Data Analytics- Project

- The aim of the study is to establish a small-scale data analytics system.

- You can use (https://www.kaggle.com/competitions) for data set and scenario selection. Or you can work on a data set you know.

- In this data analytics system;

Data transfer is carried out using Apache Kafka, which is a message distribution system, (**The decision is left to your discretion**.)

Distribution of data to the consumer in **streaming form**,

Sending all data as a stream to **Apache Spark**, which is the consumer of Kafka is expected.

Based on the utilized scenario and data, it is expected that the data will undergo specific **pre-processing stages**, followed by the selection of an appropriate **machine learning algorithm** (ML pipeline) to ensure proper data processing.

The project will be carried out **individually**.

- Any programming language can be used in the project.

- Project presentations will be made in class in the last two weeks

- The final version of the project must be uploaded to Sabis

- Projects uploaded after the upload date will not be evaluated.

- **Files to be uploaded;**

Codes of the project (.zip)

Documentation of the project (.pdf) (Summary of the project, technologies used, data analysis methods used, description of the dataset used, a general flow chart of the project, timing chart, findings obtained)