



Lembar Kerja Mahasiswa

Mata Kuliah : Data Mining
Bahasan : Ansamble Classification
Halaman : 1/3

NIM	222410101064
Nama	Jetro Sulthan Fatih Nurrafi
Kelas	A
Program Studi	Sistem Informasi
Asisten	1. Renata Sayidatul Arikha 212410101057 2. Aprodhita Nanda Eka Wijaya 212410101071

LANGKAH KERJA

1. Buat proses klasifikasi menggunakan metode Boosting khusus untuk data kategori pada dataset adult.csv. Tuliskan penjelasan analisis dan akurasinya.

```
import pandas as pd
import numpy as np
from sklearn.ensemble import AdaBoostClassifier
from sklearn.tree import DecisionTreeClassifier
from sklearn.preprocessing import LabelEncoder, StandardScaler
from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score

data = pd.read_csv("adult.csv")
data.head()

# Melakukan Clearing Data Null & Menambahkan Label Encoder

data = data.dropna()

label_encoders = {}
for column in data.select_dtypes(include=['object']).columns:
    le = LabelEncoder()
    data[column] = le.fit_transform(data[column])
    label_encoders[column] = le

# Definisi target serta standarisasi lalu membuat data training
```



Lembar Kerja Mahasiswa

Mata Kuliah : Data Mining
Bahasan : Ensemble Classification
Halaman : 2/3

```
X = data.drop('income', axis=1)
y = data['income']

scaler = StandardScaler()
X = scaler.fit_transform(X)
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,
random_state=42)

#Training Data

base_estimator = DecisionTreeClassifier(max_depth=1)
adaboost = AdaBoostClassifier(base_estimator=base_estimator, n_estimators=50,
learning_rate=1.0, random_state=42)
adaboost.fit(X_train, y_train)

y_pred = adaboost.predict(X_test)

accuracy = accuracy_score(y_test, y_pred)
print(f'Accuracy: {accuracy:.2f}')
```

Accuracy: 0.86

HASIL DAN ANALISIS DATA

Setelah melakukan klasifikasi pada dataset adult.csv menggunakan metode Boosting yang dirancang khusus untuk data kategorikal, dengan variabel target klasifikasi 'income', hasilnya menunjukkan akurasi yang cukup tinggi, yaitu 0,86. Metode ini terbukti efektif dalam menangani data kategorikal dan memberikan performa yang baik untuk klasifikasi pada kasus ini.




Lembar Kerja Mahasiswa

Mata Kuliah : Data Mining
Bahasan : Ensemble Classification
Halaman : 3/3

KESIMPULAN

Akurasi yang tinggi menunjukkan bahwa model klasifikasi ini memiliki kemampuan yang baik dalam memprediksi label pada data tersebut. Namun, performa model untuk salah satu kelas sedikit lebih rendah dibandingkan dengan kelas lainnya. Hal ini mungkin disebabkan oleh ketidakseimbangan jumlah sampel antarkelas atau kompleksitas yang lebih tinggi pada beberapa data, yang mengakibatkan penurunan akurasi prediksi untuk kelas tersebut.

Link Google Colab

 LKM 8_1064.ipynb

Link Youtube (Unlisted)

Jember,2024

Mengetahui,
Dosen Datamining

Asisten,

Fajrin Nurman Arifin, S.T., M.Eng
NIP. 198511282015041002

(Nama Jelas)
NIM.