**Data Visualization Paper**

Olumide Aluko

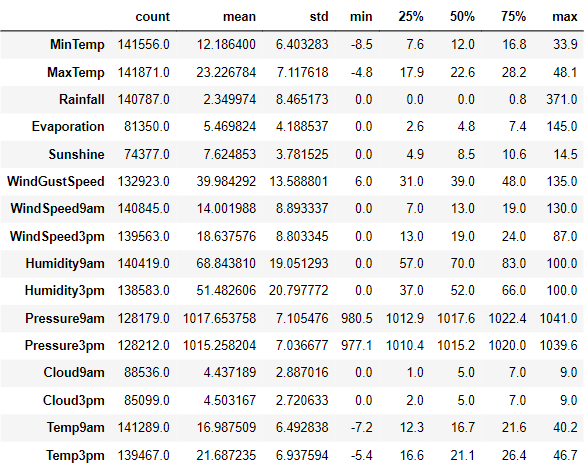
Anderson College of Business and Computing

Regis University

MSCC696: Data Science Practicum II

Prof. John Koenig

August 21, 2022

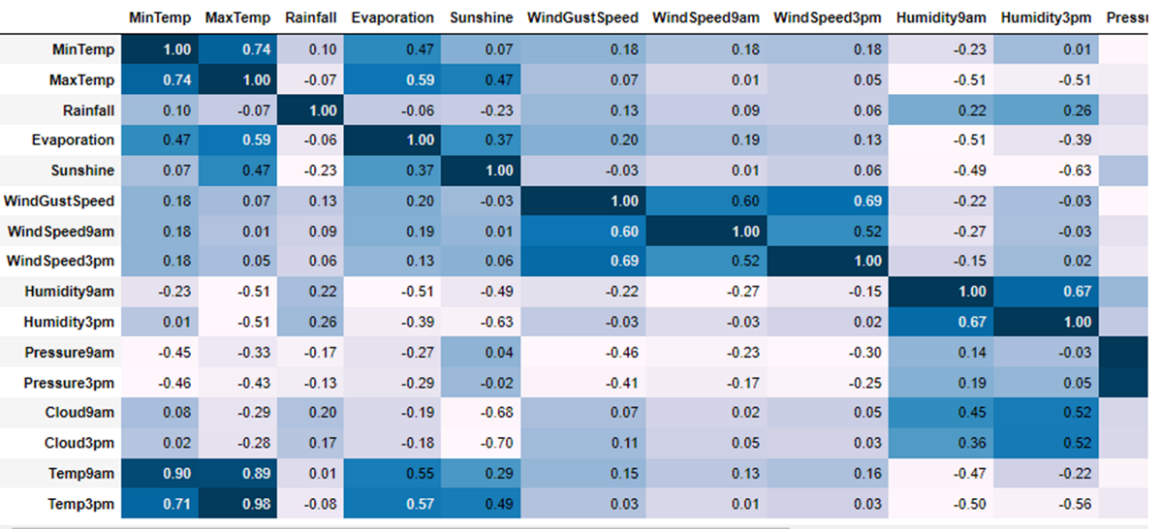


**EDA – Summary of Data**

- The average minimum temperature is 12.19 and average maximum temperature is 23.22-degree Celsius.

- The mean rainfall is 2.35 mm. -The average sunshine received is 7.62 hour.

- The average wind gust speed is 40.00 km/hr., and the median evaporation is 4.8 mm

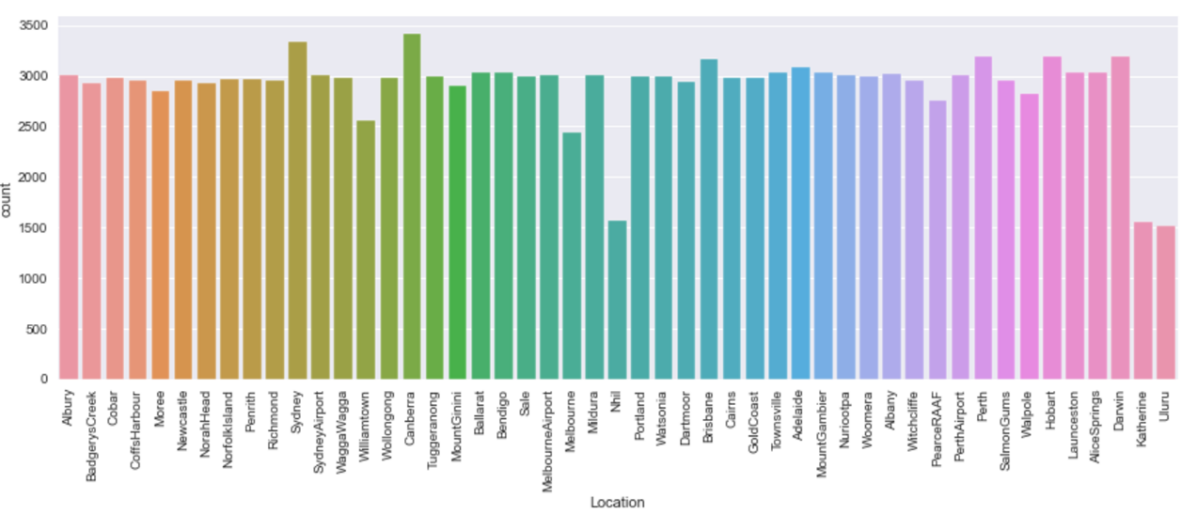


**EDA – Correlation of Data**

- Max Temp and Temp3pm have a strong positive correlation of 0.97.

- Pressure9am and Pressure3pm have a strong positive correlation of 0.96.

- Min Temp and Temp9am have a strong positive correlation of 0.90.

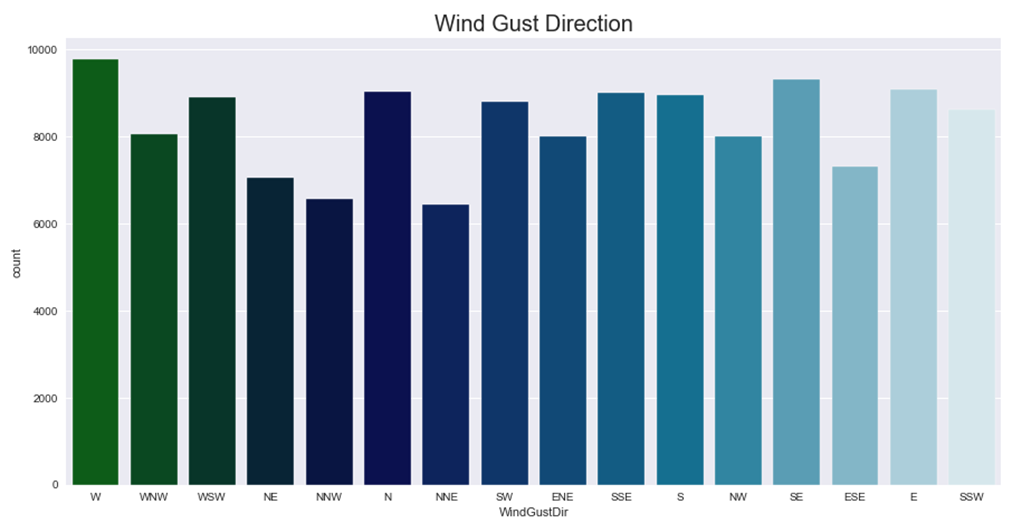


**EDA - Distribution of Location**

- Most occurred location is Canberra followed by Sydney.

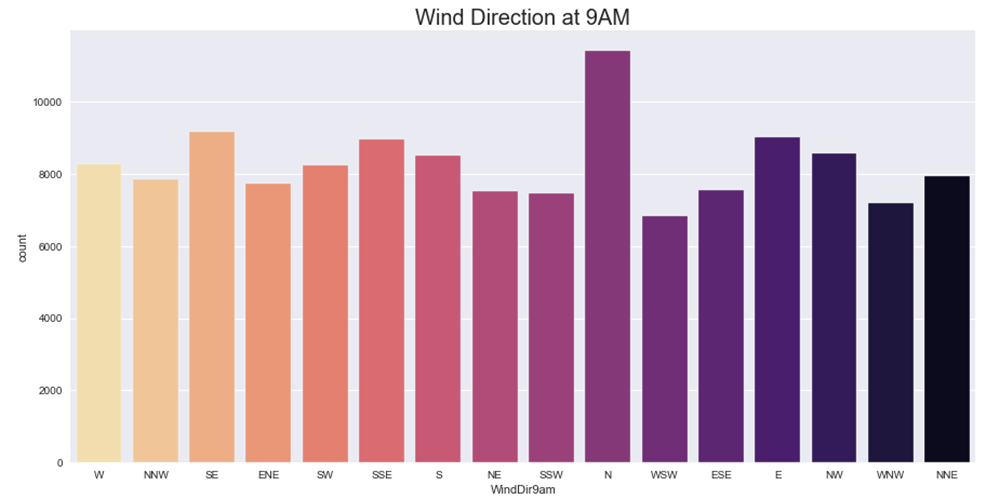
- Most of the locations have a frequency near 3000.

- Nhil, Katherine and Uluru have occurred the least.



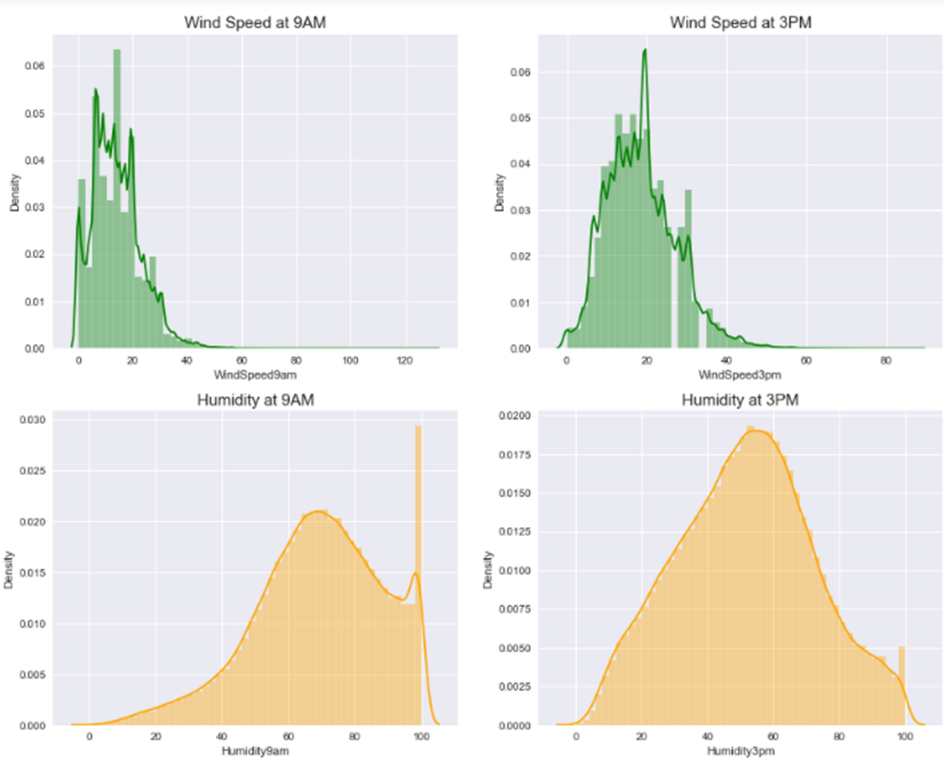
**EDA – Wind Gust Direction**

- Wind Gust Direction for maximum records(nearly 17,500) is West.



**EDA – Wind Direction at 9AM**

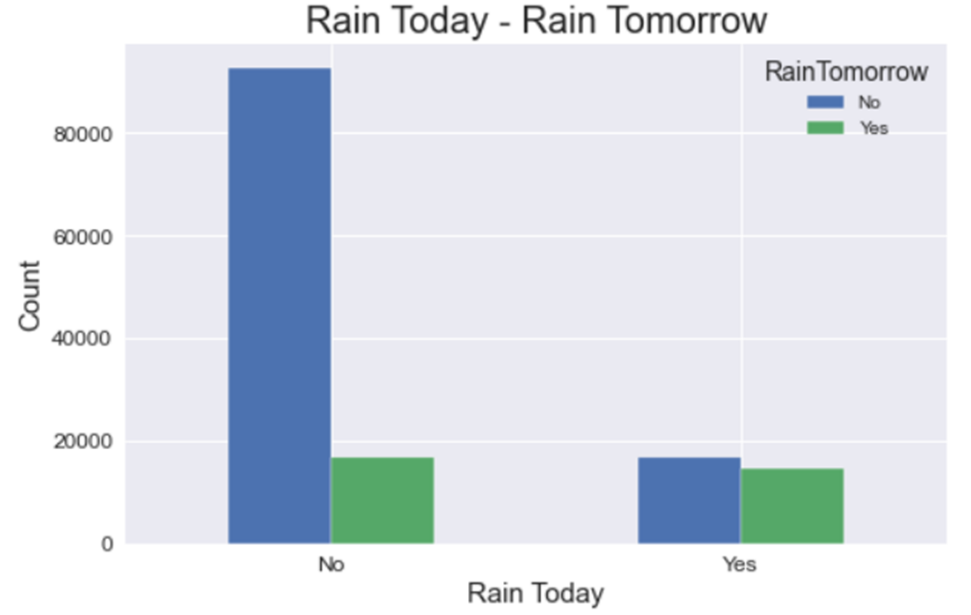
- Wind Direction at 9AM for maximum records is North followed by North-West and East.



**EDA – Wind Speed & Humidity**

- Maximum wind speed at 9AM ranges from 10 to 20 km/hr. whereas at 3PM it ranges from 15 to 22 km/hr.

- Highest concentration of points for humidity at 9AM is between 60-80% whereas at 3PM it's 40-70%.

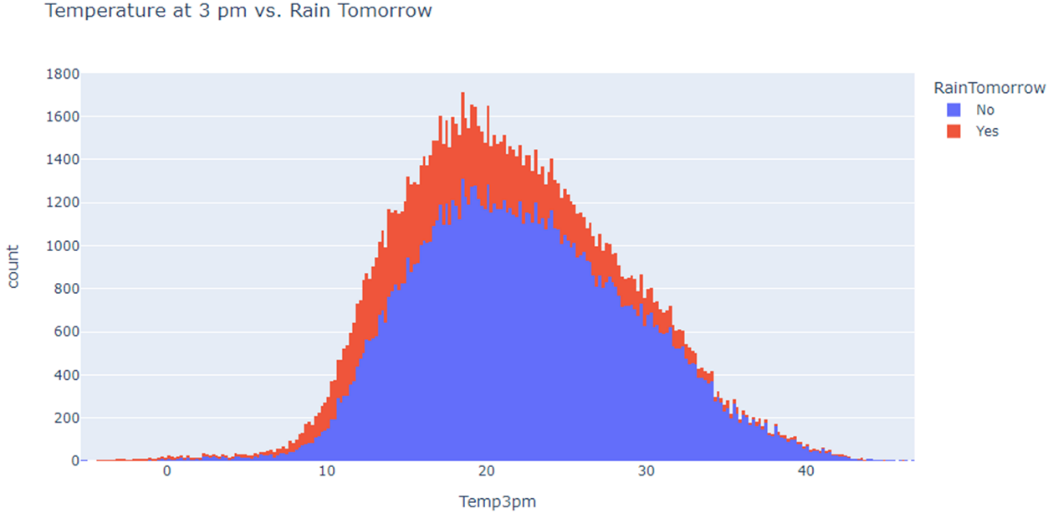


**EDA – Rain Today & Rain Tomorrow**

- For maximum records it didn't rain for both days.

- For nearly 20,000 records it didn't rain today but rained tomorrow and rained for both days.

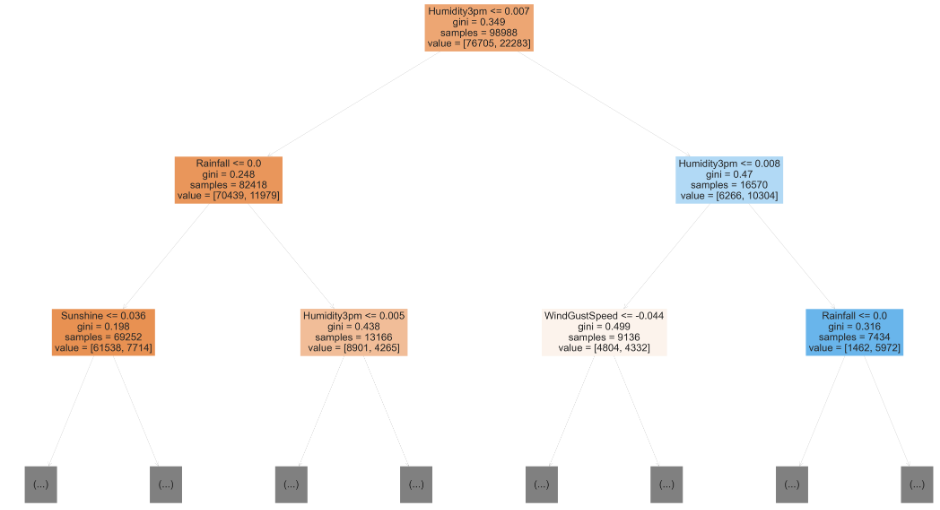
- For nearly 20,000 records it rained today but didn't rain tomorrow.



**EDA – Temperature at 3 PM VS. Rain**

- Raintomorrow with "No" has the highest count of 1311 when Temp3pm is between (18.4 – 18.5) Celsius.

- Raintomorrow with “YES" has the highest count of 401 when Temp3pm is between (18.4 – 18.5) Celsius.



**Performance of DecisionTreeClassifier**

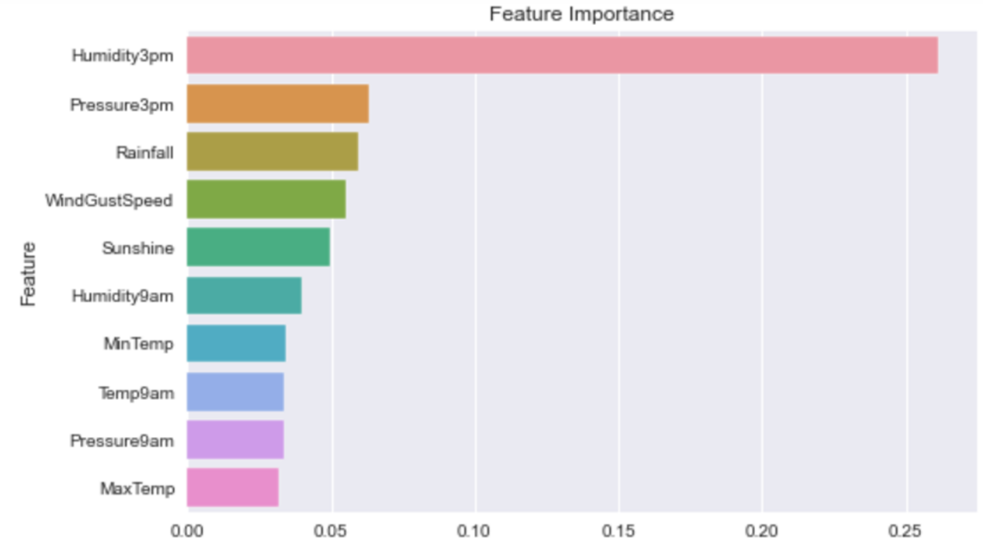
- RainTomorrow shows 78.8% 'No' and 21% 'Yes' in validation data.

- The validation accuracy is 79.28

- The training set accuracy is 99.99

- The above case was an overfitting case as tree used the (max\_depth = 48) and memorized the values

- And it failed to predict with low accuracy of 79.28% for test and validation dataset

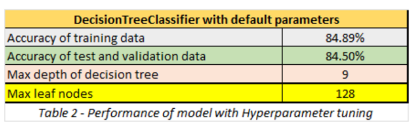
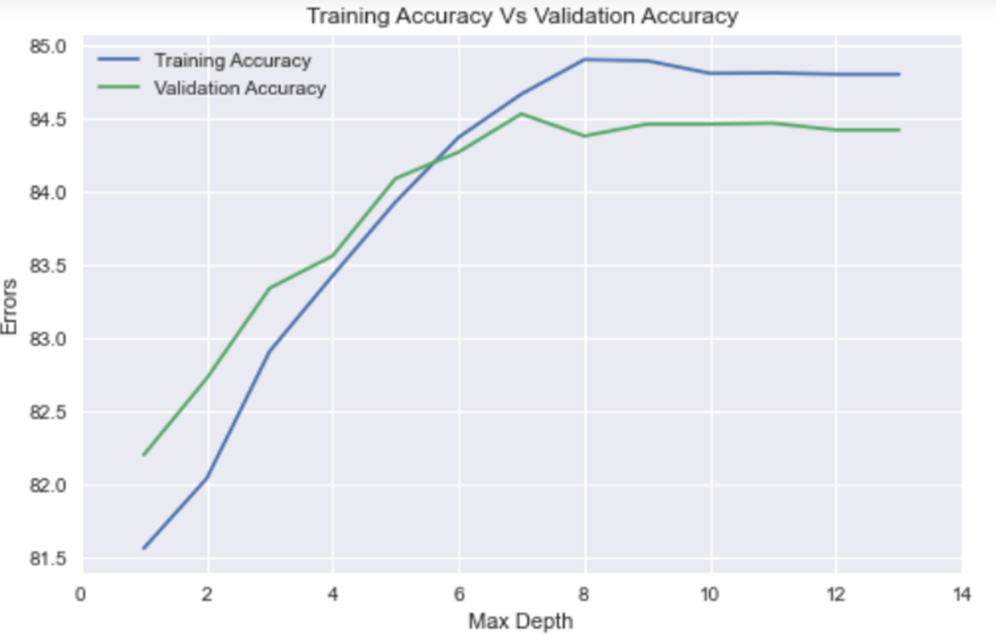
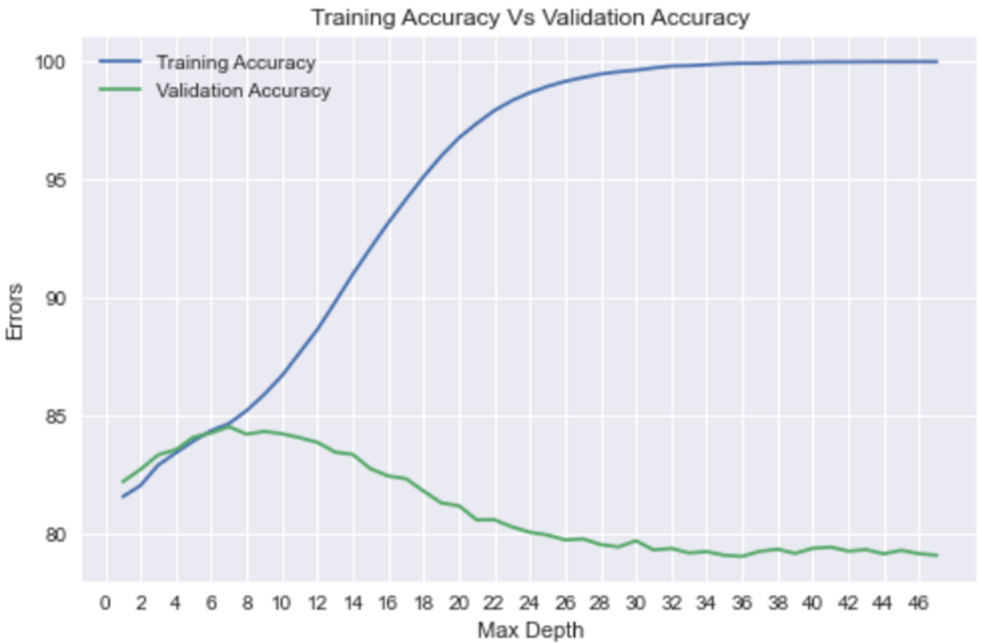


**Evaluation – Feauture Importance**

- Humidity3pm has the highest feature importance of 0.27

- Pressure3pm and Rainfall has the second highest feature importance of less than 0.10.

- MaxTemp has the lowest feature importance of less than 0.05.



**Hyperparameter Tuning – Tuning Graph 2**

- From the graph, it seems (max\_depth = 9) and (max\_leaf\_nodes = 128) is the optimal hyperparameters.

- Build Decision Tree with (max\_depth = 9), the training accuracy was 84.89

- Build Decision Tree with (max\_depth = 9), the validation accuracy was 84.46

- From the graph, performance is improved for new predictions as accuracy of training data, test data and validation data is almost the same.

**Hyperparameter Tuning – Tuning Graph 1**

- The graph shows that training accuracy increases with increase in max\_depth while validation accuracy first increases (till max\_depth = 7).

- And then decreases. Hence, optimal max\_depth is 7.

- Build Decision Tree with (max\_depth = 7), the training accuracy was 84.66

- Build Decision Tree with (max\_depth = 7), the validation accuracy was 84.53