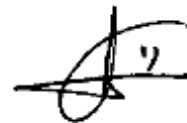


PAKTA INTEGRITAS

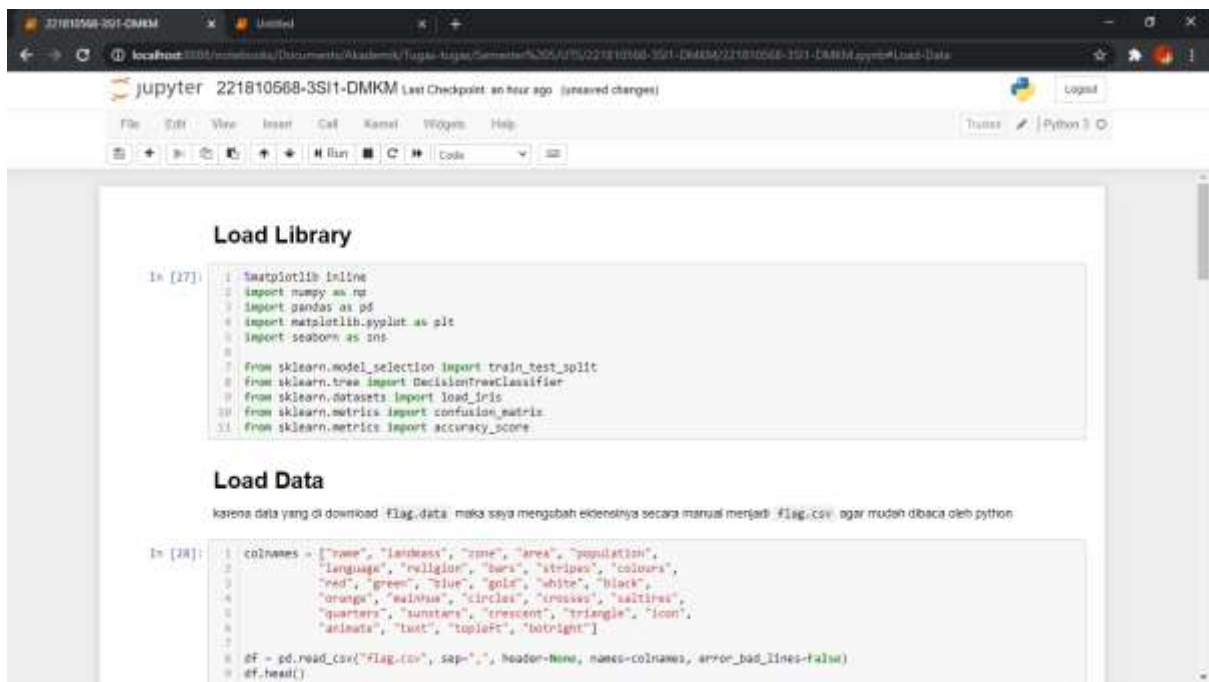
Nama : Rihadatulaisy Salsabila Nurudin
NIM : 221810568
No. Absen : 27
Kelas : 3SI1
Dosen : Pak Ibnu Santoso, SST., M.T.
Mata Kuliah : Data Mining and Knowledge Management
Hari, Tanggal : Selasa, 27 Oktober 2020

“Saya menyatakan bahwa ujian ini saya kerjakan dengan jujur sesuai kemampuan sendiri dan tidak mengutip sebagian atau seluruh pekerjaan orang lain. Jika suatu saat ditemukan saya melanggar ketentuan ujian, saya siap menerima konsekuensi yang berlaku.”



Rihadatulaisy S.N.

Load Library



A Jupyter Notebook interface with a dark theme. The browser address bar shows a local host URL. The notebook title is '221810568-3SI1-DMKM'. The code cell is titled 'Load Library' and contains 11 lines of Python code importing various libraries. Below the code cell, there is a text block titled 'Load Data' with a paragraph explaining a manual file extension change from .data to .csv.

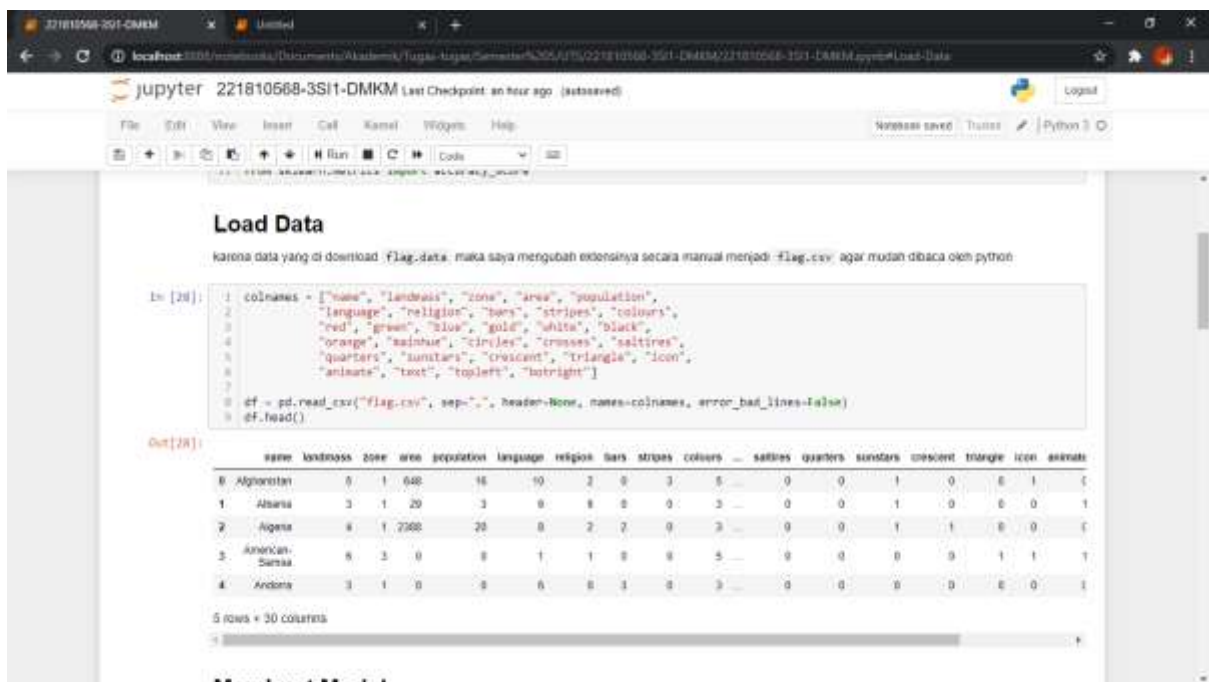
```
In [27]: 1 %matplotlib inline
2 import numpy as np
3 import pandas as pd
4 import matplotlib.pyplot as plt
5 import seaborn as sns
6
7 from sklearn.model_selection import train_test_split
8 from sklearn.tree import DecisionTreeClassifier
9 from sklearn.datasets import load_iris
10 from sklearn.metrics import confusion_matrix
11 from sklearn.metrics import accuracy_score
```

Load Data

karena data yang di download .flag.data maka saya mengubah ekstensinya secara manual menjadi .flag.csv agar mudah dibaca oleh python

```
In [28]: 1 colnames = ["name", "landmass", "zone", "area", "population",
2             "language", "religion", "bars", "stripes", "colours",
3             "red", "green", "blue", "gold", "white", "black",
4             "orange", "saltnue", "circles", "crosses", "saltires",
5             "quarters", "sunstars", "crescent", "triangle", "icon",
6             "animate", "text", "topleft", "bottomright"]
7
8 df = pd.read_csv("flag.csv", sep=";", header=None, names=colnames, error_bad_lines=False)
9 df.head()
```

Load Data



A Jupyter Notebook interface showing the same code as the previous screenshot. The output of the code cell is displayed, showing the first five rows of the loaded data as a table. The table has 20 columns and 5 rows. The first row is for 'Algeria' and the last row is for 'Andorra'.

```
In [28]: 1 colnames = ["name", "landmass", "zone", "area", "population",
2             "language", "religion", "bars", "stripes", "colours",
3             "red", "green", "blue", "gold", "white", "black",
4             "orange", "saltnue", "circles", "crosses", "saltires",
5             "quarters", "sunstars", "crescent", "triangle", "icon",
6             "animate", "text", "topleft", "bottomright"]
7
8 df = pd.read_csv("flag.csv", sep=";", header=None, names=colnames, error_bad_lines=False)
9 df.head()
```

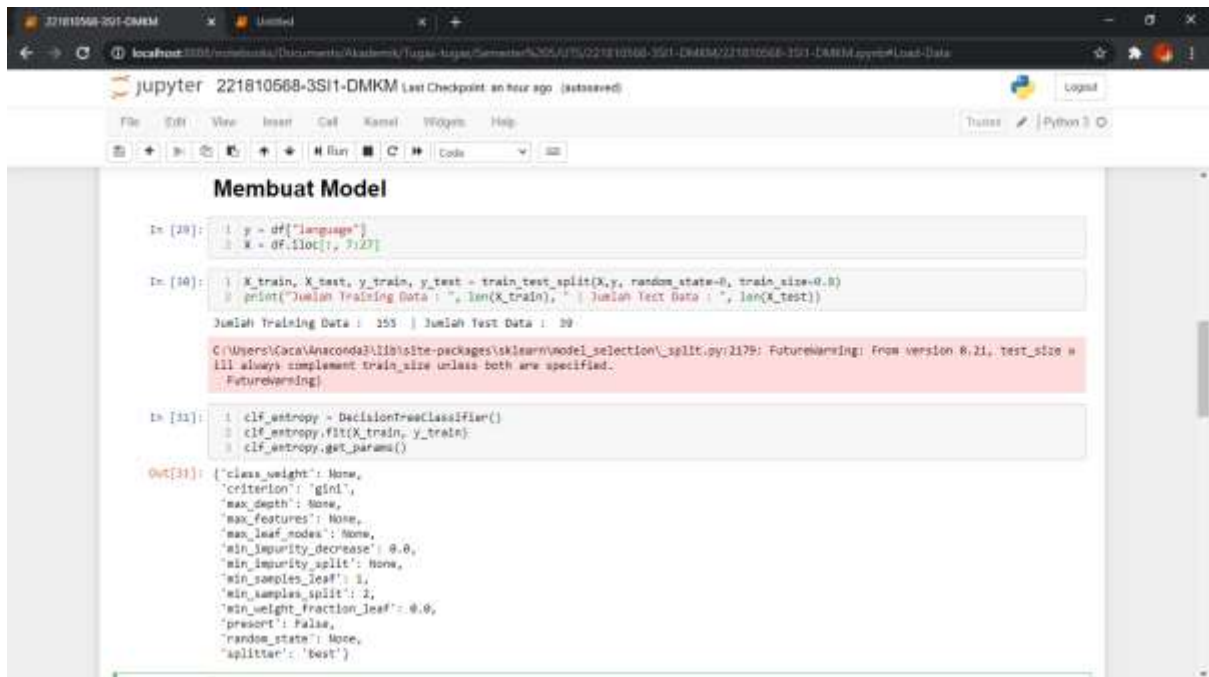
Out[28]:

	name	landmass	zone	area	population	language	religion	bars	stripes	colours	...	saltnue	quarters	sunstars	crescent	triangle	icon	animate
0	Algeria	0	1	648	16	10	2	0	3	5	...	0	0	1	0	0	1	0
1	Albania	3	1	29	3	0	8	0	0	3	...	0	0	1	0	0	0	1
2	Algeria	0	1	2368	20	0	2	2	0	3	...	0	0	1	1	0	0	0
3	American Samoa	0	3	0	0	1	1	0	0	5	...	0	0	0	0	1	1	1
4	Andorra	3	1	0	0	0	0	3	0	3	...	0	0	0	0	0	0	1

5 rows x 20 columns

Membuat Model

Membuat Model



```
In [28]: 1 y = df["language"]
        2 X = df.iloc[:, 3:27]

In [30]: 1 X_train, X_test, y_train, y_test = train_test_split(X, y, random_state=0, train_size=0.8)
        2 print("Jumlah Training Data : ", len(X_train), " | Jumlah Test Data : ", len(X_test))

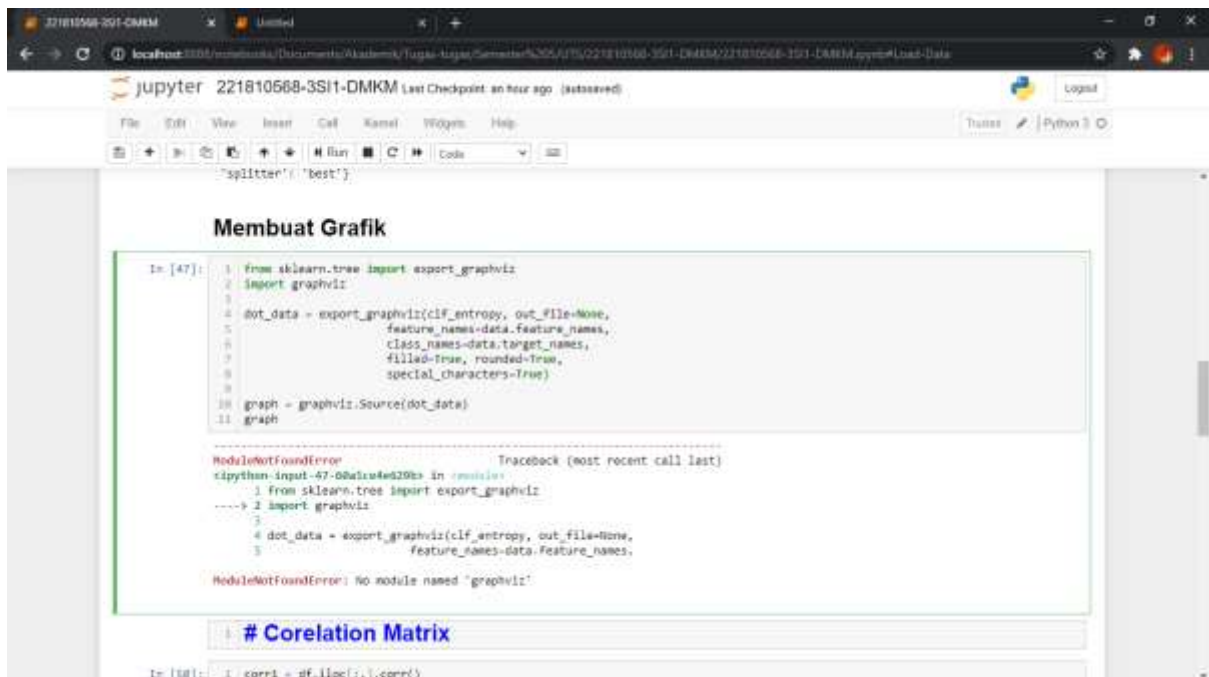
Jumlah Training Data : 255 | Jumlah Test Data : 50

C:\Users\Acara\Anaconda3\lib\site-packages\sklearn\model_selection\_split.py:2179: FutureWarning: From version 0.21, test_size =
ill always complement train_size unless both are specified.
  FutureWarning)

In [31]: 1 clf_entropy = DecisionTreeClassifier()
        2 clf_entropy.fit(X_train, y_train)
        3 clf_entropy.get_params()

Out[31]: {'class_weight': None,
          'criterion': 'gini',
          'max_depth': None,
          'max_features': None,
          'max_leaf_nodes': None,
          'min_impurity_decrease': 0.0,
          'min_impurity_split': None,
          'min_samples_leaf': 1,
          'min_samples_split': 2,
          'min_weight_fraction_leaf': 0.0,
          'presort': False,
          'random_state': None,
          'splitter': 'best'}
```

Membuat Grafik



```
In [47]: 1 from sklearn.tree import export_graphviz
        2 import graphviz
        3
        4 dot_data = export_graphviz(clf_entropy, out_file=None,
        5                          feature_names=data.feature_names,
        6                          class_names=data.target_names,
        7                          filled=True, rounded=True,
        8                          special_characters=True)
        9
       10 graph = graphviz.Source(dot_data)
       11 graph

ModuleNotFoundError: Traceback (most recent call last)
c:\python-input-47-0ba1c4e4d29e in <module>
      1 from sklearn.tree import export_graphviz
----> 2 import graphviz
      3
      4 dot_data = export_graphviz(clf_entropy, out_file=None,
      5                          feature_names=data.feature_names,
      6                          class_names=data.target_names,
      7                          filled=True, rounded=True,
      8                          special_characters=True)
      9
     10 graph = graphviz.Source(dot_data)
     11 graph

ModuleNotFoundError: No module named 'graphviz'

# Correlation Matrix

In [48]: 1 corr1 = df.iloc[:, 1].corr()
```

Menghitung Akurasi, Presisi, Sensitivity, F-1 Score

221810568-3SI1-DMKM x Untitled x

localhost:8888/~/Documents/Akademik/Fajar-Agus/Semester%205/175221810568-3SI1-DMKM/221810568-3SI1-DMKM.py?Load=Data

jupyter 221810568-3SI1-DMKM (Last Checkpoint: an hour ago) (saved changes)

File Edit View Insert Cell Kernel Windows Help Trusted Python 3

Akurasi, Presisi, Sensitivity, F-1 Score

```
In [48]: 1 from sklearn.metrics import classification_report
2
3 y_pred = clf_entropy.predict(X_test)
4
5 print(classification_report(y_test, y_pred))
```

	precision	recall	f1-score	support
1	1.00	0.46	0.63	13
2	0.00	0.00	0.00	0
3	0.00	0.00	0.00	3
4	0.00	0.00	0.00	0
5	0.00	0.00	0.00	3
6	0.17	0.20	0.18	6
7	0.00	0.00	0.00	1
8	0.20	0.33	0.25	1
9	0.00	0.00	0.00	1
10	0.40	0.40	0.40	10
micro avg	0.31	0.31	0.31	50
macro avg	0.30	0.14	0.15	50
weighted avg	0.47	0.31	0.30	50

C:\Users\Caca\Anaconda3\lib\site-packages\sklearn\metrics\classification.py:1143: UndefinedMetricWarning: Precision and F-score are ill-defined and being set to 0.0 in labels with no predicted samples.
 'precision', 'predicted', average, warn_for)
C:\Users\Caca\Anaconda3\lib\site-packages\sklearn\metrics\classification.py:1143: UndefinedMetricWarning: Recall and F-score are ill-defined and being set to 0.0 in labels with no true samples.
 'recall', 'true', average, warn_for)