AyC

EMPRESA CONTRATANTE:

Banco X

OBJETIVO DEL PROYECTO:

Asesoramiento Integral, estudio de mercado y desarrollo de modelo predictivo acerca del Riesgo Crediticio

Julio 2024

Índice

1. **INTRODUCCIÓN**
2. **OBJETIVOS**
   1. Objetivo General
   2. Objetivos Específicos
3. **FUNDAMENTACIÓN**
4. **METODOLOGIA DE TRABAJO**
   1. **Sprint 0:**
   2. **Sprint 1:**
   3. **Sprint 2:**
   4. **Sprint 3:**
5. **EQUIPO DE TRABAJO**
6. **STACK TECNOLÓGICO**
7. **DESARROLLO**
   1. Extracting, Transforming, Loading
   2. Exploratory Data Analisys
   3. Key Performance Indicators
   4. Machine Learning
   5. Deploy
8. **CONCLUSIONES**
9. **DICCIONARIO DE DATOS**
10. INTRODUCCIÓN
11. OBJETIVOS
12. FUNDAMENTACIÓN

Respecto a la elección del stack tecnológico:

El entorno de Google Colaboratory nos brinda los recursos necesarios para el análisis y el acceso a el trabajo colaborativo en conjunto con la facilidad de uso, la conexión nativa con Google Drive, y su compatibilidad con el lenguaje de programación Python (Adecuado para su uso en el análisis de datos):

Asimismo, el seguimiento del desarrollo del proyecto utilizando git y el entorno de github, en conjunto con sus funcionalidades, permiten una metodologia de trabajo unificada y cohesiva.

La herramienta power bI en conjunto con sus Herramientas asociadas (por ej. Power query) brindan la capacidad de manipulacion de los datos y su visualización.

Para terminar, contamos con medios de publicacion y despliegue de nuestro dashboard y nuestro modelo en producción, utilizando Plataformas como NovyPro o Render en conjunto con Streamlit.

Respecto al proyecto realizado:

Vertical Tecnológica: Data - BI

* Descripción

La vertical de Data-BI se enfoca en la recopilación, análisis y visualización de datos para tomar decisiones informadas y estratégicas en una organización. Esta vertical abarca desde la recopilación y limpieza de datos hasta el análisis avanzado y la presentación de insights a través de herramientas de Business Intelligence (BI) y Data Visualization.

* Objetivo de Proyecto

El objetivo del proyecto de Data-BI es utilizar datos para generar insights y conocimientos que impulsen la toma de decisiones informadas en una organización. Esto puede incluir la creación de informes ejecutivos, paneles interactivos, modelos predictivos y mucho más.

1. METODOLOGÍA DE TRABAJO

Sprint

El concepto de Sprint es el corazón de las metodologías ágiles que sirven para avanzar de forma ordenada y gradual hacia la consecución de los objetivos, mostrando el progreso de su proyecto en todos sus aspectos: funcional, estético, responsive, etc. La finalidad detrás es que el “cliente” pueda ver que el proyecto avanza y que el equipo no deje todo para el último día de la entrega final.

Cada sprint tendrá una demo con sus TLs y serán 4.

● Semana 1: Jueves.

● Semana 2: Jueves.

● Semana 3: Jueves.

● Semana 4: Último previo al Demo Day

Los principios de Lean Startup promueven la experimentación rápida y el aprendizaje validado. Esto implica estar preparados para enfrentar desafíos y conflictos, resaltando la importancia de la flexibilidad y la adaptabilidad como soft skills fundamentales para el éxito en proyectos tecnológicos.

No Country ejercerá el rol de “cliente”, esto tiene como intención simular el proceso de entorno laboral real donde el equipo debe realizar una demostración al cliente. En este contexto cada Team Leader y los diferentes canales de Slack de la comunidad representan los intereses del cliente.

MVP Mindset

Es decir, Producto Mínimo Viable (MVP, por sus siglas en inglés) como enfoque fundamental del proyecto. Esto implica la

creación de soluciones que aborden las necesidades críticas de manera efectiva, sin

complicaciones innecesarias. Durante la simulación, es crucial adoptar esta

mentalidad, enfocándonos en lo esencial para lograr resultados tangibles y valiosos.

Recuerda, el MVP es el primer paso hacia la innovación continua y la iteración basada

en el feedback del usuario.

1. EQUIPO DE TRABAJO

Nuestro equipo de trabajo se compone de los siguientes integrantes, cada uno abocado a una etapa y cumpliendo un rol esencial en el desarrollo del proyecto:

1. STACK TECNOLÓGICO

El stack tecnológico empleado está, en su mayoría, dentro del universo de paquetes y librerías de Python, el cual ofrece toda la gama de herramientas para desarrollar el ciclo completo de nuestro proyecto. Facilita la ingesta y transformación de los datos, de forma que se podrá trabajar desde el repositorio o almacén de datos que aloja los datos sin procesar, pasando por los entornos para escribir y ejecutar código hasta la generación de la data ya transformada e implementación de machine learning e inteligencia de negocios.

En la imágen se verán las herramientas, programas, lenguajes de programación y librerías contemplados para llevar a cabo cada una de las etapas de la ejecución del proyecto.

Citaremos algunas de las tecnologías, para entrar en el contexto de su funcionamiento:

* **Python:** Es un lenguaje de programación de alto nivel, interpretado y multipropósito. Python será utilizado en todas las etapas del proyecto debido a su simplicidad, versatilidad y a las poderosas bibliotecas que ofrece. Dentro de ella trabajaremos con NumPy, pandas, Matplotlib, Seaborn, Scikit-Learn, Tensor Flow.
* **Numpy:** NumPy es una biblioteca para el lenguaje de programación Python que da soporte para crear vectores y matrices grandes multidimensionales, junto con una gran colección de funciones matemáticas de alto nivel para operar con ellas
* **Pandas:** Pandas es una librería de Python especializada en la manipulación y el análisis de datos. Ofrece estructuras de datos y operaciones para manipular tablas numéricas y series temporales, es como el Excel de Python
* **Matplotlib:** Matplotlib es una [biblioteca](https://es.wikipedia.org/wiki/Biblioteca_(programaci%C3%B3n)) para la generación de gráficos en dos dimensiones, a partir de datos contenidos en listas o arrays en el lenguaje de programación [Python](https://es.wikipedia.org/wiki/Python)
* **Seaborn:** Seaborn es una biblioteca para crear gráficos estadísticos en Python . Está basado en Matplotlib y se integra con las estructuras de Pandas. Es tan poderosa como Matplotlib pero aporta simplicidad y nuevas funciones. Nos permite explorar y comprender datos rápidamente. Se pueden capturar marcos de datos completos y las funciones internas para mapeo semántico y agregación estadística permiten convertir los datos en visualizaciones gráficas.
* **Google Colaboratory:** Colab es un servicio alojado de Jupyter Notebook que no requiere configuración para su uso y proporciona acceso gratuito a recursos informáticos, incluidas GPU y TPU. Colab es especialmente adecuado para el aprendizaje automático, la ciencia de datos y la educación.
* **Jupyter Notebook:** Es un entorno informático [interactivo basado en la web](https://es.wikipedia.org/wiki/Rich_Internet_application) para crear documentos de Jupyter notebook. El término "notebook" puede hacer referencia coloquialmente a muchas entidades diferentes, principalmente la aplicación web Jupyter, el servidor web Jupyter Python o el formato de documento Jupyter según el contexto. Un documento de Jupyter Notebook es un documento [JSON](https://es.wikipedia.org/wiki/JSON), que sigue un esquema versionado y que contiene una lista ordenada de celdas de entrada/salida que pueden contener código, texto (usando [Markdown](https://es.wikipedia.org/wiki/Markdown)), matemáticas, gráficos y texto enriquecidos, generalmente terminado con la extensión ".ipynb".
* **Google Drive:** Es un servicio de alojamiento y sincronización de archivos desarrollado por Google. Lanzado el 24 de abril del 2012, el servicio permite a sus usuarios almacenar archivos en la nube, sincronizar archivos entre dispositivos y compartir archivos.

* **Power BI:** Es una plataforma de análisis que permite visualizar y compartir datos de manera efectiva para tomar decisiones informadas. Ofrece una amplia gama de herramientas para la preparación de datos, visualización de datos, análisis y colaboración en un solo lugar.
* **Streamlit**: Es una biblioteca de python que permite crear aplicaciones web interactivas para el análisis de datos y la visualización de manera rápida y sencilla. Ofrece una amplia gama de widgets para la entrada de datos, gráficos interactivos y capacidades de visualización, lo que la convierte en una herramienta poderosa para la creación rápida de prototipos y la implementación de aplicaciones de análisis y ciencia de datos.
* **Github:** GitHub es una forja (o repositorio) para alojar proyectos utilizando el sistema de control de versiones Git. Se utiliza principalmente para la creación de código fuente de programas de ordenador.
* **NovyPro:** NoviPro es una plataforma donde los desarrolladores de Business Intelligence pueden mostrar sus paneles de Live BI reales y compartirlos con su red.

1. DESARROLLO
   1. Datos
   2. Extracting, Transforming, Loading (ETL)
   3. Exploratory Data Analisys (EDA)
   4. Key Performances Indicators (KPI)
   5. Machine Learning
   6. Deploy
2. CONCLUSIONES
3. DICCIONARIO DE DATOS