

A vibrant, warm-toned photograph featuring a woman and two children. The woman, with long dark hair and a bright smile, is in the center, holding a large orange umbrella. To her left, a young girl with long brown hair is also smiling. In the foreground, a younger child wearing a yellow knit hat is looking up and smiling. The background is a solid, bright yellow, creating a cheerful and sunny atmosphere.

Machine Learning

**การทำนายฝนตกที่
ประเทศออสเตรเลีย
จากปัจจัยภายนอก**

สมาชิกกลุ่ม
" พี่สาวกะรัต "



63010524

นิชาภา พักเง้า



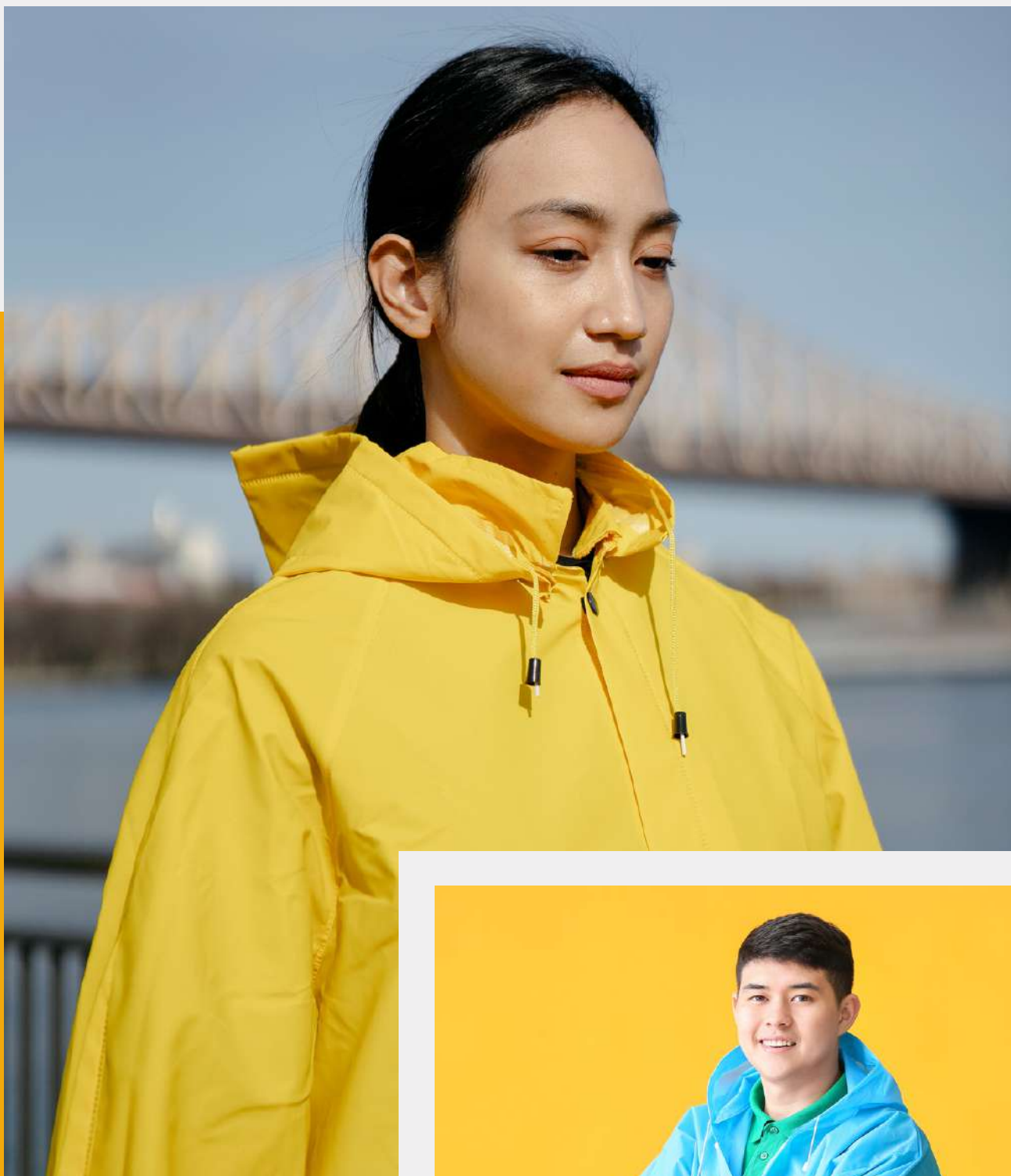
63010666

พรรณชิตา วรต่าย



จุดประสงค์

หาปัจจัยที่มีผลต่อการทำให้ฝนตก
เพื่อใช้ทำนายการตกของฝน



ข้อมูล Rain in Australia

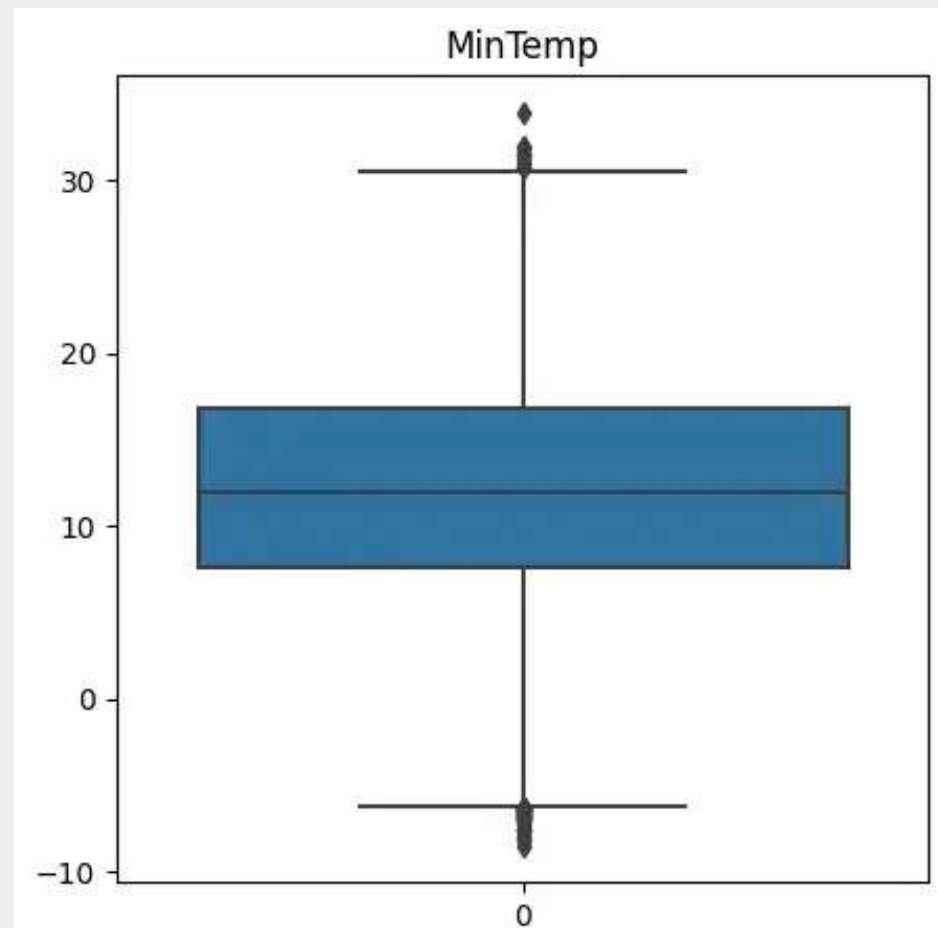
จาก <https://www.kaggle.com/datasets/jsphyg/weather-dataset-rattle-package>

- วัน/เดือน/ปี
- ชื่อรัฐทั้ง 49 รัฐ
- MinTemp
- MaxTemp
- Rainfall
- Evaporation
- Sunshine
- WindGustDir
- WindGustSpeed
- WindDir9am
- WindDir3pm
- WindSpeed9am
- WindSpeed3pm
- Humidity9am
- Humidity3pm
- Pressure9am
- Pressure3pm
- Cloud9am
- Cloud3pm
- Temp9am
- Temp3pm
- RainToday
- RainTomorrow

Clean Data

```
Location          0
MinTemp           1485
MaxTemp           1261
WindSpeed9am      1767
WindSpeed3pm      3062
Humidity9am       2654
Humidity3pm       4507
Pressure9am       15065
Pressure3pm       15028
Temp9am           1767
Temp3pm           3609
RainToday         3261
dtype: int64
```

check null



check outlier

	MinTemp	MaxTemp	WindSpeed9am	WindSpeed3pm
MinTemp	1.000000	0.734227	0.178225	0.177826
MaxTemp	0.734227	1.000000	0.016812	0.053797
WindSpeed9am	0.178225	0.016812	1.000000	0.508538
WindSpeed3pm	0.177826	0.053797	0.508538	1.000000
Humidity9am	-0.231749	-0.499133	-0.273775	-0.147255
Humidity3pm	0.005775	-0.499464	-0.032662	0.015160
Pressure9am	-0.433225	-0.319130	-0.213323	-0.275346
Pressure3pm	-0.441559	-0.406609	-0.163838	-0.237202
Temp9am	0.898957	0.881270	0.132082	0.166854
Temp3pm	0.699817	0.969617	0.006853	0.031861
RainToday	0.055775	-0.228570	0.102202	0.078528

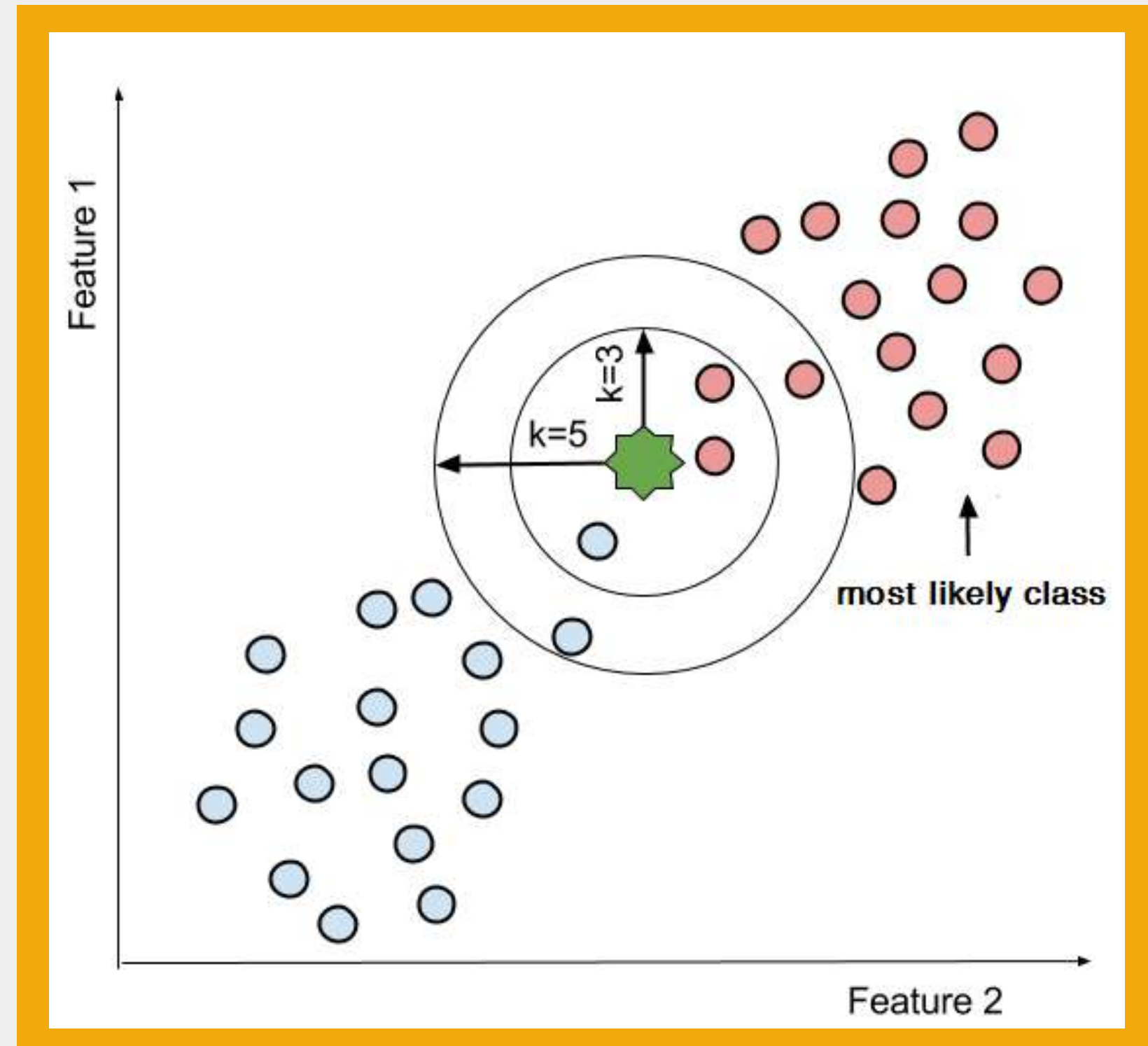
check Corr

Models for predict

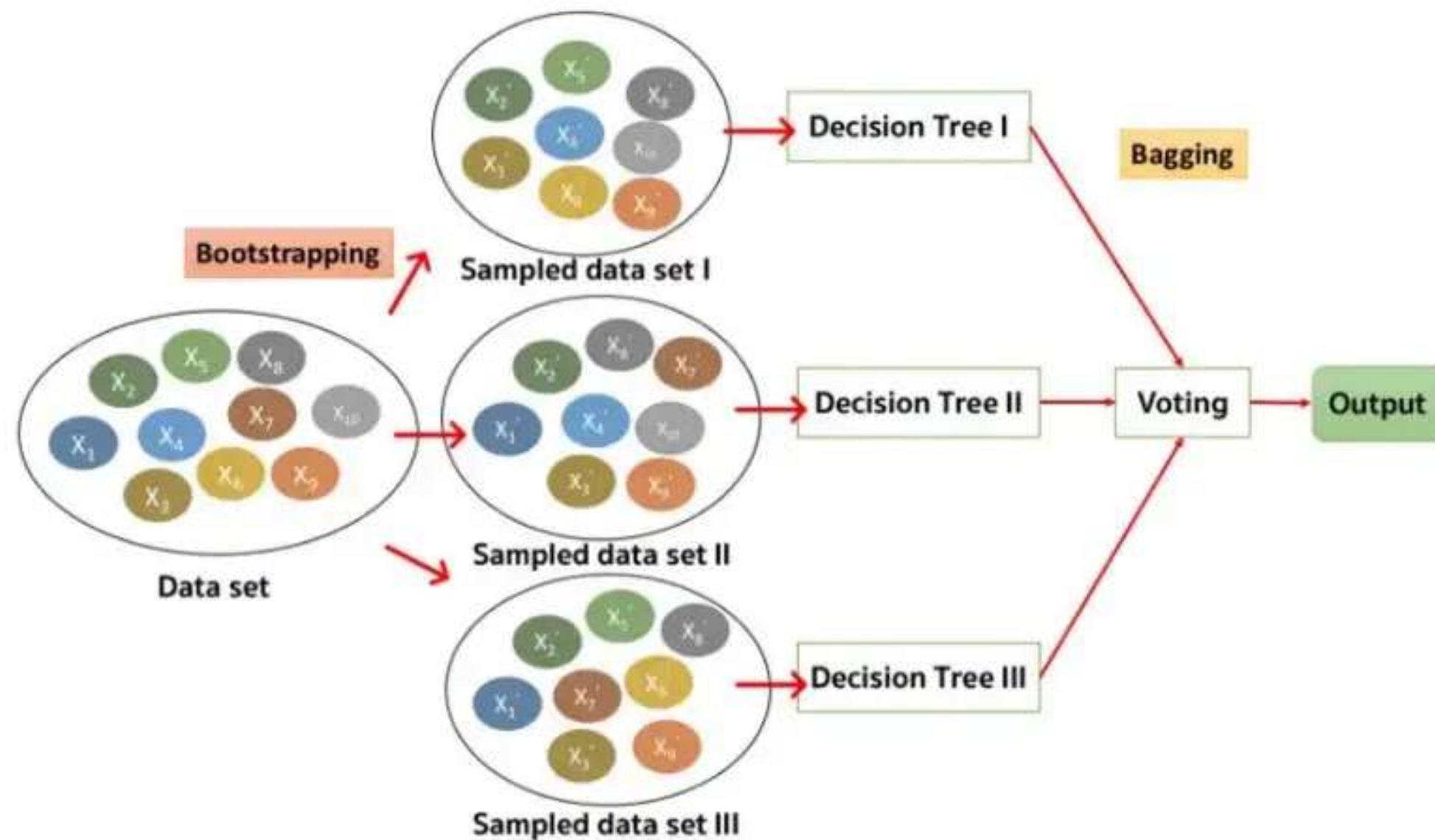
โมเดลที่ใช้ในการทำนาย

KNN

เป็นวิธีที่ใช้ในการจัดแบ่งคลาส โดยเทคนิคนี้จะตัดสินใจว่า คลาสใดที่จะแทนเงื่อนไขใหม่ๆ ได้บ้าง โดยการตรวจสอบจำนวนบางจำนวน ในขั้นตอนวิธีการเพื่อนบ้านใกล้ที่สุด



Random Forest



สร้าง model จาก Decision Tree หลายๆ model ย่อยๆ (ตั้งแต่ 10 model ถึงมากกว่า 1000 model) โดยแต่ละ model จะได้รับ dataset ไม่เหมือนกัน ซึ่งเป็น subset ของ data set ทั้งหมด ตอนทำ prediction ก็ให้แต่ละ Decision Tree ทำ prediction ของใครของมัน และคำนวณผล prediction ด้วยการ vote output ที่ถูกเลือกโดย Decision Tree มากที่สุด

วิเคราะห์ และเปรียบเทียบผลลัพธ์ที่ได้

จากการคำนวณหาค่าความสำคัญจะสามารถจัดลำดับได้ดังนี้

1.Humidity9am มีความสำคัญมากที่สุด โดยมีค่า 0.179192

2.Humidity3pm โดยมีค่า 0.154536

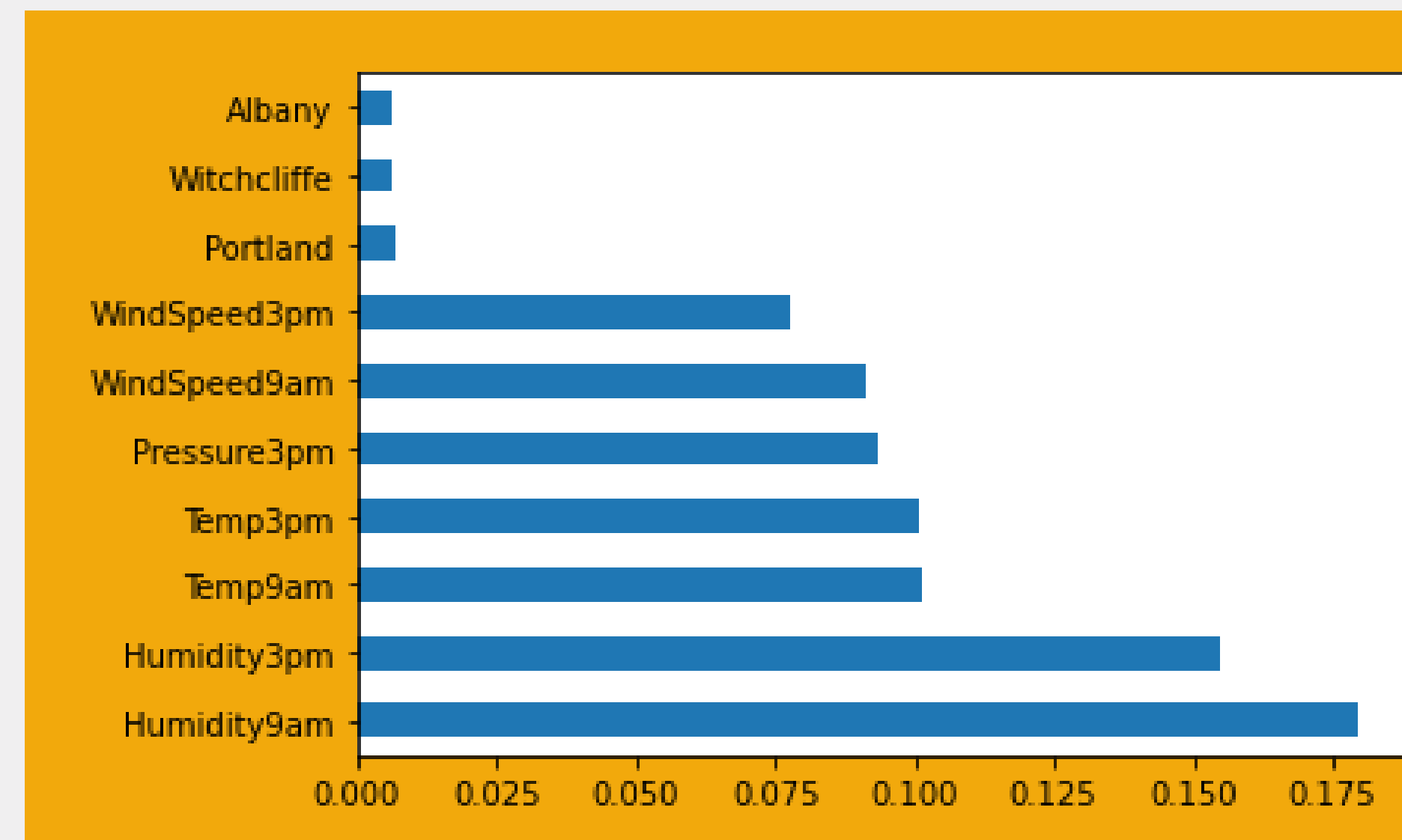
3.Temp9am โดยมีค่า 0.101056

4.Temp3pm โดยมีค่า 0.100645

5.Pressure3pm โดยมีค่า 0.093283

6.WindSpeed9am โดยมีค่า 0.091073

7.WindSpeed3pm โดยมีค่า 0.077951



นอกจากนี้จากการทำโมเดลทำนายข้างต้น 2 โมเดลได้สามารถเปรียบเทียบผลลัพธ์ได้ ดังนี้

1.accuracy score : KNN =0.820 ส่วน Random Forest = 0.839 แสดงว่าโมเดล Random Forest มีความน่าเชื่อถือมากกว่าโมเดล KNN

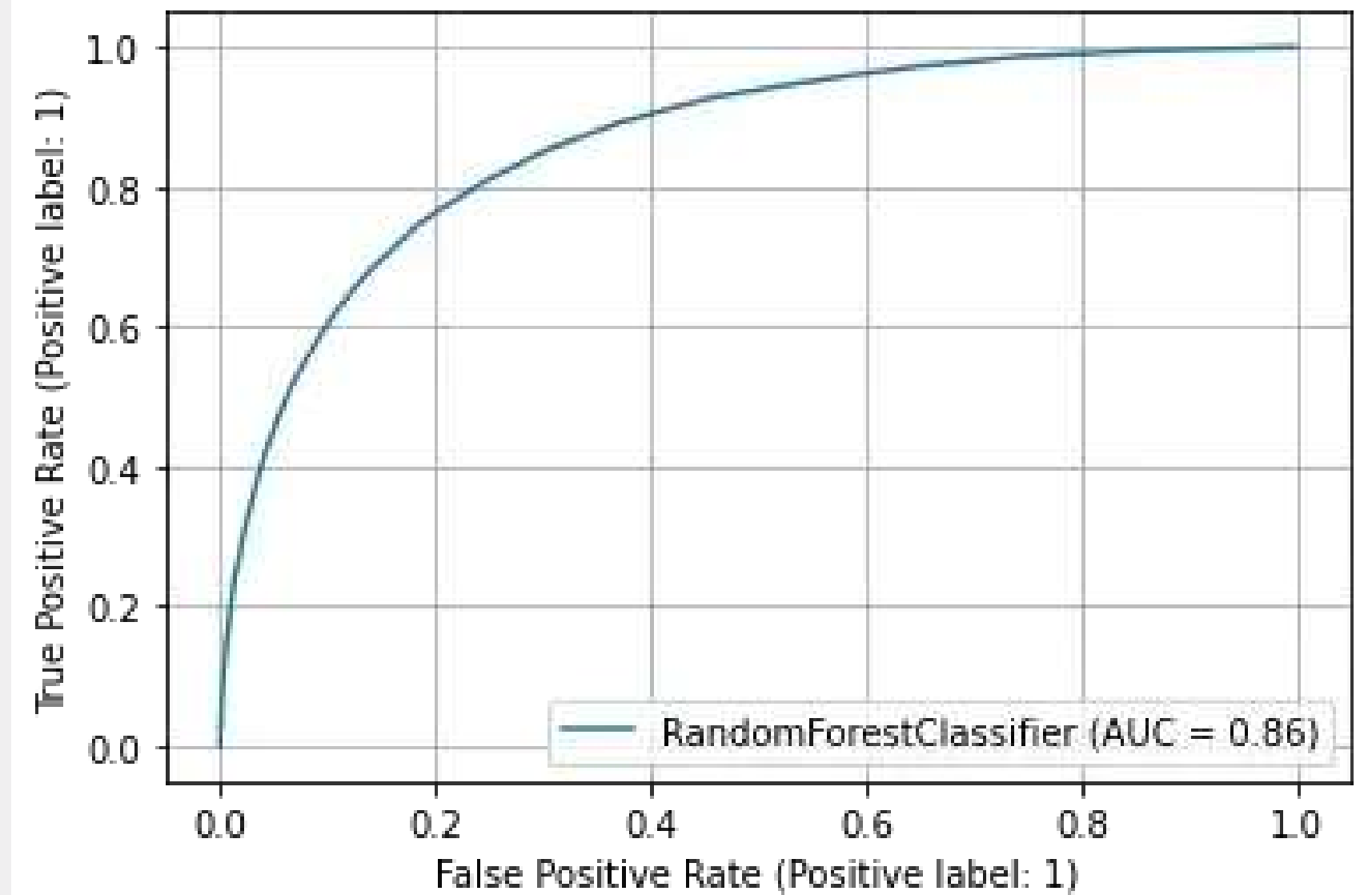
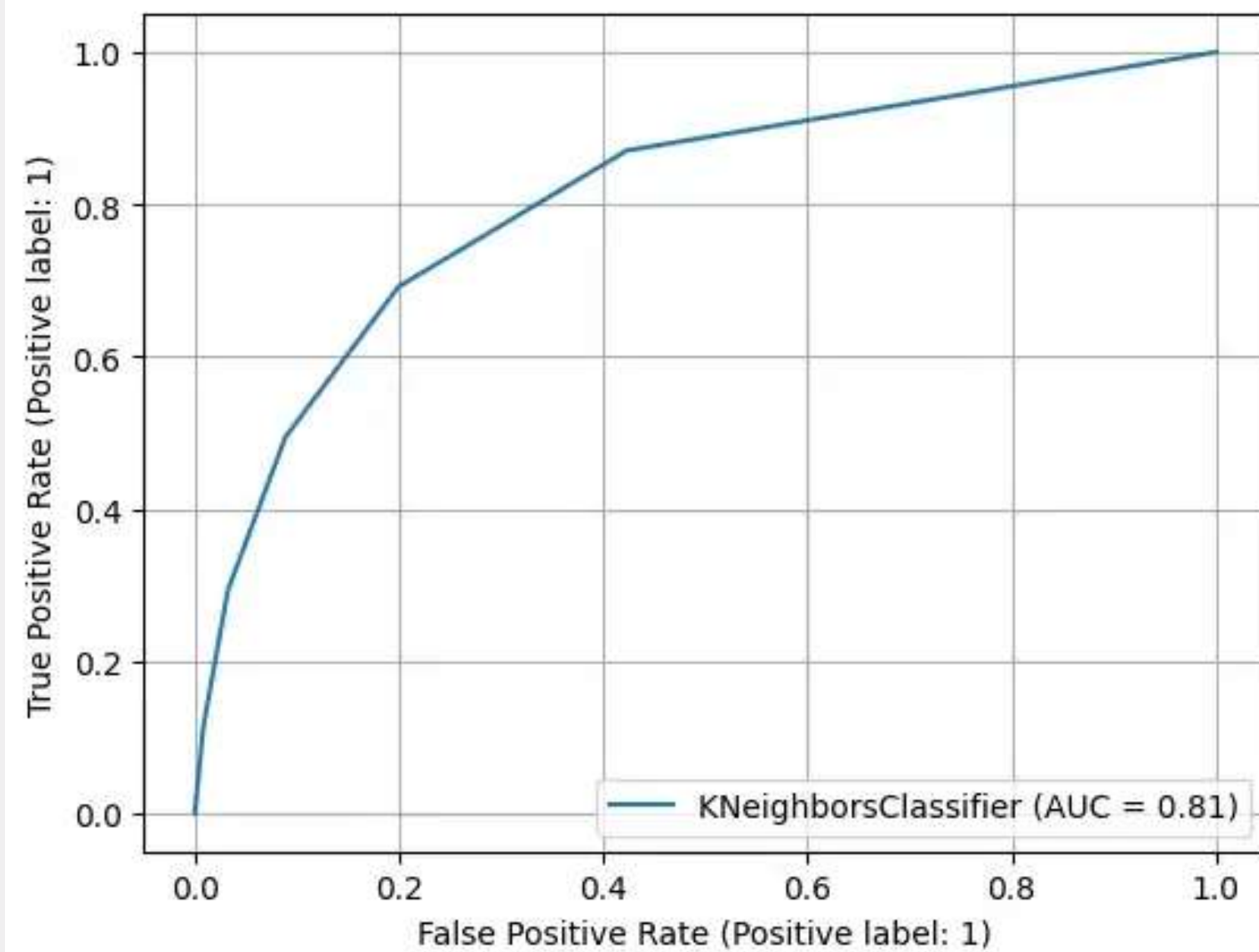
2.Mean cross validation score : KNN = 0.7228 ส่วน Random Forest = 0.7377 แสดงว่าโมเดล Random Forest มี Mean cross validation score มากกว่าโมเดล KNN

3.True Positive (TP) : KNN = 30377 ส่วน Random Forest = 31372 ซึ่งค่า TP ยิ่งมีค่ามากยิ่งดี ดังนั้นแสดงว่าโมเดล Random Forest ทำนายว่าจริง และสิ่งที่เกิดขึ้นนั้นจริง ได้ดีกว่าโมเดล KNN

4.True Negative (TN) : KNN = 4625 ส่วน Random Forest = 4451 ซึ่งค่า TN ยิ่งมีค่ามากยิ่งดี ดังนั้นแสดงว่าโมเดล KNN ทำนายว่าไม่จริงและสิ่งที่เกิดขึ้นนั้นไม่จริงได้ดีกว่าโมเดล Random Forest

5.False Positive (FP) : KNN = 4930 ส่วน Random Forest = 5104 ซึ่งค่า FP ยิ่งมีค่าน้อยยิ่งดี ดังนั้นแสดงว่าโมเดล KNN ทำนายว่าจริง แต่สิ่งที่เกิดขึ้นนั้นไม่จริงได้ดีกว่าโมเดล Random Forest

6.False Negative (FN) : KNN = 2728 ส่วน Random Forest = 1733 ซึ่งค่า FN ยิ่งมีค่าน้อยยิ่งดี ดังนั้นแสดงว่าโมเดล Random Forest ทำนายว่าไม่จริง แต่สิ่งที่เกิดขึ้นนั้นจริง ได้ดีกว่าโมเดล KNN



7.AUC : KNN = 0.81 ส่วน Random Forest = 0.86

แสดงว่าโมเดล Random Forest มี AUC มากกว่าโมเดล KNN



สรุปผลการทดสอบ

ดังนั้นจากการคำนวณหาค่าความสำคัญจะสามารถสรุปได้ว่า

- Humidity9am เป็นปัจจัยที่สำคัญ
- ปัจจัยที่สำคัญเป็นอันดับที่ 2 คือ Humidity3pm
- ปัจจัยที่สำคัญเป็นอันดับที่ 3 คือ Temp9am

ส่วนการใช้ทำนายสรุปได้ว่าโมเดล Random Forest ให้ที่ให้ผลลัพธ์ดีกว่าโมเดล KNN

Link Colabs

[https://colab.research.google.
com/drive/16TqjL9V9gJ2VPs0
kqiY2HfE4nFedU04N?
usp=sharing](https://colab.research.google.com/drive/16TqjL9V9gJ2VPs0kqiY2HfE4nFedU04N?usp=sharing)





เพิ่มเติม

https://colab.research.google.com/drive/1ZljSaDkABaWhKzJ-IcMED4pZurZUp_JN?usp=sharing

เพื่อนๆ สามารถเลือกคอลัมน์ที่จะนำไป
train/test ใน model ได้เลย