

Escuela de Programación - Python B2

Proyecto Final: Clasificación de los tipos de financiamiento

Introducción

Este documento describe el proyecto final de programación para el nivel B2 de la Escuela de Programación. El proyecto consiste en desarrollar un sistema que sugiera la clasificación de los tipos de financiamiento para un cliente específico. Esto contribuirá a mejorar la experiencia del cliente mediante el diseño de modelos que optimicen la oferta de tipos de financiamiento.

Criterios de evaluación

- Demostrar que se han asimilado correctamente los contenidos teóricos y que se aplican con precisión en una actividad de carácter práctico.
- Razonar todas las respuestas de forma analítica y sintética.
- Estructurar correctamente el código.
- Entender visualmente los datos.
- Capacidad para analizar e interpretar datos.
- Capacidad para crear y evaluar modelos o algoritmos de machine learning.

Contexto

Este ejercicio se basa en el conjunto de datos GFT Open Finance, cuyo objetivo es fomentar la competencia entre proveedores financieros, impulsar la innovación digital y promover nuevos servicios basados en datos. Este conjunto de datos se centra en la banca abierta y ofrece una valiosa oportunidad de aprendizaje sobre datos y tecnologías abiertas en el sector financiero.

La actividad propuesta permitirá a los estudiantes enfrentar la nueva realidad del intercambio de información bancaria, confrontando bases de datos de diferentes instituciones financieras, incluidas dos bancos y una compañía de seguros.

El objetivo es practicar habilidades de ingeniería de datos y ciencia de datos al desarrollar un sistema que sugiera la clasificación de los tipos de financiamiento para un cliente específico. Esto contribuirá a mejorar la experiencia del cliente mediante el diseño de modelos que optimicen la oferta de tipos de financiamiento.

Open Finance representa la evolución de Open Banking y promete traer más transparencia y autonomía a los usuarios del sistema financiero. Los datos proporcionados corresponden a la implementación de Open Finance en Brasil, donde se permite el intercambio de datos sobre seguros, inversiones y fondos de pensiones. El conjunto de datos incluye información financiera de un banco minorista (RetailBankEFG) obtenida a través de financiamiento abierto, con el consentimiento de los clientes, así como datos de un banco de inversión (InvestmentBankCDE) y una compañía de seguros (InsuranceCompanyABC). Estos conjuntos de datos contienen datos de compras anteriores de clientes, así como algunos datos demográficos.



Nota: Este ejercicio tiene una serie de pistas para que puedas desarrollar y generar una solución con todo lo visto durante el desarrollo del curso.

Descripción de las actividades

Para elaborar una solución que consolide las bases de datos de estas instituciones financieras y genere modelos de recomendación de seguros mediante algoritmos de aprendizaje automático, podemos dividir el proceso en los siguientes pasos:

- Preparación de datos: Integrar y limpiar los datos de las bases de datos de las instituciones financieras, asegurando que estén en un formato adecuado para su análisis.
- Exploración de datos: Realizar un análisis exploratorio de los datos para comprender mejor su estructura, distribución y características. Esto puede incluir la visualización de datos y la identificación de posibles patrones o relaciones.
- Ingesta de datos: Integrar las bases de datos de las instituciones financieras en un único repositorio de datos, asegurando que la información se pueda acceder de manera eficiente y segura.
- Procesamiento de datos: Realizar transformaciones adicionales en los datos según sea necesario, como la codificación de variables categóricas, la normalización de características numéricas y la manipulación de valores faltantes.
- Desarrollo de modelos para recomendar financiamiento: Seleccionar y entrenar algoritmos de clasificación adecuados para el problema de recomendación de financiamiento. Esto puede incluir técnicas como algoritmos de clasificación supervisada, como Support Vector Machines, Random Forests, o redes neuronales.
- Validación del modelo: Evaluar el rendimiento de los modelos utilizando métricas adecuadas, como precisión, recall, F1-score, etc. Utilizar técnicas como la validación cruzada para garantizar la generalización del modelo.
- Optimización del modelo: Ajustar hiperparámetros y realizar selección de características para optimizar el rendimiento de los modelos de recomendación.

Preguntas:

Durante el desarrollo del ejercicio encontrarás una o más secciones de preguntas. Por favor, responde de manera justificada y, si se solicita incluir código, investiga para responder adecuadamente.

Evaluación

La puntuación de cada pregunta es la siguiente:

- Pregunta 1 - Preparación de datos
- Pregunta 2 - Limpieza y tratamiento de Datos
- Pregunta 3 - Creación de modelos de Machine Learning
- Pregunta 4 - Optimización del modelo

Se valorará la validez de la solución y la claridad de la argumentación, en todas las preguntas.

El ejercicio tiene indicado en el enunciado su peso en la valoración final. Los criterios de evaluación para evaluar este ejercicio son los siguientes:

Pregunta	No logrado (C-)	Minimamente logrado (C+)	Logrado (B)	Logrado de forma excelente (A)
Pregunta 1	La respuesta es incorrecta o no fue desarrollada	La respuesta es parcialmente correcta y está mínimamente justificada	La respuesta es correcta y describe ciertos pasos para llegar a la solución, y está justificada	La respuesta es correcta e indica todos los pasos para llegar a la solución. Se referencia correctamente el artículo y apuntes para justificar la respuesta
Pregunta 2	La respuesta es incorrecta o no fue desarrollada	La respuesta es parcialmente correcta y está mínimamente justificada	La respuesta es correcta y describe ciertos pasos para llegar a la solución, y está justificada	La respuesta es correcta e indica todos los pasos para llegar a la solución. Se referencia correctamente el artículo y apuntes para justificar la respuesta
Pregunta 3	La respuesta es incorrecta o no fue desarrollada	La respuesta es parcialmente correcta y está mínimamente justificada	La respuesta es correcta y describe ciertos pasos para llegar a la solución, y está justificada	La respuesta es correcta e indica todos los pasos para llegar a la solución. Se referencia correctamente el artículo y apuntes para justificar la respuesta
Pregunta 4	La respuesta es incorrecta o no fue desarrollada	La respuesta es parcialmente correcta y está mínimamente justificada	La respuesta es correcta y describe ciertos pasos para llegar a la solución, y está justificada	La respuesta es correcta e indica todos los pasos para llegar a la solución. Se referencia correctamente el artículo y apuntes para justificar la respuesta

Formato

Se sugiere, con el objetivo de estandarizar el formato, la actividad siga las siguientes restricciones:

- El ejercicio se enviará en la plataforma GitHub Classroom.
- Se puede desarrollar como notebook(ipynb) o como archivo Python(py).
- Se pueden realizar varias entregas en la plataforma
- El estudiante debe notificar al Docente Colaborador un mensaje una vez que considere que el ejercicio esté listo para la calificación.
- Incluir comentarios en el código

Entrega

La entrega de la Actividad deberá realizarse en la sección:

GitHub Classroom

💡 Únicamente se calificarán aquellos proyectos que hayan enviado el correo solicitando la revisión por parte del profesor colaborador (**Miguel Ponce**) `mponce1@uoc.edu`; en caso de no cumplir con este requisito y una vez finalizado el curso, **el estudiante suspenderá de manera automática el nivel.**