

Proyecto Piloto – Data Science

Fecha de inicio: 1/11/2022

Fecha de Entrega: 18/11/2022

En equipos de entre 2 integrantes gestionarán un proyecto con las características necesarias para su implementación en el nivel técnico en una organización.

Diseñarán un modelo analítico de Data Science, acorde a la temática a tratar y su respectiva documentación. Tendrán por lo menos una iteración para la optimización del modelo construido. Todas las métricas deberán estar declaradas correctamente.

De este modo, los entregables de tu proyecto final constarán de dos partes:

- Modelo por bloques de código en Visual Studio Code (archivo .py y resguardado en Github)
- Documentación Ejecutiva del Modelo con el detalle del proyecto de análisis de datos (documento en formato de Google Slides o PDF).

Objetivo 1: Entender el problema de negocio e identificar los elementos a ser considerados para el planteamiento de un Modelo de Data Science.

Objetivo 2: Describir los datos de negocio y las relaciones entre datos mediante el Análisis Exploratorio de Datos.

Objetivo 3: Elegir el algoritmo de entrenamiento y preparar los datos para el proceso de entrenamiento del modelo.

Objetivo 4: Evaluar los indicadores de desempeño predictivo del modelo y realizar optimizaciones.

Objetivo 5: Construir una presentación ejecutiva mostrando los resultados del modelo.

User story/brief:

- El tópico sobre el que se vaya a hacer el trabajo debe ser una decisión de cada equipo. Para esta decisión, el grupo debe contar, al menos, con un "experto" o "dueño" del negocio entre los miembros del grupo que conozca el tema de interés.
- El dataset utilizado puede ser propio o público y deberá también ser validado por el profesor/a.
- El dataset debe tener una variable de target categórica y para su análisis deberán utilizarse algoritmos de clasificación.

- Las conclusiones deben basarse estrictamente en los datos obtenidos.
- La documentación ejecutiva permitirá informar a cualquier persona sobre el proyecto terminado. Este deberá ser claro para un usuario final sin conocimientos técnicos y que comprenda la construcción del modelo.

Te recomendamos incluir:

Codificación.

- Paquetes Numpy, Pandas, Matplotlib, Seaborn, Letsplot y Sklearn.
- Métrica, Funciones y Gráficos
- Comentarios
- Presentación, Storytelling y estética
- Glosario (en caso de contar con términos específicos)

Documento ejecutivo.

- Portada
- Tabla de contenidos
- Descripción del caso
- Tabla de versionado
- Objetivo del modelo
- Descripción de los datos.
- Hallazgos encontrados por el EDA
- Algoritmo elegido
- Métricas de desempeño
- Iteraciones de Optimización.
- Métricas finales del Modelo Optimizado.
- Futuras líneas
- Conclusiones.

Requisitos base

Los requisitos base serán parte de los criterios de evaluación para aprobar el proyecto.

El modelo en VSC debe contener:

- Comentarios de funciones y procesos.
- Estructura eficiente.
- Aplicación de los conocimientos vistos en el curso.
- Referencias tanto técnicas como del área investigada.

El Documento Ejecutivo debe contener:

- **Portada:** se informará el título principal del proyecto, nombre de los integrantes, institución y fecha de presentación.
- **Tabla de contenidos.**
- **Descripción del caso de negocio:** dedicación y problema abordado.
- **Tabla de versionado.**

- **Objetivos del modelo.**
- **Descripción de los datos** desde la perspectiva del Negocio:
Descripción de la base de datos usada en la que se explique: la temática, las variables que se seleccionaron y la segmentación de los registros, si aplica. Incluir además todas las decisiones que se tomaron para llegar al dataset que servirá de input para el algoritmo.
- **Algoritmo Elegido.**
- **Métricas de Desempeño** del Modelo.
- **Iteraciones de Optimización.**
- **Métricas finales del Modelo Optimizado.**
- **Futuras líneas:** breve descripción de las posibles iniciativas que se pueden llevar a cabo para complementar el proyecto.
- **Conclusiones:** contribuciones del modelo a los objetivos propuestos.

Requisitos Extra

- **Apropiación profunda del área investigada.** Puede incorporar información conocimientos a partir de lo que conocen del ámbito específico del trabajo o de la investigación para proveer mayores insights. Por ejemplo, si se hace una segmentación de clientes del ámbito hotelero, explicar un poco más de esa área de negocio, contar lo que ya se sabe, y a partir de eso brindar un contexto donde se entienda de la mejor manera posible por qué el hecho de aplicar un modelo de Data Science agrega valor a ese tipo de negocio en particular.
- **Utilización de datos propios para la resolución de un problema cotidiano** de su emprendimiento, proyecto o investigación. Por ejemplo, la empresa en la que trabaja, el side-project en el que está preparando su portfolio, etc.
- **Utilización de algoritmos de clustering** en base a variables numéricas de complejidad moderada (requiere validación y seguimiento del tutor/a).

No es necesario ni recomendado.

- **Seleccionar series de tiempo** puede complicar la llegada a resultados entendibles / aprovechables.