Analisis Machine Learning Data Lulusan

February 6, 2021

1 Analisa Machine Learning Data Lulusan

1.1 Import Data Lulusan

[305]:
```
import pandas as pd
import numpy as np
import chardet
from sklearn import metrics
from sklearn.naive_bayes import GaussianNB
from sklearn.tree import DecisionTreeClassifier
from sklearn.metrics import classification_report, confusion_matrix
file = 'data_lulusan.csv'

import pandas as pd
df = pd.read_csv('data_lulusan.csv')
df
```

[305]:
```
        IPK  \
0      2.49
1      3.01
2      3.07
3      2.78
4      2.92
…     …
1269   3.28
1270   3.03
1271   3.53
1272   3.09
1273   3.10

       Pengetahuan di bidang atau disiplin ilmu anda
(rendah)1-5(tinggi)  \
0                                                    5
1                                                    5
2                                                    5
3                                                    5
4                                                    5
…                                                  …
```

```
1269                                                    5
1270                                                    4
1271                                                    4
1272                                                    5
1273                                                    5

      Pengetahuan di luar bidang atau disiplin ilmu anda
 (rendah)1-5(tinggi)  \
0                                                       5
1                                                       5
2                                                       5
3                                                       5
4                                                       5
…                                                       …
1269                                                    5
1270                                                    5
1271                                                    4
1272                                                    5
1273                                                    5

      Pengetahuan umum                  (rendah)1-5(tinggi)  \
0                                                       5
1                                                       5
2                                                       5
3                                                       5
4                                                       5
…                                                       …
1269                                                    5
1270                                                    5
1271                                                    5
1272                                                    5
1273                                                    5

      Ketrampilan internet                        (rendah)1-5(tinggi)  \
0                                                       5
1                                                       5
2                                                       5
3                                                       5
4                                                       5
…                                                       …
1269                                                    5
1270                                                    5
1271                                                    5
1272                                                    5
1273                                                    5

      Ketrampilan komputer                        (rendah)1-5(tinggi)  \
```

```
0                                                    5
1                                                    5
2                                                    5
3                                                    5
4                                                    5
…                                                    …
1269                                                 5
1270                                                 5
1271                                                 5
1272                                                 5
1273                                                 5

      Ketrampilan                    Sistem Keamanan Komputer
(rendah)1-5(tinggi)  \
0                                                    5
1                                                    3
2                                                    3
3                                                    3
4                                                    3
…                                                    …
1269                                                 4
1270                                                 3
1271                                                 2
1272                                                 3
1273                                                 3

      Ketrampilan                    Jaringan Komputer (rendah)1-5(tinggi)  \
0                                                    5
1                                                    3
2                                                    3
3                                                    3
4                                                    3
…                                                    …
1269                                                 3
1270                                                 3
1271                                                 3
1272                                                 4
1273                                                 3

      Ketrampilan Program Aplikasi WEB                   (rendah)1-5(tinggi)  \
0                                                    3
1                                                    3
2                                                    3
3                                                    3
4                                                    3
…                                                    …
1269                                                 5
```

```
1270                                                    5
1271                                                    3
1272                                                    4
1273                                                    5

      Ketrampilan Program Aplikasi Multimedia
(rendah)1-5(tinggi)  \
0                                                       3
1                                                       3
2                                                       3
3                                                       3
4                                                       3
…                                                     …
1269                                                    5
1270                                                    5
1271                                                    3
1272                                                    5
1273                                                    5

      …  Toleransi                              (rendah)1-5(tinggi)  \
0     …                                                    5
1     …                                                    5
2     …                                                    5
3     …                                                    5
4     …                                                    5
…   …                                                    …
1269  …                                                    5
1270  …                                                    5
1271  …                                                    5
1272  …                                                    5
1273  …                                                    5

      Kemampuan adaptasi                          (rendah)1-5(tinggi)  \
0                                                       5
1                                                       5
2                                                       5
3                                                       5
4                                                       5
…                                                     …
1269                                                    5
1270                                                    5
1271                                                    3
1272                                                    5
1273                                                    5

      Loyalitas                                 (rendah)1-5(tinggi)  \
0                                                       5
```

1	5
2	5
3	5
4	5
…	…
1269	5
1270	5
1271	5
1272	5
1273	5

	Integritas f	(rendah)1-5(tinggi) \
0		5
1		5
2		5
3		5
4		5
…		…
1269		5
1270		5
1271		5
1272		5
1273		5

	Kepemimpinan	(rendah)1-5(tinggi) \
0		5
1		3
2		3
3		3
4		3
…		…
1269		4
1270		3
1271		5
1272		3
1273		5

	Kemampuan dalam memegang tanggungjawab (rendah)1-5(tinggi) \
0	5
1	5
2	5
3	5
4	5
…	…
1269	5
1270	5

```
1271                                                  5
1272                                                  5
1273                                                  5

      Inisiatif                              (rendah)1-5(tinggi)  \
0                                                     5
1                                                     5
2                                                     5
3                                                     5
4                                                     5
…                                                     …
1269                                                  5
1270                                                  5
1271                                                  5
1272                                                  5
1273                                                  5

      Kemampuan untuk memresentasikan ide/produk/laporan
(rendah)1-5(tinggi)   \
0                                                     3
1                                                     5
2                                                     5
3                                                     5
4                                                     5
…                                                     …
1269                                                  3
1270                                                  4
1271                                                  3
1272                                                  4
1273                                                  4

      Kemampuan dalam menulis laporan, memo dan dokumen
(rendah)1-5(tinggi)  \
0                                                     3
1                                                     5
2                                                     5
3                                                     5
4                                                     5
…                                                     …
1269                                                  3
1270                                                  3
1271                                                  5
1272                                                  5
1273                                                  4

         Posisi Pekerjaan
0      Networking engineer
```

```
1     System administrator
2     System administrator
3     System administrator
4     System administrator
…                     …
1269     Software engineer
1270        Web Programmer
1271    Staff Administrasi
1272    IT Quality Control
1273         Design Grafis

[1274 rows x 32 columns]
```

1.2 Distribusi Label

Label pada data ini ada pada kolom Posisi Pekerjaan dan akan dirubah menjadi kolom label untuk dipisahkan

[306]:
```
df = df.rename(columns = {'Posisi Pekerjaan': 'label'})
df['label'].value_counts()
```

[306]:
```
System administrator    214
Programmer              113
EDP Operator             97
Operator Komputer        56
Staff IT                 54
Web Designer             52
Technical engineer       51
Software engineer        48
Staff Administrasi       47
IT Support               46
Teknik Mesin             45
Sekretaris               42
IT Quality Control       41
Networking engineer      41
Sistem analis            39
Digital Marketing        36
Call Centre              36
Design Grafis            36
Staff IT Network         34
MIS Director             30
Web Programmer           29
Recepsionist             23
Customer Service         21
Sales                    16
Agent Monitoring ATM     15
Guru TIK                 12
```

```
Name: label, dtype: int64
```

1.3 Label Encoding

Label encoding mengubah setiap nilai alfabetik dalam kolom menjadi angka yang berurutan

[307]:
```
from sklearn.preprocessing import LabelEncoder
le_posisi_pekerjaan = LabelEncoder()
le_posisi_pekerjaan.fit(df["label"].astype(str))

df['label'] = le_posisi_pekerjaan.transform(df["label"].astype(str))
```

1.4 Membagi Data menjadi Set Pelatihan dan Pengujian

Untuk menilai prediksi kita, kita perlu menggunakan satu set pelatihan dan pengujian. Model tersebut belajar dari data pelatihan dan kemudian membuat prediksi pada data pengujian. Karena kami memiliki jawaban yang benar untuk data pengujian, kami dapat mengetahui seberapa baik model dapat menggeneralisasi ke data baru. Penting untuk hanya menggunakan set pengujian sekali, karena ini dimaksudkan sebagai perkiraan seberapa baik performa model pada data baru.

Kami akan menyimpan 20% contoh untuk pengujian.

[370]:
```
from sklearn.model_selection import train_test_split

RSEED =50
# Extract the labels
Y = df['label']
X = df.iloc[:,:-1].values

# 20% examples in test data
X_train, X_test, y_train, y_test = train_test_split(X, Y, test_size=0.2)
```

Imputasi nilai yang hilang Kita akan mengisi nilai yang hilang dengan mean dari kolom. Penting untuk dicatat bahwa kita mengisi nilai yang hilang di set pengujian dengan mean kolom di data pelatihan. Ini diperlukan karena jika kita mendapatkan data baru, kita harus menggunakan data pelatihan untuk mengisi nilai yang hilang.

[336]:
```
X_train.shape
```

[336]:
```
(1019, 31)
```

[337]:
```
X_test.shape
```

[337]:
```
(255, 31)
```

1.5 Random Forest

[338]:
```
from sklearn.ensemble import RandomForestClassifier

# Create the model with 100 trees
rf_model = RandomForestClassifier(n_estimators=100)

# Fit on training data
rf_model.fit(X_train, y_train)

rf_predictions = model.predict(X_test)
rf_probs = model.predict_proba(test)[:, 1]


print("Accuracy:",metrics.accuracy_score(y_test, rf_predictions))
```

```
Accuracy: 1.0
```

1.6 Adaboost Classifier

[371]:
```
from sklearn.ensemble import AdaBoostClassifier
# Create adaboost classifer object
abc = AdaBoostClassifier(n_estimators=50,
                         learning_rate=1)
# Train Adaboost Classifer
abc_model = abc.fit(X_train, y_train)

#Predict the response for test dataset
abc_pred = abc_model.predict(X_test)


print("Accuracy:",metrics.accuracy_score(y_test, abc_pred))
```

```
Accuracy: 0.8941176470588236
```

1.7 Extra Trees Classifier

[372]:
```
from sklearn.ensemble import ExtraTreesClassifier
# Create adaboost classifer object
et = ExtraTreesClassifier()
# Train Adaboost Classifer
et_model = et.fit(X_train, y_train)

#Predict the response for test dataset
et_pred = et_model.predict(X_test)


print("Accuracy:",metrics.accuracy_score(y_test, et_pred))
```

```
Accuracy: 1.0
```

[2]:
```
import sys
if "/Library/TeX/texbin" not in sys.path:
    print('adding path') # I just add this to know if the path was present or␣
 ↪not.
    sys.path.append("/Library/TeX/texbin")
```

```
adding path
```