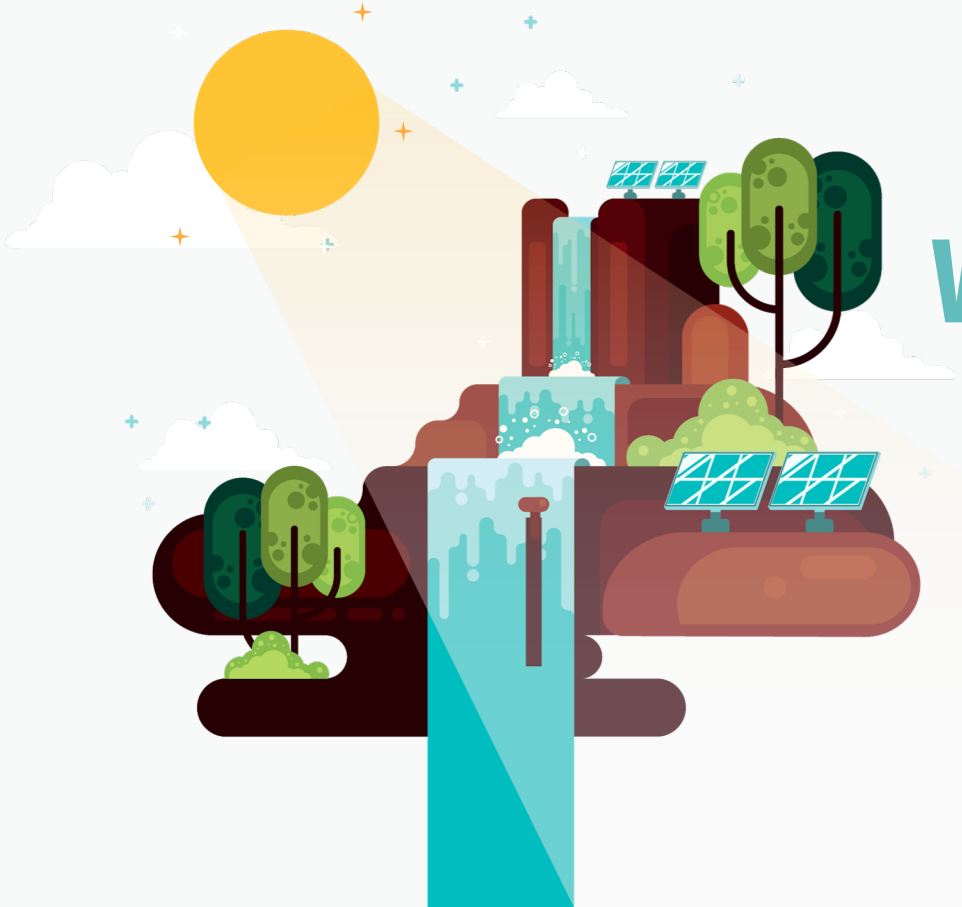


WEATHER FORECAST MACHINE LEARNING

Dicky Gemma Jacky Kary

01 BUSINESS QUESTION

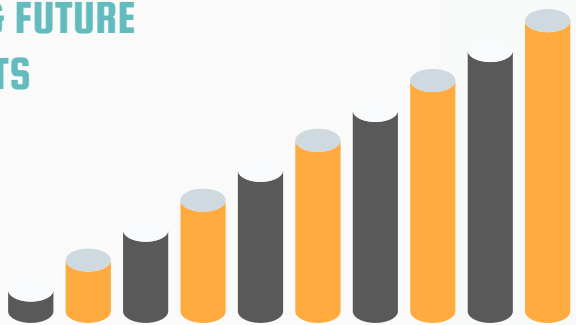
02 DATA COLLECTION

03 PREPROCESSING

04 MODELS CREATION & COMPARISON

05 MODELS EVALUATION

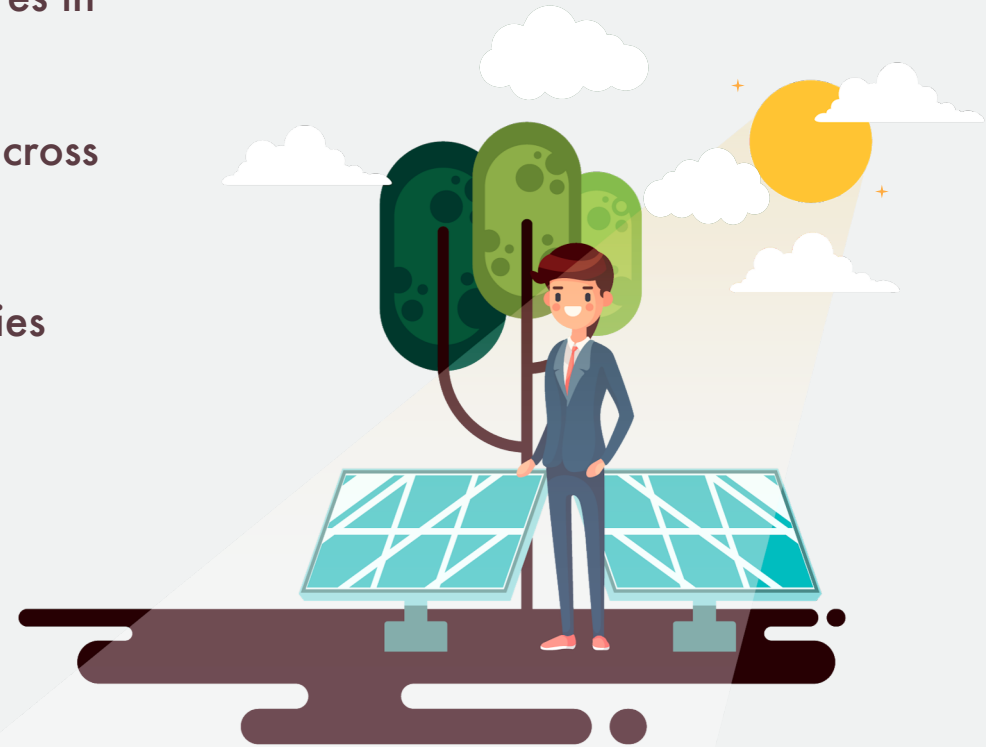
06 CONCLUSION & FUTURE IMPROVEMENTS



01

BUSINESS VALUE

- Find the most important features in predicting the chance of rain
- Compare rainfall predictors across Northern (HK) and Southern Hemisphere (Australia)
- Prevent economic loss & injuries
- Raise the public awareness



02

DATA COLLECTION

Australia Weather

2008-2017

Australian Government

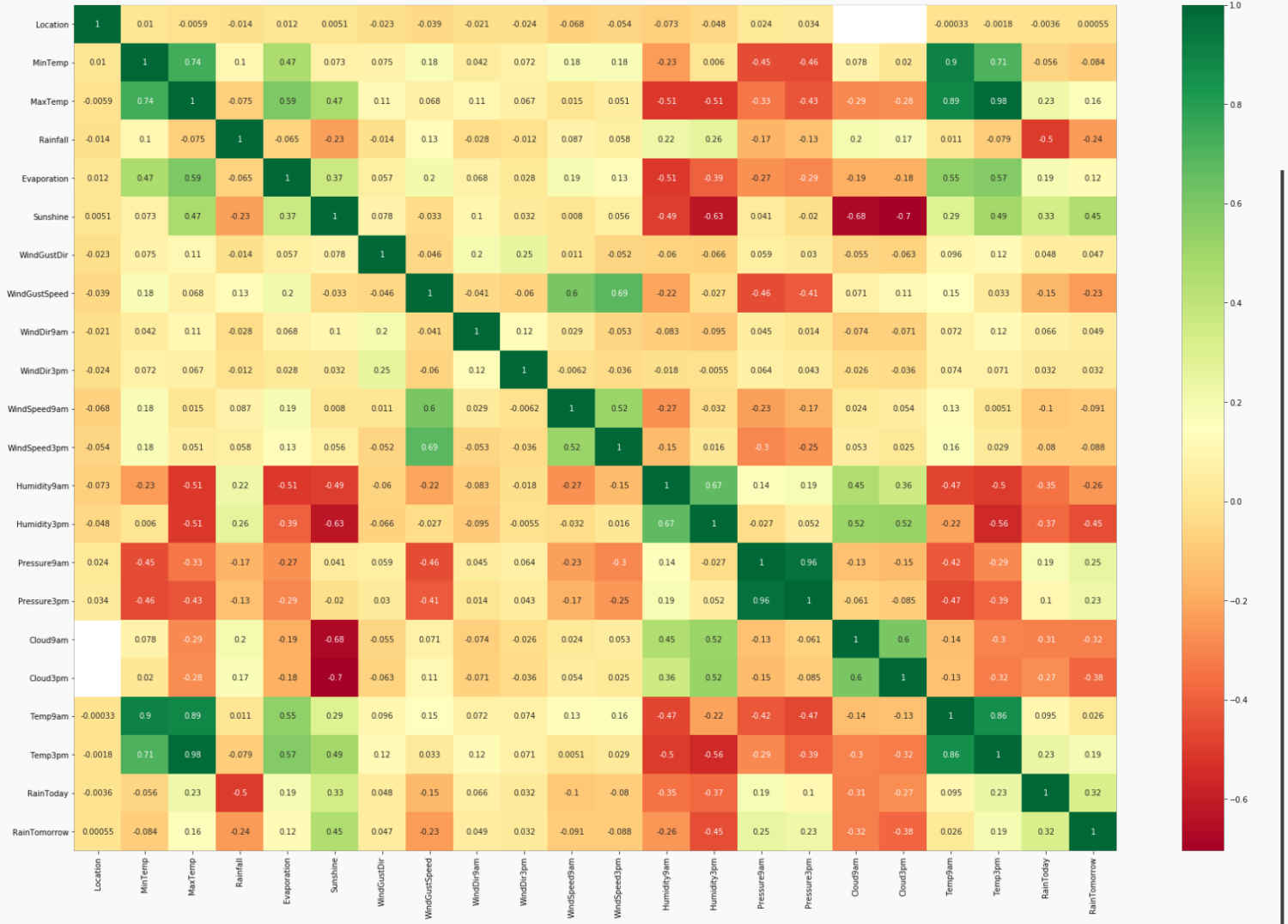
(Bureau of Meteorology)

香港天文台

HONG KONG OBSERVATORY

Hong Kong Weather

2000-2019


Location 1 0.01 -0.0059 -0.014 0.012 0.0051 -0.023 -0.039 -0.021 -0.024 -0.068 -0.054 -0.073 -0.048 0.024 0.034 -0.00033 -0.0018 -0.0036 0.00055
MinTemp 0.01 1 0.74 0.1 0.47 0.073 0.075 0.18 0.042 0.072 0.18 0.18 -0.23 0.006 -0.45 -0.46 0.078 0.02 0.9 0.71 -0.056 -0.084
MaxTemp -0.0059 0.74 1 -0.075 0.59 0.47 0.11 0.068 0.11 0.067 0.015 0.051 -0.51 -0.51 -0.33 -0.43 -0.29 -0.28 0.89 0.98 0.23 0.16
Rainfall -0.014 0.1 -0.075 1 -0.065 -0.23 -0.014 0.13 -0.028 -0.012 0.087 0.058 0.22 0.26 -0.17 -0.13 0.2 0.17 0.011 -0.079 -0.5 -0.24
Evaporation 0.012 0.47 0.59 -0.065 1 0.37 0.057 0.2 0.068 0.028 0.19 0.13 -0.51 -0.39 -0.27 -0.29 -0.19 -0.18 0.55 0.57 0.19 0.12
Sunshine 0.0051 0.073 0.47 -0.23 0.37 1 0.078 -0.033 0.1 0.032 0.008 0.056 -0.49 -0.63 0.041 -0.02 -0.68 -0.7 0.29 0.49 0.33 0.45
WindGustDir -0.023 0.075 0.11 -0.014 0.057 0.078 1 -0.046 0.2 0.25 0.011 -0.052 -0.06 -0.066 0.059 0.03 -0.055 -0.063 0.096 0.12 0.048 0.047
WindGustSpeed -0.039 0.18 0.068 0.13 0.2 -0.033 -0.046 1 -0.041 -0.06 0.6 0.69 -0.22 -0.027 -0.46 -0.41 0.071 0.11 0.15 0.033 -0.15 -0.23
WindDir9am -0.021 0.042 0.11 -0.028 0.068 0.1 0.2 -0.041 1 0.12 0.029 -0.053 -0.083 -0.095 0.045 0.014 -0.074 -0.071 0.072 0.12 0.066 0.049
WindDir3pm -0.024 0.072 0.067 -0.012 0.028 0.032 0.25 -0.06 0.12 1 -0.0062 -0.036 -0.018 -0.0055 0.064 0.043 -0.026 -0.036 0.074 0.071 0.032 0.032
WindSpeed9am -0.068 0.18 0.015 0.087 0.19 0.008 0.011 0.6 0.029 -0.0062 1 0.52 -0.27 -0.032 -0.23 -0.17 0.024 0.054 0.13 0.0051 -0.1 -0.091
WindSpeed3pm -0.054 0.18 0.051 0.058 0.13 0.056 -0.052 0.69 -0.053 -0.036 0.52 1 -0.15 0.016 -0.3 -0.25 0.053 0.025 0.16 0.029 -0.08 -0.088
Humidity9am -0.073 -0.23 -0.51 0.22 -0.51 -0.49 -0.06 -0.22 -0.083 -0.018 -0.27 -0.15 1 0.67 0.14 0.19 0.45 0.36 -0.47 -0.5 -0.35 -0.26
Humidity3pm -0.048 0.006 -0.51 0.26 -0.39 -0.63 -0.066 -0.027 -0.095 -0.0055 -0.032 0.016 0.67 1 -0.027 0.052 0.52 0.52 -0.22 -0.56 -0.37 -0.45
Pressure9am 0.024 -0.45 -0.33 -0.17 -0.27 0.041 0.059 -0.46 0.045 0.064 -0.23 -0.3 0.14 -0.027 1 0.96 -0.13 -0.15 -0.42 -0.29 0.19 0.25
Pressure3pm 0.034 -0.46 -0.43 -0.13 -0.29 -0.02 0.03 -0.41 0.014 0.043 -0.17 -0.25 0.19 0.052 0.96 1 -0.061 -0.085 -0.47 -0.39 0.1 0.23
Cloud9am 0.078 -0.29 0.2 -0.19 -0.68 -0.055 0.071 -0.074 -0.026 0.024 0.053 0.45 0.52 -0.13 -0.061 1 0.6 -0.14 -0.3 -0.31 -0.32
Cloud3pm 0.02 -0.28 0.17 -0.18 -0.7 -0.063 0.11 -0.071 -0.036 0.054 0.025 0.36 0.52 -0.15 -0.085 0.6 1 -0.13 -0.32 -0.27 -0.38
Temp9am -0.00033 0.9 0.89 0.011 0.55 0.29 0.096 0.15 0.072 0.074 0.13 0.16 -0.47 -0.22 -0.42 -0.47 -0.14 -0.13 1 0.86 0.095 0.026
Temp3pm -0.0018 0.71 0.98 -0.079 0.57 0.49 0.12 0.033 0.12 0.071 0.0051 0.029 -0.5 -0.56 -0.29 -0.39 -0.3 -0.32 0.86 1 0.23 0.19
RainToday -0.0036 -0.056 0.23 -0.5 0.19 0.33 0.048 -0.15 0.066 0.032 -0.1 -0.08 -0.35 -0.37 0.19 0.1 -0.31 -0.27 0.095 0.23 1 0.32
RainTomorrow 0.00055 -0.084 0.16 -0.24 0.12 0.45 0.047 -0.23 0.049 0.032 -0.091 -0.088 -0.26 -0.45 0.25 0.23 -0.32 -0.38 0.026 0.19 0.32 1
Location
MinTemp
MaxTemp
Rainfall
Evaporation
Sunshine
WindGustDir
WindGustSpeed
WindDir9am
WindDir3pm
WindSpeed9am
WindSpeed3pm
Humidity9am
Humidity3pm
Pressure9am
Pressure3pm
Cloud9am
Cloud3pm
Temp9am
Temp3pm
RainToday
RainTomorrow
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6

PREPROCESSING

Drop Risk_MM

Impute
Missing Values

One Hot Encoding

Drop All the
Categorical columns

Filter Common Columns
For both dataset

Split train
& test data

Remove
Outlier

Feature
Scaling

04

MODELS CREATION & COMPARISON

84%
XGBoost

82%
Random Forest

77%
Ada Boost

77%
Linear Regression

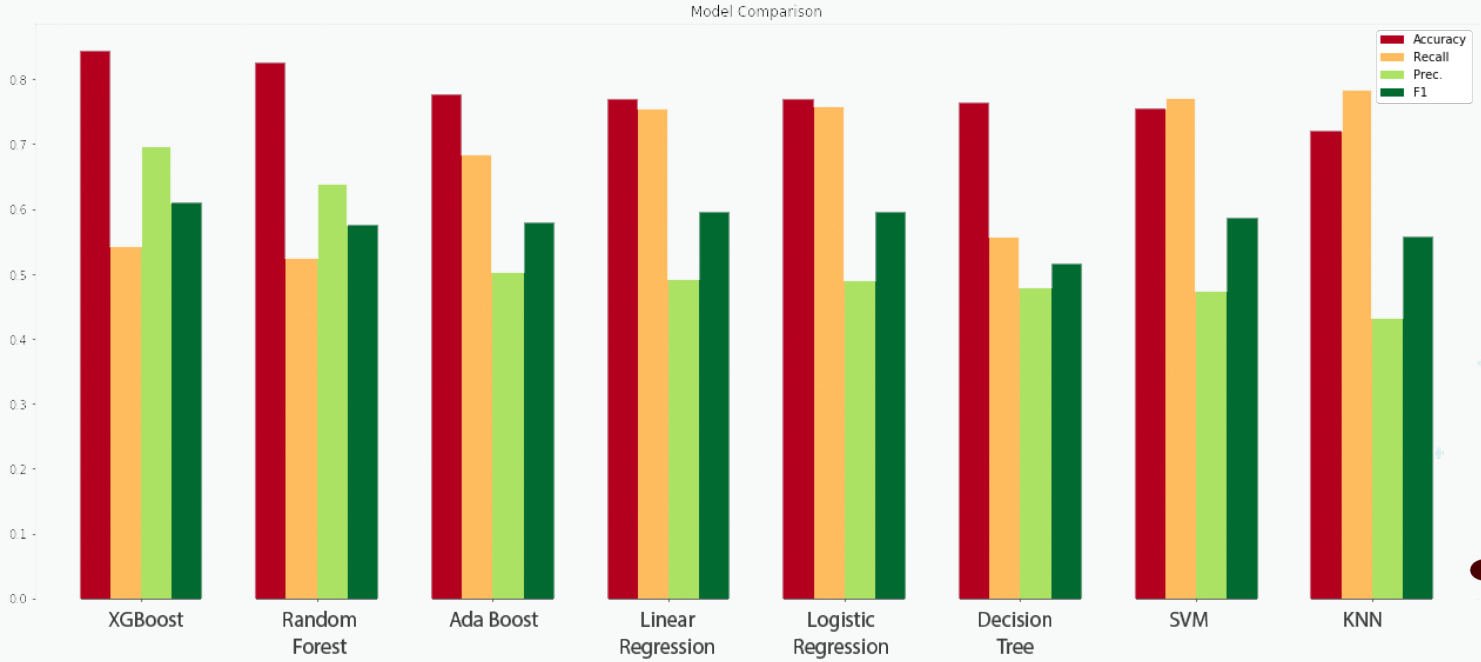
76%
Logistic Regression

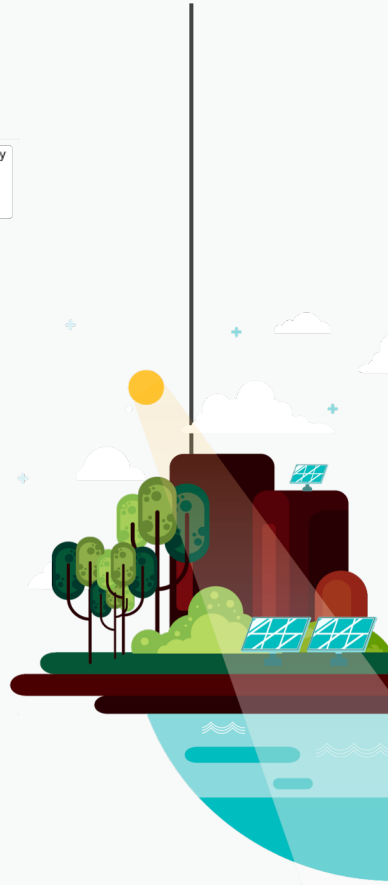
76%
Decision Tree

75%
SVM

72%
KNN

MODELS COMPARISON

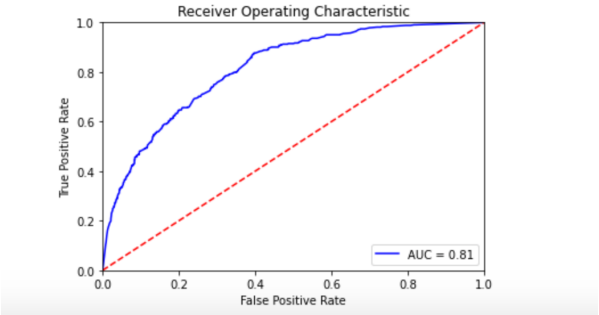




05

MODEL EVALUATION

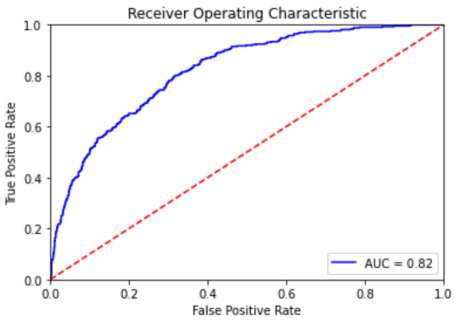
```
cm:
[[728 291]
 [111 311]]
              precision    recall  f1-score   support

           0       0.87      0.71      0.78      1019
           1       0.52      0.74      0.61       422

    accuracy                           0.72      1441
   macro avg       0.69      0.73      0.70      1441
weighted avg       0.76      0.72      0.73      1441

Accuracy: 72.10%
f1: 0.607421875
roc_auc: 0.8148856559492859
```



XGBoost

VS

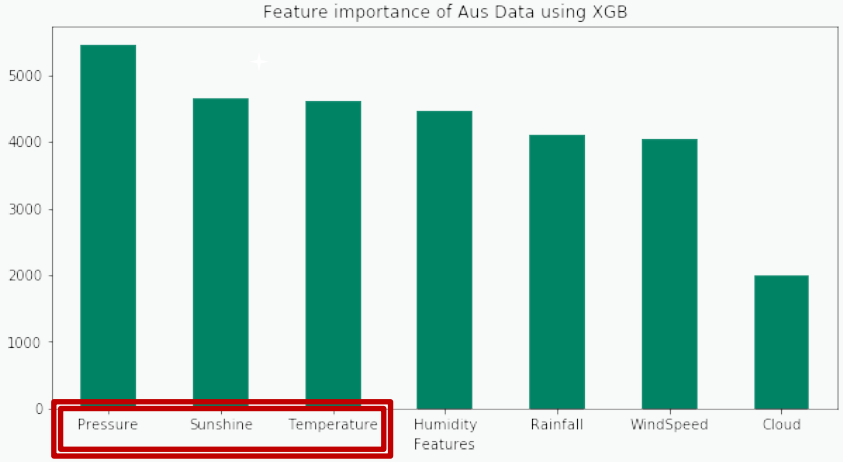
```
cm:
[[756 263]
 [121 301]]
              precision    recall  f1-score   support

           0       0.86      0.74      0.80      1019
           1       0.53      0.71      0.61       422

    accuracy                           0.73      1441
   macro avg       0.70      0.73      0.70      1441
weighted avg       0.77      0.73      0.74      1441

Accuracy: 73.35%
f1: 0.6105476673427991
roc_auc: 0.8204470510536769
```

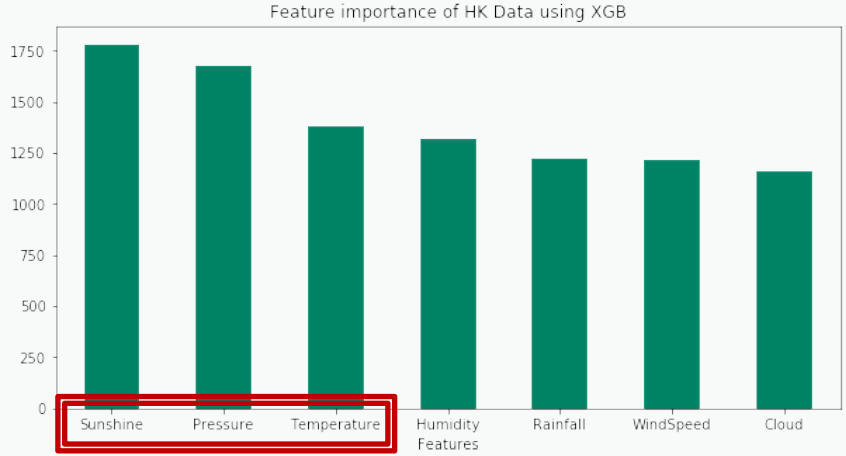


Random Forest

05

MODEL EVALUATION



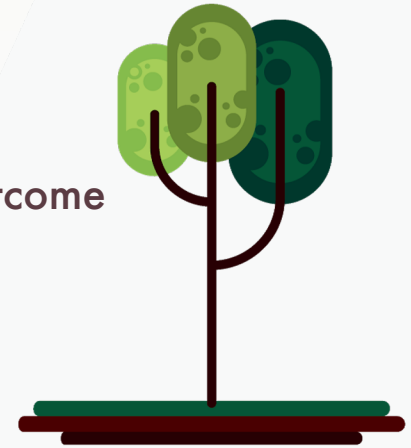


Top 3 features: Pressure, Sunshine, Temperature

06

CONCLUSION

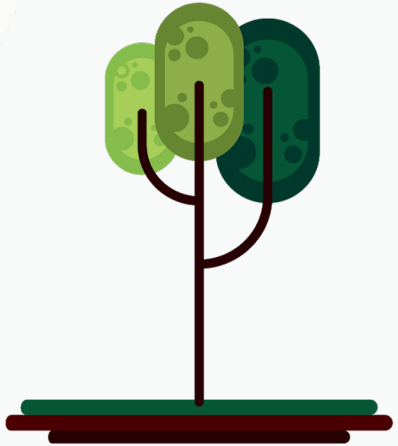
- Pressure, Temperature and Sunshine are the dominant features to predict rainfall probability in the following day
- Prediction by these features is applicable globally
- Seasonal factors do not directly correlate to the prediction outcome



06

FUTURE IMPROVEMENT

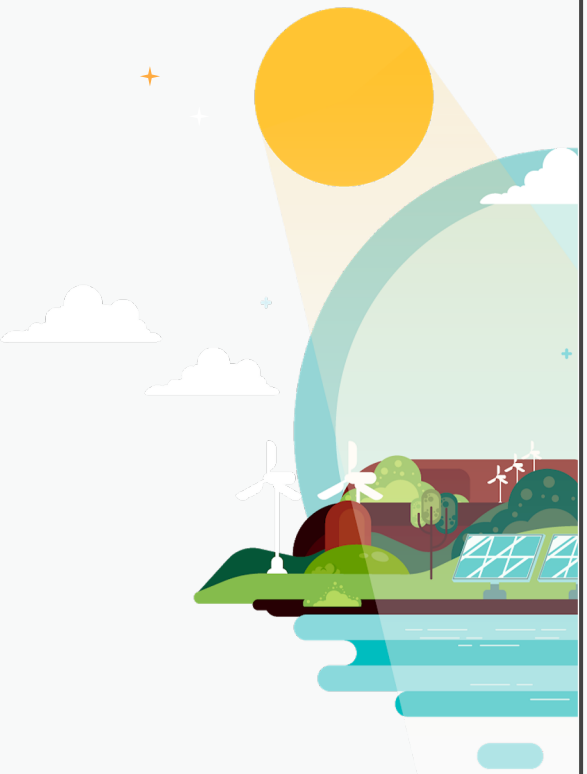
- Feature engineering: feature cross with date/month column
- Quantitative Prediction: rainfall (mm)
- Consult domain expert to have deeper analysis



SCENARIO WITH PREDICTION

TENT OF NATIONS
People Building Bridges

	Hong Kong Camping	Iceland Road Trip	Volunteer in Bethlehem
Pressure (hPa)	1013	1007	1016
Temperature (°C)	29.1	11.2	22.7
Humidity (%)	69	80	75
Cloud (%)	18	81	25
Sunshine (hour)	6.2	2.4	9.8
Windspeed (km/hr)	6	3.1	12.2
Rainfall (mm)	2	3	0
Rain Probability	0.11	0.76	0.01
Rain?	No	Yes	No



PREDICTING TODAY'S CHANCE OF RAINING!

4/10
HK WEATHER

Mean Pressure(hPa)	1009.9
Absolute Daily Max(deg. C)	31.4
Mean(deg. C)	28.4
Absolute Daily min(deg. C)	26.8
Mean Dew Point (deg. C)	24
Mean Relative Humidity(%)	78
Mean Amount of Cloud(%)	34
Total Rainfall(mm)	0
Total Bright Sunshine(hours)	9.1
Mean Wind Speed(km/h)	NaN
Location	NaN
WindDir9am	NaN
WindDir3pm	NaN
Year	2020
Month	10
Label	Yes
Score	0.6483

Q&A