisis de la deserción de clientes es un problema clave en las entidades microfinancieras, ya que perder clientes implica un alto costo de adquisición de nuevos usuarios y afecta la estabilidad financiera de la organización. Implementar una solución predictiva basada en datos ayuda a identificar patrones en la conducta de los clientes y a anticipar posibles abandonos.

2. Objetivo

El objetivo del proyecto es implementar un modelo predictivo que analice la deserción de clientes en función de diversas variables financieras y demográficas. A través del uso de algoritmos de Machine Learning, se busca proporcionar una herramienta que permita a las entidades microfinancieras mejorar sus estrategias de retención y optimizar sus recursos.

Este modelo debe ser capaz de:

- Identificar factores que influyen en la deserción de clientes.
- Predecir con una alta precisión qué clientes tienen mayor riesgo de abandonar la entidad.
- Permitir a la institución financiera tomar acciones preventivas para evitar la deserción.

3. Arquitectura del Proyecto

El proyecto se divide en dos módulos principales, diseñados para garantizar una separación clara entre la generación de datos y el análisis del modelo:

- **Generador de Datos (`generadorDatos.py`):** Este módulo se encarga de generar datos sintéticos que simulan el comportamiento de los clientes en una microfinanciera. Las variables consideradas incluyen información sobre créditos, pagos, saldos, demografía y más. Los datos generados se guardan en un archivo Excel para su posterior análisis.
- **Análisis y Modelado (`analisisDatos.py`):** Este módulo lee los datos previamente generados, los preprocesa (normalización y balanceo de clases) y entrena un modelo de Machine Learning. El objetivo de este modelo es clasificar a los clientes en función de su probabilidad de deserción, proporcionando métricas de evaluación y visualizaciones para su interpretación.

4. Librerías Necesarias

Para la correcta ejecución del proyecto, es necesario instalar las siguientes librerías, las cuales permiten manipular datos, entrenar modelos y visualizar resultados:

Las librerías mencionadas anteriormente están en el repositorio digital de github en el siguiente link: <https://github.com/vetdy03/carteraDeClientesEnMicrofi.git> como tambien se adjuntará en un formato compreso en extensión .zip.

como también se adjunta el archivo requirements.txt el cual contiene el nombre de las librerías necesarias y una vez descargada en la misma ruta se debe proceder a ejecutar el siguiente comando

- `pip install -r requirements.txt`

Ó también ejecutarse el siguiente comando en la terminal o su consola.

- `pip install pandas numpy matplotlib seaborn scikit-learn imbalanced-learn openpyxl`

Descripción de las librerías:

- **pandas**: Manipulación y análisis de datos en formato tabular.
- **numpy**: Soporte para operaciones matemáticas y manejo de arreglos.
- **matplotlib**: Creación de gráficos y visualizaciones.
- **seaborn**: Visualización de datos más avanzada y estilizada.
- **scikit-learn**: Implementación de modelos de Machine Learning y métricas de evaluación.
- **imbalanced-learn**: Métodos para balancear conjuntos de datos desbalanceados.
- **openpyxl**: Lectura y escritura de archivos Excel en Python.

5. Implementación Técnica

5.1 Generador de Datos (**generadorDatos.py**)

Este módulo simula datos relevantes para el análisis de deserción de clientes en microfinancieras. Se presta especial atención a las siguientes variables, que han sido identificadas como clave en el modelo de simulación:

- **Número de créditos vigentes**: Cantidad de créditos activos de un cliente.
- **Porcentaje de pago**: Porcentaje del monto total de la deuda que el cliente ha pagado.
- **Saldo en mora**: Monto pendiente de pago que está en mora.
- **Días de mora**: Cantidad de días que el cliente ha estado en mora.
- **Antigüedad del crédito vigente**: Tiempo en meses o años desde que el cliente obtuvo su crédito más reciente.
- **Plazo del crédito**: Duración total del crédito en meses.
- **Número de microseguros**: Cantidad de microseguros contratados por el cliente.
- **Tasa de interés**: Tasa de interés aplicada a los créditos del cliente.
- **Estado del ahorro**: Si el cliente tiene una cuenta de ahorro activa o inactiva.
- **Saldo del ahorro**: Monto actual en la cuenta de ahorro del cliente.

- **Edad:** Edad del cliente.
- **Estado civil:** Estado civil del cliente (soltero, casado, divorciado, etc.).
- **Estrato socioeconómico:** Nivel socioeconómico del cliente.
- **Segmento de relación:** Tipo de relación que el cliente tiene con la entidad financiera.

Los datos generados se guardan en un archivo Excel denominado `datos_credito.xlsx` para su posterior uso en el análisis.

5.2 Análisis y Modelado (`analisisDatos.py`)

Este módulo se encarga de realizar las siguientes tareas:

1. **Carga de datos:** Se leen los datos desde `datos_credito.xlsx`.
2. **Preprocesamiento:** Se convierten variables categóricas a numéricas, se normalizan los datos y se aplica balanceo de clases con `SMOTE` para evitar sesgos.
3. **Entrenamiento del modelo:** Se entrena un **Random Forest**, que es un modelo robusto para clasificación.
4. **Evaluación del modelo:** Se generan métricas como precisión, matriz de confusión y reporte de clasificación.
5. **Visualización de datos:** Se crean gráficos para entender mejor la distribución de clientes y sus características.

5.3 Evaluación y Resultados

El rendimiento del modelo se mide mediante las siguientes métricas:

- **Precisión:** Mide la proporción de predicciones correctas.
- **Matriz de confusión:** Analiza los aciertos y errores en la clasificación.
- **Reporte de clasificación:** Evalúa precisión, recall y f1-score de cada clase.

6. Conclusiones

El modelo desarrollado proporciona un enfoque predictivo para analizar la deserción de clientes en microfinancieras. Su rendimiento puede mejorarse mediante:

- **Ajuste de hiperparámetros:** Optimizar los valores de los parámetros del modelo para mejorar su precisión.
- **Prueba de otros modelos:** Implementar algoritmos como **XGBoost** o **Redes Neuronales**.
- **Ampliación de variables:** Incluir más características relevantes que puedan influir en la deserción.