Big Data Course

Team Project Action Plan

For students (instructor review required)

ⓒ2021 SAMSUNG. All rights reserved.

Samsung Electronics Corporate Citizenship Office holds the copyright of this document.

This document is a literary property protected by copyright law so reprint and reproduction without permission are prohibited.

To use this document other than the curriculum of Samsung innovation Campus, you must receive written consent from copyright holder.

|  |  |
| --- | --- |
| Course | Big Data Course |
| Team Name | CDC |
| Team Leader/  Members | César Agulló /  Daniel Carmona, Christina Araujo |
| Project Title | Sistema de recomendación de Universidades España |
| Goal |  |
| El objetivo es ayudar a los estudiantes seleccionar la universidad que mejor se adapte al alumno en base al área que haya elegido, sus circunstancias y gustos. | |
| Abstract |  |
| La herramienta será usada por un usuario, con el objetivo de recibir una recomendación personalizada. Inicialmente, rellenará un cuestionario indicando sus gustos, circunstancias, y el área de interés. Con los datos obtenidos y realizando una clasificación de las universidades españolas, se quiere lograr mostar un listado de las universidades más afines al estudiante para facilitar la elección de grado. | |
| Method |  |
| Se usarán las siguientes herramientas en cada una de las fases del proyecto:   1. API Twitter (Fuente de Datos) 2. Apache Kafka (Ingesta de datos) 3. MariaDB (Almacenamiento de los datos estructurados) 4. Hive (Migración desde SGBDR a HDFS) 5. Pyspark (Entrenamiento del modelo de clasificación con Mllib) 6. PowerBI/Tableau/Pyplot (Visualización de los resultados y las variables relevantes) | |
| Data |  |
| Datos en streaming:   * Tweets (datos semiestructurados) * Publicaciones académicas   Datos estructurados:   * Número matriculados(check) * **Tasa empleabilidad(check)** * Nota media * Tasa éxito(check) * **Localización de las universidades(check)** * **Rankings (reconocimiento/prestigio)(en proceso)** * **Precio matriculación y créditos()** * Facultades(check) * Movilidad * Notas de corte(no hay) * Accesibilidad para discapacitados * Preferencias del usuario | |
| Expected  Outcome |  |
| Se pretende crear una herramienta que sirva de ayuda para el estudiante en su decisión sobre su futura carrera. Para ello, la salida de la herramientas consistirá en una lista ordenada que sirva de recomendación al usuario en base a sus preferencias (que formarán parte de los datos de entrada). Estos resultados serán trazables, de modo que se pueda justificar esta afinidad, ofreciendo así toda la información relevante para el usuario final de manera centralizada. | |
| Role by  Member |  |
| Data Engineer: Daniel Carmona  Data Analyst: César Agulló (Team Leader)  Data Scientist: Christina Araujo | |
| Schedule  Summary |  |
| <Provide a summarized schedule.>  <Use the provided WBS workbook for more detailed action plan.> | |
| Comment &  Assessment |  |
| <Comment and assessment **by the instructor.**> | |