RAINFALL PREDICTION

INTRODUCTION:

An web application on predicting whether it will rain tomorrow or not by using the Rainfall in Australia dataset. This project is tested over lot of ml models like catboost, xgboost, random forest, support vector classifier, knn, naive base, logistic regression. Out of these models catboost performed very well giving an AUC score around and ROC score of 89 far better than others.

TOOLS & TECHNOLOGIES :

* Front-End: HTML, CSS, Bootstrap

* Back-End: Flask

* IDE: Jupyter notebook / Pycharm,/ Visual Studio Code

HOW TO RUN THIS APP ?

* First create a virtual environment by using this command:

* conda create -n myenv python=3.6

* Activate the environment using the below command:

* conda activate myenv

* Then install all the packages by using the following command

* pip install -r requirements.txt

* Now for the final step. Run the app

* python app.py

WORK FLOW

Data Collection:

Rainfall Prediction in Australia dataset from Kaggle

Data Preprocessing:

* Missing Values Handled by Random Sample imputation to maintain the variance

* Categorical Values like location, wind direction are handled by using Target guided encoding

* Outliers are handled using IQR and boxplot

* Feature Selection and was done but didnt perform well it can be seen in testing_notebook / Prediction.ipynb

* Feature Scaling didn’t give a lot of difference it also can be seen in testing_notebook / RainPrediction1.ipynb

* Imbalanced Dataset was handled using SMOTE

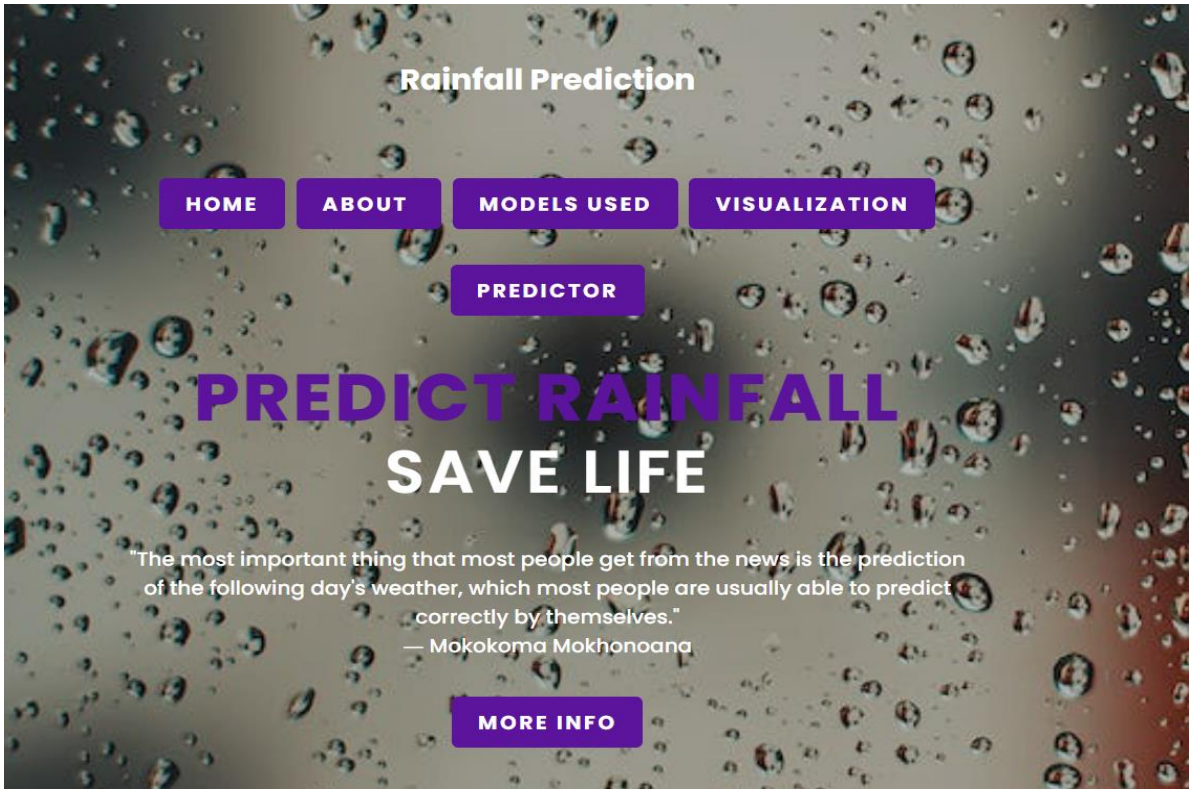
Model Creation:

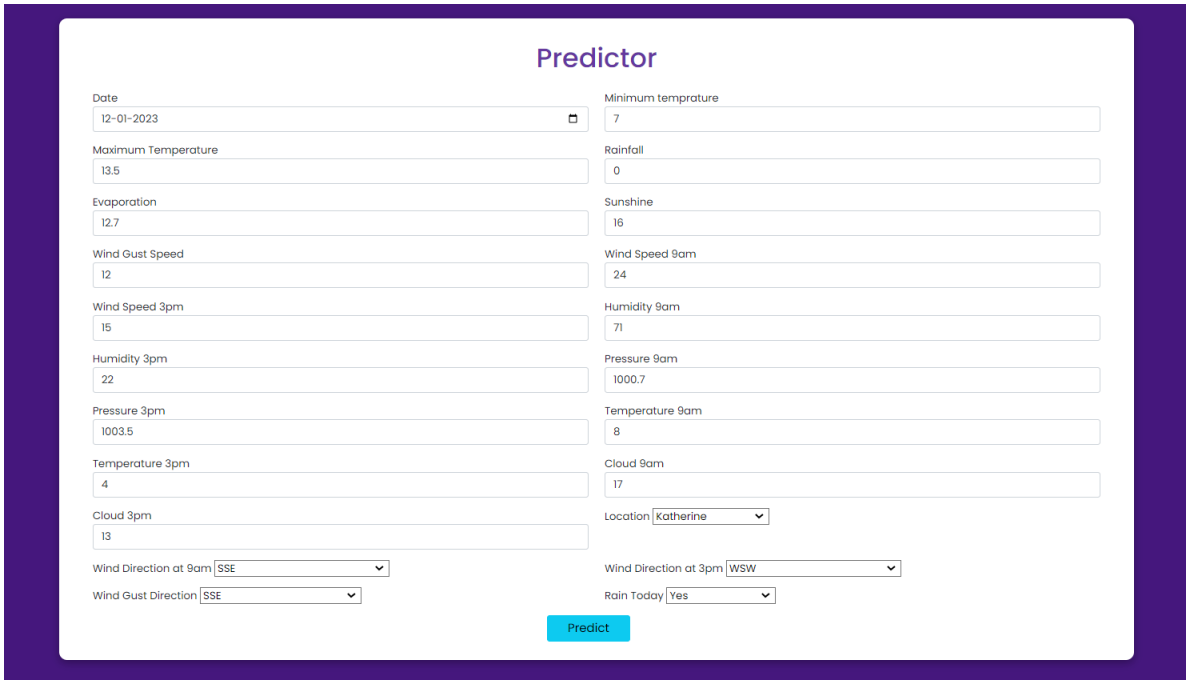
* Different types of models were tried like catboost, random forest, logistic regression, xgboost, support vector machines, knn, naive bayes.

* Out of these catboost, random forest and support vector machines were top 3

* The conclusion were made using classification metrics. ROC curve and AUC score

OUTPUT:





SUNNY DAY

Tomorrow is going to be *sunny day*. So enjoy yourselves with a cool milkshake and icecream

About Rainfall Prediction

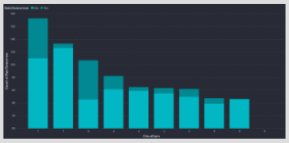
-->Rain Prediction is a web app which has a Machine Learning model running at the back.
-->The purpose of developing this app is to predict whether it will rain the next day or not.
-->This model is based on the Rain Prediction in Australia dataset.

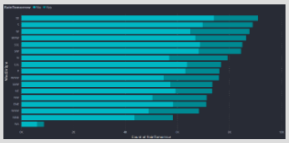
More than 80% of Australia has an annual rainfall of less than 600 mm which is less among the all continents other than Antartica which recieves less rainfall.

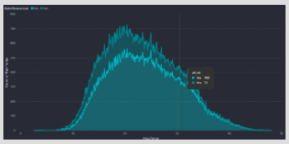
A place inland near Lake Eyre would only receive 81 mm of rain annually.The average annual rainfall in the Australian desert is low, ranging from 81 to 250 mm.Thunderstorms are relatively common in the region, with an annual average of 15 to 20 thunderstorms.The southern parts of Australia get the usual westerly winds and rain-bearing cold fronts that come when high - pressuresystems move towards northern Australia during winter. Cold snaps may bring frosts inland, though temperatures near the coast are mild or near mild all year round.Summers in southern Australia are generally dry and hot with coastal sea breezes.During a lengthy dry spell, hot and dry winds from the interior can cause bushfires in some southern and eastern states,though most commonly Victoria and New South Wales.The tropical areas of northern Australia have a wet summer because of the monsoon.During "the wet", typically October to April, humid north-westerly winds bring showers and thunderstorms.Occasionally,tropical cyclones can bring heavy rainfall to tropical coastal regions, which is also likely to reach further inland.

Dashboard

The usage of dashboards like these are to have a better understanding about the dataset and also to bring some beautiful insights. Here images of the dashboard have been attached , in order to have a good visualizations.







Models Used

Different types of models were tried like
--> catboost,
--> random forest,
--> logistic regression,
--> xgboost,
--> support vector machines,
--> knn,
--> naive bayes.
Out of these catboost, random forest and support vector machines were top 3
The conclusion were made using classification metrics. roc curve and auc score