Predicting Rain in Australia

By:
Prasuna Mannava

Introduction

Problem:

- Australia’s contrasting weather conditions from 2019 until today
- Loss of life and property due to fires and floods

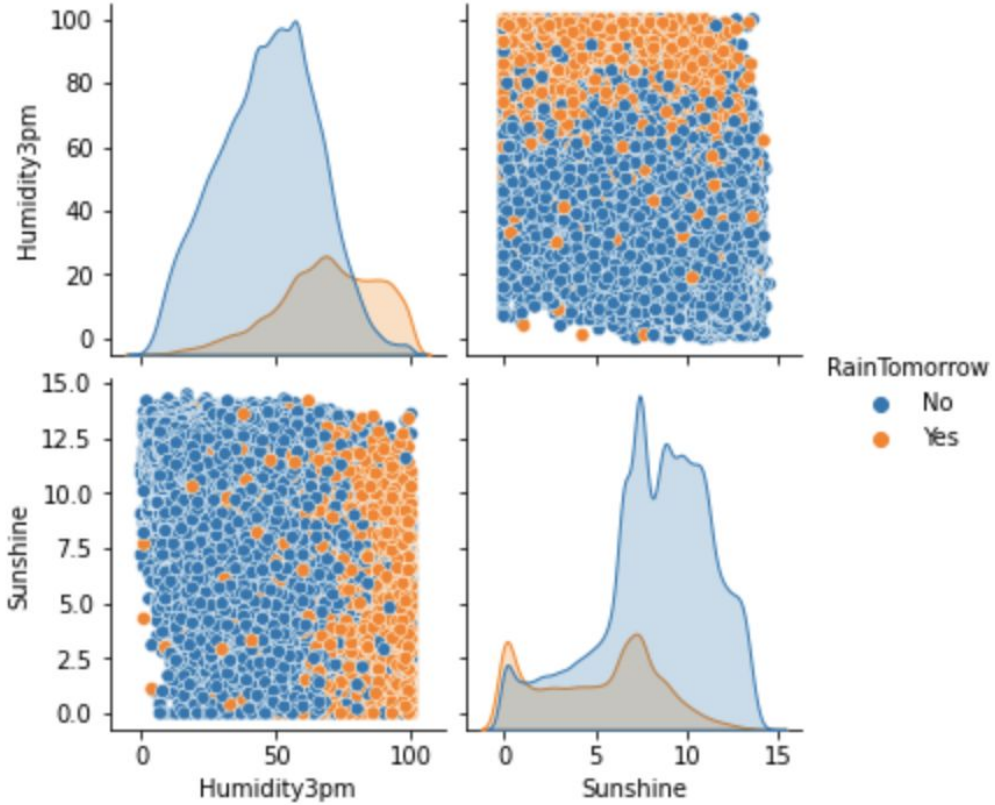
Solution:

- Classification model which can predict if it rains tomorrow or not.
- Understand features contributing towards the prediction

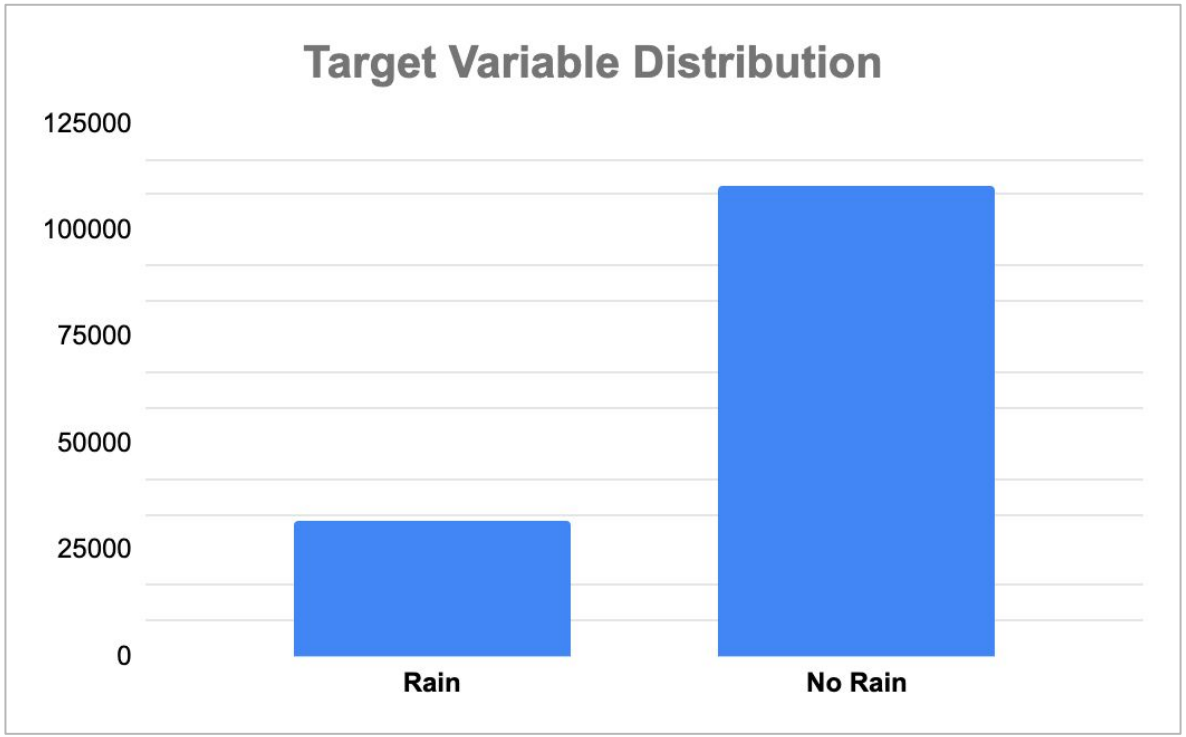
Data

- Downloaded from Kaggle- https://www.kaggle.com/jsphyg/weather-dataset-rattle-package
- 10 year rain prediction data
- 145,460 rows and 23 columns
- Features include location, min/max temperatures, pressure, humidity, windspeed, wind direction and **rain tomorrow(Y/N)**

EDA


Humidity3pm
Sunshine
Humidity3pm
Sunshine
RainTomorrow
No
Yes

EDA



Imbalanced

Metric Selection

F1 Score -

- **Precision -** Minimizes False positives
  - Predicting Rain when it won’t
- **Recall -** Minimizes False Negatives
  - Predicting No Rain when it will

Model Selection (Explainability)

- Builds trust among end users

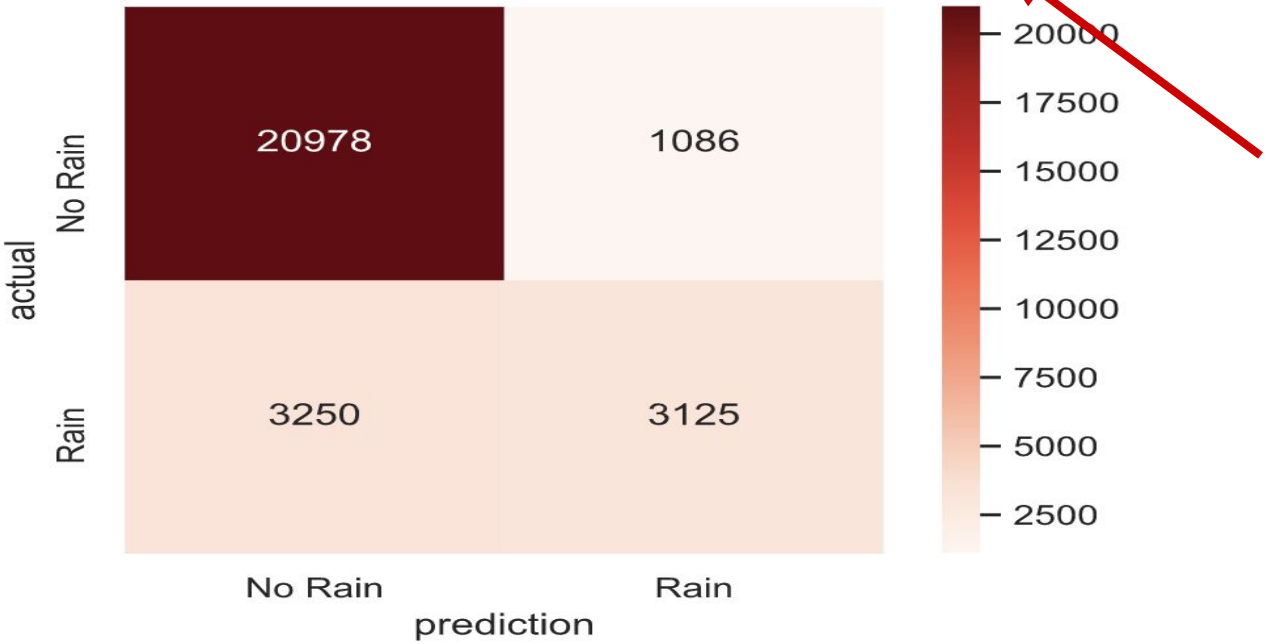
- Direction for future improvements
  - Understanding key factors to reduce customer churn

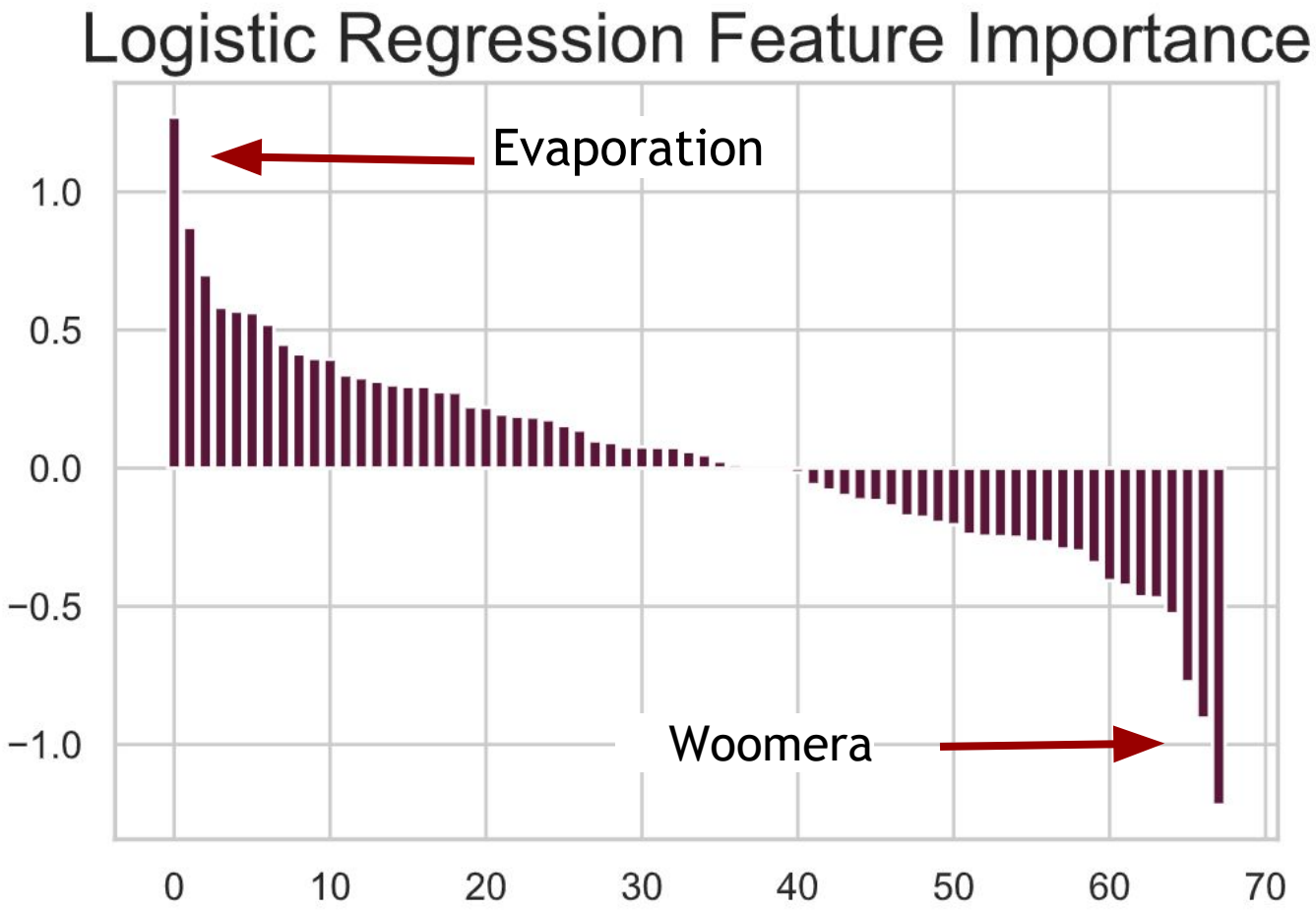
- Further data collection
  - Collecting new info from applicants may reduce Loan defaults

Logistic Regression

- C: 1
- penalty: l2

	precision	recall	f1-score	support
0.0	0.87	0.95	0.91	22064
1.0	0.74	0.49	0.59	6375
accuracy			0.85	28439
macro avg	0.80	0.72	0.75	28439
weighted avg	0.84	0.85	0.84	28439

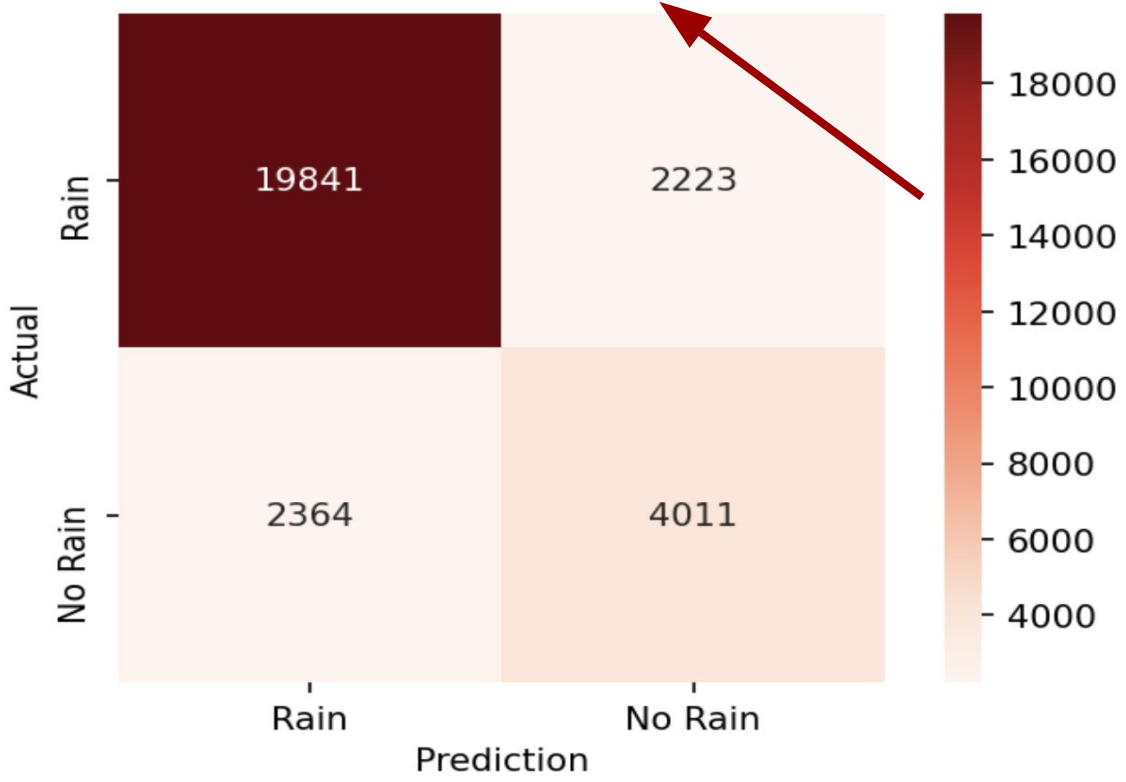



Logistic Regression Feature Importance
Evaporation
Woomera
1.0
0.5
0.0
−0.5
−1.0
0
10
20
30
40
50
60
70

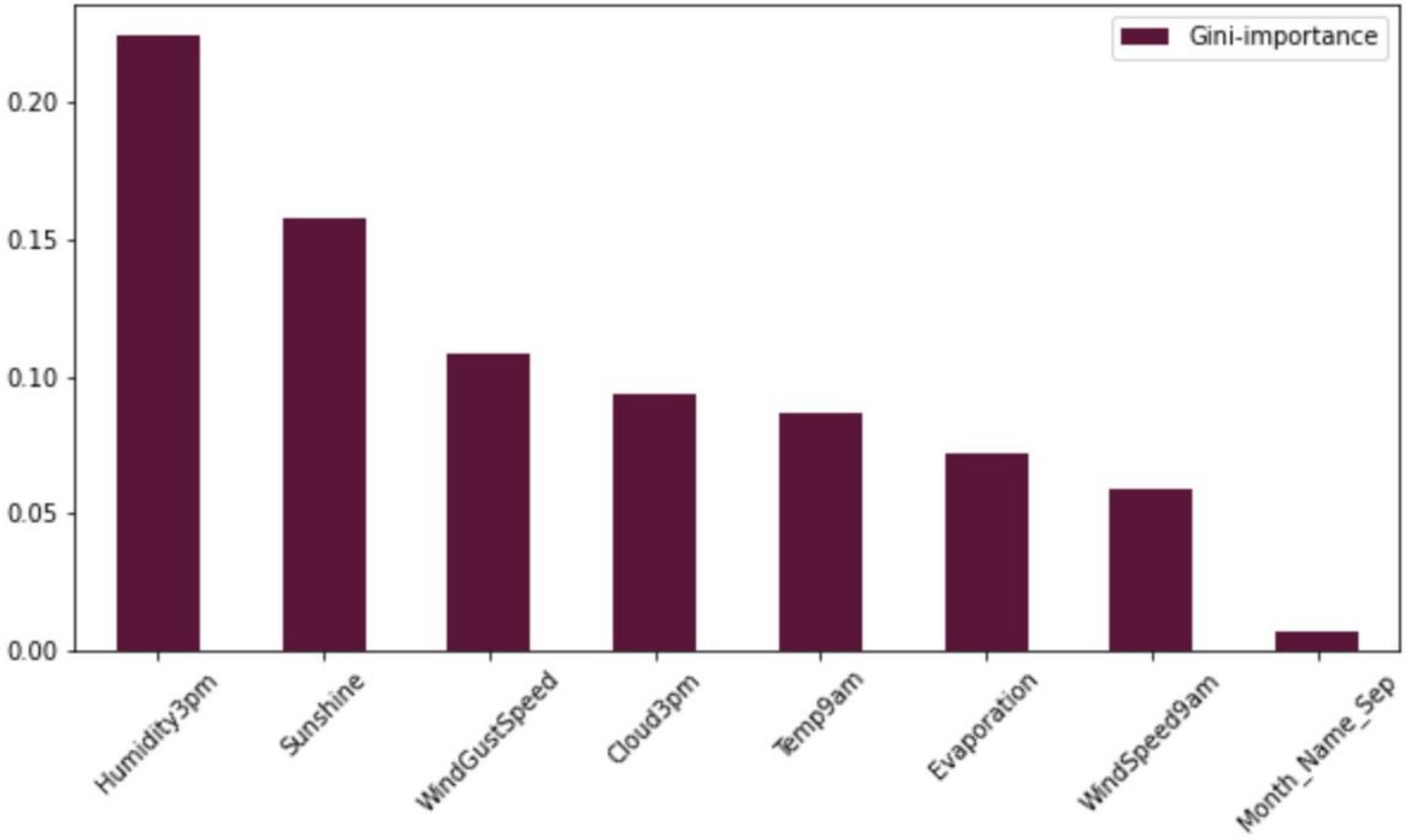
Random Forest

- max_depth: 13
- min_samples_leaf: 1
- min_samples_split: 2
- n_estimators: 100

	precision	recall	f1-score	support
0.0	0.89	0.90	0.90	22064
1.0	0.64	0.63	0.64	6375
accuracy			0.84	28439
macro avg	0.77	0.76	0.77	28439
weighted avg	0.84	0.84	0.84	28439



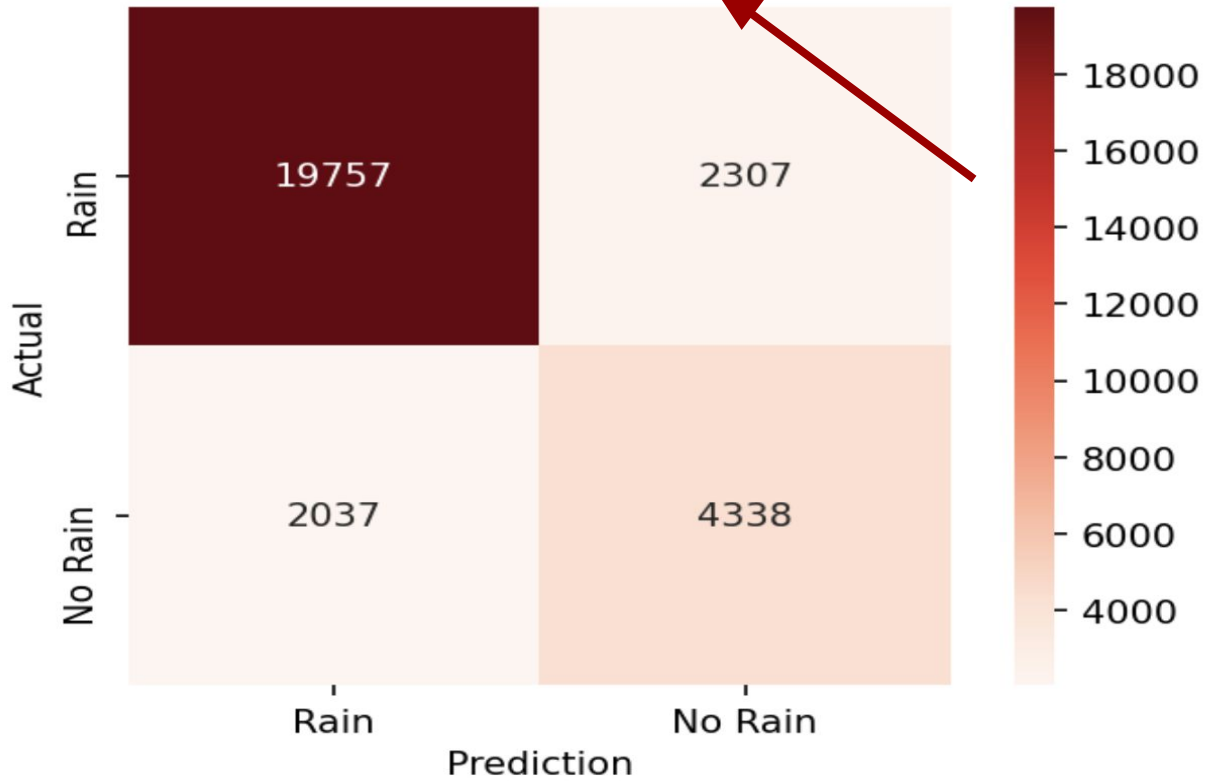
Random Forest Feature Importance



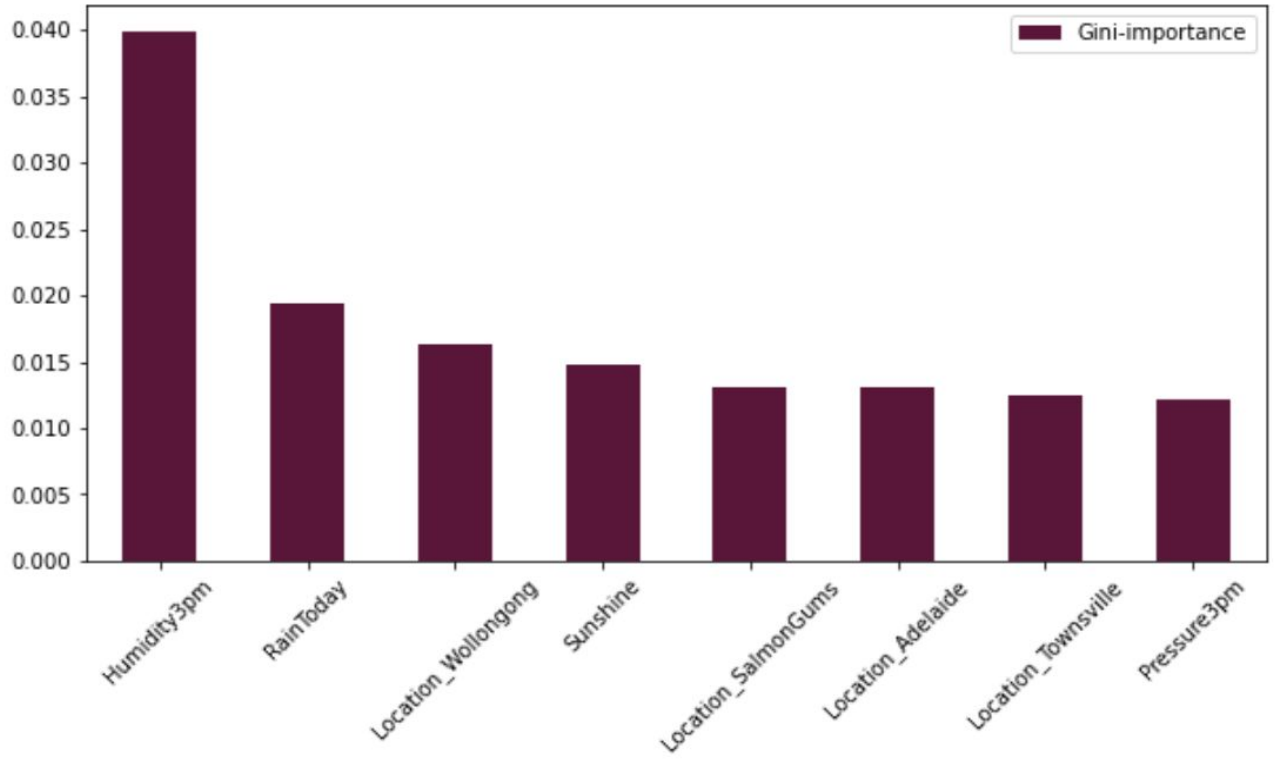
XG Boost

- max_depth: 15
- learning_rate : 0.25
- n_estimators: 100

	precision	recall	f1-score	support
0.0	0.91	0.90	0.90	22064
1.0	0.65	0.68	0.67	6375
accuracy			0.85	28439
macro avg	0.78	0.79	0.78	28439
weighted avg	0.85	0.85	0.85	28439



XGBoost Feature Importance



Future Work

- Further hyperparameter tuning
- More research and data related to population, pollution and landscape of the country
- Data for Regression models to predict the amount of rain

Tableau Dashboard

[Click here]