

1. Descripción general

El objetivo de este proyecto es desarrollar de manera conjunta entre estudiantes de la Universidad de los Andes y el Tecnológico de Monterrey Campus Santa Fé, un producto de analítica de datos sobre productividad en manufactura. Se han identificado dos usuarios finales interesados en este producto:

1. El área de producción de la empresa, interesada especialmente en los tiempos y tasas de producción.
2. El área de recursos humanos, interesada especialmente en las condiciones y desempeño de los empleados de la planta.

Seleccione uno de los dos usuarios finales y diseñe su producto pensando especialmente en ese usuario. Esto quiere decir que su desarrollo debe ser completo, pero debe estar especialmente enfocado en los intereses de este usuario.

2. Roles

Para la realización de este proyecto se han contemplado los siguientes roles:

1. Ingeniería de datos.
2. Análisis de datos.
3. Ciencia de datos.
4. Análisis de negocio.

Cada miembro del grupo debe seleccionar al menos 1 rol, y realizar las tareas asociadas a los roles seleccionados. La calificación de cada miembro del equipo estará asociada no solo al resultado global del proyecto, sino a las tareas específicas de los roles tomados.

3. Preguntas de negocio y plan de acción

Tarea 1

Determine la pregunta (o preguntas) de negocio que quiere resolver para su cliente seleccionado. Identifique cómo puede resolver estas preguntas a través de visualizaciones de los datos (descriptivo) y un modelo predictivo (idealmente de regresión, pero también puede ser de clasificación).

Roles involucrados: todos, con liderazgo de Análisis de negocio.

4. Datos

Para desarrollar este producto usted debe hacer uso de los datos adjuntos al enunciado del proyecto.

Tarea 2 - Limpieza y alistamiento de datos

Cargue los datos en python, explore los datos disponibles y realice una limpieza cuidadosa. Identifique datos faltantes, inconsistencias, datos duplicados y decida una estrategia para su gestión. Asegúrese de que los datos queden en un formato que permita su posterior análisis. Documente los procedimientos realizados.

Roles involucrados: Ingeniería de datos.

Tarea 3 - Exploración de datos

Realice un análisis exploratorio que permita describir estadística y visualmente el comportamiento de las variables a considerar. Calcule estadísticas descriptivas, realice histogramas, diagramas de caja, diagramas de dispersión, diagramas de violín y otros que permitan comprender cómo se comportan las variables. Documente el análisis realizado.

Roles involucrados: Análisis de datos.

5. Modelos

Tras explorar en detenimiento los datos y tener claras las preguntas de negocio, es hora de pasar a construir los modelos. Se espera que sean modelos de regresión (Simple, Múltiple, Logística), aunque también pueden ser de clasificación. Por lo menos uno de los modelos debe ser basado en redes neuronales (regresión o clasificación). Tenga presente lo aprendido en la exploración de datos, así como el usuario final seleccionado y las preguntas a resolver.

Tarea 4 - Modelamiento

Aquí deberá explorar los diferentes modelos y diferentes configuraciones de los mismos, realizar ingeniería de características (variables), emplear diferentes métodos de estimación, comparar y seleccionar las mejores alternativas, considerando diferentes métricas. Consulte bibliografía que le permita contar con elementos para proponer los modelos. No es necesario emplear todas las variables disponibles, pero todas las variables incluidas y sus relaciones deben estar correctamente justificadas. Como hay un buen número de variables, se espera que el modelo desarrollado por cada equipo sea único. Evalúe su modelo usando métricas apropiadas. Documente el modelamiento realizado.

Roles involucrados: Ciencia de datos.

6. Entregables

Como resultado de las tareas anteriores deberá entregar los siguientes resultados y soportes:

1. Resultado: reporte de **máximo 7 páginas** con la documentación de las tareas, los resultados principales del análisis exploratorio de datos y la modelización, y conclusiones.
2. Reporte de trabajo en equipo: incluya un pequeño reporte de cómo se dividieron los roles entre los miembros del equipo.
3. Soporte 1: fuentes de limpieza (cuadernos de jupyter o archivos .py con la limpieza de datos).
4. Soporte 2: fuentes de análisis (cuadernos de jupyter o archivos .py con el análisis exploratorio).
5. Soporte 3: fuentes de modelización (cuadernos de jupyter o archivos .py con la modelización desarrollada). Aquí debe incluir los pasos de entrenamiento, prueba y evaluación del modelo.

Nota 1: los soportes son parte fundamental de la entrega. Su no entrega lleva a una alta penalización.

Nota 2: si bien el trabajo es en equipo, la nota es individual, luego es necesario que cada miembro del equipo demuestre su contribución al proyecto, por lo cual en el reporte es importante especificar la contribución de cada miembro.

Nota 3: La entrega la deben realizar de manera individual en Bloque Neón (Estudiantes Uniandes) y en Canvas (Estudiantes TEC). Adicional a los entregables, de manera individual, cada estudiante evaluará su desempeño y el de sus compañeros en el desarrollo del taller.

7. Recomendaciones

1. El objetivo del proyecto es lograr un buen producto, bien soportado y claramente desarrollado. Justifique adecuadamente sus decisiones, observaciones y conclusiones.
2. Sea conciso y eficiente con el espacio. El reporte no debe ser largo. Al contrario, en un buen reporte cada gráfica y afirmación importa.
3. Es un trabajo en equipo. Defina los ítems de trabajo, asígnelos entre los miembros del equipo, defina fechas de entrega y revisión interna. Discuta los resultados, observaciones y conclusiones. Priorice tareas y resultados a incluir.
4. Empiece a trabajar prontamente y discuta con el instructor su avance y resultados.

Fecha de entrega: jueves 16 de octubre de 2024, antes de las 2:00 pm Hora Colombia, ó 1:00 pm Hora CDMX