**Pokok Bahasan X**

**Analisis Group Means Titanic Dataset**

**Kode Pokok Bahasan**: TIK.RPL03.001.006.01

**Deskripsi Pokok Bahasan**:

Membahas tentang Studi Kasus Data Mining pada Python dengan Titanic Dataset

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| No | Elemen Kompetensi | Indikator Kinerja | Jml Jam | Hal |
| 1 | Memahami proses analisis group means | Mampu Lakukan analisis group means berdasarkan nilai survival sebagai grouping variable pada Python | 1 | 12 |
| 2 | Menerapkan decision tree untuk membangun model | Mampu melakukan pemodelan prediksi dengan decision tree | 2 | 15 |

**TUGAS PENDAHULUAN**

Hal yang harus dilakukan dan acuan yang harus dibaca sebelum praktikum :

1. Menginstal Python pada PC masing-masing praktikan.

2. Menginstal Python Studio pada PC masing-masing praktikan.

**DAFTAR PERTANYAAN**

1. Apa tujuan melakukan analisis group means?

Analisis Diskriminan adalah salah satu tehnik analisa Statistika dependensi yang memiliki kegunaan untuk mengklasifikasikan objek beberapa kelompok. Pengelompokan dengan analisis diskriminan ini terjadi karena ada pengaruh satu atau lebih variabel lain yang merupakan variabel independen.

2. apa keunggulan decision tree?

Memahami kasus dan seluruh aspek yang terkait. Menggambarkan kerangka berfikir yang sistematis. Menggambarkan struktur pengambilan keputusan yang dilakukan desicion maker sepanjang tahapan atau urutan waktu termasuk seluruh kemungkinan keputusan dan outcome. 1) Mudah dibaca dan ditafsirkan tanpa perlu pengetahuan statistik; 2) Mudah disiapkan tanpa harus menghitung dengan perhitungan yang rumit; 3) Proses Data Cleaning cenderung lebih rapih

**Penjelasan Dataset**

**Training set** harus digunakan untuk membuat model machine learning Anda. Untuk the training set, kami memberikan hasil (juga dikenal sebagai “ground truth ) untuk setiap penumpang. Model Anda akan didasarkan pada "fitur" seperti jenis kelamin dan kelas penumpang. Anda juga dapat menggunakan rekayasa fitur untuk membuat fitur baru.

**Test set** harus digunakan untuk melihat seberapa baik performa model Anda pada data yang tidak terlihat. Untuk test set, kami tidak memberikan ground truth untuk setiap penumpang. itu tugas Anda untuk memprediksi hasil ini. Untuk setiap penumpang dalam test set, gunakan model yang Anda latih untuk memprediksi apakah mereka selamat atau tidak tenggelamnya Titanic.

Kami juga menyertakan gender\_submission.csv, sekumpulan prediksi yang mengasumsikan semua dan hanya penumpang wanita yang selamat, sebagai contoh bagaimana seharusnya tampilan file pengiriman.

### Data Dictionary

**VariableDefinitionKey**

Survival 0 = No, 1 = Yes,

pclass Ticket class 1 = 1st, 2 = 2nd, 3 = 3rd,

Sex,

Age in years,

sibsp # of siblings / spouses aboard,

the Titanic parch # of parents / children aboard

the Titanic Ticket number fare Passenger

Source : https://www.kaggle.com/c/titanic/data?select=train.csv

**LAB SETUP**

Hal yang harus disiapkan dan dilakukan oleh praktikan untuk menjalankan praktikum modul ini.

1. Menginstall library yang dibutuhkan untuk mengerjakan modul.

2. Menjalankan Python.

**ELEMEN KOMPETENSI I**

**Deskripsi:**

Memahami proses analisis group means

**Kompetensi Dasar**:

Mampu Lakukan analisis group means berdasarkan nilai survival sebagai grouping variable pada Python

**Latihan 1.1.1**

**Penjelasan Singkat :**

Pada latihan ini anda akan diminta untuk Lakukan analisis group means berdasarkan nilai survival sebagai grouping variable.

**Langkah-Langkah Praktikum:**

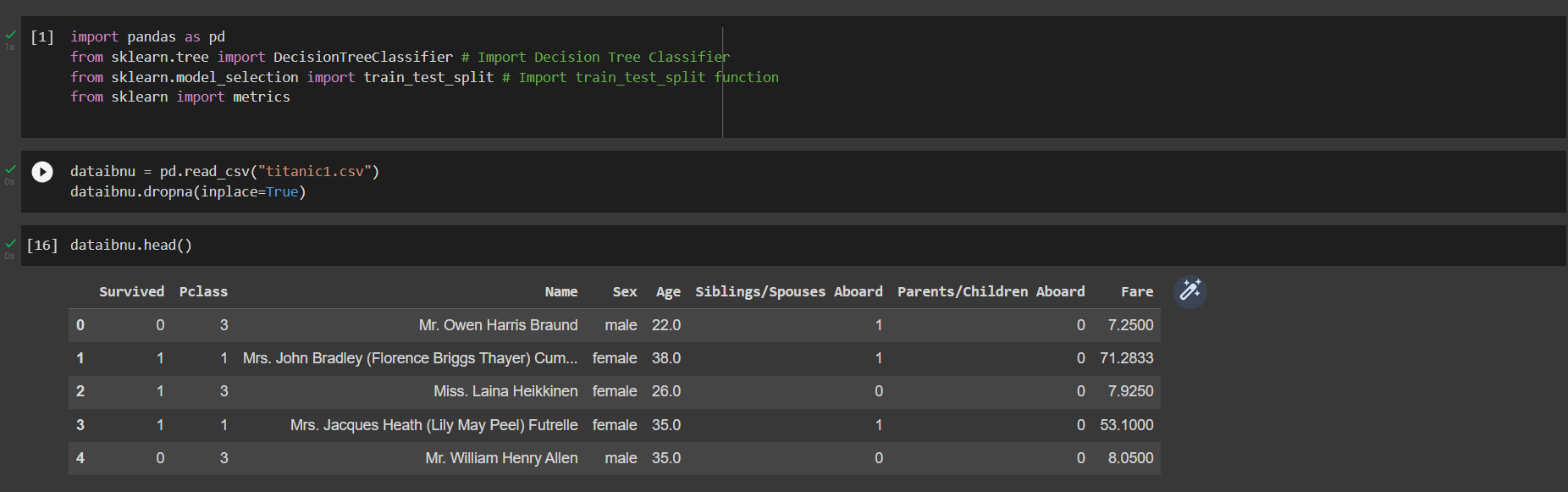
Target (class) : *Survival (1=survived; 0= not survived)*

Data :*titanic.csv*,

|  |
| --- |
| import pandas as pd  from sklearn.tree import DecisionTreeClassifier # Import Decision Tree Classifier  from sklearn.model\_selection import train\_test\_split # Import train\_test\_split function  from sklearn import metrics  data\_namapraktikan = pd.read\_csv("titanic.csv")  data\_namapraktikan.head()  data\_namapraktikan.dropna(inplace=True)  data\_namapraktikan [['Age', 'Survived']].groupby(['Survived'], as\_index=False).mean().sort\_values(by='Survived', ascending=False)  data\_namapraktikan [['Fare', 'Survived']].groupby(['Survived'], as\_index=False).mean().sort\_values(by='Survived', ascending=False) |

pd.crosstab(data['Sex'], data['Survived'])

pd.crosstab(data['Pclass'], data['Survived'])



A screenshot of a computer

Description automatically generated with medium confidence

Graphical user interface, application

Description automatically generated

Lakukan analisis boxplot untuk setiap variable predictor yang bersifat numerik. Jelaskan maknanya

**Boxplot bertujuan untuk melakukan analisis dari dua data yang disatukan, bagaimana dari kedua data itu bisa kita temukan hasil yang ingin kita lihat, sepertii dalam kasus titanic ini kita lihat umur dari seluruh penumpang dan apakah mereka selamat atau tidak, begitu juga dengan fare(tarif) apakah mempengaruhi keselamatan dari penumpang.**

import seaborn as sns

sns.boxplot(x ='Survived', y ='Age', data = data\_namapraktikan)

sns.boxplot(x ='Survived', y ='Fare', data = data\_namapraktikan)

Chart, box and whisker chart

Description automatically generated

Chart, box and whisker chart

Description automatically generated

**ELEMEN KOMPETENSI II**

**Deskripsi:**

Menerapkan decision tree untuk membangun model

**Kompetensi Dasar**:

Mampu melakukan pemodelan prediksi dengan decision tree

**Latihan 1.2.1**

**Penjelasan Singkat :**

Pada latihan ini anda akan diminta untuk Lakukan analisis group means berdasarkan nilai survival sebagai grouping variable untuk pemodelan prediksi dengan decision tree.

**Langkah-Langkah Praktikum:**

Buatlah decision tree menggunakan data training untuk membangun model yang dapat digunakan untuk memprediksi kelas survive.

!pip install six

import pandas as pd

from sklearn.tree import DecisionTreeClassifier # Import Decision Tree Classifier

from sklearn.model\_selection import train\_test\_split # Import train\_test\_split function

from sklearn import metrics

data = pd.read\_csv("titanic.csv")

data.head()

data.dropna(inplace=True)

data.replace({"male":0, "female":1}, inplace=True)

|  |
| --- |
| feature\_cols = ['Sex', 'Age', 'Fare', 'Pclass']  x = data[feature\_cols]  y = data.Survived  X\_train, X\_test, y\_train, y\_test = train\_test\_split(x, y, test\_size=0.3, random\_state=1)  clf = DecisionTreeClassifier()  # Train Decision Tree Classifer  clf = clf.fit(X\_train,y\_train)  #Predict the response for test dataset  y\_pred = clf.predict(X\_test)  feature\_cols\_new = X\_train.columns.values.tolist()  from sklearn.tree import export\_graphviz  from six import StringIO  from IPython.display import Image  import pydotplus  dot\_data = StringIO()  export\_graphviz(clf, out\_file=dot\_data,                  filled=True, rounded=True,                  special\_characters=True,feature\_names = feature\_cols\_new,class\_names=['Not Survived','Survived'])  graph = pydotplus.graph\_from\_dot\_data(dot\_data.getvalue())  Image(graph.create\_png()) |

Output :

|  |
| --- |
|  |

print("Accuracy:",metrics.accuracy\_score(y\_test, y\_pred))

pd.crosstab(y\_test, y\_pred)

Output :

|  |
| --- |
|  |

**CEK LIST**

|  |  |  |  |
| --- | --- | --- | --- |
| Elemen Kompetensi | No Latihan | Penyelesaian | |
| Selesai | Tidak selesai |
| 1 | 1.1.1 |  |  |
| 2 | 1.2.1 |  |  |

**FORM UMPAN BALIK**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Elemen Kompetensi** | **Tingkat Kesulitan** | | | **Tingkat Ketertarikan** | | | **Waktu Penyelesaian dalam menit** |
| Memahami proses analisis group means |  |  |  |  |  |  |  |
|  |  | Sangat Mudah |  |  | Tidak Tertarik | 20 |
|  |  |  |  |  |  |  |
|  |  | Mudah |  |  | Cukup Tertarik |  |
|  |  |  |  |  |  |  |
|  |  | * Biasa |  |  | Tertarik |  |
|  |  |  |  |  |  |  |
|  |  | Sulit |  |  | * Sangat Tertarik |  |
|  |  |  |  |  |  |  |
|  |  | Sangat Sulit |  |  |  |  |
|  |  |  |  |  |  |  |
| Menerapkan decision tree untuk membangun model |  |  |  |  |  |  |  |
|  |  | Sangat Mudah |  |  | Tidak Tertarik | 20 |
|  |  |  |  |  |  |  |
|  |  | Mudah |  |  | Cukup Tertarik |  |
|  |  |  |  |  |  |  |
|  |  | * Biasa |  |  | Tertarik |  |
|  |  |  |  |  |  |  |
|  |  | Sulit |  |  | * Sangat Tertarik |  |
|  |  |  |  |  |  |  |
|  |  | Sangat Sulit |  |  |  |  |