

Proyecto Data Engineering

¡Bienvenidos al proyecto! Os acaban de pedir una PoC (Proof of Concept) de un proyecto de Data Science para el departamento de Marketing. Al ser una PoC no se busca que esté perfecta, sino que mostréis el potencial del proyecto.



Objetivos

1. Dar solución a un caso de uso real.
2. Conocer las necesidades de otros equipos y hacer accesibles nuestras soluciones.
3. Afianzar el desarrollo y despliegue de nuestras aplicaciones en cloud.
4. Abordar un proyecto de principio a fin.
5. Ampliar el portfolio de proyectos.

Problema

Los tamaños de los bootcamps pueden variar mucho, dependiendo de algunos factores internos, como inversión en campañas de marketing, como externos, momentos del año, tendencias... Esto afecta a la reactividad para poder anticiparse en cuanto a la organización de las aulas dependiendo del número de alumnos o la contratación de personal.

Si podemos saber con antelación los picos altos y bajos de

alumnos que van a cursar el bootcamp, podemos anticiparnos y tomar unas mejores decisiones para dar una mejor experiencia y mejorar métricas claves de negocio.

Solución

Desarrollar y desplegar un modelo que prediga el número de alumnos que solicitan información en la [página de la escuela](#), utilizando el histórico de datos hasta la fecha que podemos obtener de GAnalytics.

Requisitos

1. Documentación del código y de la API
2. **Control de versiones (Github)**
3. Presentación: De acuerdo al problema (equipo de marketing) se expondrá la solución, donde se mostrará una demo en directo, también es importante presentar cómo lo habéis desarrollado y los detalles técnicos. (20 minutos)
4. Gestión de proyectos: Es recomendable utilizar una herramienta de gestión de proyectos como Trello/Asana/Miró.

Requisitos técnicos

Nos dividiremos en 3 equipos (2 grupos de 7 alumnos y un grupo de 8 alumnos) con 3 propuestas diferentes.

Dentro de cada uno de ellos nos dividiremos en 3 subequipos que se coordinarán en el desarrollo de la solución.

Sub Equipo Modelo:

- Un modelo que realice la predicción de los usuarios que visitan la página de forma diaria del próximo mes o semana (input del usuario).
- Crear una base de datos local con los datos descargados de

Analytics desde la cual se alimentará el modelo. Encapsular en función/script que cree la bbdd e ingeste los datos.

- El modelo se tendrá que poder volver a entrenar de forma mensual y utilizando los últimos datos recogidos. Encapsular en función/script.
- El modelo tendrá que tener un MAPE en test máximo del 20%.

Sub Equipo API:

- Tendrá que desarrollarse una API con Flask con una serie de endpoints y su documentación.
- Endpoints necesarios (puede haber más):
 - Endpoint home con un frontend sencillo para su uso.
 - Endpoint para actualizar la base de datos con los nuevos datos recogidos el último mes y que monitorice el rendimiento del modelo con los nuevos datos
 - Endpoint que reentrene el modelo con un grid search con los nuevos datos si el rendimiento ha bajado para tener un modelo nuevo con mejor rendimiento.
 - Endpoint que devuelva la predicción del modelo.

Sub Equipo Cloud:

- Desplegar la base de datos en RDS de AWS
- Desplegar la app con CI/CD en EBS + CodePipeline de AWS con Github.

Entrega y presentación

La entrega constará de un zip con la presentación y un txt con enlace de GitHub por cada equipo, que se subirá al classroom.