Rainfall Prediction - Weather Forecasting

Weather forecasting is the application of science and technology to predict the conditions of the atmosphere for a given location and time. Weather forecasts are made by collecting quantitative data about the current state of the atmosphere at a given place and using meteorology to project how the atmosphere will change.

Rain Dataset is to predict whether or not it will rain tomorrow. The Dataset contains about 10 years of daily weather observations of different locations in Australia.

1. Problem Statement:

1. Design a predictive model with the use of machine learning algorithms to forecast whether or not it will rain tomorrow.
2. Design a predictive model with the use of machine learning algorithms to predict how much rainfall could be there.

2. Data Analysis:

1. I used to see the size of our data which is 8425 rows and 23 columns. After that I use these Steps:-
2. Firstly, I use describe method to know some basic details like mean and standard deviation. In that I found that there is so much error their like in rainfall the mean is 2.8 but the standard deviation is 10.4 which is more than mean which is wrong always remember that standard deviation is always minimum ands very much less than it’s mean.
3. Same type of error is found in many column and some of them are having same data like Evaporation column where mean and standard deviation are same.
4. After that we found that around 90% columns are having null values present and some of the columns have so much null values and if we use mean, mode, or median to fill those missing data then there will be a chance that our model prediction will be gone to wrong way. So, now what should we do.
5. There is only 2 option left first the columns which has not more than 100 null values we can use mean or mode to fill those columns values and second one is we can remove those which has more than 100 null values.
6. After removing all null values now it’s time to convert float and string into integers. For example:- Location, WindGustDir.
7. After that now it’s time to EDA Concluding Process.

3. EDA Concluding Remark:-

1. In EDA process we first use distplot to align the graph and see the Flow of the data in pyramid flow and we see that Min Temperature, Max temp and many more are having wrong flow.
2. In the above process we have remove all null values now it’s time to fill some data using mean and mode to fill out those wrong values. For example:- Wind speed:- It can never be zero. So, we use mean to fill out those and some places we use mode.
3. Now we use Boxplot to check the outliers of the model. After that we found so many outliers.
4. Now, What should we do, we have only 2 option First we use them or second one is we drop them.
5. These 2 option are the only option that we have in this type of error.
6. The best way is to remove those outliers so that our prediction go on the right way.
7. Otherwise, Our prediction can be wrong and it will be a huge disaster for everyone.
8. We use Quantile to remove those outliers.

4. Pre-Processing Pipeline:-

1. Now, Our model is ready to train and test the data, Our first step to separate our features and labels.
2. Remember that we have to predict 2 things:- 1. Tomorrow Rainfall, 2. Amount of rainfall.
3. So , for that we have use both regression and classification models to predict those values.
4. For regression we have Linear Regression model but for classification we have Knn model, Decision Tree, Logistic Regression and many more.
5. First we will train and test the Linear Model and for that we separate our features and label amount of rainfall.
6. The second one is classification model for that we use Logistic Regression Model to train and test our data.

5. Building Machine Learning Models:-

1. First we train and test our Classification model we use standard scaler to convert all our data into equal amount of wattage.
2. After that we train and test our data, we use fit method to train our data and then we use predict method to predict the data.
3. After that we use accuracy score to find out the accuracy of our data.
4. The final part is to check the confusion matrix to check False Positive and False Negative.
5. Second part we train and test our regression model we use standard scaler to convert all our data into equal amount of wattage.
6. After that we train and test our data, we use fit method to train our data and then we use predict method to predict the data.
7. After that we use mean absolute error and mean squared error to check the model.

6. Concluding Remarks:- In this model we predict both regression and classification model and both of them are very interesting and very fun to solve them.

At some point we can find that there so much much error in the data but that’s the interesting part to solve them.

Hence, we conclude that our Classification model predict with 88% accuracy and our regression model with 75% accuracy.