

# Weather Prediction using Machine Learning Report

Prepared by: **Yusuf Okunlola**

The goal of this task is to compare the performance of five different AI models for the prediction of weather condition.

## Data Overview

Weather data which have been tracked on hourly basis since April 1st, 2006 – September 9th, 2016 was sourced from [Kaggle](#) and used.

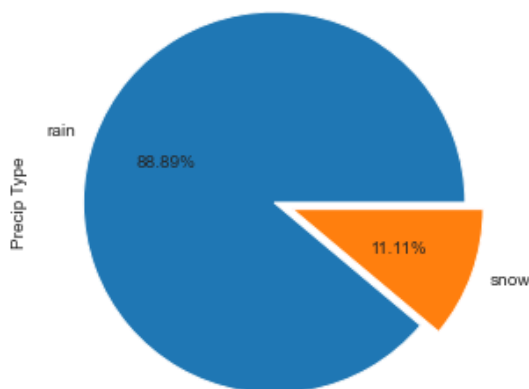
Dataset Columns:

- Formatted Date
- Summary
- Precip Type
- Temperature (C)
- Apparent Temperature (C)
- Humidity
- Wind Speed (km/h)
- Wind Bearing (degrees)
- Visibility (km)
- Loud Cover
- Pressure (millibars)
- Daily Summary

## Data Pre-processing

- Loaded dataset into pandas data frames
- The "Precip Type" was treated for missing value.
- The feature, 'Loud Cover' was dropped because it has low cardinality.
- The features 'Formatted Date' and 'Daily Summary' were dropped due to high-cardinality.
- The feature, 'Apparent Temperature (C)' was dropped due to multi-collinearity.
- Percentage of 'Precip Type' feature was visualized using Pie chart

Percentage of Precipitation Type



## Modeling

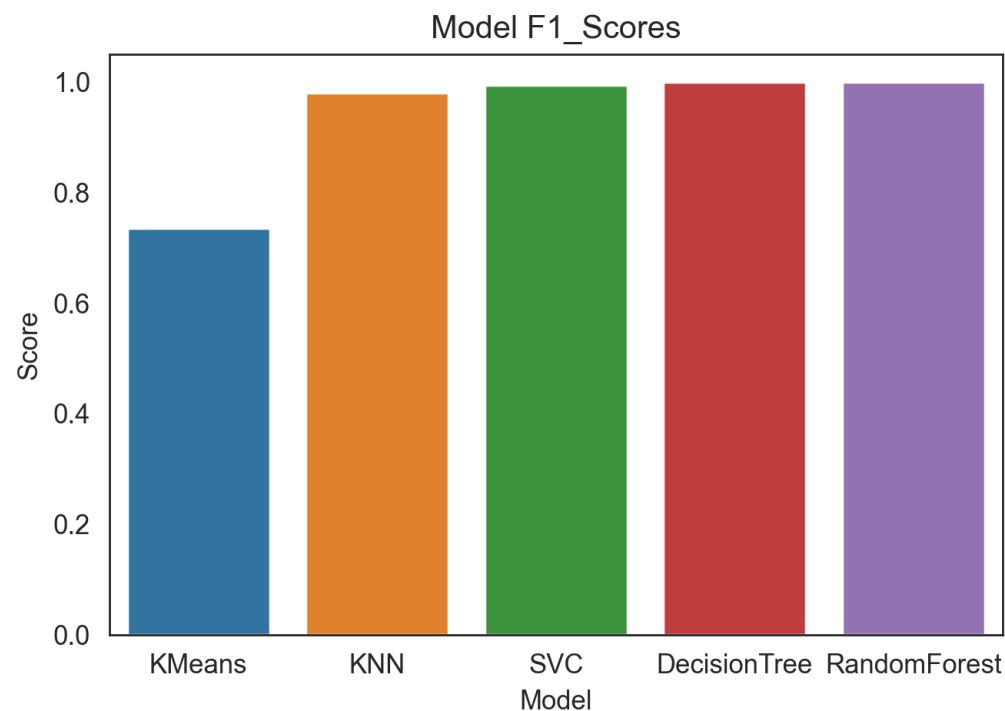
- Prepared master data set (Mapping and Ordinal encoding)
- Feature used for modelling include:  
Summary, Precip Type, Temperature (C), Humidity, Wind Speed (km/h), Wind Bearing (degrees), Visibility (km), Pressure (millibars)
- Saved prepared dataset to folder
- Scaled data (StandardScaler)
- Train-Test-Split data (80:20)
- Model Selection - Classifiers used include: K Nearest neighbor classifier, Support vector classifier, Decision Tree classifier, Random Forest classifier, K-Means Clustering

## Model Evaluation

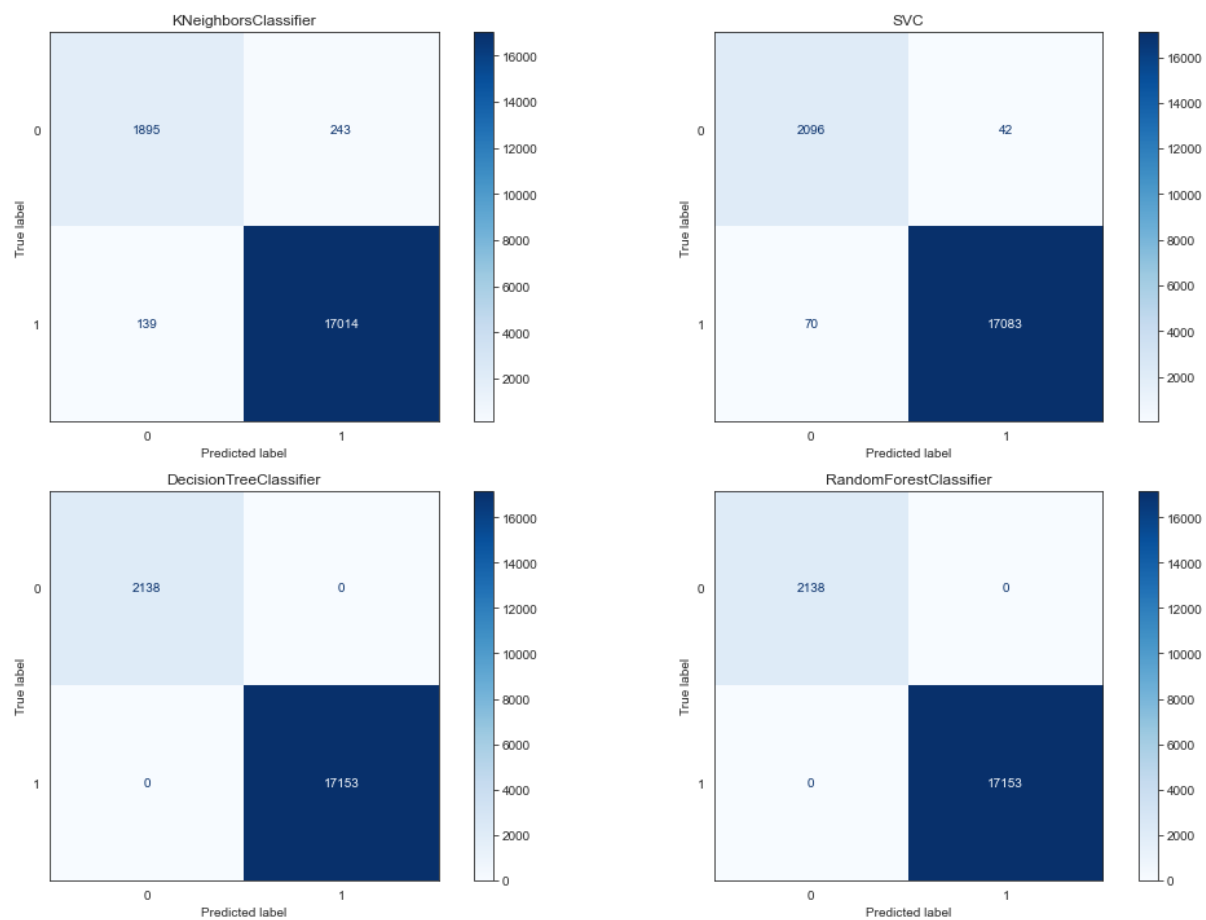
### (a) F1\_Score evaluation metric

Decision tree and Random Forest classifiers ranked 100% using F1\_Score evaluation metric.

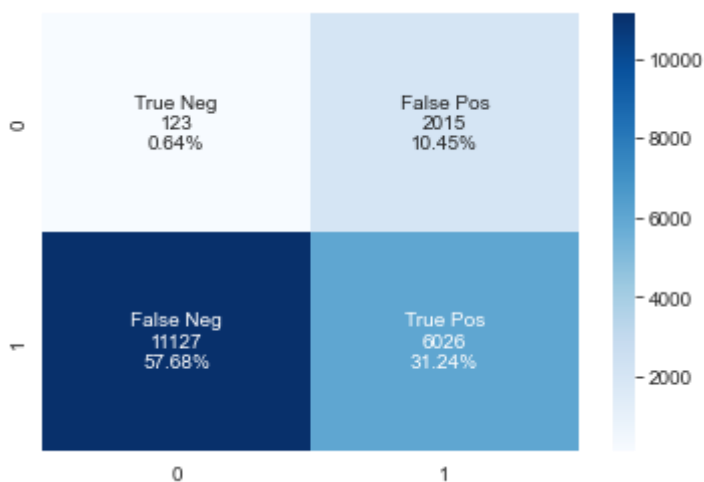
The different test accuracies achieved by each of the implemented models is illustrated in the bar chart below.



## (b) Confusion matrix



## KMeans Clustering Classification plot of Confusion matrix



## Conclusion

The results obtained by these models will influence decisions on agricultural activities such as planting and harvesting for the future as well as mitigating the dangerous effects of climate change.