Tugas Prapraktikum

Tugas Prapraktikum dikerjakan dengan *dataset* Rain in Australia. Tanpa meninjau waktu (`date`), prediksi status hujan pada keesokan harinya (`RainTomorrow`). Berikan nilai `1` jika diprediksi hujan pada keesokan harinya, `0` jika tidak.

Tugas dikerjakan secara berkelompok. Setiap kelompok terdiri atas 2 (dua) mahasiswa. Kumpulkan paling lambat pada Minggu, 16 April 2023, pukul 23:59 WIB melalui Edunex.

0. Persiapan Data and Pustaka

Install and import dependencies

```
!pip install imblearn
!pip install pandas
!pip install scikit-learn
!pip install matplotlib
```

```
Collecting imblearn
  Downloading imblearn-0.0-py2.py3-none-any.whl (1.9 kB)
Collecting imbalanced-learn
  Downloading imbalanced_learn-0.10.1-py3-none-any.whl (226 kB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 226.0/226.0 KB 10.8 MB/s eta
0:00:00
ent already satisfied: scipy>=1.3.2 in
/shared-libs/python3.9/py/lib/python3.9/site-packages (from
imbalanced-learn->imblearn) (1.9.3)
Requirement already satisfied: numpy>=1.17.3 in
/shared-libs/python3.9/py/lib/python3.9/site-packages (from
imbalanced-learn->imblearn) (1.23.4)
Requirement already satisfied: threadpoolctl>=2.0.0 in
/shared-libs/python3.9/py/lib/python3.9/site-packages (from
imbalanced-learn->imblearn) (3.1.0)
Requirement already satisfied: joblib>=1.1.1 in
/shared-libs/python3.9/py/lib/python3.9/site-packages (from
imbalanced-learn->imblearn) (1.2.0)
Requirement already satisfied: scikit-learn>=1.0.2 in
/shared-libs/python3.9/py/lib/python3.9/site-packages (from
imbalanced-learn->imblearn) (1.1.2)
Installing collected packages: imbalanced-learn, imblearn
Successfully installed imbalanced-learn-0.10.1 imblearn-0.0
WARNING: You are using pip version 22.0.4; however, version 23.1 is
available.
You should consider upgrading via the '/root/venv/bin/python -m pip
install --upgrade pip' command.
Requirement already satisfied: pandas in
/shared-libs/python3.9/py/lib/python3.9/site-packages (1.2.5)
Requirement already satisfied: pytz>=2017.3 in
```

```
/shared-libs/python3.9/py/lib/python3.9/site-packages (from pandas)
(2022.5)
Requirement already satisfied: python-dateutil>=2.7.3 in /shared-
libs/python3.9/py-core/lib/python3.9/site-packages (from pandas)
(2.8.2)
Requirement already satisfied: numpy>=1.16.5 in
/shared-libs/python3.9/py/lib/python3.9/site-packages (from pandas)
(1.23.4)
Requirement already satisfied: six>=1.5 in /shared-libs/python3.9/py-
core/lib/python3.9/site-packages (from python-dateutil>=2.7.3->pandas)
(1.16.0)
WARNING: You are using pip version 22.0.4; however, version 23.1 is
available.
You should consider upgrading via the '/root/venv/bin/python -m pip
install --upgrade pip' command.
Requirement already satisfied: scikit-learn in
/shared-libs/python3.9/py/lib/python3.9/site-packages (1.1.2)
Requirement already satisfied: threadpoolctl>=2.0.0 in
/shared-libs/python3.9/py/lib/python3.9/site-packages (from scikit-
learn) (3.1.0)
Requirement already satisfied: joblib>=1.0.0 in
/shared-libs/python3.9/py/lib/python3.9/site-packages (from scikit-
learn) (1.2.0)
Requirement already satisfied: numpy>=1.17.3 in
/shared-libs/python3.9/py/lib/python3.9/site-packages (from scikit-
learn) (1.23.4)
Requirement already satisfied: scipy>=1.3.2 in
/shared-libs/python3.9/py/lib/python3.9/site-packages (from scikit-
learn) (1.9.3)
WARNING: You are using pip version 22.0.4; however, version 23.1 is
available.
You should consider upgrading via the '/root/venv/bin/python -m pip
install --upgrade pip' command.
Requirement already satisfied: matplotlib in
/shared-libs/python3.9/py/lib/python3.9/site-packages (3.6.0)
Requirement already satisfied: contourpy>=1.0.1 in
/shared-libs/python3.9/py/lib/python3.9/site-packages (from
matplotlib) (1.0.5)
Requirement already satisfied: numpy>=1.19 in
/shared-libs/python3.9/py/lib/python3.9/site-packages (from
matplotlib) (1.23.4)
Requirement already satisfied: pillow>=6.2.0 in
/shared-libs/python3.9/py/lib/python3.9/site-packages (from
matplotlib) (9.2.0)
Requirement already satisfied: python-dateutil>=2.7 in
/shared-libs/python3.9/py-core/lib/python3.9/site-packages (from
matplotlib) (2.8.2)
Requirement already satisfied: packaging>=20.0 in
/shared-libs/python3.9/py-core/lib/python3.9/site-packages (from
matplotlib) (21.3)
```

```
Requirement already satisfied: fonttools>=4.22.0 in
/shared-libs/python3.9/py/lib/python3.9/site-packages (from
matplotlib) (4.37.4)
Requirement already satisfied: cycler>=0.10 in
/shared-libs/python3.9/py/lib/python3.9/site-packages (from
matplotlib) (0.11.0)
Requirement already satisfied: pyparsing>=2.2.1 in
/shared-libs/python3.9/py-core/lib/python3.9/site-packages (from
matplotlib) (3.0.9)
Requirement already satisfied: kiwisolver>=1.0.1 in
/shared-libs/python3.9/py/lib/python3.9/site-packages (from
matplotlib) (1.4.4)
Requirement already satisfied: six>=1.5 in /shared-libs/python3.9/py-
core/lib/python3.9/site-packages (from python-dateutil>=2.7-
>matplotlib) (1.16.0)
WARNING: You are using pip version 22.0.4; however, version 23.1 is
available.
You should consider upgrading via the '/root/venv/bin/python -m pip
install --upgrade pip' command.
```

```python
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import LabelEncoder, MinMaxScaler
from sklearn.linear_model import LogisticRegression
from sklearn.svm import SVC
from sklearn.tree import DecisionTreeClassifier
from sklearn.metrics import accuracy_score, precision_score,
recall_score, f1_score, confusion_matrix
from sklearn.model_selection import cross_validate, GridSearchCV
from sklearn.ensemble import VotingClassifier, StackingClassifier
from imblearn.over_sampling import SMOTE
from imblearn.under_sampling import RandomUnderSampler
import matplotlib.pyplot as plt
```

Read the data

```python
df = pd.read_csv('weatherAUS.csv')
df = df.drop(['Date'], axis=1)
df
```

```
{"column_count":22,"columns":[{"dtype":"object","name":"Location"},
{"dtype":"float64","name":"MinTemp"},
{"dtype":"float64","name":"MaxTemp"},
{"dtype":"float64","name":"Rainfall"},
{"dtype":"float64","name":"Evaporation"},
{"dtype":"float64","name":"Sunshine"},
{"dtype":"object","name":"WindGustDir"},
{"dtype":"float64","name":"WindGustSpeed"},
{"dtype":"object","name":"WindDir9am"},
{"dtype":"object","name":"WindDir3pm"},
{"dtype":"float64","name":"WindSpeed9am"},
{"dtype":"float64","name":"WindSpeed3pm"},
{"dtype":"float64","name":"Humidity9am"},
{"dtype":"float64","name":"Humidity3pm"},
{"dtype":"float64","name":"Pressure9am"},
{"dtype":"float64","name":"Pressure3pm"},
{"dtype":"float64","name":"Cloud9am"},
{"dtype":"float64","name":"Cloud3pm"},
{"dtype":"float64","name":"Temp9am"},
{"dtype":"float64","name":"Temp3pm"},
{"dtype":"object","name":"RainToday"},
{"dtype":"object","name":"RainTomorrow"},
{"dtype":"int64","name":"_deepnote_index_column"}],"row_count":145460,
"rows":
[{"Cloud3pm":"nan","Cloud9am":"8.0","Evaporation":"nan","Humidity3pm":
"22.0","Humidity9am":"71.0","Location":"Albury","MaxTemp":"22.9","MinT
emp":"13.4","Pressure3pm":"1007.1","Pressure9am":"1007.7","RainToday":
"No","RainTomorrow":"No","Rainfall":"0.6","Sunshine":"nan","Temp3pm":"
21.8","Temp9am":"16.9","WindDir3pm":"WNW","WindDir9am":"W","WindGustDi
r":"W","WindGustSpeed":"44.0","WindSpeed3pm":"24.0","WindSpeed9am":"20
.0","_deepnote_index_column":"0"},
{"Cloud3pm":"nan","Cloud9am":"nan","Evaporation":"nan","Humidity3pm":"
25.0","Humidity9am":"44.0","Location":"Albury","MaxTemp":"25.1","MinTe
mp":"7.4","Pressure3pm":"1007.8","Pressure9am":"1010.6","RainToday":"N
o","RainTomorrow":"No","Rainfall":"0.0","Sunshine":"nan","Temp3pm":"24
.3","Temp9am":"17.2","WindDir3pm":"WSW","WindDir9am":"NNW","WindGustDi
r":"WNW","WindGustSpeed":"44.0","WindSpeed3pm":"22.0","WindSpeed9am":"
4.0","_deepnote_index_column":"1"},
{"Cloud3pm":"2.0","Cloud9am":"nan","Evaporation":"nan","Humidity3pm":"
30.0","Humidity9am":"38.0","Location":"Albury","MaxTemp":"25.7","MinTe
mp":"12.9","Pressure3pm":"1008.7","Pressure9am":"1007.6","RainToday":"
No","RainTomorrow":"No","Rainfall":"0.0","Sunshine":"nan","Temp3pm":"2
3.2","Temp9am":"21.0","WindDir3pm":"WSW","WindDir9am":"W","WindGustDir
":"WSW","WindGustSpeed":"46.0","WindSpeed3pm":"26.0","WindSpeed9am":"1
9.0","_deepnote_index_column":"2"},
{"Cloud3pm":"nan","Cloud9am":"nan","Evaporation":"nan","Humidity3pm":"
16.0","Humidity9am":"45.0","Location":"Albury","MaxTemp":"28.0","MinTe
mp":"9.2","Pressure3pm":"1012.8","Pressure9am":"1017.6","RainToday":"N
o","RainTomorrow":"No","Rainfall":"0.0","Sunshine":"nan","Temp3pm":"26
.5","Temp9am":"18.1","WindDir3pm":"E","WindDir9am":"SE","WindGustDir":
"NE","WindGustSpeed":"24.0","WindSpeed3pm":"9.0","WindSpeed9am":"11.0"
,"_deepnote_index_column":"3"},
{"Cloud3pm":"8.0","Cloud9am":"7.0","Evaporation":"nan","Humidity3pm":"
33.0","Humidity9am":"82.0","Location":"Albury","MaxTemp":"32.3","MinTe
mp":"17.5","Pressure3pm":"1006.0","Pressure9am":"1010.8","RainToday":"
No","RainTomorrow":"No","Rainfall":"1.0","Sunshine":"nan","Temp3pm":"2
9.7","Temp9am":"17.8","WindDir3pm":"NW","WindDir9am":"ENE","WindGustDi
r":"W","WindGustSpeed":"41.0","WindSpeed3pm":"20.0","WindSpeed9am":"7.
0","_deepnote_index_column":"4"},
{"Cloud3pm":"nan","Cloud9am":"nan","Evaporation":"nan","Humidity3pm":"
23.0","Humidity9am":"55.0","Location":"Albury","MaxTemp":"29.7","MinTe
mp":"14.6","Pressure3pm":"1005.4","Pressure9am":"1009.2","RainToday":"
No","RainTomorrow":"No","Rainfall":"0.2","Sunshine":"nan","Temp3pm":"2
8.9","Temp9am":"20.6","WindDir3pm":"W","WindDir9am":"W","WindGustDir":
"WNW","WindGustSpeed":"56.0","WindSpeed3pm":"24.0","WindSpeed9am":"19.
0","_deepnote_index_column":"5"},
{"Cloud3pm":"nan","Cloud9am":"1.0","Evaporation":"nan","Humidity3pm":"
19.0","Humidity9am":"49.0","Location":"Albury","MaxTemp":"25.0","MinTe
mp":"14.3","Pressure3pm":"1008.2","Pressure9am":"1009.6","RainToday":"
No","RainTomorrow":"No","Rainfall":"0.0","Sunshine":"nan","Temp3pm":"2
4.6","Temp9am":"18.1","WindDir3pm":"W","WindDir9am":"SW","WindGustDir"
:"W","WindGustSpeed":"50.0","WindSpeed3pm":"24.0","WindSpeed9am":"20.0
","_deepnote_index_column":"6"},
{"Cloud3pm":"nan","Cloud9am":"nan","Evaporation":"nan","Humidity3pm":"
19.0","Humidity9am":"48.0","Location":"Albury","MaxTemp":"26.7","MinTe
mp":"7.7","Pressure3pm":"1010.1","Pressure9am":"1013.4","RainToday":"N
o","RainTomorrow":"No","Rainfall":"0.0","Sunshine":"nan","Temp3pm":"25
.5","Temp9am":"16.3","WindDir3pm":"W","WindDir9am":"SSE","WindGustDir"
:"W","WindGustSpeed":"35.0","WindSpeed3pm":"17.0","WindSpeed9am":"6.0"
,"_deepnote_index_column":"7"},
{"Cloud3pm":"nan","Cloud9am":"nan","Evaporation":"nan","Humidity3pm":"
9.0","Humidity9am":"42.0","Location":"Albury","MaxTemp":"31.9","MinTem
p":"9.7","Pressure3pm":"1003.6","Pressure9am":"1008.9","RainToday":"No
","RainTomorrow":"Yes","Rainfall":"0.0","Sunshine":"nan","Temp3pm":"30
.2","Temp9am":"18.3","WindDir3pm":"NW","WindDir9am":"SE","WindGustDir"
:"NNW","WindGustSpeed":"80.0","WindSpeed3pm":"28.0","WindSpeed9am":"7.
0","_deepnote_index_column":"8"},
{"Cloud3pm":"nan","Cloud9am":"nan","Evaporation":"nan","Humidity3pm":"
27.0","Humidity9am":"58.0","Location":"Albury","MaxTemp":"30.1","MinTe
mp":"13.1","Pressure3pm":"1005.7","Pressure9am":"1007.0","RainToday":"
Yes","RainTomorrow":"No","Rainfall":"1.4","Sunshine":"nan","Temp3pm":"
28.2","Temp9am":"20.1","WindDir3pm":"SSE","WindDir9am":"S","WindGustDi
r":"W","WindGustSpeed":"28.0","WindSpeed3pm":"11.0","WindSpeed9am":"15
.0","_deepnote_index_column":"9"}]}
```

I. Pemahaman Data

Tujuan dari bagian ini adalah peserta dapat memahami kualitas dari data yang diberikan. Hal yang diliputi adalah sebagai berikut:

1. Ukuran data
2. Statistik dari tiap fitur
3. Pencilan (*outlier*)
4. Korelasi
5. Distribusi

I.1

Carilah:

1. Ukuran dari data (instansi dan fitur)
2. Tipe dari setiap fitur
3. Banyak nilai unik dari fitur yang bertipe kategorikal
4. Nilai minimum, maksimum, rata-rata, median, dan standar deviasi dari fitur nonkategorikal

```
# I.1 Kode di sini.

# 1. Ukuran dari data (instansi dan fitur)
df.shape
```

```
(145460, 22)
```

```
# 2. Tipe dari setiap fitur
df.dtypes
```

```
Location          object
MinTemp          float64
MaxTemp          float64
Rainfall         float64
Evaporation      float64
Sunshine         float64
WindGustDir       object
WindGustSpeed    float64
WindDir9am        object
WindDir3pm        object
WindSpeed9am     float64
WindSpeed3pm     float64
Humidity9am      float64
Humidity3pm      float64
Pressure9am      float64
Pressure3pm      float64
Cloud9am         float64
Cloud3pm         float64
Temp9am          float64
Temp3pm          float64
RainToday         object
RainTomorrow      object
dtype: object
```

```
# 3. Banyak nilai unik dari fitur yang bertipe kategorikal
df.select_dtypes(include=['object']).nunique()
```

```
Location        49
WindGustDir     16
WindDir9am      16
WindDir3pm      16
RainToday        2
RainTomorrow     2
dtype: int64
```

```
# 4. Nilai minimum, maksimum, rata-rata, median, dan standar deviasi
dari fitur nonkategorikal
df.select_dtypes(include=['float64']).describe()
```

```
{"column_count":16,"columns":
[{"dtype":"float64","name":"MinTemp","stats":{"histogram":
[{"bin_end":14389.85,"bin_start":-8.5,"count":7},
{"bin_end":28788.2,"bin_start":14389.85,"count":0},
{"bin_end":43186.55,"bin_start":28788.2,"count":0},
{"bin_end":57584.9,"bin_start":43186.55,"count":0},
{"bin_end":71983.25,"bin_start":57584.9,"count":0},
{"bin_end":86381.6,"bin_start":71983.25,"count":0},
{"bin_end":100779.95,"bin_start":86381.6,"count":0},
{"bin_end":115178.3,"bin_start":100779.95,"count":0},
{"bin_end":129576.65000000001,"bin_start":115178.3,"count":0},
{"bin_end":143975,"bin_start":129576.65000000001,"count":1}],"max":"14
3975.0","min":"-8.5","nan_count":0,"unique_count":8}},
{"dtype":"float64","name":"MaxTemp","stats":{"histogram":
[{"bin_end":14415.58,"bin_start":-4.8,"count":7},
{"bin_end":28835.96,"bin_start":14415.58,"count":0},
{"bin_end":43256.34,"bin_start":28835.96,"count":0},
{"bin_end":57676.719999999994,"bin_start":43256.34,"count":0},
{"bin_end":72097.09999999999,"bin_start":57676.719999999994,"count":0}
,{"bin_end":86517.48,"bin_start":72097.09999999999,"count":0},
{"bin_end":100937.85999999999,"bin_start":86517.48,"count":0},
{"bin_end":115358.23999999999,"bin_start":100937.85999999999,"count":0
},{"bin_end":129778.62,"bin_start":115358.23999999999,"count":0},
{"bin_end":144199,"bin_start":129778.62,"count":1}],"max":"144199.0","
min":"-4.8","nan_count":0,"unique_count":8}},
{"dtype":"float64","name":"Rainfall","stats":{"histogram":
[{"bin_end":14219.9,"bin_start":0,"count":7},
{"bin_end":28439.8,"bin_start":14219.9,"count":0},
{"bin_end":42659.7,"bin_start":28439.8,"count":0},
{"bin_end":56879.6,"bin_start":42659.7,"count":0},
{"bin_end":71099.5,"bin_start":56879.6,"count":0},
{"bin_end":85319.4,"bin_start":71099.5,"count":0},
{"bin_end":99539.3,"bin_start":85319.4,"count":0},
{"bin_end":113759.2,"bin_start":99539.3,"count":0},
{"bin_end":127979.09999999999,"bin_start":113759.2,"count":0},
{"bin_end":142199,"bin_start":127979.09999999999,"count":1}],"max":"14
2199.0","min":"0.0","nan_count":0,"unique_count":6}},
{"dtype":"float64","name":"Evaporation","stats":{"histogram":
[{"bin_end":8267,"bin_start":0,"count":7},
{"bin_end":16534,"bin_start":8267,"count":0},
{"bin_end":24801,"bin_start":16534,"count":0},
{"bin_end":33068,"bin_start":24801,"count":0},
{"bin_end":41335,"bin_start":33068,"count":0},
{"bin_end":49602,"bin_start":41335,"count":0},
{"bin_end":57869,"bin_start":49602,"count":0},
{"bin_end":66136,"bin_start":57869,"count":0},
{"bin_end":74403,"bin_start":66136,"count":0},
{"bin_end":82670,"bin_start":74403,"count":1}],"max":"82670.0","min":"
0.0","nan_count":0,"unique_count":8}},
{"dtype":"float64","name":"Sunshine","stats":{"histogram":
[{"bin_end":7562.5,"bin_start":0,"count":7},
{"bin_end":15125,"bin_start":7562.5,"count":0},
{"bin_end":22687.5,"bin_start":15125,"count":0},
{"bin_end":30250,"bin_start":22687.5,"count":0},
{"bin_end":37812.5,"bin_start":30250,"count":0},
{"bin_end":45375,"bin_start":37812.5,"count":0},
{"bin_end":52937.5,"bin_start":45375,"count":0},
{"bin_end":60500,"bin_start":52937.5,"count":0},
{"bin_end":68062.5,"bin_start":60500,"count":0},
{"bin_end":75625,"bin_start":68062.5,"count":1}],"max":"75625.0","min"
:"0.0","nan_count":0,"unique_count":8}},
{"dtype":"float64","name":"WindGustSpeed","stats":{"histogram":
[{"bin_end":13525.1,"bin_start":6,"count":7},
{"bin_end":27044.2,"bin_start":13525.1,"count":0},
{"bin_end":40563.3,"bin_start":27044.2,"count":0},
{"bin_end":54082.4,"bin_start":40563.3,"count":0},
{"bin_end":67601.5,"bin_start":54082.4,"count":0},
{"bin_end":81120.6,"bin_start":67601.5,"count":0},
{"bin_end":94639.7,"bin_start":81120.6,"count":0},
{"bin_end":108158.8,"bin_start":94639.7,"count":0},
{"bin_end":121677.90000000001,"bin_start":108158.8,"count":0},
{"bin_end":135197,"bin_start":121677.90000000001,"count":1}],"max":"13
5197.0","min":"6.0","nan_count":0,"unique_count":8}},
{"dtype":"float64","name":"WindSpeed9am","stats":{"histogram":
[{"bin_end":14369.3,"bin_start":0,"count":7},
{"bin_end":28738.6,"bin_start":14369.3,"count":0},
{"bin_end":43107.899999999994,"bin_start":28738.6,"count":0},
{"bin_end":57477.2,"bin_start":43107.899999999994,"count":0},
{"bin_end":71846.5,"bin_start":57477.2,"count":0},
{"bin_end":86215.79999999999,"bin_start":71846.5,"count":0},
{"bin_end":100585.09999999999,"bin_start":86215.79999999999,"count":0}
,{"bin_end":114954.4,"bin_start":100585.09999999999,"count":0},
{"bin_end":129323.7,"bin_start":114954.4,"count":0},
{"bin_end":143693,"bin_start":129323.7,"count":1}],"max":"143693.0","m
in":"0.0","nan_count":0,"unique_count":8}},
{"dtype":"float64","name":"WindSpeed3pm","stats":{"histogram":
[{"bin_end":14239.8,"bin_start":0,"count":7},
{"bin_end":28479.6,"bin_start":14239.8,"count":0},
{"bin_end":42719.399999999994,"bin_start":28479.6,"count":0},
{"bin_end":56959.2,"bin_start":42719.399999999994,"count":0},
{"bin_end":71199,"bin_start":56959.2,"count":0},
{"bin_end":85438.79999999999,"bin_start":71199,"count":0},
{"bin_end":99678.59999999999,"bin_start":85438.79999999999,"count":0},
{"bin_end":113918.4,"bin_start":99678.59999999999,"count":0},
{"bin_end":128158.2,"bin_start":113918.4,"count":0},
{"bin_end":142398,"bin_start":128158.2,"count":1}],"max":"142398.0","m
```

```
in":"0.0","nan_count":0,"unique_count":8}},
{"dtype":"float64","name":"Humidity9am","stats":{"histogram":
[{"bin_end":14280.6,"bin_start":0,"count":7},
{"bin_end":28561.2,"bin_start":14280.6,"count":0},
{"bin_end":42841.8,"bin_start":28561.2,"count":0},
{"bin_end":57122.4,"bin_start":42841.8,"count":0},
{"bin_end":71403,"bin_start":57122.4,"count":0},
{"bin_end":85683.6,"bin_start":71403,"count":0},
{"bin_end":99964.2,"bin_start":85683.6,"count":0},
{"bin_end":114244.8,"bin_start":99964.2,"count":0},
{"bin_end":128525.40000000001,"bin_start":114244.8,"count":0},
{"bin_end":142806,"bin_start":128525.40000000001,"count":1}],"max":"14
2806.0","min":"0.0","nan_count":0,"unique_count":8}},
{"dtype":"float64","name":"Humidity3pm","stats":{"histogram":
[{"bin_end":14095.3,"bin_start":0,"count":7},
{"bin_end":28190.6,"bin_start":14095.3,"count":0},
{"bin_end":42285.899999999994,"bin_start":28190.6,"count":0},
{"bin_end":56381.2,"bin_start":42285.899999999994,"count":0},
{"bin_end":70476.5,"bin_start":56381.2,"count":0},
{"bin_end":84571.79999999999,"bin_start":70476.5,"count":0},
{"bin_end":98667.09999999999,"bin_start":84571.79999999999,"count":0},
{"bin_end":112762.4,"bin_start":98667.09999999999,"count":0},
{"bin_end":126857.7,"bin_start":112762.4,"count":0},
{"bin_end":140953,"bin_start":126857.7,"count":1}],"max":"140953.0","m
in":"0.0","nan_count":0,"unique_count":8}},
{"dtype":"float64","name":"Pressure9am","stats":{"histogram":
[{"bin_end":13045.89587725877,"bin_start":7.106530287523087,"count":7}
,
{"bin_end":26084.685224230016,"bin_start":13045.89587725877,"count":0}
,
{"bin_end":39123.47457120127,"bin_start":26084.685224230016,"count":0}
,
{"bin_end":52162.26391817251,"bin_start":39123.47457120127,"count":0},
{"bin_end":65201.05326514375,"bin_start":52162.26391817251,"count":0},
{"bin_end":78239.84261211501,"bin_start":65201.05326514375,"count":0},
{"bin_end":91278.63195908625,"bin_start":78239.84261211501,"count":0},
{"bin_end":104317.4213060575,"bin_start":91278.63195908625,"count":0},
{"bin_end":117356.21065302874,"bin_start":104317.4213060575,"count":0}
,
{"bin_end":130395,"bin_start":117356.21065302874,"count":1}],"max":"13
0395.0","min":"7.106530287523087","nan_count":0,"unique_count":8}},
{"dtype":"float64","name":"Pressure3pm","stats":{"histogram":
[{"bin_end":13049.533672427351,"bin_start":7.037413808167775,"count":7
},
{"bin_end":26092.029931046534,"bin_start":13049.533672427351,"count":0
},
{"bin_end":39134.52618966572,"bin_start":26092.029931046534,"count":0}
,{"bin_end":52177.0224482849,"bin_start":39134.52618966572,"count":0},
{"bin_end":65219.518706904084,"bin_start":52177.0224482849,"count":0},
{"bin_end":78262.01496552327,"bin_start":65219.518706904084,"count":0}
```

```
,
{"bin_end":91304.51122414245,"bin_start":78262.01496552327,"count":0},
{"bin_end":104347.00748276163,"bin_start":91304.51122414245,"count":0}
,
{"bin_end":117389.50374138082,"bin_start":104347.00748276163,"count":0
},
{"bin_end":130432,"bin_start":117389.50374138082,"count":1}],"max":"13
0432.0","min":"7.037413808167775","nan_count":0,"unique_count":8}},
{"dtype":"float64","name":"Cloud9am","stats":{"histogram":
[{"bin_end":8957.2,"bin_start":0,"count":7},
{"bin_end":17914.4,"bin_start":8957.2,"count":0},
{"bin_end":26871.600000000002,"bin_start":17914.4,"count":0},
{"bin_end":35828.8,"bin_start":26871.600000000002,"count":0},
{"bin_end":44786,"bin_start":35828.8,"count":0},
{"bin_end":53743.200000000004,"bin_start":44786,"count":0},
{"bin_end":62700.40000000001,"bin_start":53743.200000000004,"count":0}
,{"bin_end":71657.6,"bin_start":62700.40000000001,"count":0},
{"bin_end":80614.8,"bin_start":71657.6,"count":0},
{"bin_end":89572,"bin_start":80614.8,"count":1}],"max":"89572.0","min"
:"0.0","nan_count":0,"unique_count":8}},
{"dtype":"float64","name":"Cloud3pm","stats":{"histogram":
[{"bin_end":8610.2,"bin_start":0,"count":7},
{"bin_end":17220.4,"bin_start":8610.2,"count":0},
{"bin_end":25830.600000000002,"bin_start":17220.4,"count":0},
{"bin_end":34440.8,"bin_start":25830.600000000002,"count":0},
{"bin_end":43051,"bin_start":34440.8,"count":0},
{"bin_end":51661.200000000004,"bin_start":43051,"count":0},
{"bin_end":60271.40000000001,"bin_start":51661.200000000004,"count":0}
,{"bin_end":68881.6,"bin_start":60271.40000000001,"count":0},
{"bin_end":77491.8,"bin_start":68881.6,"count":0},
{"bin_end":86102,"bin_start":77491.8,"count":1}],"max":"86102.0","min"
:"0.0","nan_count":0,"unique_count":8}},
{"dtype":"float64","name":"Temp9am","stats":{"histogram":
[{"bin_end":14362.82,"bin_start":-7.2,"count":7},
{"bin_end":28732.84,"bin_start":14362.82,"count":0},
{"bin_end":43102.86,"bin_start":28732.84,"count":0},
{"bin_end":57472.880000000005,"bin_start":43102.86,"count":0},
{"bin_end":71842.90000000001,"bin_start":57472.880000000005,"count":0}
,{"bin_end":86212.92,"bin_start":71842.90000000001,"count":0},
{"bin_end":100582.94,"bin_start":86212.92,"count":0},
{"bin_end":114952.96,"bin_start":100582.94,"count":0},
{"bin_end":129322.98000000001,"bin_start":114952.96,"count":0},
{"bin_end":143693,"bin_start":129322.98000000001,"count":1}],"max":"14
3693.0","min":"-7.2","nan_count":0,"unique_count":8}},
{"dtype":"float64","name":"Temp3pm","stats":{"histogram":
[{"bin_end":14180.24,"bin_start":-5.4,"count":7},
{"bin_end":28365.879999999997,"bin_start":14180.24,"count":0},
{"bin_end":42551.52,"bin_start":28365.879999999997,"count":0},
{"bin_end":56737.159999999996,"bin_start":42551.52,"count":0},
{"bin_end":70922.8,"bin_start":56737.159999999996,"count":0},
{"bin_end":85108.44,"bin_start":70922.8,"count":0},
{"bin_end":99294.08,"bin_start":85108.44,"count":0},
{"bin_end":113479.72,"bin_start":99294.08,"count":0},
{"bin_end":127665.36,"bin_start":113479.72,"count":0},
{"bin_end":141851,"bin_start":127665.36,"count":1}],"max":"141851.0","
min":"-5.4","nan_count":0,"unique_count":8}},
{"dtype":"object","name":"_deepnote_index_column"}],"row_count":8,"row
s":
[{"Cloud3pm":"86102.0","Cloud9am":"89572.0","Evaporation":"82670.0","H
umidity3pm":"140953.0","Humidity9am":"142806.0","MaxTemp":"144199.0","
MinTemp":"143975.0","Pressure3pm":"130432.0","Pressure9am":"130395.0",
"Rainfall":"142199.0","Sunshine":"75625.0","Temp3pm":"141851.0","Temp9
am":"143693.0","WindGustSpeed":"135197.0","WindSpeed3pm":"142398.0","W
indSpeed9am":"143693.0","_deepnote_index_column":"count"},
{"Cloud3pm":"4.509930082924903","Cloud9am":"4.4474612602152455","Evapo
ration":"5.468231522922462","Humidity3pm":"51.5391158755046","Humidity
9am":"68.88083133761887","MaxTemp":"23.22134827564685","MinTemp":"12.1
9403438096892","Pressure3pm":"1015.2558888309618","Pressure9am":"1017.
6499397983052","Rainfall":"2.3609181499166656","Sunshine":"7.611177520
6611565","Temp3pm":"21.68339031800974","Temp9am":"16.990631415587398",
"WindGustSpeed":"40.03523007167319","WindSpeed3pm":"18.662656778887342
","WindSpeed9am":"14.043425914971502","_deepnote_index_column":"mean"}
,
{"Cloud3pm":"2.7203573103324588","Cloud9am":"2.887158853517239","Evapo
ration":"4.1937040941517845","Humidity3pm":"20.795901656021154","Humid
ity9am":"19.02916445184414","MaxTemp":"7.119048845979134","MinTemp":"6
.398494975663227","Pressure3pm":"7.037413808167775","Pressure9am":"7.1
06530287523087","Rainfall":"8.478059737726372","Sunshine":"3.785482964
7401664","Temp3pm":"6.936650460035525","Temp9am":"6.488753140877267","
WindGustSpeed":"13.607062267381362","WindSpeed3pm":"8.809800021251492"
,"WindSpeed9am":"8.915375322679528","_deepnote_index_column":"std"},
{"Cloud3pm":"0.0","Cloud9am":"0.0","Evaporation":"0.0","Humidity3pm":"
0.0","Humidity9am":"0.0","MaxTemp":"-4.8","MinTemp":"-
8.5","Pressure3pm":"977.1","Pressure9am":"980.5","Rainfall":"0.0","Sun
shine":"0.0","Temp3pm":"-5.4","Temp9am":"-
7.2","WindGustSpeed":"6.0","WindSpeed3pm":"0.0","WindSpeed9am":"0.0","
_deepnote_index_column":"min"},
{"Cloud3pm":"2.0","Cloud9am":"1.0","Evaporation":"2.6","Humidity3pm":"
37.0","Humidity9am":"57.0","MaxTemp":"17.9","MinTemp":"7.6","Pressure3
pm":"1010.4","Pressure9am":"1012.9","Rainfall":"0.0","Sunshine":"4.8",
"Temp3pm":"16.6","Temp9am":"12.3","WindGustSpeed":"31.0","WindSpeed3pm
":"13.0","WindSpeed9am":"7.0","_deepnote_index_column":"25%"},
{"Cloud3pm":"5.0","Cloud9am":"5.0","Evaporation":"4.8","Humidity3pm":"
52.0","Humidity9am":"70.0","MaxTemp":"22.6","MinTemp":"12.0","Pressure
3pm":"1015.2","Pressure9am":"1017.6","Rainfall":"0.0","Sunshine":"8.4"
,"Temp3pm":"21.1","Temp9am":"16.7","WindGustSpeed":"39.0","WindSpeed3p
m":"19.0","WindSpeed9am":"13.0","_deepnote_index_column":"50%"},
{"Cloud3pm":"7.0","Cloud9am":"7.0","Evaporation":"7.4","Humidity3pm":"
66.0","Humidity9am":"83.0","MaxTemp":"28.2","MinTemp":"16.9","Pressure
3pm":"1020.0","Pressure9am":"1022.4","Rainfall":"0.8","Sunshine":"10.6
```

","Temp3pm":"26.4","Temp9am":"21.6","WindGustSpeed":"48.0","WindSpeed3
pm":"24.0","WindSpeed9am":"19.0","_deepnote_index_column":"75%"},
{"Cloud3pm":"9.0","Cloud9am":"9.0","Evaporation":"145.0","Humidity3pm"
:"100.0","Humidity9am":"100.0","MaxTemp":"48.1","MinTemp":"33.9","Pres
sure3pm":"1039.6","Pressure9am":"1041.0","Rainfall":"371.0","Sunshine"
:"14.5","Temp3pm":"46.7","Temp9am":"40.2","WindGustSpeed":"135.0","Win
dSpeed3pm":"87.0","WindSpeed9am":"130.0","_deepnote_index_column":"max
"}]}

I.2

Carilah:

1. Nilai hilang (*missing*) dari setiap fitur
2. Nilai pencilan (*outlier*) dari setiap fitur

```
# I.2 Kode di sini.

# 1. Nilai hilang (missing) dari setiap fitur
df.isnull().sum()
```

```
Location             0
MinTemp           1485
MaxTemp           1261
Rainfall          3261
Evaporation      62790
Sunshine         69835
WindGustDir      10326
WindGustSpeed    10263
WindDir9am       10566
WindDir3pm        4228
WindSpeed9am      1767
WindSpeed3pm      3062
Humidity9am       2654
Humidity3pm       4507
Pressure9am      15065
Pressure3pm      15028
Cloud9am         55888
Cloud3pm         59358
Temp9am           1767
Temp3pm           3609
RainToday         3261
RainTomorrow      3267
dtype: int64
```

```
# 2. Nilai pencilan (outlier) dari setiap fitur

q1 = df.select_dtypes(include=['float64']).quantile(0.25)
q3 = df.select_dtypes(include=['float64']).quantile(0.75)
iqr = q3 - q1
```

```
minimum_range = q1 - 1.5 * iqr
maximum_range = q3 + 1.5 * iqr

# Nilai outlier cuman bisa di type number
((df.select_dtypes(include=['float64']) < minimum_range) |
(df.select_dtypes(include=['float64']) > maximum_range)).sum()
```

```
MinTemp              54
MaxTemp             489
Rainfall          25578
Evaporation        1995
Sunshine              0
WindGustSpeed      3092
WindSpeed9am       1817
WindSpeed3pm       2523
Humidity9am        1425
Humidity3pm           0
Pressure9am        1191
Pressure3pm         919
Cloud9am              0
Cloud3pm              0
Temp9am             262
Temp3pm             764
dtype: int64
```

I.3

Lakukan:

1. Pencarian korelasi antarfitur
2. Visualisasi distribusi setiap fitur (kategorikal dan kontinu)
3. Visualisasi distribusi setiap fitur per target (`RainTomorrow`)

```
# I.3 Kode di sini.

# 1. Pencarian korelasi antar fitur
df.select_dtypes(include=['float64']).corr()
```

```
{"column_count":16,"columns":
[{"dtype":"float64","name":"MinTemp","stats":{"histogram":
[{"bin_end":-0.31516255327547643,"bin_start":-
0.46129172586164047,"count":2},{"bin_end":-
0.1690333806893124,"bin_start":-0.31516255327547643,"count":1},
{"bin_end":-2.290420810314836e-2,"bin_start":-
0.1690333806893124,"count":0},
{"bin_end":0.12322496448301568,"bin_start":-2.290420810314836e-
2,"count":5},
{"bin_end":0.26935413706917966,"bin_start":0.12322496448301568,"count"
:3},
{"bin_end":0.41548330965534375,"bin_start":0.26935413706917966,"count"
:0},
{"bin_end":0.5616124822415078,"bin_start":0.41548330965534375,"count":
1},
{"bin_end":0.7077416548276718,"bin_start":0.5616124822415078,"count":0
},
{"bin_end":0.8538708274138358,"bin_start":0.7077416548276718,"count":2
},
{"bin_end":1,"bin_start":0.8538708274138358,"count":2}],"max":"1.0","m
in":"-0.46129172586164047","nan_count":0,"unique_count":16}},
{"dtype":"float64","name":"MaxTemp","stats":{"histogram":[{"bin_end":-
0.3579698779985809,"bin_start":-0.5088554199984232,"count":3},
{"bin_end":-0.20708433599873854,"bin_start":-
0.3579698779985809,"count":3},{"bin_end":-5.619879399889621e-
2,"bin_start":-0.20708433599873854,"count":1},
{"bin_end":9.468674800094612e-2,"bin_start":-5.619879399889621e-
2,"count":3},
{"bin_end":0.2455722900007884,"bin_start":9.468674800094612e-
2,"count":0},
{"bin_end":0.3964578320006308,"bin_start":0.2455722900007884,"count":0
},
{"bin_end":0.5473433740004732,"bin_start":0.3964578320006308,"count":1
},
{"bin_end":0.6982289160003154,"bin_start":0.5473433740004732,"count":1
},
{"bin_end":0.8491144580001577,"bin_start":0.6982289160003154,"count":1
},
{"bin_end":1,"bin_start":0.8491144580001577,"count":3}],"max":"1.0","m
in":"-0.5088554199984232","nan_count":0,"unique_count":16}},
{"dtype":"float64","name":"Rainfall","stats":{"histogram":
[{"bin_end":-0.10479371798370837,"bin_start":-
0.22754857553745375,"count":3},{"bin_end":1.7961139570037005e-
2,"bin_start":-0.10479371798370837,"count":4},
{"bin_end":0.14071599712378238,"bin_start":1.7961139570037005e-
2,"count":4},
{"bin_end":0.26347085467752773,"bin_start":0.14071599712378238,"count"
:4},
{"bin_end":0.3862257122312731,"bin_start":0.26347085467752773,"count":
0},
{"bin_end":0.5089805697850185,"bin_start":0.3862257122312731,"count":0
},
{"bin_end":0.6317354273387639,"bin_start":0.5089805697850185,"count":0
},
{"bin_end":0.7544902848925092,"bin_start":0.6317354273387639,"count":0
},
{"bin_end":0.8772451424462546,"bin_start":0.7544902848925092,"count":0
},
{"bin_end":1,"bin_start":0.8772451424462546,"count":1}],"max":"1.0","m
in":"-0.22754857553745375","nan_count":0,"unique_count":16}},
{"dtype":"float64","name":"Evaporation","stats":{"histogram":
[{"bin_end":-0.35368302751716,"bin_start":-
0.5040922527968444,"count":2},{"bin_end":-
```

```
0.2032738022374756,"bin_start":-0.35368302751716,"count":2},
{"bin_end":-5.2864576957791176e-2,"bin_start":-
0.2032738022374756,"count":3},{"bin_end":9.754464832189325e-
2,"bin_start":-5.2864576957791176e-2,"count":0},
{"bin_end":0.24795387360157772,"bin_start":9.754464832189325e-
2,"count":3},
{"bin_end":0.3983630988812621,"bin_start":0.24795387360157772,"count":
1},
{"bin_end":0.5487723241609465,"bin_start":0.3983630988812621,"count":2
},
{"bin_end":0.6991815494406309,"bin_start":0.5487723241609465,"count":2
},
{"bin_end":0.8495907747203154,"bin_start":0.6991815494406309,"count":0
},
{"bin_end":1,"bin_start":0.8495907747203154,"count":1}],"max":"1.0","m
in":"-0.5040922527968444","nan_count":0,"unique_count":16}},
{"dtype":"float64","name":"Sunshine","stats":{"histogram":
[{"bin_end":-0.5335366703379754,"bin_start":-
0.7039296337088615,"count":3},{"bin_end":-
0.36314370696708925,"bin_start":-0.5335366703379754,"count":1},
{"bin_end":-0.19275074359620314,"bin_start":-
0.36314370696708925,"count":1},{"bin_end":-2.235778022531698e-
2,"bin_start":-0.19275074359620314,"count":1},
{"bin_end":0.14803518314556918,"bin_start":-2.235778022531698e-
2,"count":5},
{"bin_end":0.31842814651645523,"bin_start":0.14803518314556918,"count"
:1},
{"bin_end":0.4888211098873415,"bin_start":0.31842814651645523,"count":
2},
{"bin_end":0.6592140732582276,"bin_start":0.4888211098873415,"count":1
},
{"bin_end":0.8296070366291136,"bin_start":0.6592140732582276,"count":0
},
{"bin_end":1,"bin_start":0.8296070366291136,"count":1}],"max":"1.0","m
in":"-0.7039296337088615","nan_count":0,"unique_count":16}},
{"dtype":"float64","name":"WindGustSpeed","stats":{"histogram":
[{"bin_end":-0.31286979622109845,"bin_start":-
0.45874421802344273,"count":2},{"bin_end":-
0.16699537441875417,"bin_start":-0.31286979622109845,"count":1},
{"bin_end":-2.1120952616409894e-2,"bin_start":-
0.16699537441875417,"count":2},
{"bin_end":0.12475346918593438,"bin_start":-2.1120952616409894e-
2,"count":4},
{"bin_end":0.2706278909882786,"bin_start":0.12475346918593438,"count":
4},
{"bin_end":0.41650231279062294,"bin_start":0.2706278909882786,"count":
0},
{"bin_end":0.5623767345929673,"bin_start":0.41650231279062294,"count":
0},
{"bin_end":0.7082511563953116,"bin_start":0.5623767345929673,"count":2
```

```
},
{"bin_end":0.8541255781976558,"bin_start":0.7082511563953116,"count":0
},
{"bin_end":1,"bin_start":0.8541255781976558,"count":1}],"max":"1.0","m
in":"-0.45874421802344273","nan_count":0,"unique_count":16}},
{"dtype":"float64","name":"WindSpeed9am","stats":{"histogram":
[{"bin_end":-0.14377254923108654,"bin_start":-
0.2708583880345406,"count":3},{"bin_end":-1.668671042763248e-
2,"bin_start":-0.14377254923108654,"count":1},
{"bin_end":0.11039912837582161,"bin_start":-1.668671042763248e-
2,"count":6},
{"bin_end":0.23748496717927564,"bin_start":0.11039912837582161,"count"
:3},
{"bin_end":0.3645708059827297,"bin_start":0.23748496717927564,"count":
0},
{"bin_end":0.4916566447861838,"bin_start":0.3645708059827297,"count":0
},
{"bin_end":0.6187424835896378,"bin_start":0.4916566447861838,"count":2
},
{"bin_end":0.7458283223930919,"bin_start":0.6187424835896378,"count":0
},
{"bin_end":0.8729141611965461,"bin_start":0.7458283223930919,"count":0
},
{"bin_end":1,"bin_start":0.8729141611965461,"count":1}],"max":"1.0","m
in":"-0.2708583880345406","nan_count":0,"unique_count":16}},
{"dtype":"float64","name":"WindSpeed3pm","stats":{"histogram":
[{"bin_end":-0.16671605089667244,"bin_start":-
0.2963511676629694,"count":2},{"bin_end":-3.70809341303755e-
2,"bin_start":-0.16671605089667244,"count":1},
{"bin_end":9.255418263592147e-2,"bin_start":-3.70809341303755e-
2,"count":7},
{"bin_end":0.22218929940221838,"bin_start":9.255418263592147e-
2,"count":3},
{"bin_end":0.3518244161685153,"bin_start":0.22218929940221838,"count":
0},
{"bin_end":0.4814595329348123,"bin_start":0.3518244161685153,"count":0
},
{"bin_end":0.6110946497011092,"bin_start":0.4814595329348123,"count":1
},
{"bin_end":0.7407297664674062,"bin_start":0.6110946497011092,"count":1
},
{"bin_end":0.8703648832337032,"bin_start":0.7407297664674062,"count":0
},
{"bin_end":1,"bin_start":0.8703648832337032,"count":1}],"max":"1.0","m
in":"-0.2963511676629694","nan_count":0,"unique_count":16}},
{"dtype":"float64","name":"Humidity9am","stats":{"histogram":
[{"bin_end":-0.35369882452457346,"bin_start":-
0.5041098050273038,"count":5},{"bin_end":-
0.20328784402184308,"bin_start":-0.35369882452457346,"count":3},
{"bin_end":-5.287686351911269e-2,"bin_start":-
```

0.20328784402184308,"count":1},{"bin_end":9.753411698361769e-
2,"bin_start":-5.287686351911269e-2,"count":0},
{"bin_end":0.24794509748634808,"bin_start":9.753411698361769e-
2,"count":3},
{"bin_end":0.39835607798907846,"bin_start":0.24794509748634808,"count"
:1},
{"bin_end":0.5487670584918087,"bin_start":0.39835607798907846,"count":
1},
{"bin_end":0.6991780389945392,"bin_start":0.5487670584918087,"count":1
},
{"bin_end":0.8495890194972697,"bin_start":0.6991780389945392,"count":0
},
{"bin_end":1,"bin_start":0.8495890194972697,"count":1}],"max":"1.0","m
in":"-0.5041098050273038","nan_count":0,"unique_count":16}},
{"dtype":"float64","name":"Humidity3pm","stats":{"histogram":
[{"bin_end":-0.46621732398507154,"bin_start":-
0.6291303599834128,"count":3},{"bin_end":-
0.30330428798673026,"bin_start":-0.46621732398507154,"count":1},
{"bin_end":-0.14039125198838898,"bin_start":-
0.30330428798673026,"count":1},{"bin_end":2.2521784009952306e-
2,"bin_start":-0.14039125198838898,"count":5},
{"bin_end":0.1854348200082936,"bin_start":2.2521784009952306e-
2,"count":1},
{"bin_end":0.34834785600663487,"bin_start":0.1854348200082936,"count":
1},
{"bin_end":0.5112608920049762,"bin_start":0.34834785600663487,"count":
0},
{"bin_end":0.6741739280033174,"bin_start":0.5112608920049762,"count":3
},
{"bin_end":0.8370869640016587,"bin_start":0.6741739280033174,"count":0
},
{"bin_end":1,"bin_start":0.8370869640016587,"count":1}],"max":"1.0","m
in":"-0.6291303599834128","nan_count":0,"unique_count":16}},
{"dtype":"float64","name":"Pressure9am","stats":{"histogram":
[{"bin_end":-0.31286979622109845,"bin_start":-
0.45874421802344273,"count":4},{"bin_end":-
0.16699537441875417,"bin_start":-0.31286979622109845,"count":5},
{"bin_end":-2.1120952616409894e-2,"bin_start":-
0.16699537441875417,"count":3},
{"bin_end":0.12475346918593438,"bin_start":-2.1120952616409894e-
2,"count":1},
{"bin_end":0.2706278909882786,"bin_start":0.12475346918593438,"count":
1},
{"bin_end":0.41650231279062294,"bin_start":0.2706278909882786,"count":
0},
{"bin_end":0.5623767345929673,"bin_start":0.41650231279062294,"count":
0},
{"bin_end":0.7082511563953116,"bin_start":0.5623767345929673,"count":0
},
{"bin_end":0.8541255781976558,"bin_start":0.7082511563953116,"count":0

```
},
{"bin_end":1,"bin_start":0.8541255781976558,"count":2}],"max":"1.0","m
in":"-0.45874421802344273","nan_count":0,"unique_count":16}},
{"dtype":"float64","name":"Pressure3pm","stats":{"histogram":
[{"bin_end":-0.3231681599433327,"bin_start":-
0.4701868443814808,"count":5},{"bin_end":-
0.17614947550518462,"bin_start":-0.3231681599433327,"count":2},
{"bin_end":-2.9130791067036532e-2,"bin_start":-
0.17614947550518462,"count":4},
{"bin_end":0.11788789337111155,"bin_start":-2.9130791067036532e-
2,"count":2},
{"bin_end":0.2649065778092596,"bin_start":0.11788789337111155,"count":
1},
{"bin_end":0.4119252622474077,"bin_start":0.2649065778092596,"count":0
},
{"bin_end":0.5589439466855559,"bin_start":0.4119252622474077,"count":0
},
{"bin_end":0.7059626311237039,"bin_start":0.5589439466855559,"count":0
},
{"bin_end":0.852981315561852,"bin_start":0.7059626311237039,"count":0}
,
{"bin_end":1,"bin_start":0.852981315561852,"count":2}],"max":"1.0","mi
n":"-0.4701868443814808","nan_count":0,"unique_count":16}},
{"dtype":"float64","name":"Cloud9am","stats":{"histogram":
[{"bin_end":-0.5077909787349589,"bin_start":-
0.6753233097055099,"count":1},{"bin_end":-
0.34025864776440795,"bin_start":-0.5077909787349589,"count":0},
{"bin_end":-0.17272631679385697,"bin_start":-
0.34025864776440795,"count":3},{"bin_end":-5.193985823305991e-
3,"bin_start":-0.17272631679385697,"count":3},
{"bin_end":0.162338345147245,"bin_start":-5.193985823305991e-
3,"count":4},
{"bin_end":0.32987067611779597,"bin_start":0.162338345147245,"count":1
},
{"bin_end":0.49740300708834695,"bin_start":0.32987067611779597,"count"
:1},
{"bin_end":0.6649353380588979,"bin_start":0.49740300708834695,"count":
2},
{"bin_end":0.8324676690294489,"bin_start":0.6649353380588979,"count":0
},
{"bin_end":1,"bin_start":0.8324676690294489,"count":1}],"max":"1.0","m
in":"-0.6753233097055099","nan_count":0,"unique_count":16}},
{"dtype":"float64","name":"Cloud3pm","stats":{"histogram":
[{"bin_end":-0.5335366703379754,"bin_start":-
0.7039296337088615,"count":1},{"bin_end":-
0.36314370696708925,"bin_start":-0.5335366703379754,"count":0},
{"bin_end":-0.19275074359620314,"bin_start":-
0.36314370696708925,"count":2},{"bin_end":-2.235778022531698e-
2,"bin_start":-0.19275074359620314,"count":4},
{"bin_end":0.14803518314556918,"bin_start":-2.235778022531698e-
```

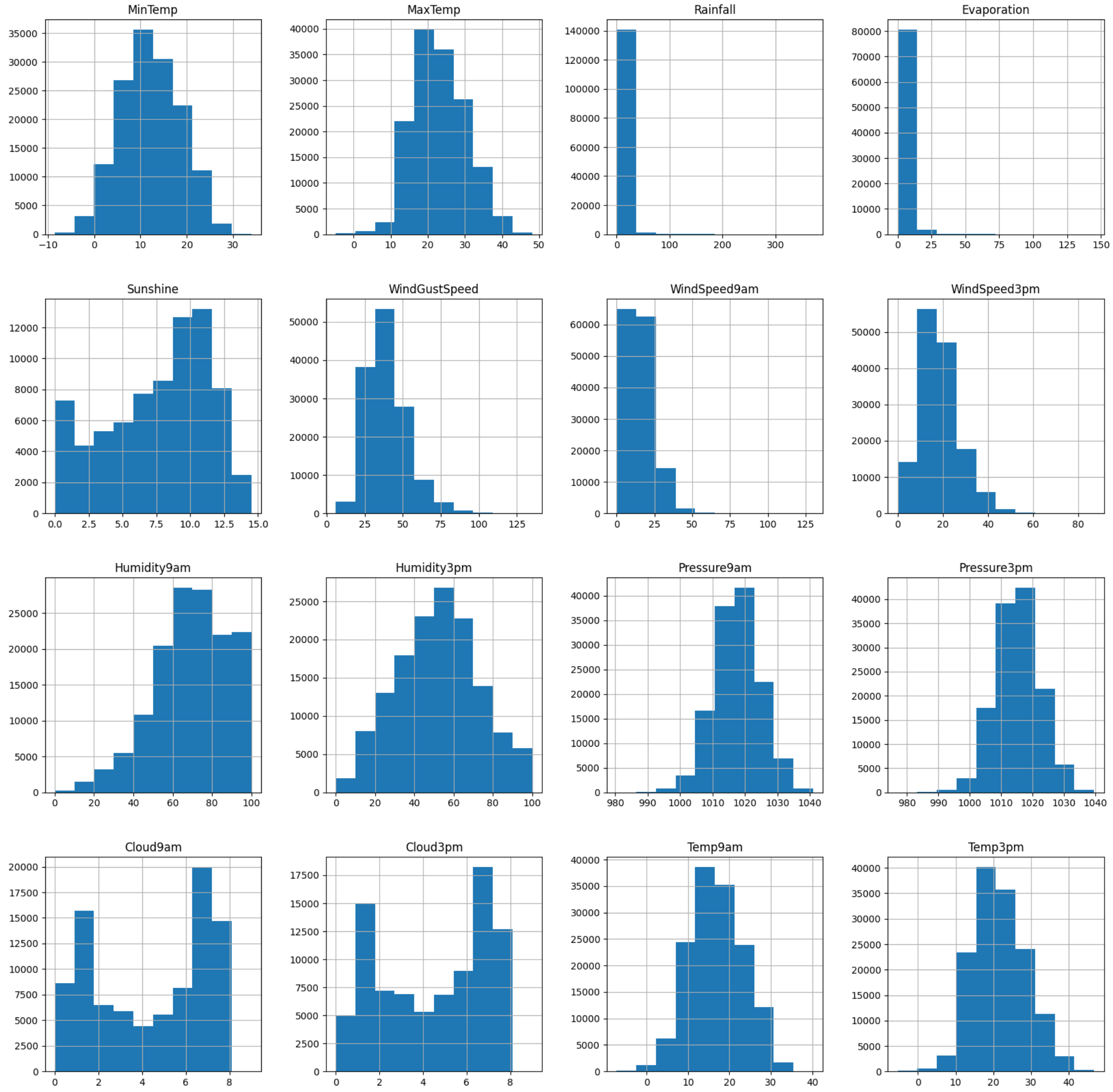
```
2,"count":4},
{"bin_end":0.31842814651645523,"bin_start":0.14803518314556918,"count"
:1},
{"bin_end":0.4888211098873415,"bin_start":0.31842814651645523,"count":
1},
{"bin_end":0.6592140732582276,"bin_start":0.4888211098873415,"count":2
},
{"bin_end":0.8296070366291136,"bin_start":0.6592140732582276,"count":0
},
{"bin_end":1,"bin_start":0.8296070366291136,"count":1}],"max":"1.0","m
in":"-0.7039296337088615","nan_count":0,"unique_count":16}},
{"dtype":"float64","name":"Temp9am","stats":{"histogram":[{"bin_end":-
0.3242185459849106,"bin_start":-0.47135393998323405,"count":3},
{"bin_end":-0.17708315198658725,"bin_start":-
0.3242185459849106,"count":1},{"bin_end":-2.994775798826388e-
2,"bin_start":-0.17708315198658725,"count":2},
{"bin_end":0.11718763601005955,"bin_start":-2.994775798826388e-
2,"count":1},
{"bin_end":0.264323030008383,"bin_start":0.11718763601005955,"count":3
},
{"bin_end":0.4114584240067063,"bin_start":0.264323030008383,"count":1}
,
{"bin_end":0.5585938180050298,"bin_start":0.4114584240067063,"count":1
},
{"bin_end":0.7057292120033531,"bin_start":0.5585938180050298,"count":0
},
{"bin_end":0.852864606001676,"bin_start":0.7057292120033531,"count":0
},
{"bin_end":1,"bin_start":0.8528646060016765,"count":4}],"max":"1.0","m
in":"-0.47135393998323405","nan_count":0,"unique_count":16}},
{"dtype":"float64","name":"Temp3pm","stats":{"histogram":[{"bin_end":-
0.4020569024944216,"bin_start":-0.5578410027715796,"count":2},
{"bin_end":-0.2462728022172636,"bin_start":-
0.4020569024944216,"count":4},{"bin_end":-9.048870194010561e-
2,"bin_start":-0.2462728022172636,"count":0},
{"bin_end":6.529539833705233e-2,"bin_start":-9.048870194010561e-
2,"count":4},
{"bin_end":0.22107949861421028,"bin_start":6.529539833705233e-
2,"count":0},
{"bin_end":0.37686359889136833,"bin_start":0.22107949861421028,"count"
:0},
{"bin_end":0.5326476991685262,"bin_start":0.37686359889136833,"count":
1},
{"bin_end":0.6884317994456842,"bin_start":0.5326476991685262,"count":1
},
{"bin_end":0.8442158997228423,"bin_start":0.6884317994456842,"count":1
},
{"bin_end":1,"bin_start":0.8442158997228423,"count":3}],"max":"1.0","m
in":"-0.5578410027715796","nan_count":0,"unique_count":16}},
{"dtype":"object","name":"_deepnote_index_column"}],"row_count":16,"ro
```

```
ws":
[{"Cloud3pm":"0.021604685038888648","Cloud9am":"0.0787538283427066","E
vaporation":"0.4669925569214256","Humidity3pm":"0.006088664224423575",
"Humidity9am":"-
0.23289924559073164","MaxTemp":"0.736554879947358","MinTemp":"1.0","Pr
essure3pm":"-0.46129172586164047","Pressure9am":"-
0.4509699734544065","Rainfall":"0.10393795340859532","Sunshine":"0.072
58586808397571","Temp3pm":"0.7089055392798455","Temp9am":"0.9018205621
778996","WindGustSpeed":"0.1774145260976753","WindSpeed3pm":"0.1751726
3108767825","WindSpeed9am":"0.17506364073869057","_deepnote_index_colu
mn":"MinTemp"},{"Cloud3pm":"-0.27792100295511346","Cloud9am":"-
0.2893701949239778","Evaporation":"0.5879322271019112","Humidity3pm":"
-0.5088554199984232","Humidity9am":"-
0.5041098050273038","MaxTemp":"1.0","MinTemp":"0.736554879947358","Pre
ssure3pm":"-0.4271668835749745","Pressure9am":"-
0.33206130275331036","Rainfall":"-
0.07499209497993575","Sunshine":"0.4701557064131683","Temp3pm":"0.9845
028310904481","Temp9am":"0.8872096927504229","WindGustSpeed":"0.067614
77569098283","WindSpeed3pm":"0.050300397194580494","WindSpeed9am":"0.0
1444964588964599","_deepnote_index_column":"MaxTemp"},
{"Cloud3pm":"0.17240289926478258","Cloud9am":"0.19852828605297007","Ev
aporation":"-
0.06435117292604013","Humidity3pm":"0.2557546575244556","Humidity9am":
"0.22440497587716363","MaxTemp":"-
0.07499209497993575","MinTemp":"0.10393795340859532","Pressure3pm":"-
0.1265338436909568","Pressure9am":"-
0.16815364149847833","Rainfall":"1.0","Sunshine":"-
0.22754857553745375","Temp3pm":"-
0.0796570726164772","Temp9am":"0.011191601489034678","WindGustSpeed":"
0.13365939753988504","WindSpeed3pm":"0.05788711972355656","WindSpeed9a
m":"0.08733769468151066","_deepnote_index_column":"Rainfall"},
{"Cloud3pm":"-0.18261841213978056","Cloud9am":"-
0.18379325900830737","Evaporation":"1.0","Humidity3pm":"-
0.39024259071039363","Humidity9am":"-
0.5040922527968444","MaxTemp":"0.5879322271019112","MinTemp":"0.466992
5569214256","Pressure3pm":"-0.2935805340204502","Pressure9am":"-
0.2703621404798875","Rainfall":"-
0.06435117292604013","Sunshine":"0.36560215022270703","Temp3pm":"0.572
8932109454876","Temp9am":"0.5451152126470746","WindGustSpeed":"0.20302
143800515285","WindSpeed3pm":"0.12940001633420806","WindSpeed9am":"0.1
9308448307770443","_deepnote_index_column":"Evaporation"},
{"Cloud3pm":"-0.7039296337088615","Cloud9am":"-
0.6753233097055099","Evaporation":"0.36560215022270703","Humidity3pm":
"-0.6291303599834128","Humidity9am":"-
0.49081885139829456","MaxTemp":"0.4701557064131683","MinTemp":"0.07258
586808397571","Pressure3pm":"-
0.019719031024735247","Pressure9am":"0.041969831552005885","Rainfall":
"-
0.22754857553745375","Sunshine":"1.0","Temp3pm":"0.4905006212075918","
Temp9am":"0.29118794175836554","WindGustSpeed":"-
```

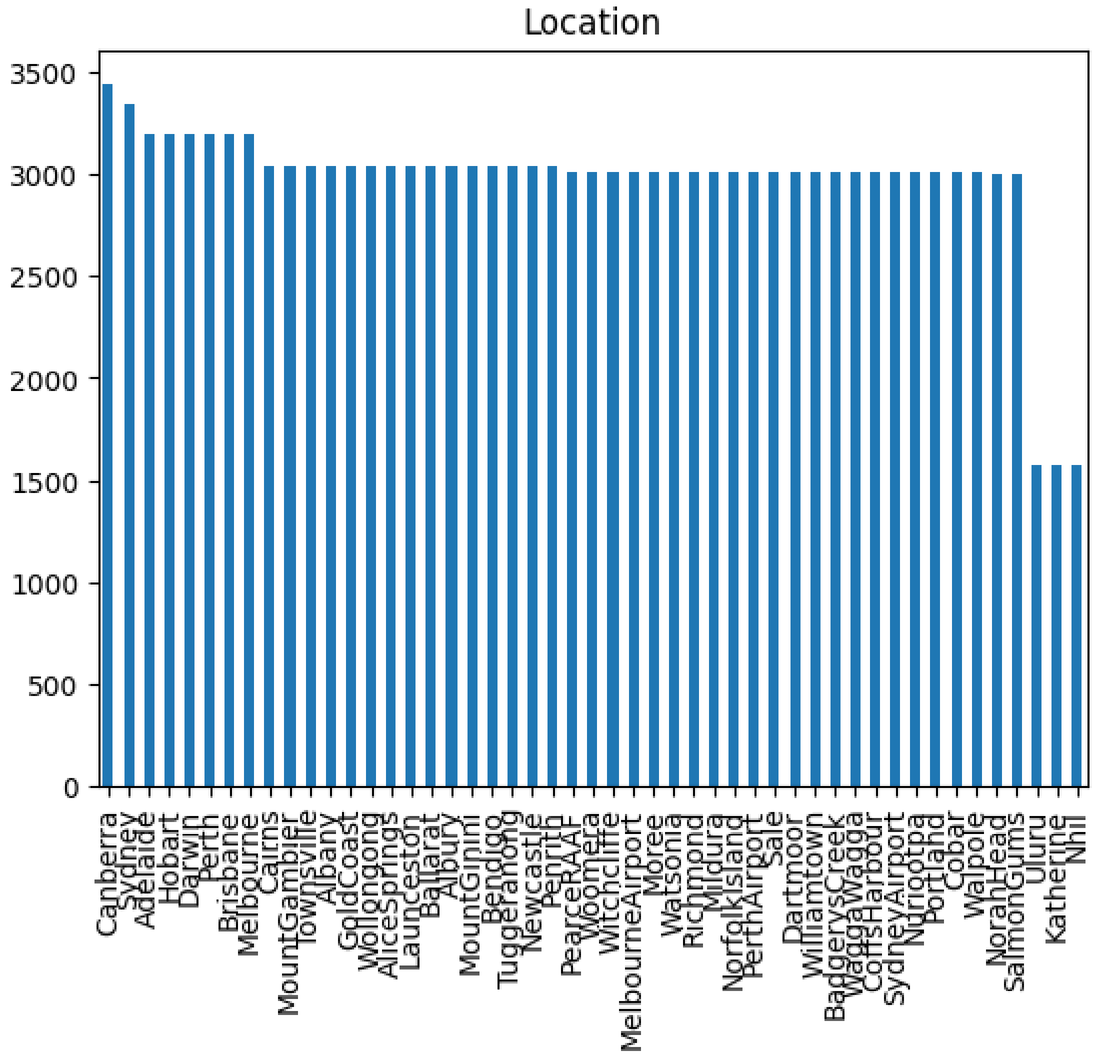
```
0.03475045685199232","WindSpeed3pm":"0.053834024723800285","WindSpeed9
am":"0.005498574825312206","_deepnote_index_column":"Sunshine"},
{"Cloud3pm":"0.1091675240686803","Cloud9am":"0.0717355819801594","Evap
oration":"0.20302143800515285","Humidity3pm":"-
0.026327121063789812","Humidity9am":"-
0.21506968756795294","MaxTemp":"0.06761477569098283","MinTemp":"0.1774
145260976753","Pressure3pm":"-0.41374904863463996","Pressure9am":"-
0.45874421802344273","Rainfall":"0.13365939753988504","Sunshine":"-
0.03475045685199232","Temp3pm":"0.03274777925600875","Temp9am":"0.1501
5009761950066","WindGustSpeed":"1.0","WindSpeed3pm":"0.686306743470663
","WindSpeed9am":"0.6053032893809175","_deepnote_index_column":"WindGu
stSpeed"},
{"Cloud3pm":"0.054638619717694493","Cloud9am":"0.02511225384648038","E
vaporation":"0.19308448307770443","Humidity3pm":"-
0.0316138259484374","Humidity9am":"-
0.2708583880345406","MaxTemp":"0.014449645889645 99","MinTemp":"0.17506
364073869057","Pressure3pm":"-0.17581712628835547","Pressure9am":"-
0.22874305315934174","Rainfall":"0.08733769468151066","Sunshine":"0.00
5498574825312206","Temp3pm":"0.004568833457939502","Temp9am":"0.128544
66322246128","WindGustSpeed":"0.6053032893809175","WindSpeed3pm":"0.51
9546834492561","WindSpeed9am":"1.0","_deepnote_index_column":"WindSpee
d9am"},
{"Cloud3pm":"0.025395671222374603","Cloud9am":"0.05333665419063742","E
vaporation":"0.12940001633420806","Humidity3pm":"0.016432471350722434"
,"Humidity9am":"-
0.145525126140691","MaxTemp":"0.05030039719458049 4","MinTemp":"0.17517
263108767825","Pressure3pm":"-0.25543861805056034","Pressure9am":"-
0.2963511676629694","Rainfall":"0.05788711972355656","Sunshine":"0.053
834024723800285","Temp3pm":"0.027777592344805757","Temp9am":"0.1630299
6588669325","WindGustSpeed":"0.686306743470663","WindSpeed3pm":"1.0","
WindSpeed9am":"0.519546834492561","_deepnote_index_column":"WindSpeed3
pm"},
{"Cloud3pm":"0.3573261614347828","Cloud9am":"0.4522969906998 3464","Eva
poration":"-
0.5040922527968444","Humidity3pm":"0.6669485686855433","Humidity9am":"
1.0","MaxTemp":"-0.5041098050273038","MinTemp":"-
0.23289924559073164","Pressure3pm":"0.18685839923670616","Pressure9am"
:"0.13944214914517747","Rainfall":"0.22440497587716363","Sunshine":"-
0.49081885139829456","Temp3pm":"-0.49839946252911915","Temp9am":"-
0.47135393998323405","WindGustSpeed":"-
0.21506968756795294","WindSpeed3pm":"-
0.145525126140691","WindSpeed9am":"-
0.2708583880345406","_deepnote_index_column":"Humidity9am"},
{"Cloud3pm":"0.5231199766631821","Cloud9am":"0.517119947699489","Evapo
ration":"-
0.39024259071039363","Humidity3pm":"1.0","Humidity9am":"0.666948568685
5433","MaxTemp":"-
0.5088554199984232","MinTemp":"0.006088664224423575","Pressure3pm":"0.
05199662593498018","Pressure9am":"-
0.02754442712799566","Rainfall":"0.2557546575244556","Sunshine":"-
```

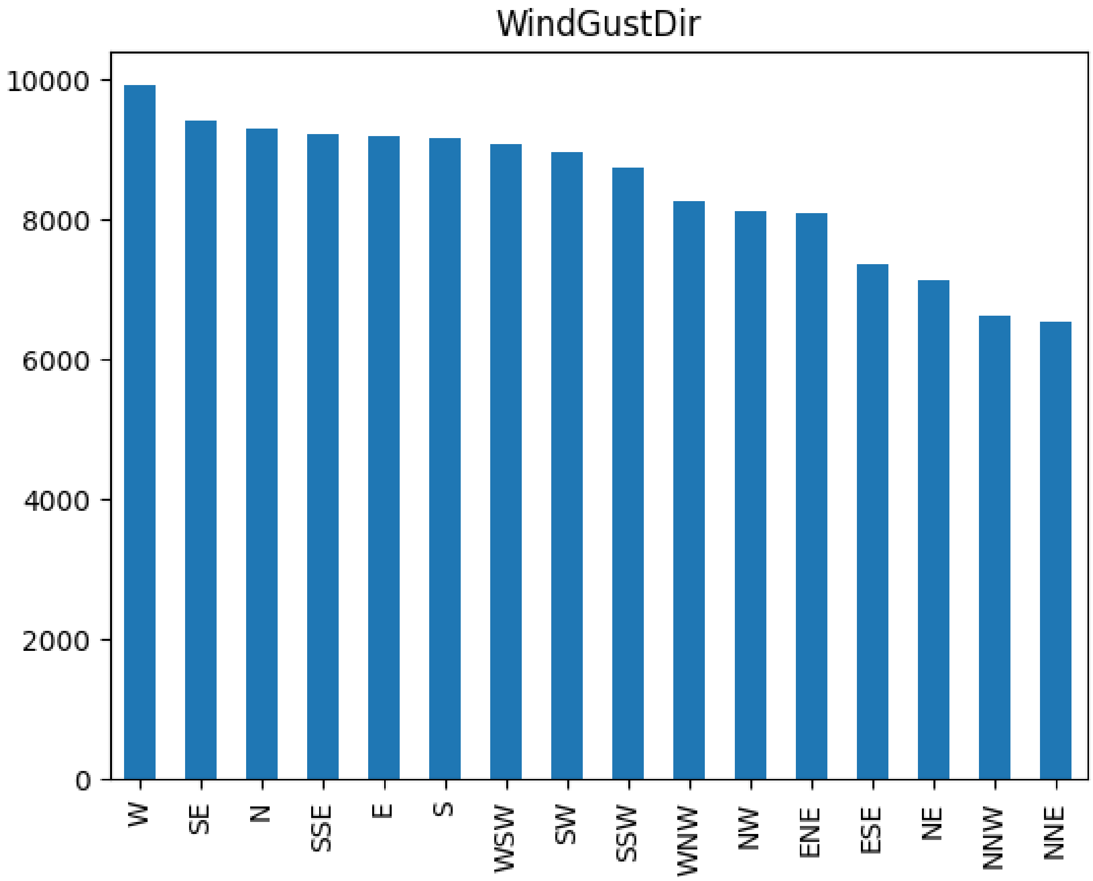
0.6291303599834128","Temp3pm":"-0.5578410027715796","Temp9am":"-0.22101938347112143","WindGustSpeed":"-0.026327121063789812","WindSpeed3pm":"0.016432471350722434","WindSpeed9am":"-0.0316138259484374","_deepnote_index_column":"Humidity3pm"}]}

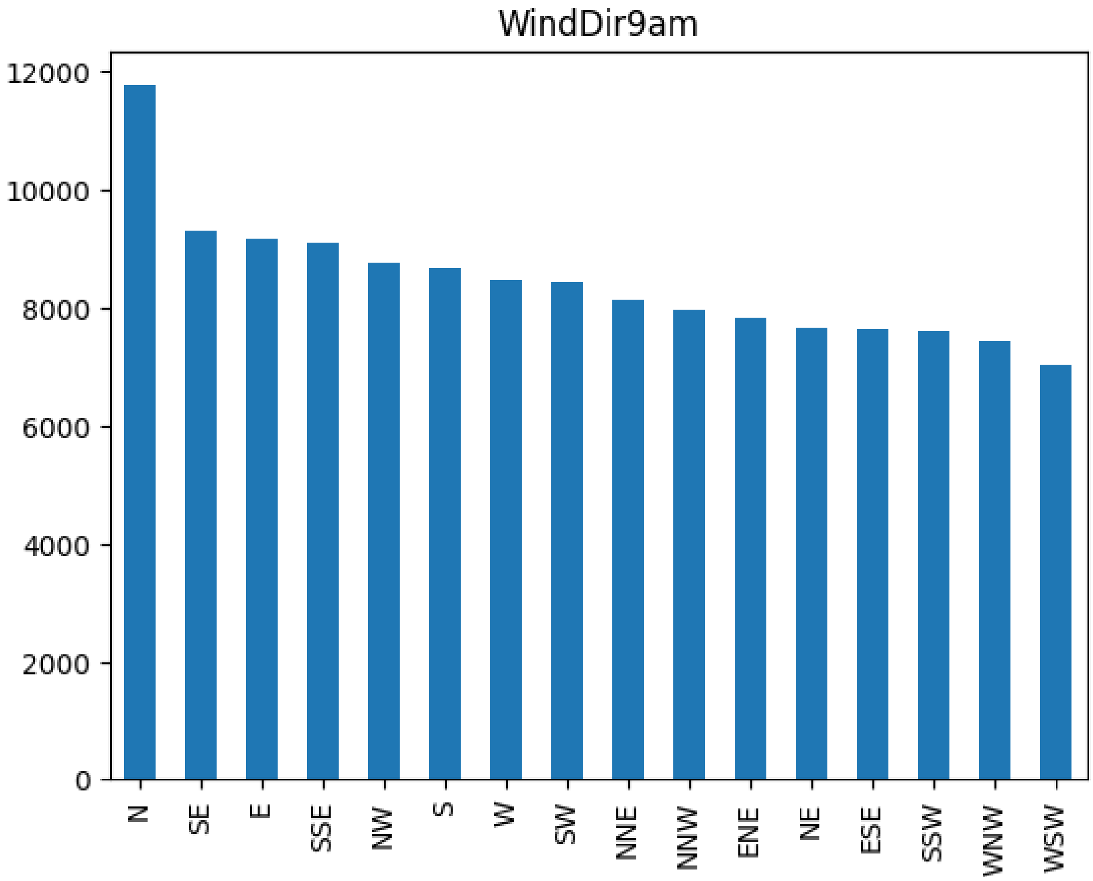
```
# 2. Visualisasi distribusi setiap fitur (kategorikal dan kontinu)

# Kontinu
df.hist(figsize=(20, 20))
plt.show()
```

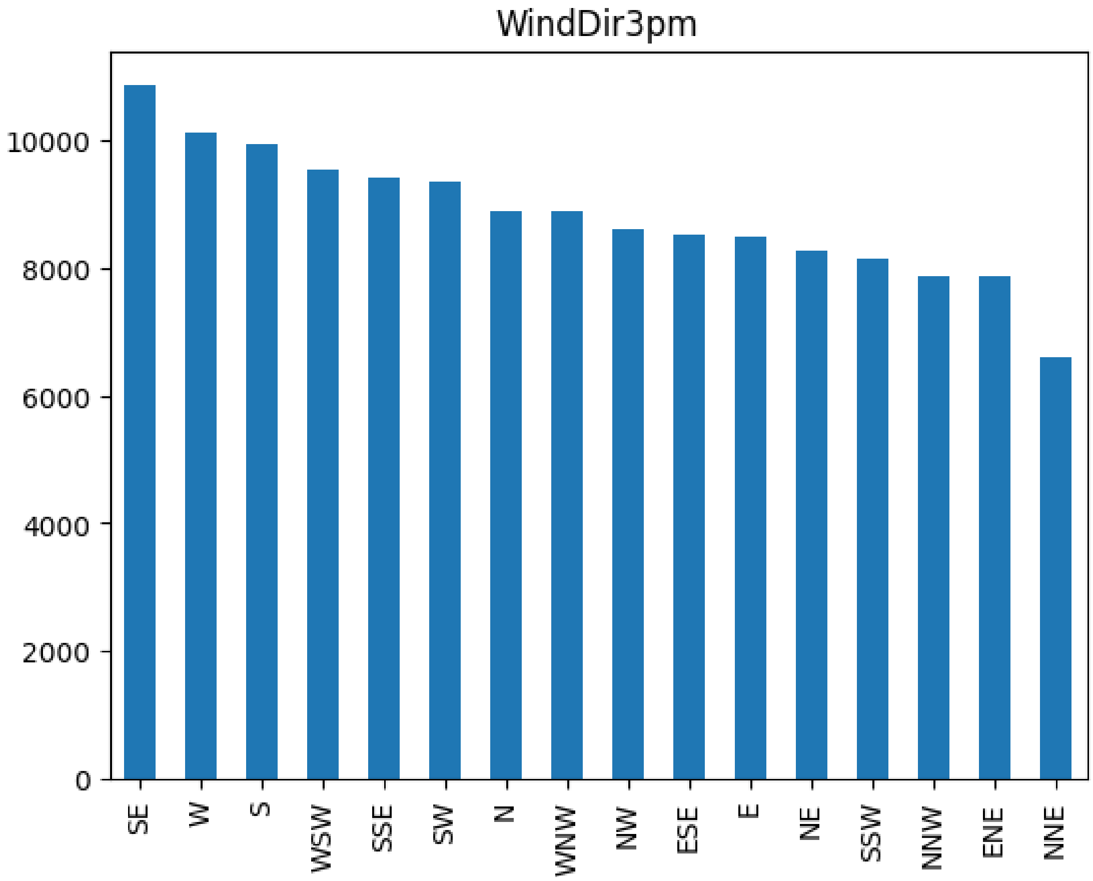


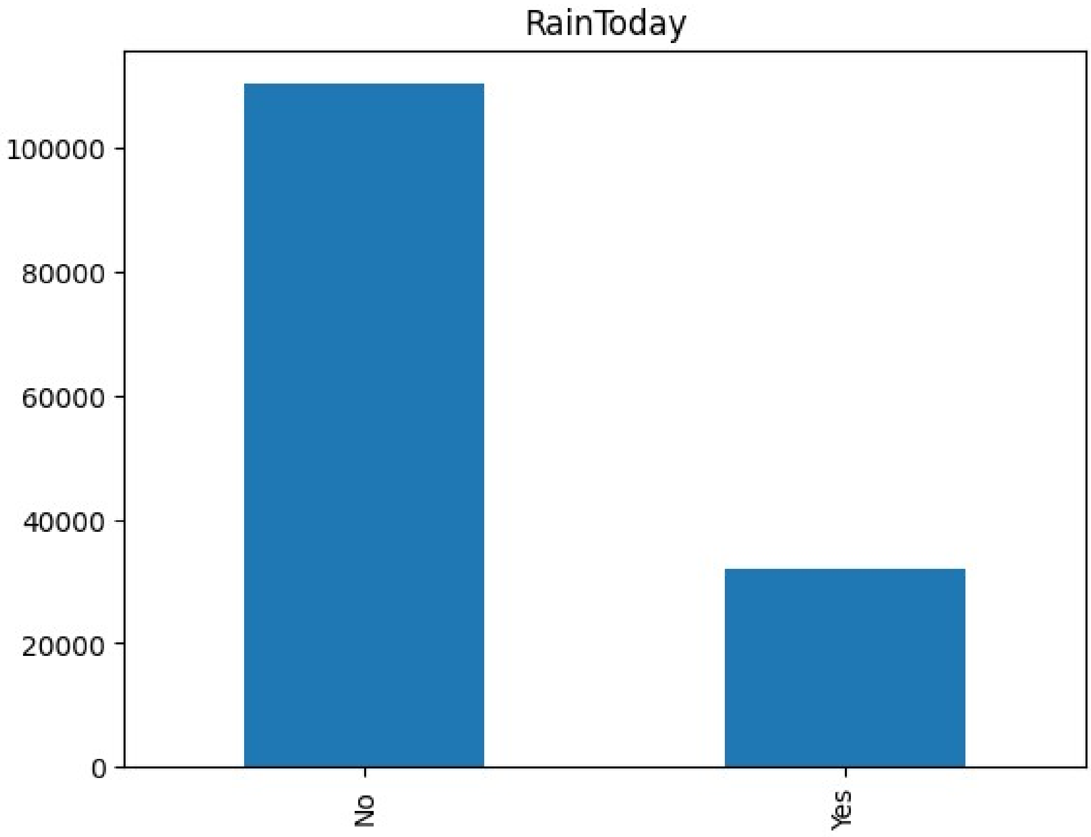
```
# Kategorikal
for col in df.select_dtypes(include=['object']):
    df[col].value_counts().plot(kind='bar')
    plt.title(col)
    plt.show()
```

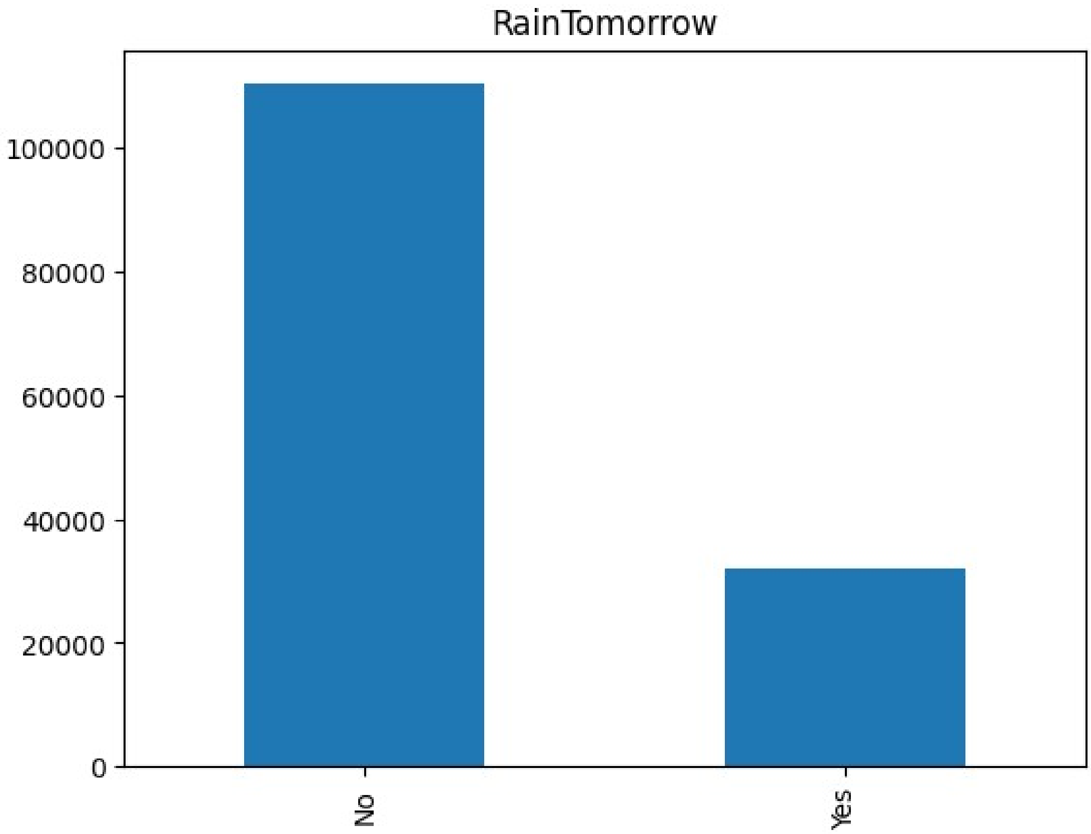

Location
3500
3000
2500
2000
1500
1000
500
0
Canberra
Sydney
Adelaide
Hobart
Darwin
Perth
Brisbane
Melbourne
Cairns
MountGambier
Townsville
Albany
GoldCoast
Wollongong
AliceSprings
Launceston
Ballarat
Albury
MountGinini
Bendigo
Tuggeranong
Newcastle
Penrith
PearceRAAF
Woomera
Witchcliffe
MelbourneAirport
Moree
Watsonia
Richmond
Mildura
NorfolkIsland
PerthAirport
Sale
Dartmoor
Williamtown
BadgerysCreek
WaggaWagga
CoffsHarbour
SydneyAirport
Nuriootpa
Portland
Cobar
Walpole
NorahHead
SalmonGums
Uluru
Katherine
Nhil


WindGustDir
10000
8000
6000
4000
2000
0
W
SE
N
SSE
E
S
WSW
SW
SSW
WNW
NW
ENE
ESE
NE
NNW
NNE


WindDir9am
12000
10000
8000
6000
4000
2000
0
N
SE
E
SSE
NW
S
W
SW
NNE
NNW
ENE
NE
ESE
SSW
WNW
WSW


WindDir3pm
10000
8000
6000
4000
2000
0
SE
W
S
WSW
SSE
SW
N
WNW
NW
ESE
E
NE
SSW
NNW
ENE
NNE


RainToday
100000
80000
60000
40000
20000
0
No
Yes



I.4

Lakukanlah analisis lebih lanjut jika diperlukan, kemudian lakukan hal berikut:

1. Penambahan fitur jika memungkinkan
2. Pembuangan fitur yang menurut kalian tidak dibutuhkan
3. Penanganan nilai hilang
4. Transformasi data kategorikal menjadi numerikal (*encoding*)
5. *Scaling* dengan `MinMaxScaler`

```python
# I.4 Put your code here

# 1. Penambahan fitur jika memungkinkan
# Tidak perlu adanya penambahan fitur

# 2. Pembuangan fitur yang menurut kalian tidak dibutuhkan
# Hujan biasanya tidak dipengaruhi lokasi karena tanda tanda hujan di
semua tempat seharusnya sama
df = df.drop(['Location'], axis=1)

# Drop data dengan null yang banyak
for col in df:
    if df[col].isnull().sum() > 50000:
        df = df.drop([col], axis=1)
```

```
print(df.shape)
df
```

```
(145460, 17)
```

```
{"column_count":17,"columns":[{"dtype":"float64","name":"MinTemp"},
{"dtype":"float64","name":"MaxTemp"},
{"dtype":"float64","name":"Rainfall"},
{"dtype":"object","name":"WindGustDir"},
{"dtype":"float64","name":"WindGustSpeed"},
{"dtype":"object","name":"WindDir9am"},
{"dtype":"object","name":"WindDir3pm"},
{"dtype":"float64","name":"WindSpeed9am"},
{"dtype":"float64","name":"WindSpeed3pm"},
{"dtype":"float64","name":"Humidity9am"},
{"dtype":"float64","name":"Humidity3pm"},
{"dtype":"float64","name":"Pressure9am"},
{"dtype":"float64","name":"Pressure3pm"},
{"dtype":"float64","name":"Temp9am"},
{"dtype":"float64","name":"Temp3pm"},
{"dtype":"object","name":"RainToday"},
{"dtype":"object","name":"RainTomorrow"},
{"dtype":"int64","name":"_deepnote_index_column"}],"row_count":145460,
"rows":
[{"Humidity3pm":"22.0","Humidity9am":"71.0","MaxTemp":"22.9","MinTemp"
:"13.4","Pressure3pm":"1007.1","Pressure9am":"1007.7","RainToday":"No"
,"RainTomorrow":"No","Rainfall":"0.6","Temp3pm":"21.8","Temp9am":"16.9
","WindDir3pm":"WNW","WindDir9am":"W","WindGustDir":"W","WindGustSpeed
":"44.0","WindSpeed3pm":"24.0","WindSpeed9am":"20.0","_deepnote_index_
column":"0"},
{"Humidity3pm":"25.0","Humidity9am":"44.0","MaxTemp":"25.1","MinTemp":
"7.4","Pressure3pm":"1007.8","Pressure9am":"1010.6","RainToday":"No","
RainTomorrow":"No","Rainfall":"0.0","Temp3pm":"24.3","Temp9am":"17.2",
"WindDir3pm":"WSW","WindDir9am":"NNW","WindGustDir":"WNW","WindGustSpe
ed":"44.0","WindSpeed3pm":"22.0","WindSpeed9am":"4.0","_deepnote_index
_column":"1"},
{"Humidity3pm":"30.0","Humidity9am":"38.0","MaxTemp":"25.7","MinTemp":
"12.9","Pressure3pm":"1008.7","Pressure9am":"1007.6","RainToday":"No",
"RainTomorrow":"No","Rainfall":"0.0","Temp3pm":"23.2","Temp9am":"21.0"
,"WindDir3pm":"WSW","WindDir9am":"W","WindGustDir":"WSW","WindGustSpee
d":"46.0","WindSpeed3pm":"26.0","WindSpeed9am":"19.0","_deepnote_index
_column":"2"},
{"Humidity3pm":"16.0","Humidity9am":"45.0","MaxTemp":"28.0","MinTemp":
"9.2","Pressure3pm":"1012.8","Pressure9am":"1017.6","RainToday":"No","
RainTomorrow":"No","Rainfall":"0.0","Temp3pm":"26.5","Temp9am":"18.1",
"WindDir3pm":"E","WindDir9am":"SE","WindGustDir":"NE","WindGustSpeed":
"24.0","WindSpeed3pm":"9.0","WindSpeed9am":"11.0","_deepnote_index_col
umn":"3"},
{"Humidity3pm":"33.0","Humidity9am":"82.0","MaxTemp":"32.3","MinTemp":
```

```
"17.5","Pressure3pm":"1006.0","Pressure9am":"1010.8","RainToday":"No",
"RainTomorrow":"No","Rainfall":"1.0","Temp3pm":"29.7","Temp9am":"17.8"
,"WindDir3pm":"NW","WindDir9am":"ENE","WindGustDir":"W","WindGustSpeed
":"41.0","WindSpeed3pm":"20.0","WindSpeed9am":"7.0","_deepnote_index_c
olumn":"4"},
{"Humidity3pm":"23.0","Humidity9am":"55.0","MaxTemp":"29.7","MinTemp":
"14.6","Pressure3pm":"1005.4","Pressure9am":"1009.2","RainToday":"No",
"RainTomorrow":"No","Rainfall":"0.2","Temp3pm":"28.9","Temp9am":"20.6"
,"WindDir3pm":"W","WindDir9am":"W","WindGustDir":"WNW","WindGustSpeed"
:"56.0","WindSpeed3pm":"24.0","WindSpeed9am":"19.0","_deepnote_index_c
olumn":"5"},
{"Humidity3pm":"19.0","Humidity9am":"49.0","MaxTemp":"25.0","MinTemp":
"14.3","Pressure3pm":"1008.2","Pressure9am":"1009.6","RainToday":"No",
"RainTomorrow":"No","Rainfall":"0.0","Temp3pm":"24.6","Temp9am":"18.1"
,"WindDir3pm":"W","WindDir9am":"SW","WindGustDir":"W","WindGustSpeed":
"50.0","WindSpeed3pm":"24.0","WindSpeed9am":"20.0","_deepnote_index_co
lumn":"6"},
{"Humidity3pm":"19.0","Humidity9am":"48.0","MaxTemp":"26.7","MinTemp":
"7.7","Pressure3pm":"1010.1","Pressure9am":"1013.4","RainToday":"No","
RainTomorrow":"No","Rainfall":"0.0","Temp3pm":"25.5","Temp9am":"16.3",
"WindDir3pm":"W","WindDir9am":"SSE","WindGustDir":"W","WindGustSpeed":
"35.0","WindSpeed3pm":"17.0","WindSpeed9am":"6.0","_deepnote_index_col
umn":"7"},
{"Humidity3pm":"9.0","Humidity9am":"42.0","MaxTemp":"31.9","MinTemp":"
9.7","Pressure3pm":"1003.6","Pressure9am":"1008.9","RainToday":"No","R
ainTomorrow":"Yes","Rainfall":"0.0","Temp3pm":"30.2","Temp9am":"18.3",
"WindDir3pm":"NW","WindDir9am":"SE","WindGustDir":"NNW","WindGustSpeed
":"80.0","WindSpeed3pm":"28.0","WindSpeed9am":"7.0","_deepnote_index_c
olumn":"8"},
{"Humidity3pm":"27.0","Humidity9am":"58.0","MaxTemp":"30.1","MinTemp":
"13.1","Pressure3pm":"1005.7","Pressure9am":"1007.0","RainToday":"Yes"
,"RainTomorrow":"No","Rainfall":"1.4","Temp3pm":"28.2","Temp9am":"20.1
","WindDir3pm":"SSE","WindDir9am":"S","WindGustDir":"W","WindGustSpeed
":"28.0","WindSpeed3pm":"11.0","WindSpeed9am":"15.0","_deepnote_index_
column":"9"}]}
```

```python
# 3. Penanganan nilai hilang
# Yang kategorikal pake modus
for col in df.select_dtypes(include=['object']).columns:
    df[col] = df[col].fillna(df[col].mode()[0])

# Yang numerikal pake mean
for col in df.select_dtypes(include=['float64']).columns:
    df[col] = df[col].fillna(df[col].mean())

df
```

```
{"column_count":17,"columns":[{"dtype":"float64","name":"MinTemp"},
{"dtype":"float64","name":"MaxTemp"},
{"dtype":"float64","name":"Rainfall"},
{"dtype":"object","name":"WindGustDir"},
{"dtype":"float64","name":"WindGustSpeed"},
{"dtype":"object","name":"WindDir9am"},
{"dtype":"object","name":"WindDir3pm"},
{"dtype":"float64","name":"WindSpeed9am"},
{"dtype":"float64","name":"WindSpeed3pm"},
{"dtype":"float64","name":"Humidity9am"},
{"dtype":"float64","name":"Humidity3pm"},
{"dtype":"float64","name":"Pressure9am"},
{"dtype":"float64","name":"Pressure3pm"},
{"dtype":"float64","name":"Temp9am"},
{"dtype":"float64","name":"Temp3pm"},
{"dtype":"object","name":"RainToday"},
{"dtype":"object","name":"RainTomorrow"},
{"dtype":"int64","name":"_deepnote_index_column"}],"row_count":145460,
"rows":
[{"Humidity3pm":"22.0","Humidity9am":"71.0","MaxTemp":"22.9","MinTemp"
:"13.4","Pressure3pm":"1007.1","Pressure9am":"1007.7","RainToday":"No"
,"RainTomorrow":"No","Rainfall":"0.6","Temp3pm":"21.8","Temp9am":"16.9
","WindDir3pm":"WNW","WindDir9am":"W","WindGustDir":"W","WindGustSpeed
":"44.0","WindSpeed3pm":"24.0","WindSpeed9am":"20.0","_deepnote_index_
column":"0"},
{"Humidity3pm":"25.0","Humidity9am":"44.0","MaxTemp":"25.1","MinTemp":
"7.4","Pressure3pm":"1007.8","Pressure9am":"1010.6","RainToday":"No","
RainTomorrow":"No","Rainfall":"0.0","Temp3pm":"24.3","Temp9am":"17.2",
"WindDir3pm":"WSW","WindDir9am":"NNW","WindGustDir":"WNW","WindGustSpe
ed":"44.0","WindSpeed3pm":"22.0","WindSpeed9am":"4.0","_deepnote_index
_column":"1"},
{"Humidity3pm":"30.0","Humidity9am":"38.0","MaxTemp":"25.7","MinTemp":
"12.9","Pressure3pm":"1008.7","Pressure9am":"1007.6","RainToday":"No",
"RainTomorrow":"No","Rainfall":"0.0","Temp3pm":"23.2","Temp9am":"21.0"
,"WindDir3pm":"WSW","WindDir9am":"W","WindGustDir":"WSW","WindGustSpee
d":"46.0","WindSpeed3pm":"26.0","WindSpeed9am":"19.0","_deepnote_index
_column":"2"},
{"Humidity3pm":"16.0","Humidity9am":"45.0","MaxTemp":"28.0","MinTemp":
"9.2","Pressure3pm":"1012.8","Pressure9am":"1017.6","RainToday":"No","
RainTomorrow":"No","Rainfall":"0.0","Temp3pm":"26.5","Temp9am":"18.1",
"WindDir3pm":"E","WindDir9am":"SE","WindGustDir":"NE","WindGustSpeed":
"24.0","WindSpeed3pm":"9.0","WindSpeed9am":"11.0","_deepnote_index_col
umn":"3"},
{"Humidity3pm":"33.0","Humidity9am":"82.0","MaxTemp":"32.3","MinTemp":
"17.5","Pressure3pm":"1006.0","Pressure9am":"1010.8","RainToday":"No",
"RainTomorrow":"No","Rainfall":"1.0","Temp3pm":"29.7","Temp9am":"17.8"
,"WindDir3pm":"NW","WindDir9am":"ENE","WindGustDir":"W","WindGustSpeed
":"41.0","WindSpeed3pm":"20.0","WindSpeed9am":"7.0","_deepnote_index_c
olumn":"4"},
{"Humidity3pm":"23.0","Humidity9am":"55.0","MaxTemp":"29.7","MinTemp":
"14.6","Pressure3pm":"1005.4","Pressure9am":"1009.2","RainToday":"No",
"RainTomorrow":"No","Rainfall":"0.2","Temp3pm":"28.9","Temp9am":"20.6"
,"WindDir3pm":"W","WindDir9am":"W","WindGustDir":"WNW","WindGustSpeed"
:"56.0","WindSpeed3pm":"24.0","WindSpeed9am":"19.0","_deepnote_index_c
olumn":"5"},
{"Humidity3pm":"19.0","Humidity9am":"49.0","MaxTemp":"25.0","MinTemp":
"14.3","Pressure3pm":"1008.2","Pressure9am":"1009.6","RainToday":"No",
"RainTomorrow":"No","Rainfall":"0.0","Temp3pm":"24.6","Temp9am":"18.1"
,"WindDir3pm":"W","WindDir9am":"SW","WindGustDir":"W","WindGustSpeed":
"50.0","WindSpeed3pm":"24.0","WindSpeed9am":"20.0","_deepnote_index_co
lumn":"6"},
{"Humidity3pm":"19.0","Humidity9am":"48.0","MaxTemp":"26.7","MinTemp":
"7.7","Pressure3pm":"1010.1","Pressure9am":"1013.4","RainToday":"No","
RainTomorrow":"No","Rainfall":"0.0","Temp3pm":"25.5","Temp9am":"16.3",
"WindDir3pm":"W","WindDir9am":"SSE","WindGustDir":"W","WindGustSpeed":
"35.0","WindSpeed3pm":"17.0","WindSpeed9am":"6.0","_deepnote_index_col
umn":"7"},
{"Humidity3pm":"9.0","Humidity9am":"42.0","MaxTemp":"31.9","MinTemp":"
9.7","Pressure3pm":"1003.6","Pressure9am":"1008.9","RainToday":"No","R
ainTomorrow":"Yes","Rainfall":"0.0","Temp3pm":"30.2","Temp9am":"18.3",
"WindDir3pm":"NW","WindDir9am":"SE","WindGustDir":"NNW","WindGustSpeed
":"80.0","WindSpeed3pm":"28.0","WindSpeed9am":"7.0","_deepnote_index_c
olumn":"8"},
{"Humidity3pm":"27.0","Humidity9am":"58.0","MaxTemp":"30.1","MinTemp":
"13.1","Pressure3pm":"1005.7","Pressure9am":"1007.0","RainToday":"Yes"
,"RainTomorrow":"No","Rainfall":"1.4","Temp3pm":"28.2","Temp9am":"20.1
","WindDir3pm":"SSE","WindDir9am":"S","WindGustDir":"W","WindGustSpeed
":"28.0","WindSpeed3pm":"11.0","WindSpeed9am":"15.0","_deepnote_index_
column":"9"}]}
```

```
# 4. Transformasi data kategorikal menjadi numerikal (encoding)

encoder = LabelEncoder()
obj = df.select_dtypes(include=['object']).columns

for col in df.select_dtypes(include=['object']).columns:
    df[col] = encoder.fit_transform(df[col])

df
```

```
{"column_count":17,"columns":[{"dtype":"float64","name":"MinTemp"},
{"dtype":"float64","name":"MaxTemp"},
{"dtype":"float64","name":"Rainfall"},
{"dtype":"int64","name":"WindGustDir"},
{"dtype":"float64","name":"WindGustSpeed"},
{"dtype":"int64","name":"WindDir9am"},
{"dtype":"int64","name":"WindDir3pm"},
{"dtype":"float64","name":"WindSpeed9am"},
{"dtype":"float64","name":"WindSpeed3pm"},
{"dtype":"float64","name":"Humidity9am"},
{"dtype":"float64","name":"Humidity3pm"},
{"dtype":"float64","name":"Pressure9am"},
{"dtype":"float64","name":"Pressure3pm"},
{"dtype":"float64","name":"Temp9am"},
{"dtype":"float64","name":"Temp3pm"},
{"dtype":"int64","name":"RainToday"},
{"dtype":"int64","name":"RainTomorrow"},
{"dtype":"int64","name":"_deepnote_index_column"}],"row_count":145460,
"rows":
[{"Humidity3pm":"22.0","Humidity9am":"71.0","MaxTemp":"22.9","MinTemp"
:"13.4","Pressure3pm":"1007.1","Pressure9am":"1007.7","RainToday":"0",
"RainTomorrow":"0","Rainfall":"0.6","Temp3pm":"21.8","Temp9am":"16.9",
"WindDir3pm":"14","WindDir9am":"13","WindGustDir":"13","WindGustSpeed"
:"44.0","WindSpeed3pm":"24.0","WindSpeed9am":"20.0","_deepnote_index_c
olumn":"0"},
{"Humidity3pm":"25.0","Humidity9am":"44.0","MaxTemp":"25.1","MinTemp":
"7.4","Pressure3pm":"1007.8","Pressure9am":"1010.6","RainToday":"0","R
ainTomorrow":"0","Rainfall":"0.0","Temp3pm":"24.3","Temp9am":"17.2","W
indDir3pm":"15","WindDir9am":"6","WindGustDir":"14","WindGustSpeed":"4
4.0","WindSpeed3pm":"22.0","WindSpeed9am":"4.0","_deepnote_index_colum
n":"1"},
{"Humidity3pm":"30.0","Humidity9am":"38.0","MaxTemp":"25.7","MinTemp":
"12.9","Pressure3pm":"1008.7","Pressure9am":"1007.6","RainToday":"0","
RainTomorrow":"0","Rainfall":"0.0","Temp3pm":"23.2","Temp9am":"21.0","
WindDir3pm":"15","WindDir9am":"13","WindGustDir":"15","WindGustSpeed":
"46.0","WindSpeed3pm":"26.0","WindSpeed9am":"19.0","_deepnote_index_co
lumn":"2"},
{"Humidity3pm":"16.0","Humidity9am":"45.0","MaxTemp":"28.0","MinTemp":
"9.2","Pressure3pm":"1012.8","Pressure9am":"1017.6","RainToday":"0","R
ainTomorrow":"0","Rainfall":"0.0","Temp3pm":"26.5","Temp9am":"18.1","W
indDir3pm":"0","WindDir9am":"9","WindGustDir":"4","WindGustSpeed":"24.
0","WindSpeed3pm":"9.0","WindSpeed9am":"11.0","_deepnote_index_column"
:"3"},
{"Humidity3pm":"33.0","Humidity9am":"82.0","MaxTemp":"32.3","MinTemp":
"17.5","Pressure3pm":"1006.0","Pressure9am":"1010.8","RainToday":"0","
RainTomorrow":"0","Rainfall":"1.0","Temp3pm":"29.7","Temp9am":"17.8","
WindDir3pm":"7","WindDir9am":"1","WindGustDir":"13","WindGustSpeed":"4
1.0","WindSpeed3pm":"20.0","WindSpeed9am":"7.0","_deepnote_index_colum
n":"4"},
{"Humidity3pm":"23.0","Humidity9am":"55.0","MaxTemp":"29.7","MinTemp":
"14.6","Pressure3pm":"1005.4","Pressure9am":"1009.2","RainToday":"0","
RainTomorrow":"0","Rainfall":"0.2","Temp3pm":"28.9","Temp9am":"20.6","
WindDir3pm":"13","WindDir9am":"13","WindGustDir":"14","WindGustSpeed":
"56.0","WindSpeed3pm":"24.0","WindSpeed9am":"19.0","_deepnote_index_co
lumn":"5"},
{"Humidity3pm":"19.0","Humidity9am":"49.0","MaxTemp":"25.0","MinTemp":
"14.3","Pressure3pm":"1008.2","Pressure9am":"1009.6","RainToday":"0","
RainTomorrow":"0","Rainfall":"0.0","Temp3pm":"24.6","Temp9am":"18.1","
WindDir3pm":"13","WindDir9am":"12","WindGustDir":"13","WindGustSpeed":
"50.0","WindSpeed3pm":"24.0","WindSpeed9am":"20.0","_deepnote_index_co
lumn":"6"},
{"Humidity3pm":"19.0","Humidity9am":"48.0","MaxTemp":"26.7","MinTemp":
"7.7","Pressure3pm":"1010.1","Pressure9am":"1013.4","RainToday":"0","R
ainTomorrow":"0","Rainfall":"0.0","Temp3pm":"25.5","Temp9am":"16.3","W
indDir3pm":"13","WindDir9am":"10","WindGustDir":"13","WindGustSpeed":"
35.0","WindSpeed3pm":"17.0","WindSpeed9am":"6.0","_deepnote_index_colu
mn":"7"},
{"Humidity3pm":"9.0","Humidity9am":"42.0","MaxTemp":"31.9","MinTemp":"
9.7","Pressure3pm":"1003.6","Pressure9am":"1008.9","RainToday":"0","Ra
inTomorrow":"1","Rainfall":"0.0","Temp3pm":"30.2","Temp9am":"18.3","Wi
ndDir3pm":"7","WindDir9am":"9","WindGustDir":"6","WindGustSpeed":"80.0
","WindSpeed3pm":"28.0","WindSpeed9am":"7.0","_deepnote_index_column":
"8"},
{"Humidity3pm":"27.0","Humidity9am":"58.0","MaxTemp":"30.1","MinTemp":
"13.1","Pressure3pm":"1005.7","Pressure9am":"1007.0","RainToday":"1","
RainTomorrow":"0","Rainfall":"1.4","Temp3pm":"28.2","Temp9am":"20.1","
WindDir3pm":"10","WindDir9am":"8","WindGustDir":"13","WindGustSpeed":"
28.0","WindSpeed3pm":"11.0","WindSpeed9am":"15.0","_deepnote_index_col
umn":"9"}]}
```

```python
# 5. Scaling dengan MinMaxScaler

scaler = MinMaxScaler()
df_scaled = pd.DataFrame(scaler.fit_transform(df), columns=df.columns)
df_scaled
```

```
{"column_count":17,"columns":[{"dtype":"float64","name":"MinTemp"},
{"dtype":"float64","name":"MaxTemp"},
{"dtype":"float64","name":"Rainfall"},
{"dtype":"float64","name":"WindGustDir"},
{"dtype":"float64","name":"WindGustSpeed"},
{"dtype":"float64","name":"WindDir9am"},
{"dtype":"float64","name":"WindDir3pm"},
{"dtype":"float64","name":"WindSpeed9am"},
{"dtype":"float64","name":"WindSpeed3pm"},
{"dtype":"float64","name":"Humidity9am"},
{"dtype":"float64","name":"Humidity3pm"},
{"dtype":"float64","name":"Pressure9am"},
{"dtype":"float64","name":"Pressure3pm"},
{"dtype":"float64","name":"Temp9am"},
{"dtype":"float64","name":"Temp3pm"},
{"dtype":"float64","name":"RainToday"},
{"dtype":"float64","name":"RainTomorrow"},
{"dtype":"int64","name":"_deepnote_index_column"}],"row_count":145460,
"rows":
[{"Humidity3pm":"0.22","Humidity9am":"0.71","MaxTemp":"0.5236294896030
246","MinTemp":"0.5165094339622642","Pressure3pm":"0.4800000000000022"
,"Pressure9am":"0.44958677685950477","RainToday":"0.0","RainTomorrow":
"0.0","Rainfall":"0.0016172506738544475","Temp3pm":"0.5220729366602687
","Temp9am":"0.5084388185654009","WindDir3pm":"0.9333333333333333","Wi
ndDir9am":"0.8666666666666667","WindGustDir":"0.8666666666666667","Win
dGustSpeed":"0.2945736434108527","WindSpeed3pm":"0.27586206896551724",
"WindSpeed9am":"0.15384615384615385","_deepnote_index_column":"0"},
{"Humidity3pm":"0.25","Humidity9am":"0.44","MaxTemp":"0.56521739130434
79","MinTemp":"0.375","Pressure3pm":"0.4912000000000097","Pressure9am
":"0.49752066115702576","RainToday":"0.0","RainTomorrow":"0.0","Rainfa
ll":"0.0","Temp3pm":"0.5700575815738963","Temp9am":"0.5147679324894514
","WindDir3pm":"1.0","WindDir9am":"0.4","WindGustDir":"0.9333333333333
333","WindGustSpeed":"0.2945736434108527","WindSpeed3pm":"0.2528735632
1839083","WindSpeed9am":"0.03076923076923077","_deepnote_index_column"
:"1"},
{"Humidity3pm":"0.3","Humidity9am":"0.38","MaxTemp":"0.576559546313799
6","MinTemp":"0.5047169811320755","Pressure3pm":"0.5055999999999994","
Pressure9am":"0.4479338842975231","RainToday":"0.0","RainTomorrow":"0.
0","Rainfall":"0.0","Temp3pm":"0.5489443378119001","Temp9am":"0.594936
7088607594","WindDir3pm":"1.0","WindDir9am":"0.8666666666666667","Wind
GustDir":"1.0","WindGustSpeed":"0.31007751937984496","WindSpeed3pm":"0
.29885057471264365","WindSpeed9am":"0.14615384615384616","_deepnote_in
dex_column":"2"},
{"Humidity3pm":"0.16","Humidity9am":"0.45","MaxTemp":"0.62003780718336
48","MinTemp":"0.4174528301886793","Pressure3pm":"0.5711999999999993",
"Pressure9am":"0.6132231404958688","RainToday":"0.0","RainTomorrow":"0
.0","Rainfall":"0.0","Temp3pm":"0.6122840690978886","Temp9am":"0.53375
52742616034","WindDir3pm":"0.0","WindDir9am":"0.6","WindGustDir":"0.26
666666666666666","WindGustSpeed":"0.13953488372093023","WindSpeed3pm":
"0.10344827586206896","WindSpeed9am":"0.08461538461538462","_deepnote_
index_column":"3"},
{"Humidity3pm":"0.33","Humidity9am":"0.8200000000000001","MaxTemp":"0.
7013232514177693","MinTemp":"0.6132075471698114","Pressure3pm":"0.4624
000000000006","Pressure9am":"0.5008264462809926","RainToday":"0.0","Ra
inTomorrow":"0.0","Rainfall":"0.0026954177897574125","Temp3pm":"0.6737
04414587332","Temp9am":"0.5274261603375527","WindDir3pm":"0.4666666666
666667","WindDir9am":"0.06666666666666667","WindGustDir":"0.8666666666
666667","WindGustSpeed":"0.2713178294573644","WindSpeed3pm":"0.2298850
5747126436","WindSpeed9am":"0.05384615384615385","_deepnote_index_colu
mn":"4"},
{"Humidity3pm":"0.23","Humidity9am":"0.55","MaxTemp":"0.65217391304347
83","MinTemp":"0.544811320754717","Pressure3pm":"0.4528000000000165",
"Pressure9am":"0.4743801652892578 6","RainToday":"0.0","RainTomorrow":"
0.0","Rainfall":"0.0005390835579514825","Temp3pm":"0.6583493282149712"
,"Temp9am":"0.5864978902953587","WindDir3pm":"0.8666666666666667","Win
dDir9am":"0.8666666666666667","WindGustDir":"0.9333333333333333","Wind
GustSpeed":"0.3875968992248062","WindSpeed3pm":"0.27586206896551724","
WindSpeed9am":"0.14615384615384616","_deepnote_index_column":"5"},
{"Humidity3pm":"0.19","Humidity9am":"0.49","MaxTemp":"0.56332703213610
59","MinTemp":"0.5377358490566039","Pressure3pm":"0.4976000000000026"
,"Pressure9am":"0.4809917355371915 4","RainToday":"0.0","RainTomorrow":
"0.0","Rainfall":"0.0","Temp3pm":"0.575815738963531 7","Temp9am":"0.533
755274261603 4","WindDir3pm":"0.8666666666666667","WindDir9am":"0.8","W
indGustDir":"0.8666666666666667","WindGustSpeed":"0.3410852713178295",
"WindSpeed3pm":"0.27586206896551724","WindSpeed9am":"0.153846153846153
85","_deepnote_index_column":"6"},
{"Humidity3pm":"0.19","Humidity9am":"0.48","MaxTemp":"0.59546313799621
92","MinTemp":"0.38207547169811323","Pressure3pm":"0.5280000000000005"
,"Pressure9am":"0.5438016528925615","RainToday":"0.0","RainTomorrow":"
0.0","Rainfall":"0.0","Temp3pm":"0.5930902111324377","Temp9am":"0.4957
8059071729963","WindDir3pm":"0.8666666666666667","WindDir9am":"0.66666
66666666666","WindGustDir":"0.8666666666666667","WindGustSpeed":"0.224
80620155038758","WindSpeed3pm":"0.1954022988505747","WindSpeed9am":"0.
046153846153846156","_deepnote_index_column":"7"},
{"Humidity3pm":"0.09","Humidity9am":"0.42","MaxTemp":"0.69376181474480
15","MinTemp":"0.42924528301886794","Pressure3pm":"0.42400000000000126
","Pressure9am":"0.4694214876033058","RainToday":"0.0","RainTomorrow":
"1.0","Rainfall":"0.0","Temp3pm":"0.6833013435700576","Temp9am":"0.537
9746835443038","WindDir3pm":"0.4666666666666667","WindDir9am":"0.6","W
indGustDir":"0.4","WindGustSpeed":"0.5736434108527132","WindSpeed3pm":
"0.3218390804597701","WindSpeed9am":"0.05384615384615385","_deepnote_i
ndex_column":"8"},
{"Humidity3pm":"0.27","Humidity9am":"0.58","MaxTemp":"0.65973534971644
61","MinTemp":"0.5094339622641509","Pressure3pm":"0.4576000000000011",
"Pressure9am":"0.4380165289256226","RainToday":"1.0","RainTomorrow":"0
.0","Rainfall":"0.0037735849056603774","Temp3pm":"0.6449136276391554",
"Temp9am":"0.5759493670886077","WindDir3pm":"0.6666666666666666","Wind
Dir9am":"0.5333333333333333","WindGustDir":"0.8666666666666667","WindG
ustSpeed":"0.17054263565891473","WindSpeed3pm":"0.12643678160919541","
WindSpeed9am":"0.11538461538461539","_deepnote_index_column":"9"}]}
```

II. Desain Eksperimen

Tujuan dari bagian ini adalah peserta dapat memahami cara melakukan eksperimen mencari metode terbaik dengan benar. Hal yang diliputi adalah sebagai berikut:

1. Pembuatan model
2. Proses validasi
3. *Hyperparameter tuning*

II.1

Tentukanlah metrik yang akan digunakan pada eksperimen kali ini. Metrik yang dapat lebih dari satu jenis.

(Tuliskan jawaban bagian II.1 di sini.)

> Matriks yang akan digunakan adalah nilai akurasi, precision, recall, dan juga nilai F1

II.2

Bagi data dengan perbandingan 0,8 untuk data latih dan 0,2 untuk data validasi.

```
# II.2 Kode di sini
```

```
X = df_scaled.drop(['RainTomorrow'], axis=1)
y = df_scaled['RainTomorrow']
X_train, X_test, y_train, y_test = train_test_split(X, y, 
test_size=0.2, random_state=10)
print(X_train, y_train)
print(X_test, y_test)
```

```
         MinTemp   MaxTemp  Rainfall  WindGustDir  WindGustSpeed  
WindDir9am  \
96775   0.400943  0.527410  0.000000     0.400000       0.286822   
0.333333   
96840   0.622642  0.508507  0.000000     0.733333       0.310078   
0.733333   
66795   0.462264  0.421550  0.001078     0.933333       0.356589   
0.466667   
104598  0.285377  0.508507  0.000000     0.533333       0.271318   
0.600000   
137064  0.372642  0.536862  0.000000     0.666667       0.271318   
0.600000   
...          ...       ...       ...          ...            ...   
...   
9372    0.540094  0.559546  0.000000     0.866667       0.263839   
0.466667   
105595  0.830189  0.916824  0.000000     0.600000       0.294574   
0.600000   
93553   0.617925  0.631380  0.000000     0.066667       0.193798   
0.733333   
94735   0.804245  0.737240  0.030728     1.000000       0.310078   
0.866667   
83209   0.233491  0.393195  0.000539     0.600000       0.255814   
0.200000   


        WindDir3pm  WindSpeed9am  WindSpeed3pm  Humidity9am  
Humidity3pm  \
96775     0.400000      0.115385      0.298851     0.450000   
0.230000   
96840     0.800000      0.130769      0.275862     0.640000   
0.400000   
66795     1.000000      0.215385      0.321839     0.550000   
0.470000   
104598    0.800000      0.130769      0.218391     0.540000   
0.320000   
137064    0.600000      0.215385      0.229885     0.160000   
0.140000   
...            ...           ...           ...          ...          .
..   
9372      0.266667      0.115385      0.252874     0.610000   
0.560000   
105595    0.800000      0.184615      0.080460     0.370000   
0.120000   
```

```
93553     0.266667      0.069231      0.229885     0.710000
0.570000
94735     0.866667      0.146154      0.172414     0.650000
0.470000
83209     0.533333      0.000000      0.218391     0.688808
0.515391

        Pressure9am  Pressure3pm   Temp9am   Temp3pm  RainToday
96775      0.525620       0.4912  0.521097  0.537428        0.0
96840      0.618182       0.6832  0.544304  0.518234        0.0
66795      0.606612       0.6480  0.462025  0.422265        0.0
104598     0.699174       0.6800  0.432489  0.512476        0.0
137064     0.676033       0.6480  0.544304  0.550864        0.0
...             ...          ...       ...       ...        ...
9372       0.697521       0.6928  0.613924  0.568138        0.0
105595     0.457851       0.4464  0.786920  0.898273        0.0
93553      0.601653       0.5920  0.668776  0.623800        0.0
94735      0.294215       0.2928  0.757384  0.746641        1.0
83209      0.963636       0.9520  0.310127  0.381958        0.0

[116368 rows x 16 columns] 96775     0.0
96840     0.0
66795     0.0
104598    0.0
137064    0.0
         ...
9372      0.0
105595    0.0
93553     0.0
94735     0.0
83209     0.0
Name: RainTomorrow, Length: 116368, dtype: float64
         MinTemp   MaxTemp  Rainfall  WindGustDir  WindGustSpeed
WindDir9am  \
124601  0.589623  0.603025  0.002156     0.066667       0.155039
0.000000
19713   0.455189  0.387524  0.000000     0.733333       0.325581
0.800000
98025   0.535377  0.531191  0.000000     0.266667       0.224806
0.266667
58336   0.240566  0.277883  0.000000     0.266667       0.240310
0.066667
116285  0.391509  0.495274  0.030189     0.000000       0.271318
0.000000
...          ...       ...       ...          ...            ...
...
65552   0.304245  0.521739  0.000000     0.200000       0.410853
0.200000
59794   0.308962  0.421550  0.008086     0.533333       0.193798
0.333333
```

```
4308    0.351415  0.423440  0.000000     0.666667       0.240310
0.800000
101795  0.375000  0.372401  0.001617     0.866667       0.356589
0.400000
96871   0.558962  0.567108  0.000000     1.000000       0.286822
0.666667

        WindDir3pm  WindSpeed9am  WindSpeed3pm  Humidity9am
Humidity3pm  \
124601    0.066667      0.100000      0.080460         0.79
0.48
19713     0.533333      0.153846      0.425287         0.41
0.47
98025     0.200000      0.115385      0.172414         0.57
0.44
58336     0.266667      0.130769      0.229885         0.88
0.81
116285    0.666667      0.184615      0.103448         0.48
0.33
...            ...           ...           ...          ...          .
..
65552     0.200000      0.284615      0.379310         0.42
0.26
59794     1.000000      0.015385      0.229885         0.89
0.49
4308      0.733333      0.146154      0.149425         0.76
0.91
101795    0.933333      0.169231      0.229885         0.96
0.72
96871     0.800000      0.084615      0.275862         0.50
0.40

        Pressure9am  Pressure3pm   Temp9am   Temp3pm  RainToday
124601     0.614049     0.610494  0.565401  0.596929        0.0
19713      0.631405     0.681600  0.430380  0.397313        0.0
98025      0.631405     0.619200  0.540084  0.543186        0.0
58336      0.550413     0.518400  0.267932  0.266795        0.0
116285     0.725620     0.705600  0.457806  0.499040        1.0
...             ...          ...       ...       ...        ...
65552      0.723967     0.707200  0.455696  0.497121        0.0
59794      0.752066     0.752000  0.394515  0.433781        1.0
4308       0.823140     0.819200  0.402954  0.343570        0.0
101795     0.636364     0.622400  0.362869  0.366603        0.0
96871      0.695868     0.670400  0.516878  0.531670        0.0

[29092 rows x 16 columns] 124601    0.0
19713     0.0
98025     1.0
58336     0.0
116285    0.0
```

```
         ...
65552     0.0
59794     0.0
4308      1.0
101795    1.0
96871     0.0
Name: RainTomorrow, Length: 29092, dtype: float64
```

II.3

Lakukan hal berikut:

1. Prediksi dengan menggunakan model *logistic regression* sebagai *baseline*.
2. Tampilkan evaluasi dari model yang dibangun dari metrik yang ditentukan pada II.1
3. Tampilkan *confusion matrix*.

```
# II.3 Kode di sini

model = LogisticRegression(max_iter=10000).fit(X_train, y_train)
y_pred = model.predict(X_test)

print('Accuracy: ', accuracy_score(y_test, y_pred))
print('Precision: ', precision_score(y_test, y_pred))
print('Recall: ', recall_score(y_test, y_pred))
print('F1: ', f1_score(y_test, y_pred))
print(confusion_matrix(y_test, y_pred))
```

```
Accuracy:  0.8415028186443009
Precision:  0.7131669114047968
Recall:  0.45868093813946165
F1:  0.5582910240444486
[[21567  1172]
 [ 3439  2914]]
```

II.4

Lakukanlah:

1. Pembelajaran dengan model lain
2. *Hyperparameter tuning* untuk model yang dipakai dengan menggunakan *grid search* (perhatikan *random factor* pada beberapa algoritma model)
3. Validasi dengan *cross validation*

```
# II.4 Kode di sini.

# Menggunakan decision tree
decisionTree = DecisionTreeClassifier()

param = {
    "criterion": ["gini", "entropy", "log_loss"],
    "splitter": ["best"],
    "max_depth": [2,4,6,8,10],
    "min_samples_leaf": [2, 4, 8]
}

grid_search = GridSearchCV(decisionTree, param, cv=5,
scoring='accuracy').fit(X_train, y_train)
best_params = grid_search.best_params_
best_models = grid_search.best_estimator_

best_models.fit(X_train, y_train)
y_pred = best_models.predict(X_test)

print('Accuracy: ', accuracy_score(y_test, y_pred))
print('Precision: ', precision_score(y_test, y_pred))
print('Recall: ', recall_score(y_test, y_pred))
print('F1: ', f1_score(y_test, y_pred))
print(confusion_matrix(y_test, y_pred))
```

```
Accuracy:  0.8415715660662725
Precision:  0.7110358180058083
Recall:  0.46245868093813947
F1:  0.5604196471149261
[[21545  1194]
 [ 3415  2938]]
```

```
scores = cross_validate(best_models, X_test, y_test, cv=5,
scoring=('accuracy', 'precision', 'recall', 'f1'))

print('Accuracy: ', scores['test_accuracy'].mean())
print('Precision: ', scores['test_precision'].mean())
print('Recall: ', scores['test_recall'].mean())
print('F1: ', scores['test_f1'].mean())
```

```
Accuracy:  0.8352122475944576
Precision:  0.6836683985084694
Recall:  0.4564755880731274
F1:  0.5471962790452917
```

III. Improvement

Pada bagian ini, kalian diharapkan dapat:

1. melakukan pelatihan dengan data hasil *oversampling* / *undersampling*, disertai dengan validasi yang benar; serta
2. menerapkan beberapa metode untuk menggabungkan beberapa model.

Kedua hal ini adalah contoh metode untuk meningkatkan kinerja dari model.

III.1

Lakukanlah:

1. *Oversampling* pada kelas minoritas pada data latih
2. *Undersampling* pada kelas mayoritas pada data latih

Pada setiap tahap, latih dengan model *baseline* (II.3), dan validasi dengan data validasi. Data latih dan validasi adalah data yang disusun pada bagian II.2.

```
# III.1 Kode di sini.

# Oversampling
smote = SMOTE(random_state=12)
OX_train, Oy_train = smote.fit_resample(X_train, y_train)
print(OX_train.shape)

model = LogisticRegression(max_iter=10000).fit(OX_train, Oy_train)
y_pred = model.predict(X_test)

print('Accuracy: ', accuracy_score(y_test, y_pred))
print('Precision: ', precision_score(y_test, y_pred))
print('Recall: ', recall_score(y_test, y_pred))
print('F1: ', f1_score(y_test, y_pred))
print(confusion_matrix(y_test, y_pred))
```

```
(181688, 16)
Accuracy:  0.7785645538292314
Precision:  0.4953871669949207
Recall:  0.752243034786715
F1:  0.597375
[[17871  4868]
 [ 1574  4779]]
```

```
# Validasi
scores = cross_validate(model, X_test, y_test, cv=5,
scoring=('accuracy', 'precision', 'recall', 'f1'))

print('Accuracy: ', scores['test_accuracy'].mean())
print('Precision: ', scores['test_precision'].mean())
print('Recall: ', scores['test_recall'].mean())
print('F1: ', scores['test_f1'].mean())
```

```
Accuracy:  0.8408156422184978
Precision:  0.7147639405148151
Recall:  0.4509702199892205
F1:  0.5529381370573242
```

```
# Undersampling
rus = RandomUnderSampler(random_state=12)
UX_train, Uy_train = rus.fit_resample(X_train, y_train)
```

```
print(UX_train.shape)

model = LogisticRegression(max_iter=10000).fit(UX_train, Uy_train)
y_pred = model.predict(X_test)

print('Accuracy: ', accuracy_score(y_test, y_pred))
print('Precision: ', precision_score(y_test, y_pred))
print('Recall: ', recall_score(y_test, y_pred))
print('F1: ', f1_score(y_test, y_pred))
print(confusion_matrix(y_test, y_pred))
```

```
(51048, 16)
Accuracy:  0.7796301388697924
Precision:  0.49698168193172354
Recall:  0.7516134109869353
F1:  0.5983334377545266
[[17906  4833]
 [ 1578  4775]]
```

```
scores = cross_validate(model, X_test, y_test, cv=5,
scoring=('accuracy', 'precision', 'recall', 'f1'))

print('Accuracy: ', scores['test_accuracy'].mean())
print('Precision: ', scores['test_precision'].mean())
print('Recall: ', scores['test_recall'].mean())
print('F1: ', scores['test_f1'].mean())
```

```
Accuracy:  0.8408156422184978
Precision:  0.7147639405148151
Recall:  0.4509702199892205
F1:  0.5529381370573242
```

III.2

Lakukanlah:

1. Eksplorasi *soft voting*, *hard voting*, dan *stacking*.
2. Buatlah model *logistic regression* dan SVM.
3. Lakukanlah *soft voting* dari model-model yang dibangun pada poin 2.
4. Lakukan *hard voting* dari model-model yang dibangun pada poin 2.
5. Lakukanlah *stacking* dengan *final classifier* adalah *logistic regression* dari model-model yang dibangun pada poin 2.
6. Lakukan validasi dengan metrics yang telah ditentukan untuk poin 3, 4, dan 5.

(Tuliskan hasil eksplorasi III.2 poin 1 di sini.)

```
# III.2 Kode di sini.
lr = LogisticRegression(max_iter=1000)
svm = SVC(probability=True, max_iter=100)
```

```
# Soft Voting
model = VotingClassifier(estimators=[('lr', lr), ('svm', svm)],
voting='soft')
model.fit(X_train, y_train)
y_pred = model.predict(X_test)

print('Accuracy: ', accuracy_score(y_test, y_pred))
print('Precision: ', precision_score(y_test, y_pred))
print('Recall: ', recall_score(y_test, y_pred))
print('F1: ', f1_score(y_test, y_pred))
print(confusion_matrix(y_test, y_pred))

scores = cross_validate(model, X_test, y_test, cv=5,
scoring=('accuracy', 'precision', 'recall', 'f1'))

print('Accuracy: ', scores['test_accuracy'].mean())
print('Precision: ', scores['test_precision'].mean())
print('Recall: ', scores['test_recall'].mean())
print('F1: ', scores['test_f1'].mean())
```

```
/shared-libs/python3.9/py/lib/python3.9/site-packages/sklearn/svm/
_base.py:301: ConvergenceWarning: Solver terminated early
(max_iter=100).  Consider pre-processing your data with StandardScaler
or MinMaxScaler.
  warnings.warn(
Accuracy:  0.8297126357761584
Precision:  0.7995717344753748
Recall:  0.2938769085471431
F1:  0.42978821362799263
[[22271   468]
 [ 4486  1867]]
/shared-libs/python3.9/py/lib/python3.9/site-packages/sklearn/svm/
_base.py:301: ConvergenceWarning: Solver terminated early
(max_iter=100).  Consider pre-processing your data with StandardScaler
or MinMaxScaler.
  warnings.warn(
/shared-libs/python3.9/py/lib/python3.9/site-packages/sklearn/svm/
_base.py:301: ConvergenceWarning: Solver terminated early
(max_iter=100).  Consider pre-processing your data with StandardScaler
or MinMaxScaler.
  warnings.warn(
/shared-libs/python3.9/py/lib/python3.9/site-packages/sklearn/svm/
_base.py:301: ConvergenceWarning: Solver terminated early
(max_iter=100).  Consider pre-processing your data with StandardScaler
or MinMaxScaler.
  warnings.warn(
/shared-libs/python3.9/py/lib/python3.9/site-packages/sklearn/svm/
_base.py:301: ConvergenceWarning: Solver terminated early
(max_iter=100).  Consider pre-processing your data with StandardScaler
or MinMaxScaler.
```

```
  warnings.warn(
/shared-libs/python3.9/py/lib/python3.9/site-packages/sklearn/svm/
_base.py:301: ConvergenceWarning: Solver terminated early
(max_iter=100).  Consider pre-processing your data with StandardScaler
or MinMaxScaler.
  warnings.warn(
Accuracy:  0.8104982368600719
Precision:  0.8110993661373588
Recall:  0.1692194750243159
F1:  0.26908772007071013
```

```
# Hard Voting
model = VotingClassifier(estimators=[('lr', lr), ('svm', svm)],
voting='hard')
model.fit(X_train, y_train)
y_pred = model.predict(X_test)

print('Accuracy: ', accuracy_score(y_test, y_pred))
print('Precision: ', precision_score(y_test, y_pred))
print('Recall: ', recall_score(y_test, y_pred))
print('F1: ', f1_score(y_test, y_pred))
print(confusion_matrix(y_test, y_pred))

scores = cross_validate(model, X_test, y_test, cv=5,
scoring=('accuracy', 'precision', 'recall', 'f1'))

print('Accuracy: ', scores['test_accuracy'].mean())
print('Precision: ', scores['test_precision'].mean())
print('Recall: ', scores['test_recall'].mean())
print('F1: ', scores['test_f1'].mean())
```

```
/shared-libs/python3.9/py/lib/python3.9/site-packages/sklearn/svm/
_base.py:301: ConvergenceWarning: Solver terminated early
(max_iter=100).  Consider pre-processing your data with StandardScaler
or MinMaxScaler.
  warnings.warn(
Accuracy:  0.7916265640038499
Precision:  0.6644067796610169
Recall:  0.09255469856760586
F1:  0.16247582205029013
[[22442   297]
 [ 5765   588]]
/shared-libs/python3.9/py/lib/python3.9/site-packages/sklearn/svm/
_base.py:301: ConvergenceWarning: Solver terminated early
(max_iter=100).  Consider pre-processing your data with StandardScaler
or MinMaxScaler.
  warnings.warn(
/shared-libs/python3.9/py/lib/python3.9/site-packages/sklearn/svm/
_base.py:301: ConvergenceWarning: Solver terminated early
(max_iter=100).  Consider pre-processing your data with StandardScaler
```

```
or MinMaxScaler.
  warnings.warn(
/shared-libs/python3.9/py/lib/python3.9/site-packages/sklearn/svm/
_base.py:301: ConvergenceWarning: Solver terminated early
(max_iter=100).  Consider pre-processing your data with StandardScaler
or MinMaxScaler.
  warnings.warn(
/shared-libs/python3.9/py/lib/python3.9/site-packages/sklearn/svm/
_base.py:301: ConvergenceWarning: Solver terminated early
(max_iter=100).  Consider pre-processing your data with StandardScaler
or MinMaxScaler.
  warnings.warn(
/shared-libs/python3.9/py/lib/python3.9/site-packages/sklearn/svm/
_base.py:301: ConvergenceWarning: Solver terminated early
(max_iter=100).  Consider pre-processing your data with StandardScaler
or MinMaxScaler.
  warnings.warn(
Accuracy:  0.7887394342604397
Precision:  0.6709149184323178
Recall:  0.06217449215386235
F1:  0.11248526303084021
```

```python
model = StackingClassifier(estimators=[('lr', lr), ('svm', svm)],
final_estimator=LogisticRegression(max_iter=10000))
model.fit(X_train, y_train)
y_pred = model.predict(X_test)

print('Accuracy: ', accuracy_score(y_test, y_pred))
print('Precision: ', precision_score(y_test, y_pred))
print('Recall: ', recall_score(y_test, y_pred))
print('F1: ', f1_score(y_test, y_pred))
print(confusion_matrix(y_test, y_pred))

scores = cross_validate(model, X_test, y_test, cv=5,
scoring=('accuracy', 'precision', 'recall', 'f1'))

print('Accuracy: ', scores['test_accuracy'].mean())
print('Precision: ', scores['test_precision'].mean())
print('Recall: ', scores['test_recall'].mean())
print('F1: ', scores['test_f1'].mean())
```

```
/shared-libs/python3.9/py/lib/python3.9/site-packages/sklearn/svm/
_base.py:301: ConvergenceWarning: Solver terminated early
(max_iter=100).  Consider pre-processing your data with StandardScaler
or MinMaxScaler.
  warnings.warn(
/shared-libs/python3.9/py/lib/python3.9/site-packages/sklearn/svm/
_base.py:301: ConvergenceWarning: Solver terminated early
(max_iter=100).  Consider pre-processing your data with StandardScaler
or MinMaxScaler.
```

```
  warnings.warn(
/shared-libs/python3.9/py/lib/python3.9/site-packages/sklearn/svm/
_base.py:301: ConvergenceWarning: Solver terminated early
(max_iter=100).  Consider pre-processing your data with StandardScaler
or MinMaxScaler.
  warnings.warn(
/shared-libs/python3.9/py/lib/python3.9/site-packages/sklearn/svm/
_base.py:301: ConvergenceWarning: Solver terminated early
(max_iter=100).  Consider pre-processing your data with StandardScaler
or MinMaxScaler.
  warnings.warn(
/shared-libs/python3.9/py/lib/python3.9/site-packages/sklearn/svm/
_base.py:301: ConvergenceWarning: Solver terminated early
(max_iter=100).  Consider pre-processing your data with StandardScaler
or MinMaxScaler.
  warnings.warn(
/shared-libs/python3.9/py/lib/python3.9/site-packages/sklearn/svm/
_base.py:301: ConvergenceWarning: Solver terminated early
(max_iter=100).  Consider pre-processing your data with StandardScaler
or MinMaxScaler.
  warnings.warn(
Accuracy:  0.8413309500893716
Precision:  0.7112624665531501
Recall:  0.46025499763891076
F1:  0.5588685015290519
[[21552  1187]
 [ 3429  2924]]
/shared-libs/python3.9/py/lib/python3.9/site-packages/sklearn/svm/
_base.py:301: ConvergenceWarning: Solver terminated early
(max_iter=100).  Consider pre-processing your data with StandardScaler
or MinMaxScaler.
  warnings.warn(
/shared-libs/python3.9/py/lib/python3.9/site-packages/sklearn/svm/
_base.py:301: ConvergenceWarning: Solver terminated early
(max_iter=100).  Consider pre-processing your data with StandardScaler
or MinMaxScaler.
  warnings.warn(
/shared-libs/python3.9/py/lib/python3.9/site-packages/sklearn/svm/
_base.py:301: ConvergenceWarning: Solver terminated early
(max_iter=100).  Consider pre-processing your data with StandardScaler
or MinMaxScaler.
  warnings.warn(
/shared-libs/python3.9/py/lib/python3.9/site-packages/sklearn/svm/
_base.py:301: ConvergenceWarning: Solver terminated early
(max_iter=100).  Consider pre-processing your data with StandardScaler
or MinMaxScaler.
  warnings.warn(
/shared-libs/python3.9/py/lib/python3.9/site-packages/sklearn/svm/
_base.py:301: ConvergenceWarning: Solver terminated early
(max_iter=100).  Consider pre-processing your data with StandardScaler
```

```
or MinMaxScaler.
  warnings.warn(
/shared-libs/python3.9/py/lib/python3.9/site-packages/sklearn/svm/
_base.py:301: ConvergenceWarning: Solver terminated early
(max_iter=100).  Consider pre-processing your data with StandardScaler
or MinMaxScaler.
  warnings.warn(
/shared-libs/python3.9/py/lib/python3.9/site-packages/sklearn/svm/
_base.py:301: ConvergenceWarning: Solver terminated early
(max_iter=100).  Consider pre-processing your data with StandardScaler
or MinMaxScaler.
  warnings.warn(
/shared-libs/python3.9/py/lib/python3.9/site-packages/sklearn/svm/
_base.py:301: ConvergenceWarning: Solver terminated early
(max_iter=100).  Consider pre-processing your data with StandardScaler
or MinMaxScaler.
  warnings.warn(
/shared-libs/python3.9/py/lib/python3.9/site-packages/sklearn/svm/
_base.py:301: ConvergenceWarning: Solver terminated early
(max_iter=100).  Consider pre-processing your data with StandardScaler
or MinMaxScaler.
  warnings.warn(
/shared-libs/python3.9/py/lib/python3.9/site-packages/sklearn/svm/
_base.py:301: ConvergenceWarning: Solver terminated early
(max_iter=100).  Consider pre-processing your data with StandardScaler
or MinMaxScaler.
  warnings.warn(
/shared-libs/python3.9/py/lib/python3.9/site-packages/sklearn/svm/
_base.py:301: ConvergenceWarning: Solver terminated early
(max_iter=100).  Consider pre-processing your data with StandardScaler
or MinMaxScaler.
  warnings.warn(
/shared-libs/python3.9/py/lib/python3.9/site-packages/sklearn/svm/
_base.py:301: ConvergenceWarning: Solver terminated early
(max_iter=100).  Consider pre-processing your data with StandardScaler
or MinMaxScaler.
  warnings.warn(
/shared-libs/python3.9/py/lib/python3.9/site-packages/sklearn/svm/
_base.py:301: ConvergenceWarning: Solver terminated early
(max_iter=100).  Consider pre-processing your data with StandardScaler
or MinMaxScaler.
  warnings.warn(
/shared-libs/python3.9/py/lib/python3.9/site-packages/sklearn/svm/
_base.py:301: ConvergenceWarning: Solver terminated early
(max_iter=100).  Consider pre-processing your data with StandardScaler
or MinMaxScaler.
  warnings.warn(
/shared-libs/python3.9/py/lib/python3.9/site-packages/sklearn/svm/
_base.py:301: ConvergenceWarning: Solver terminated early
(max_iter=100).  Consider pre-processing your data with StandardScaler
```

```
or MinMaxScaler.
  warnings.warn(
/shared-libs/python3.9/py/lib/python3.9/site-packages/sklearn/svm/
_base.py:301: ConvergenceWarning: Solver terminated early
(max_iter=100).  Consider pre-processing your data with StandardScaler
or MinMaxScaler.
  warnings.warn(
/shared-libs/python3.9/py/lib/python3.9/site-packages/sklearn/svm/
_base.py:301: ConvergenceWarning: Solver terminated early
(max_iter=100).  Consider pre-processing your data with StandardScaler
or MinMaxScaler.
  warnings.warn(
/shared-libs/python3.9/py/lib/python3.9/site-packages/sklearn/svm/
_base.py:301: ConvergenceWarning: Solver terminated early
(max_iter=100).  Consider pre-processing your data with StandardScaler
or MinMaxScaler.
  warnings.warn(
/shared-libs/python3.9/py/lib/python3.9/site-packages/sklearn/svm/
_base.py:301: ConvergenceWarning: Solver terminated early
(max_iter=100).  Consider pre-processing your data with StandardScaler
or MinMaxScaler.
  warnings.warn(
/shared-libs/python3.9/py/lib/python3.9/site-packages/sklearn/svm/
_base.py:301: ConvergenceWarning: Solver terminated early
(max_iter=100).  Consider pre-processing your data with StandardScaler
or MinMaxScaler.
  warnings.warn(
/shared-libs/python3.9/py/lib/python3.9/site-packages/sklearn/svm/
_base.py:301: ConvergenceWarning: Solver terminated early
(max_iter=100).  Consider pre-processing your data with StandardScaler
or MinMaxScaler.
  warnings.warn(
/shared-libs/python3.9/py/lib/python3.9/site-packages/sklearn/svm/
_base.py:301: ConvergenceWarning: Solver terminated early
(max_iter=100).  Consider pre-processing your data with StandardScaler
or MinMaxScaler.
  warnings.warn(
/shared-libs/python3.9/py/lib/python3.9/site-packages/sklearn/svm/
_base.py:301: ConvergenceWarning: Solver terminated early
(max_iter=100).  Consider pre-processing your data with StandardScaler
or MinMaxScaler.
  warnings.warn(
/shared-libs/python3.9/py/lib/python3.9/site-packages/sklearn/svm/
_base.py:301: ConvergenceWarning: Solver terminated early
(max_iter=100).  Consider pre-processing your data with StandardScaler
or MinMaxScaler.
  warnings.warn(
/shared-libs/python3.9/py/lib/python3.9/site-packages/sklearn/svm/
_base.py:301: ConvergenceWarning: Solver terminated early
(max_iter=100).  Consider pre-processing your data with StandardScaler
```

```
or MinMaxScaler.
  warnings.warn(
/shared-libs/python3.9/py/lib/python3.9/site-packages/sklearn/svm/
_base.py:301: ConvergenceWarning: Solver terminated early
(max_iter=100).  Consider pre-processing your data with StandardScaler
or MinMaxScaler.
  warnings.warn(
/shared-libs/python3.9/py/lib/python3.9/site-packages/sklearn/svm/
_base.py:301: ConvergenceWarning: Solver terminated early
(max_iter=100).  Consider pre-processing your data with StandardScaler
or MinMaxScaler.
  warnings.warn(
/shared-libs/python3.9/py/lib/python3.9/site-packages/sklearn/svm/
_base.py:301: ConvergenceWarning: Solver terminated early
(max_iter=100).  Consider pre-processing your data with StandardScaler
or MinMaxScaler.
  warnings.warn(
/shared-libs/python3.9/py/lib/python3.9/site-packages/sklearn/svm/
_base.py:301: ConvergenceWarning: Solver terminated early
(max_iter=100).  Consider pre-processing your data with StandardScaler
or MinMaxScaler.
  warnings.warn(
/shared-libs/python3.9/py/lib/python3.9/site-packages/sklearn/svm/
_base.py:301: ConvergenceWarning: Solver terminated early
(max_iter=100).  Consider pre-processing your data with StandardScaler
or MinMaxScaler.
  warnings.warn(
Accuracy:  0.840712502180627
Precision:  0.7102763653138782
Recall:  0.45695112658517995
F1:  0.5560731342657197
```

IV. Analisis

Bandingkan hasil dari hal-hal berikut:

1. Model *baseline* (II.3)
2. Model lain (II.4)
3. Hasil *undersampling*
4. Hasil *oversampling*
5. Hasil *soft voting*
6. Hasil *hard voting*
7. Hasil *stacking*

(Tuliskan jawaban bagian IV di sini.)

IV.1

Model Baseline digunakan sebagai perbandingan untuk model model yang lain. Model ini menggunakan logistic regression sebagai baseline. Model Baseline memiliki metrik berikut

```
Accuracy:  0.8415028186443009
Precision:  0.7131669114047968
Recall:  0.45868093813946165
F1:  0.5582910240444486
```

Dan berikut adalah confusion matrixnya

```
[[21567  1172]
 [ 3439  2914]]
```

IV.2

Model lain (Decision Tree Classifier) memiliki metrik berikut. Dapat dilihat bahwa model dengan menggunakan DTC lebih baik karena nilai Accuracy dan F1 nya meningkat jauh lebih banyak dibandingkan dengan model baseline

```
Accuracy:  0.8415715660662725
Precision:  0.7110358180058083
Recall:  0.46245868093813947
F1:  0.5604196471149261
```

Dan ketika di cross validate, hasil Accuracy dan F1 menurun sehingga hal ini terjadi karena terdapat overfitting pada modelnya yang terlalu fit dengan data training yang diberikan

```
Accuracy:  0.8352122475944576
Precision:  0.6836683985084694
Recall:  0.4564755880731274
F1:  0.5471962790452917
```

IV.3

Hasil undersampling memiliki metrik berikut. undersampling menunjukan akurasi yang lebih kecil dibandingkan dengan baseline. namun memiliki recall yang lebih tinggi

```
Accuracy:  0.7796301388697924
Precision:  0.49698168193172354
Recall:  0.7516134109869353
F1:  0.5983334377545266
```

Dan ketika di cross validate, hasilnya meningkatkan akurasi dan precision namun menurunkan recall dan F1. Karena recall turun otomatis F1 juga turun. Recall dapat turun karena banyak sampling yang bersifat true positif banyak dibuang

```
Accuracy:  0.8408156422184978
Precision:  0.7147639405148151
Recall:  0.4509702199892205
F1:  0.5529381370573242
```

IV.4

Hasil oversampling memiliki metrik berikut. Dapat dilihat oversampling memilki nilai akursai dan precision memiliki nilai yang lebih jelek dibandingkan yang lain namun sisanya meningkat

```
Accuracy:  0.7785645538292314
Precision:  0.4953871669949207
Recall:  0.752243034786715
F1:  0.597375
```

Dan ketika di cross validate, hasilnya adalah terjadi peningkatan di segala aspek kecuali recall dan F1

```
Accuracy:  0.8408156422184978
Precision:  0.7147639405148151
Recall:  0.4509702199892205
F1:  0.5529381370573242
```

IV.5

Hasil soft voting memiliki metrik berikut. metrik memiliki nilai yang lebih rendah di segala aspek kecuali akurasi dan

```
Accuracy:  0.8297126357761584
Precision:  0.7995717344753748
Recall:  0.2938769085471431
F1:  0.42978821362799263
```

Dan ketika di cross validate, hasilnya adalah

```
Accuracy:  0.8104982368600719
Precision:  0.8110993661373588
Recall:  0.1692194750243159
F1:  0.26908772007071013
```

IV.6

Hasil hard voting memiliki metrik berikut. Nilai recall dan F1 sangat kecil karena belum sampai convergence

```
Accuracy:  0.7916265640038499
Precision:  0.6644067796610169
Recall:  0.09255469856760586
F1:  0.16247582205029013
```

Dan ketika di cross validate, hasilnya menurun dibandingkan modelnya karena terjadi overfitting pada model

```
Accuracy:  0.7887394342604397
Precision:  0.6709149184323178
```

```
Recall:  0.06217449215386235
F1:  0.11248526303084021
```

IV.7

Hasil stacking memiliki metrik berikut

```
Accuracy:  0.8413309500893716
Precision:  0.7112624665531501
Recall:  0.46025499763891076
F1:  0.5588685015290519
```

Dan ketika di cross validate, hasilnya adalah tidak banyak jauh berubah dibandingkan yang lain sehingga model ini sudah tepat

```
Accuracy:  0.840712502180627
Precision:  0.7102763653138782
Recall:  0.45695112658517995
F1:  0.5560731342657197
```

Created in Deepnote