

Technical Report Machine Learning



OLEH:

M.Daffa.Satria

(1103204018)

PROGRAM STUDI TEKNIK KOMPUTER

FAKULTAS TEKNIK ELEKTRO

UNIVERSITAS TELKOM

2023

Tujuan dari laporan ini adalah untuk menganalisis dan memodelkan dataset kanker payudara menggunakan Scikit Learn dan Seaborn framework di Google Colab. Dataset ini berisi fitur medis dari pasien yang telah didiagnosis dengan kanker payudara. Tujuan dari analisis ini adalah untuk memprediksi apakah pasien tersebut menderita kanker jinak atau ganas.

Langkah pertama dalam analisis adalah untuk memuat dataset menggunakan Scikit Learn. Dataset tersebut kemudian diubah menjadi dataframe menggunakan Pandas sehingga dapat dengan mudah diakses dan dimanipulasi. Kemudian, dilakukan visualisasi data dengan menggunakan Seaborn. Terdapat banyak cara untuk memvisualisasikan data, dan salah satu yang paling umum adalah dengan menggunakan pairplot, seperti yang ditunjukkan pada contoh kode di atas.

Setelah visualisasi data selesai, dilakukan eksplorasi data menggunakan beberapa algoritma machine learning. Dalam laporan ini, digunakan algoritma Decision Tree untuk memprediksi apakah pasien tersebut menderita kanker jinak atau ganas. Langkah-langkah yang diambil dalam membangun model ini adalah memisahkan dataset menjadi data latih dan data uji menggunakan `train_test_split`, menginisialisasi model Decision Tree Classifier, melatih model pada data latih, dan akhirnya melakukan prediksi pada data uji. Hasil akurasi model kemudian dicetak pada output.

Kemudian, eksplorasi data dilanjutkan dengan menggunakan algoritma Random Forest dan Self-Training. Metode ini digunakan untuk mengevaluasi performa model. Algoritma Random Forest digunakan untuk membangun banyak pohon keputusan dan mengambil keputusan berdasarkan mayoritas dari semua pohon. Sedangkan Self-Training adalah teknik semi-supervised learning di mana model di awal dilatih pada sejumlah sampel yang diketahui dengan pasti, dan kemudian model tersebut menggunakan sampel lain yang kemungkinan besar benar.

Kesimpulan yang dapat diambil dari eksplorasi data ini adalah bahwa algoritma Decision Tree, Random Forest, dan Self-Training dapat memberikan akurasi prediksi yang cukup baik pada dataset kanker payudara. Meskipun hasil prediksi ini dapat bervariasi tergantung pada data yang digunakan, namun penggunaan algoritma machine learning tersebut dapat membantu dalam mendeteksi kanker payudara lebih awal dan meningkatkan kesembuhan pasien.

Source code

```
!pip install scikit-learn seaborn pandas numpy

from sklearn.datasets import load_breast_cancer
import pandas as pd

data = load_breast_cancer()
df = pd.DataFrame(data.data, columns=data.feature_names)
df['target'] = data.target

import seaborn as sns
sns.pairplot(df, hue='target', vars=['mean radius', 'mean texture', 'mean
perimeter', 'mean area'])

from sklearn.tree import DecisionTreeClassifier
from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score
```

```

X = df.drop('target', axis=1)
y = df['target']

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3)

model = DecisionTreeClassifier()
model.fit(X_train, y_train)

y_pred = model.predict(X_test)
accuracy = accuracy_score(y_test, y_pred)

print('Accuracy:', accuracy)

```

output

