Visualization and Analysis of Rainfall

Team Enigma:

Utkarsh Patel (B00883322)

Anuj Upadhyay (B00862821)

Fenil Shah (B00871894)

Gaurav Jariwala (B00888116)

Overview

The sheer volume of big data has created tremendous prospects for weather forecasting and analysis. Data visualization is a common occurrence in everyday life. Using data visualization tools, several charts and graphs are utilized to explain the practical approach to rainfall classification. Because of the complexities of analyzing rainfall and the potential to learn from previous datasets, data visualization techniques have made it easy to produce graphs for better weather knowledge of a place. In this project, we have taken the weather dataset of Australian cities. The dataset includes the data patterns such as the highest, lowest temperature, and average rainfall etc. of different cities.

There would be numerous phases involved in the display of data to estimate rainfall in a certain area. We would visualize data in four major processes, from examining the dataset to showing it on a graph: analyzing the problem, cleaning the datasets, data preprocessing, and algorithm and methodologies. There are a number of factors that will be considered, as well as an unknown variable. Taking into account several factors, the goal of this project is to provide a better understanding of whether it will rain the next day.

Goals

1. Analysis and predicting rainfall for cities in Australia.
2. Perform Exploratory Data Analysis to get a better insight of features.
3. Calculating and visualizing correlation between dataset features.
4. Building interactive website to display the visualization of selected features and display the rainfall prediction.
5. Get better comparison between features like humidity and temperatures through graphs.

Dataset

Source: Kaggle [1]

Link: https://www.kaggle.com/jsphyg/weather-dataset-rattle-package

CSV Name: weatherAUS.csv

The dataset contains a csv with 23 columns showing weather of different cities of Australia during different days of the year. The data contains fields like location, temperature, wind speed, wind direction, humidity, rainfall etc. at different time of the day.

There are two binary columns named RainToday and RainTomorrow, which shows if there was actual rainfall based on the weather conditions.

Implementation Approach

1. Data cleaning and pre-processing in order to remove null values from the data, remove unwanted features (columns).
2. Data transformation in order to make it meaningful and fit for our machine learning model (Cyclic encoding of datetime).
3. Preliminary Data Analysis on processed data.
4. Finding outliers and removing unwanted rows (data).
5. Data Visualization on the transformed data to find insights of the data and patterns.
6. Apply machine learning algorithm (ANN) to predict next day rainfall.

Minimum Functionalities

1. Data cleaning and pre-processing
2. Data transformation
3. Hypothesis confirmation based on EDA
4. Building predictive model (ANN) for rainfall prediction
5. Visualization of the result using D3.JS library

Expected Functionalities

1. Building another machine learning model apart from ANN to compare and visualize the results obtained from both the models.
2. Correlation maps between feature/attribute columns such as temperature vs humidity, temperature vs pressure etc.

Additional Functionalities

1. Predicting rainfall for a specific location for the next 7 days
2. Creating heat map of Australia of a specific day showing possibilities of rainfall at each location.

Dashboard Mockup

So far, we have planned to implement a web portal with three tabs that provides different visualizations and analysis to implement the proposed functionalities.

As shown in the below figure, the first tab will provide the trend analysis feature to allow the users to select the features and visualize the trends of the feature variables over time.

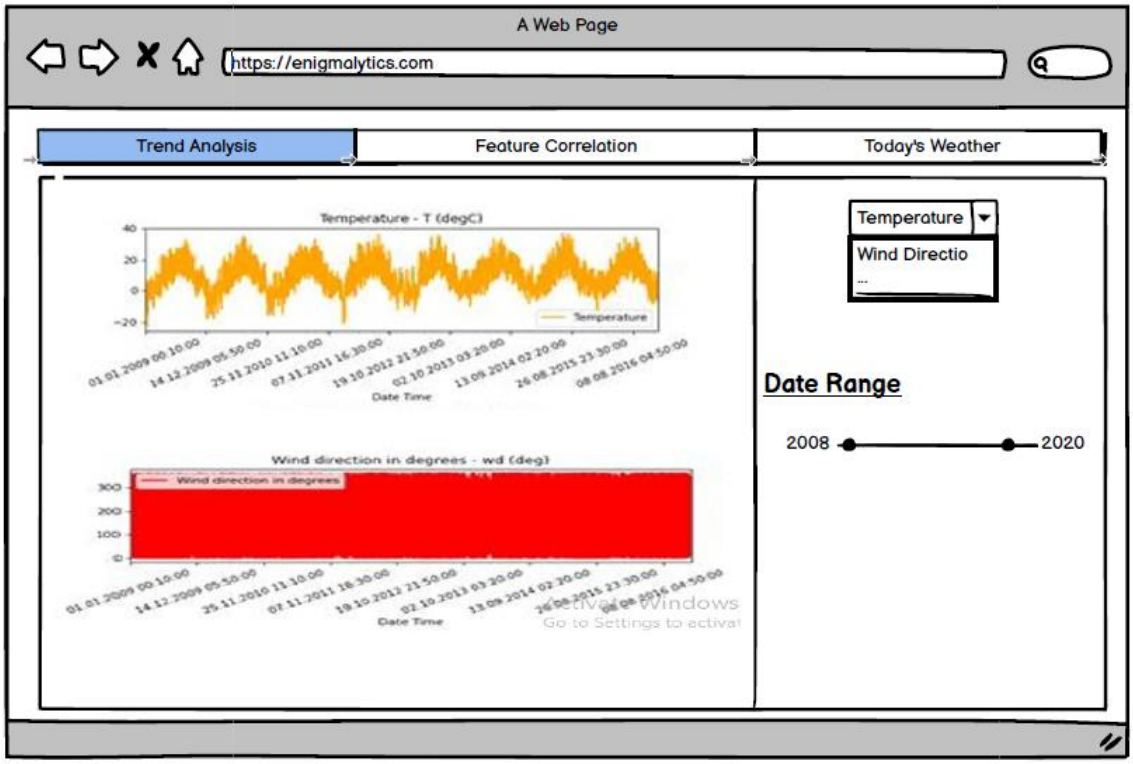

Fig. 1: Trend Analysis [2]

As presented in the figure2, the second tab will provide the correlation between all the features to allow the users to visualize how all the features are correlated with each other can choose from 2 options, one for the matrix and other for visualization in the form of cluster.

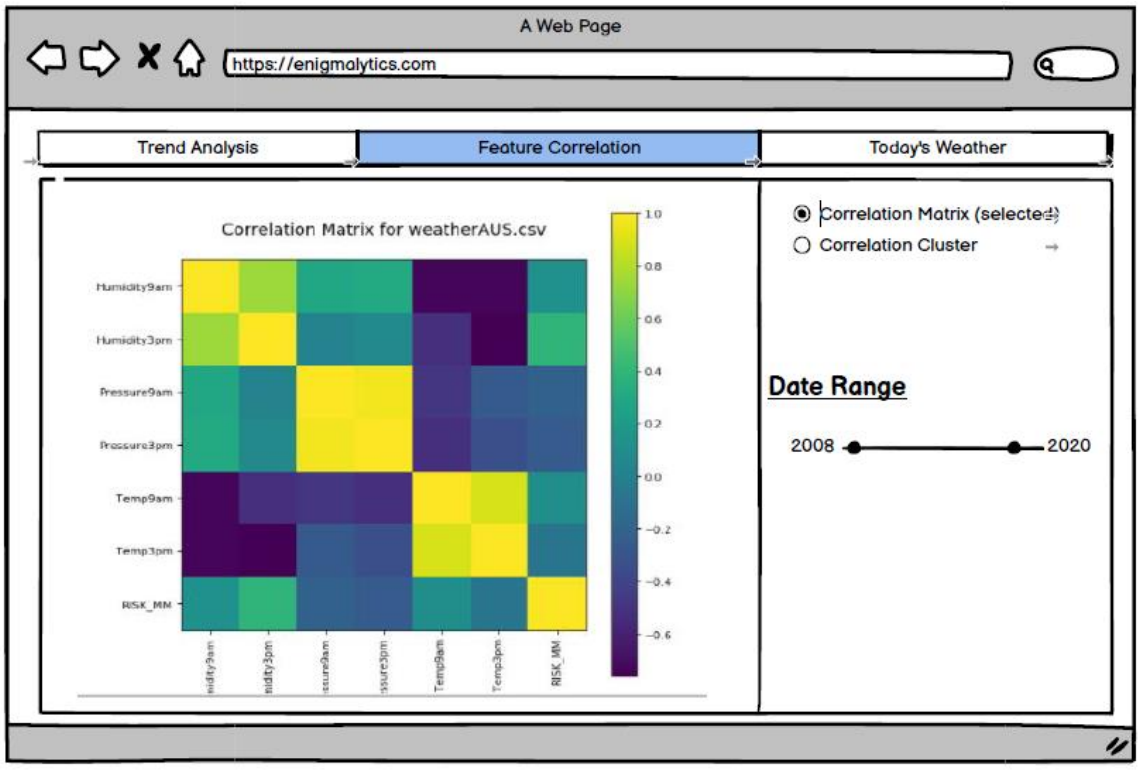

Fig. 2: Feature Correlation Matrix [2]

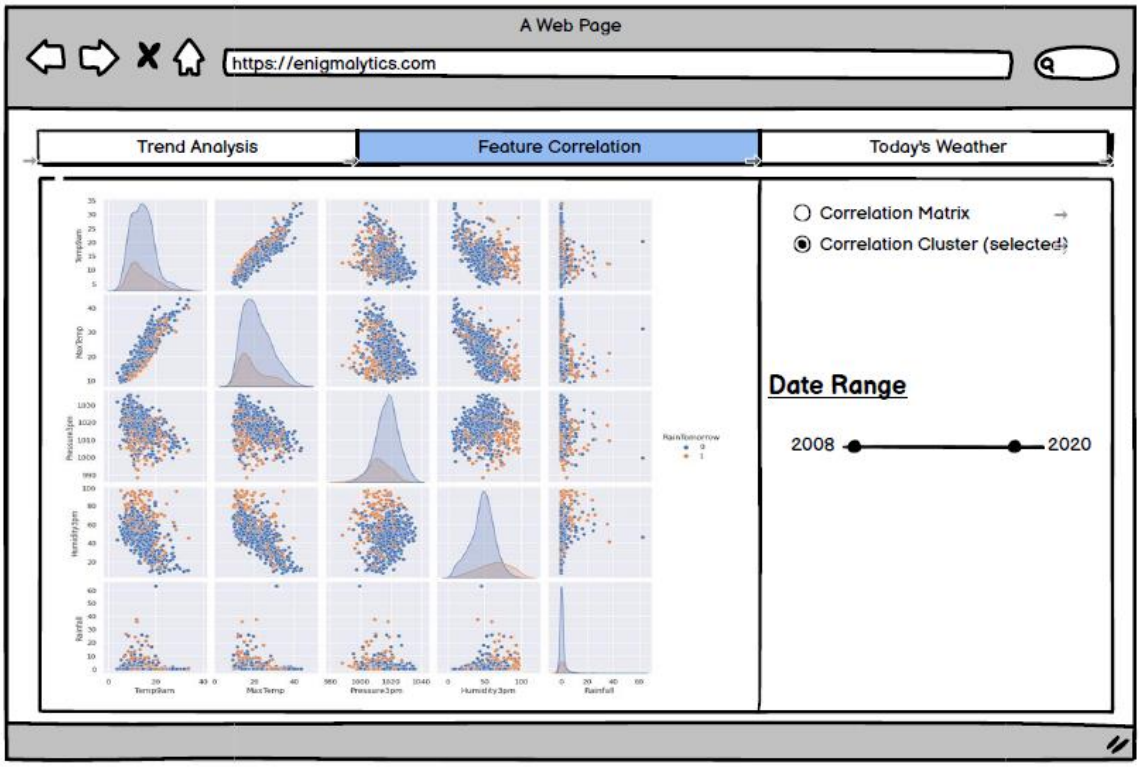

Fig. 3: Feature Correlation Cluster [2]

Figure4 displays the functionality of third tab, which provides the rainfall/ temperature prediction for the day along with the probable future predictions based on the dataset and the features.

User can also view the predictions for different cities for the country. [2]

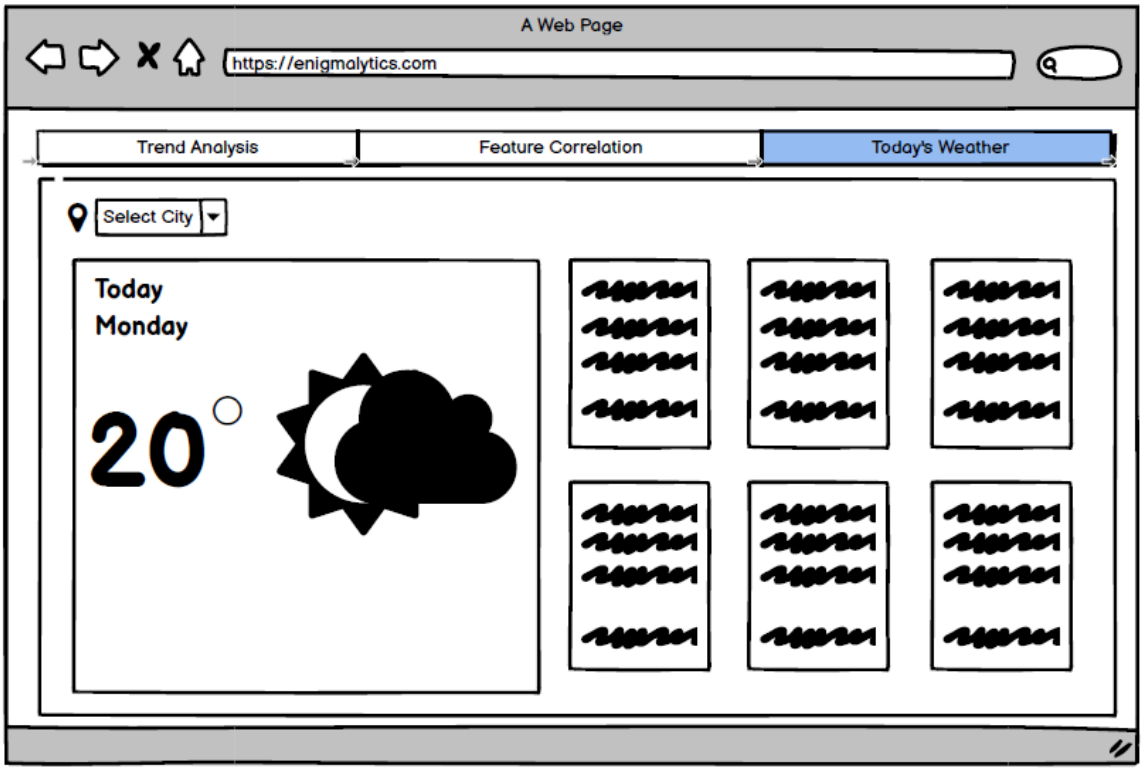

Fig. 4: Weather Prediction [2]

Timeline

	Oct 1 - Oct 7	Oct-8 - Oct 14	Oct 15 - Oct 21	Oct 22 - Oct 28	Oct 29 - Nov 4	Nov 5 - Nov 11	Nov 12 - Nov 18	Nov 19 - Nov 25	Nov 26 - Dec 2
Data Cleaning and Preprocessing									
Exploratory Data Analysis									
Feature Engineering									
Predictive Modeling									
Data Visualization									
Building Website and Integration									
Basic Functionality Implementation									
Additional Functionality Implementation									
Deploy and Project Report									

References

[1] "Rain in Australia | Kaggle," [Online]. Available: https://www.kaggle.com/jsphyg/weather-dataset-rattle-package. [Accessed 27 09 2021].

[2] "Balsamiq Wireframes - Industry Standard Low-Fidelity Wireframing Software | Balsamiq," 27 09 2021. [Online]. Available: https://balsamiq.com/wireframes/.