Regis University
MSDS696 Data Science
Practicum II

Olumide Aluko

Contents

- Problem or Situation
- Research Question
- Data
- Proposed Methodology
- Project Timeline

Problem or Situation

- Rainfall Prediction is a complex and uncertain task that significantly impacts human society.

- Well-timed and accurate forecasting can proactively help reduce human and financial loss.

- This project presents a set of experiments that use standard machine learning methods to build models that can predict whether it will rain tomorrow or not based on the weather data for that day in major cities in Australia.

Research Question

- Can we predict whether it will rain tomorrow or not using data?

- ***Solution:*** Design a predictive classification model (*Decision Tree and Logistics Regression*) using machine learning algorithms to forecast whether or not it will rain tomorrow in Australia.

Data

- Dataset Source: https://www.kaggle.com/code/ankitjoshi97/rainfall-in-australia-eda-prediction-89-acc/data

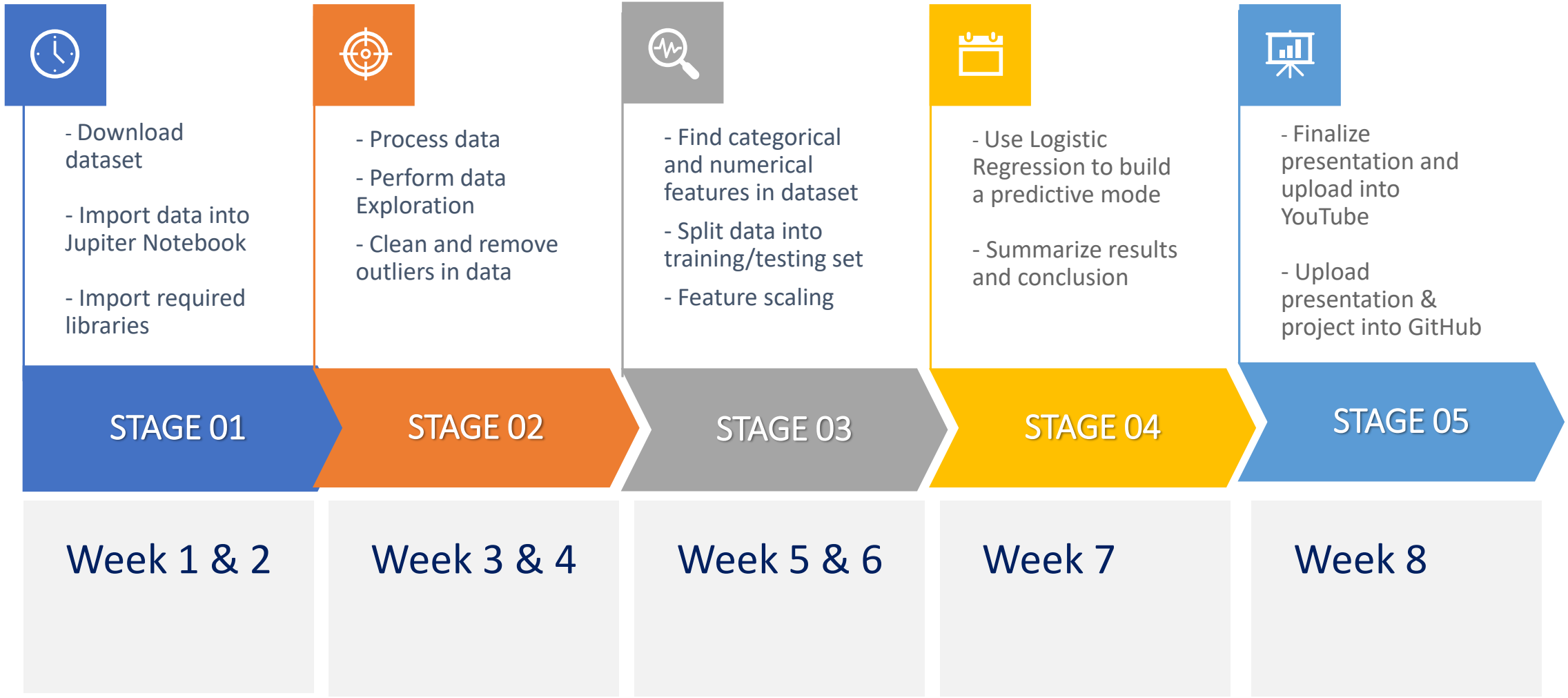
- The dataset is taken from Kaggle and contains about 10 years of daily weather observations from many locations across Australia.

- **Dataset Description:**
- Number of columns: 23
- Number of rows: 145460
- Number of Independent Columns: 22
- Number of Dependent Column: 1

Proposed Methodology

- Download dataset
- Import data into Jupiter Notebook
- Import required libraries
- Process data
- Perform data Exploration
- Clean and remove outliers in data
- Find categorical and numerical features in dataset
- Split data into training and testing set
- Perform feature scaling
- Use Decision Tree & Logistic Regression to build a predictive model whether or not it will rain tomorrow.
- Summarize results and conclusion

Project Timeline – MSDS696 Data Science Practicum II



Mid-Project Check-in & Updates

- The project is going on well as planned, see slide no.7.
- I added Decision Tree alongside Logistic Regression to build a predictive model whether or not it will rain tomorrow in Australia.
- I have completed the following:

- Download dataset

- Import data into Jupiter Notebook & Import required libraries

- Process data & perform data exploration

- Clean and remove outliers in data & find categorical/numerical features in dataset

- In the next 2 to 3 weeks:

- I plan to Split data into training and testing set.

- Use Decision Tree & Logistic Regression to build a predictive model whether or not it will rain tomorrow then finalize my result and conclusion