Klasifikasi Dataset Wine Menggunakan Algoritma Visualisasi Decision Tree

By Muhammad Aziz Habiburrahim
DSF 33.0 Data Science – dibimbing.id

Table of contents

```python
# import library
import pandas as pd
import matplotlib.pyplot as mlp
from sklearn.metrics import accuracy_score
from sklearn.tree import DecisionTreeClassifier, plot_tree
from sklearn.model_selection import train_test_split
from google.colab import sheets
from matplotlib import pyplot as plt
from sklearn.datasets import load_wine
# load dataset and convert to dataframe
data = load_wine()
# convert data load_wine to dataframe
df = pd.DataFrame(data.data, columns=data.feature_names)
df['target'] = data.target
df.head(10)
```

01
Import Dataset

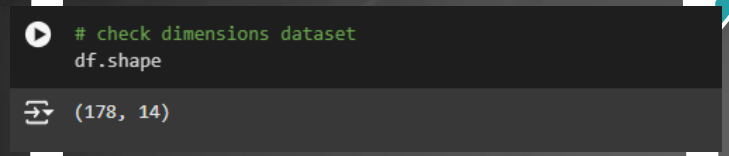
Mengambil Dataset Wine menggunakan Import sklearn

```
# convert dataset to sheet
from google.colab import sheets
sheet = sheets.InteractiveSheet(df=df)
```

02

Spreadsheet Dataset

Mengubah Dataset Wine menjadi Spreadsheets



03
Dimensions Dataset

Dimensi Dataset Wine sebesar 178 baris dan 14 kolom

```
# convert format view decimal number
pd.options.display.float_format = '{:,.5f}'.format

# view decription statistics
df.describe()
```

04
Float Dataset

Mengubah Dataset Wine menjadi Float Dataset

```
[ ] # info dataset
    df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 178 entries, 0 to 177
Data columns (total 14 columns):
 #   Column                        Non-Null Count  Dtype  
---  ------                        --------------  -----  
 0   alcohol                       178 non-null    float64
 1   malic_acid                    178 non-null    float64
 2   ash                           178 non-null    float64
 3   alcalinity_of_ash             178 non-null    float64
 4   magnesium                     178 non-null    float64
 5   total_phenols                 178 non-null    float64
 6   flavanoids                    178 non-null    float64
 7   nonflavanoid_phenols          178 non-null    float64
 8   proanthocyanins               178 non-null    float64
 9   color_intensity               178 non-null    float64
 10  hue                           178 non-null    float64
 11  od280/od315_of_diluted_wines  178 non-null    float64
 12  proline                       178 non-null    float64
 13  target                        178 non-null    int64  
dtypes: float64(13), int64(1)
memory usage: 19.6 KB
```

05
Null Dataset

Mengecek Data Kosong pada Dataset Wine

```
[ ]  from sklearn.model_selection import train_test_split

     # define features and target
     x = df[data.feature_names]
     y = df['target']

     # split dataset into training and test sets
     x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.2, random_state=42)
```

```
[ ]  from sklearn.tree import DecisionTreeClassifier, plot_tree

     # initialize decision tree classifier
     dt_classifier = DecisionTreeClassifier(random_state=42)

     # train the model
     dt_classifier.fit(x_train, y_train)
```

```
DecisionTreeClassifier
DecisionTreeClassifier(random_state=42)
```

06 DTC Dataset

Evaluasi Dataset Wine Menggunakan DTC (Decision Tree Classifier)

```python
from sklearn.metrics import accuracy_score

# predict on the test set
y_pred = dt_classifier.predict(x_test)

# calculate accuracy
accuracy = accuracy_score(y_test, y_pred)
print(accuracy*100)
```

```
94.44444444444444
```

07
Accuracy Dataset

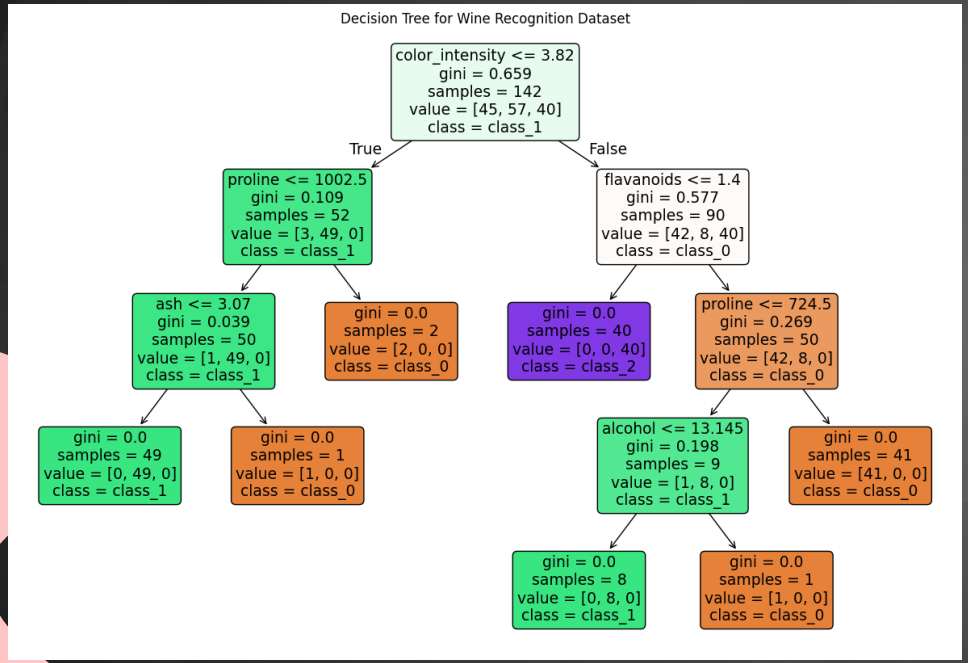
Menghitung Akurasi Dataset Wine

```
import matplotlib.pyplot as mlp

# visualize the decision tree
mlp.figure(figsize=(15, 10))
plot_tree(dt_classifier, filled=True, feature_names=data.feature_names, class_names=data.target_names, rounded=True)
mlp.title("Decision Tree for Wine Recognition Dataset")
mlp.show()
```

08
Roadmap Dataset

Menampilkan roadmap Dataset Winemenggunakan DTC



09
Conclusion

Pada gambar tersebut menjelaskan nilai kebenaran/kesalahan Dataset Wine menggunakan Decision Tree Classifier

Thanks!

Do you have any questions?

azizhabibrahim@gmail.com
https://www.linkedin.com/in/mhabibr02/