

1. Descripción general

El objetivo de este proyecto es desarrollar un producto de analítica de datos sobre los resultados de las pruebas Saber 11 en el país. Se han identificado dos usuarios finales interesados en este producto:

1. Ciudadanos interesados en conocer el estado de la educación, en su municipio o departamento, o en el país.
2. Secretarías de educación de alcaldías y gobernaciones interesadas en conocer cómo sus instituciones se comparan entre sí y cómo se comparan con las de otros municipios y departamentos. Unidades del Ministerio de Educación interesados en las mismas comparaciones a nivel país.

Seleccione uno de los dos usuarios finales y diseñe su producto pensando en ese usuario. El nivel esperado de desarrollo de este producto es de **prototipo funcional**.

2. Datos

La fuente de datos es el repositorio de Datos Abiertos, específicamente el conjunto de datos disponible en: <https://www.datos.gov.co/Educaci-n/Resultados-nicos-Saber-11/kgxf-xxbe>.

Tarea 1

La primera tarea a realizar consiste en explorar y extraer los datos relevantes. Como el conjunto de datos original es relativamente grande emplee una solución como AWS Glue y AWS Athena para extraer subconjuntos de datos relevantes para un análisis exploratorio. Documente su procedimiento de exploración y extracción en su reporte (queries realizados, visualizaciones, observaciones, conclusiones).

Tarea 2

Con los resultados de la exploración realizada defina claramente el producto mínimo viable (MVP) que implementará. Para esto defina los siguientes ítems:

1. Objetivo del MVP: ¿qué quiere lograr con el producto de analítica?
2. Subconjunto de datos objetivo: explique en qué datos decidió enfocarse (fecha, ubicación, valor de alguna otra variable).
3. Visualizaciones: 3 visualizaciones de los datos seleccionados que son relevantes para el objetivo planteado.
4. Modelo: 1 modelo predictivo (de redes bayesianas u otro que considere relevante) que sea relevante para el objetivo planteado.

Tarea 3

Realice las visualizaciones definidas, asegúrese de que el estilo de éstas transmita adecuadamente el mensaje que quiere llevar a su audiencia.

3. Modelos

Tarea 4

Entrene el modelo predictivo definido arriba. Defina claramente el modelo planteado, las características empleadas, la variable objetivo, el proceso de entrenamiento, los datos de entrenamiento y prueba, y reporte las métricas de desempeño adecuadas.

4. Producto

Tarea 5

El producto de analítica debe tener el formato de un tablero donde el usuario pueda acceder a la información descriptiva (visualizaciones) así como realizar las predicciones que permita el modelo. El tablero debe quedar desplegado en AWS en una máquina de EC2, y emplear una base de datos en AWS RDS como soporte para las visualizaciones. Además, los modelos que se desplieguen deben estar serializados.

Puede emplear código usado previamente en los proyectos anteriores, así como otros aprendizajes que hayan surgido durante su desarrollo.

5. Entregables

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

1. **(30 puntos)** Resultado 1: reporte de máximo 10 páginas con los resultados de la exploración de datos, así como el entrenamiento y evaluación de modelos.
2. **(25 puntos)** Resultado 2: presentación de no más de 10 minutos con los resultados principales de la modelización y el tablero desarrollado.
3. **(40 puntos)** Resultado 3: tablero desarrollado en Dash y desplegado en la nube empleando EC2 y RDS.
4. **(5 puntos)** Reporte de trabajo en equipo: incluya un pequeño reporte de cómo se dividieron las tareas entre los miembros del equipo.
5. Soporte 1: fuentes de análisis (scripts .sql con las consultas de extracción realizadas, cuadernos de jupyter o archivos .py con las visualizaciones y la modelización desarrollada, así como la evaluación de los modelos).

6. Soporte 2: fuentes del tablero (archivos .py del tablero desarrollado y .sql de la base de datos).
7. Soporte 3: repositorio Git en Github, con un historial de commits que claramente refleje el aporte de cada miembro del grupo.

Nota: si bien el trabajo es en grupo (de 2 o 3 personas), la nota es individual, luego es necesario que cada miembro del equipo demuestre su contribución al proyecto, tanto a través de los commits en el repositorio, como a través del reporte de trabajo en equipo y la sustentación.

6. 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. Ni el reporte ni la presentación deben ser largos. Al contrario, en un buen reporte cada gráfica y afirmación importa, y en una buena presentación cada diapositiva cuenta.
3. Es un trabajo en equipo. Defina los ítems de trabajo, asígneles 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 1 de junio, hasta las 23:59.