Project Proposal

Roushan Kumar
Department of Computer Science, Stevens Institute of Technology
CS 513-B: Knowledge Disc. & Data Mining
Prof. Khashayar Dehnad

Problem Statement:

Rain plays a vital role in our lives. Clouds are responsible for bringing rain to humans. In order to forecast when it will rain, the weather department tries to do some forecasting.

So, I am interested to predict next-day rain using different classificaton model / machine learning algorithm / data mining techniques from 10 years of daily weather observations of many locations across Australia. In addition, I am interested in investigating the possible causes of rain that falls the next day.

DataSet:

There are 23 features in the dataset that are represented in the form of columns, out of which I may choose to use the most important features during the implementation .I may use different feature reductions techniques to select the important features.

The daily observations are available from : http://www.bom.gov.au/climate/data

An example of latest weather observations in Canberra:
http://www.bom.gov.au/climate/dwo/IDCJDW2801.latest.shtml

Definitions adapted from : http://www.bom.gov.au/climate/dwo/IDCJDW0000.shtml

Data source: http://www.bom.gov.au/climate/dwo/ and http://www.bom.gov.au/climate/data

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

Implementation Strategy and algorithms Used:

I have decided to implement and compare 6 different models

1. Random Forest
2. Logistic Regression
3. K-means
4. Decision Tree
5. Naive Bayes
6. Support vector machine

Model metrics and Evaluation:

Evaluation of different models used in project uses confusion matrix for understanding the features patterns with Correlation Matrix and EDA and ROC plot for accuracy comparison and k-fold cross validation for minimizing overfitting.