

Modul Praktikum Kecerdasan Buatan



Rolly Maulana Awangga
0410118609

Applied Bachelor of Informatics Engineering
Program Studi D4 Teknik Informatika

Applied Bachelor Program of Informatics Engineering
Politeknik Pos Indonesia

Bandung 2019

‘Jika Kamu tidak dapat menahan lelahnya belajar,
Maka kamu harus sanggup menahan perihnya Kebodohan.’
Imam Syafi’i

Acknowledgements

Pertama-tama kami panjatkan puji dan syukur kepada Allah SWT yang telah memberikan rahmat dan hidayah-Nya sehingga Buku Pedoman Tingkat Akhir ini dapat diselesaikan.

Abstract

Buku Pedoman ini dibuat dengan tujuan memberikan acuan, bagi mahasiswa Tingkat Akhir dan dosen Pembimbing. Pada intinya buku ini menjelaskan secara lengkap tentang Standar pengerjaan Intership dan Tugas Akhir di Program Studi D4 Teknik Informatika, dan juga mengatur mekanisme, teknik penulisan, serta penilaiannya. Dengan demikian diharapkan semua pihak yang terlibat dalam aktivitas Bimbingan Mahasiswa Tingkat Akhir berjalan lancar dan sesuai dengan standar.

Contents

1	Mengenal Kecerdasan Buatan dan Scikit-Learn	1
1.1	Teori	1
1.2	Instalasi	2
1.3	Penanganan Error	2
1.4	Teori/Mhd Zulfikar Akram Nasution/1164081	2
1.5	Jesron Marudut Hatuan/1164077	5
1.5.1	Teori	5
1.5.2	Instalasi	7
1.5.2.1	Instalasi Library Scikit dari Anaconda	7
1.6	Teori/Puad Hamdani/1164084	8
1.7	Instalasi/Mhd Zulfikar Akram Nasution/1164081	13
1.7.1	Instalasi	15
1.7.1.1	Loading an Example Datasets	15
1.8	Learning and Predicting	15
1.8.0.1	Model Persistence	16
1.8.0.2	Conventions	17
1.9	Penanganan Error	19
2	Related Works	24
2.1	puad hamdani/1164084	24
2.1.1	binary classification	24
2.1.2	supervised learning dan unsupervised learning	24
2.1.3	evaluasi dan akurasi	25
2.1.4	bagaimana cara membuat dan membaca confusion matrix, buat confusion matrix	25
2.1.5	bagaimana K-fold cross validation bekerja dengan gambar ilus- trasi	25
2.1.6	decision tree	26

2.1.7	Information Gain	26
2.2	Puad Hamdani/ 1164084	26
2.2.1	Scikit-Learn	26
2.3	Penanganan Error	32
2.4	Same Topics	32
2.4.1	Topic 1	32
2.4.2	Topic 2	32
2.5	Same Method	32
2.5.1	Method 1	32
2.5.2	Method 2	33
3	Methods	34
3.1	The data	34
3.2	Method 1	34
3.3	Method 2	34
4	Experiment and Result	35
4.1	Experiment	35
4.2	Result	35
5	Conclusion	36
5.1	Conclusion of Problems	36
5.2	Conclusion of Method	36
5.3	Conclusion of Experiment	36
5.4	Conclusion of Result	36
6	Discussion	37
7	Discussion	38
8	Discussion	39
9	Discussion	40
10	Discussion	41
11	Discussion	42
12	Discussion	43

13 Discussion	44
14 Discussion	45
A Form Penilaian Jurnal	46
B FAQ	49
Bibliography	51

List of Figures

1.1	Install Scikit-Learn Conda	4
1.2	Install Scikit-Learn ke Python	4
1.3	Kompilasi Kode	5
1.4	Import Datasets	5
1.5	Buat variable iris	5
1.6	Buat variable digits	6
1.7	Applikasi Anaconda.	7
1.8	Versi Anaconda.	7
1.9	Instalasi.	8
1.10	Langkah instalasi anaconda.	8
1.11	Langkah terakhir.	9
1.12	Proses Instalasi	11
1.13	Gabung Conda dan Python	11
1.14	Kompilasi Kode	12
1.15	Variable Digits	12
1.16	13
1.17	13
1.18	14
1.19	14
1.20	15
1.21	15
1.22	16
1.23	16
1.24	17
1.25	17
1.26	Import file svm	18
1.27	Buat variable Classifier	18
1.28	Lihat array baru dengan syntac Python	19

1.29	Lihat classifier array	19
1.30	Import file	20
1.31	Variable classifier	20
1.32	Variable iris	20
1.33	Penyesuaian Classifier	20
1.34	Import Pickle	20
1.35	Import numpy	20
1.36	Variable rng	20
1.37	Variable X dan hasil random	20
1.38	Variable transformer random	20
1.39	Variable X new type pada transformer	21
1.40	Hasil dari X new type pada transformer	21
1.41	Screenshoot Error	21
1.42	Install Joblib	21
1.43	Solusi Error	21
1.44	Perintah sklearn import datasets	21
1.45	Perintah Variabel Iris	22
1.46	Perintah Variabel Digits	22
1.47	Error Import	22
1.48	Install Library Joblib	22
1.49	Berhasil Import Library Joblib	23
2.1	Load Dataset	26
2.2	Generate Binary Label	27
2.3	Pemanggilan get dummies	27
2.4	Mendefinisikan pembagian data	28
2.5	Membuktikan pengujian	28
2.6	Gambaran decision tree	29
2.7	Library Graphviz	29
2.8	Menampilkan hasil perhitungan 2 parameter	30
2.9	Mendefinisikan library sklearn	30
2.10	Menampilkan hasil fungsi max depth dan accuracy	31
2.11	Menjelaskan variable kari	31
2.12	Menjelaskan dan menampilkan gambar grafik	32
A.1	Form nilai bagian 1.	47
A.2	form nilai bagian 2.	48

Chapter 1

Mengenai Kecerdasan Buatan dan Scikit-Learn

Buku umum yang digunakan adalah [2] dan untuk sebelum UTS menggunakan buku *Python Artificial Intelligence Projects for Beginners*[1]. Dengan praktek menggunakan python 3 dan editor anaconda dan library python scikit-learn. Tujuan pembelajaran pada pertemuan pertama antara lain:

1. Mengerti definisi kecerdasan buatan, sejarah kecerdasan buatan, perkembangan dan penggunaan di perusahaan
2. Memahami cara instalasi dan pemakaian sci-kit learn
3. Memahami cara penggunaan variabel explorer di spyder

Tugas dengan cara dikumpulkan dengan pull request ke github dengan menggunakan latex pada repo yang dibuat oleh asisten riset.

1.1 Teori

Praktek teori penunjang yang dikerjakan :

1. Buat Resume Definisi, Sejarah dan perkembangan Kecerdasan Buatan, dengan bahasa yang mudah dipahami dan dimengerti. Buatan sendiri bebas plagiat[hari ke 1](10)
2. Buat Resume mengenai definisi supervised learning, klasifikasi, regresi dan unsupervised learning. Data set, training set dan testing set.[hari ke 1](10)

1.2 Instalasi

Membuka <https://scikit-learn.org/stable/tutorial/basic/tutorial.html>. Dengan menggunakan bahasa yang mudah dimengerti dan bebas plagiat. Dan wajib skrinsut dari komputer sendiri.

1. Instalasi library scikit dari anaconda, mencoba kompilasi dan uji coba ambil contoh kode dan lihat variabel explorer[hari ke 1](10)
2. Mencoba Loading an example dataset, menjelaskan maksud dari tulisan tersebut dan mengartikan per baris[hari ke 1](10)
3. Mencoba Learning and predicting, menjelaskan maksud dari tulisan tersebut dan mengartikan per baris[hari ke 2](10)
4. mencoba Model persistence, menjelaskan maksud dari tulisan tersebut dan mengartikan per baris[hari ke 2](10)
5. Mencoba Conventions, menjelaskan maksud dari tulisan tersebut dan mengartikan per baris[hari ke 2](10)

1.3 Penanganan Error

Dari percobaan yang dilakukan di atas, apabila mendapatkan error maka:

1. skrinsut error[hari ke 2](10)
2. Tuliskan kode eror dan jenis errornya [hari ke 2](10)
3. Solusi pemecahan masalah error tersebut[hari ke 2](10)

iiiiii HEAD

1.4 Teori/Mhd Zulfikar Akram Nasution/1164081

1. Definisi, Sejarah dan Perkembangan Kecerdasan Buatan

- Definisi

Kecerdasan Buatan adalah kecerdasan yang ditambahkan kepada suatu sistem yang bisa diatur dalam konteks ilmiah yang berhubungan dengan pemanfaatan mesin untuk memecahkan persoalan yang rumit dengan cara yang lebih manusiawi.

- Sejarah dan Perkembangan

Sejarah dan perkembangan kecerdasan buatan terjadi pada musim panas tahun 1956 tercatat adanya seminar mengenai AI di Darmouth College. Seminar pada waktu itu dihadiri oleh sejumlah pakar komputer dan membahas potensi komputer dalam meniru kepandaian manusia. Akan tetapi perkembangan yang sering terjadi semenjak diciptakannya LISP, yaitu bahasa kecerdasan buatan yang dibuat tahun 1960 oleh John McCarthy. Istilah pada kecerdasan buatan atau Artificial Intelligence diambil dari Marvin Minsky dari MIT. Dia menulis karya ilmiah berjudul Step towards Artificial Intelligence, The Institute of radio Engineers Proceedings 49, January 1961.

2. Definisi Supervised Learning, Unsupervised Learning, Klasifikasi, Regresi, Data Set, Training Set dan Testing Set

- Supervised Learning dan Unsupervised Learning

Supervised learning merupakan sebuah pendekatan dimana sudah terdapat data yang dilatih, dan terdapat variable yang ditargetkan sehingga tujuan dari pendekatan ini adalah mengelompokkan suatu data ke data yang sudah ada. Sedangkan unsupervised learning tidak memiliki data latih, sehingga dari data yang ada, kita mengelompokkan data tersebut menjadi 2 bagian atau 3 bagian dan seterusnya.

- Klasifikasi

Klasifikasi adalah salah satu topik utama dalam data mining atau machine learning. Klasifikasi yaitu suatu pengelompokan data dimana data yang digunakan tersebut mempunyai kelas label atau target.

- Regresi

Regresi adalah Supervised learning tidak hanya mempelajari classifier, tetapi juga mempelajari fungsi yang dapat memprediksi suatu nilai numerik. Contoh, ketika diberi foto seseorang, kita ingin memprediksi umur, tinggi, dan berat orang yang ada pada foto tersebut.

- Data Set

Data set adalah cabang aplikasi dari Artificial Intelligence/Kecerdasan Buatan yang fokus pada pengembangan sebuah sistem yang mampu belajar sendiri tanpa harus berulang kali di program oleh manusia.

- Training Set

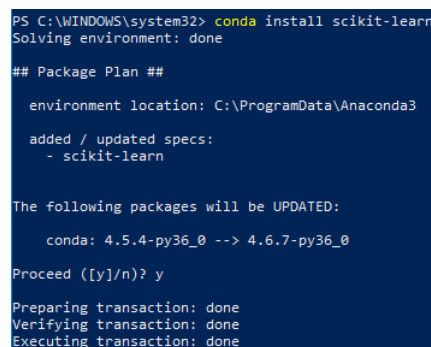
Training set yaitu jika pasangan objek, dan kelas yang menunjuk pada objek tersebut adalah suatu contoh yang telah diberi label akan menghasilkan suatu algoritma pembelajaran.

- Testing Set

Testing set digunakan untuk mengukur sejauh mana classifier berhasil melakukan klasifikasi dengan benar.

3. Instalasi Scikit-Learn dari Anaconda

- Pertama install Anaconda di pc masing-masing
- Kemudian buka cmd untuk menginstall scikit-learn
- Ketik perintah "conda install scikit-learn" dan pilih "y"



```
PS C:\WINDOWS\system32> conda install scikit-learn
Solving environment: done

## Package Plan ##

  environment location: C:\ProgramData\Anaconda3

  added / updated specs:
    - scikit-learn

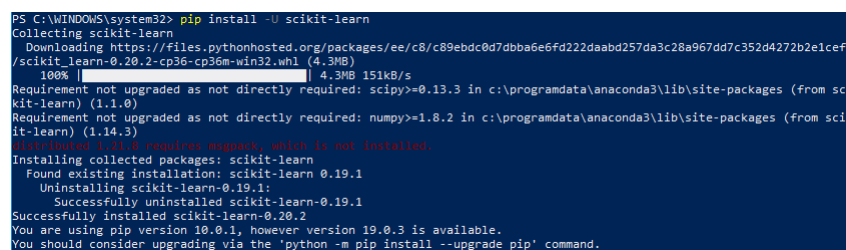
The following packages will be UPDATED:

    conda: 4.5.4-py36_0 --> 4.6.7-py36_0

Proceed ([y]/n)? y
Preparing transaction: done
Verifying transaction: done
Executing transaction: done
```

Figure 1.1: Install Scikit-Learn Conda

- Lalu ketik "pip install -U scikit-learn" untuk memasukkan anaconda ke python



```
PS C:\WINDOWS\system32> pip install -U scikit-learn
Collecting scikit-learn
  Downloading https://files.pythonhosted.org/packages/ee/c8/c89ebdc0d7dbba6e6fd222daabd257da3c28a967dd7c352d4272b2e1cef6/scikit_learn-0.20.2-cp36-cp36m-win32.whl (4.3MB)
    100% |#####| 4.3MB 151kB/s
Requirement not upgraded as not directly required: scipy>=0.13.3 in c:\programdata\anaconda3\lib\site-packages (from scikit-learn) (1.1.0)
Requirement not upgraded as not directly required: numpy>=1.8.2 in c:\programdata\anaconda3\lib\site-packages (from scikit-learn) (1.14.3)
Installing collected packages: scikit-learn
  Found existing installation: scikit-learn 0.19.1
    Uninstalling scikit-learn-0.19.1:
      Successfully uninstalled scikit-learn-0.19.1
Successfully installed scikit-learn-0.20.2
You are using pip version 10.0.1, however version 19.0.3 is available.
You should consider upgrading via the 'python -m pip install --upgrade pip' command.
```

Figure 1.2: Install Scikit-Learn ke Python

- Setelah itu, kompilasi kode di dalam python dengan ketik "python", lalu "print('Zulfikar')", maka akan menghasilkan seperti gambar berikut.

```
PS C:\WINDOWS\system32> python
Python 3.6.5 [Anaconda, Inc.] (default, Mar 29 2018, 13:23:52) [MSC v.1900 32 bit (Intel)] on win32
Type "help", "copyright", "credits" or "license" for more information.
>>> print('Zulfikar')
Zulfikar
```

Figure 1.3: Kompilasi Kode

4. Loading an Example Dataset

- Ketik perintah berikut "from sklearn import datasets" untuk mengimport dataset dari sklearn.

```
C:\Users\user>python
Python 3.6.5 [Anaconda, Inc.] (default, Mar 29 2018, 13:23:52) [MSC v.1900 32 bit (Intel)] on win32
Type "help", "copyright", "credits" or "license" for more information.
>>> from sklearn import datasets
```

Figure 1.4: Import Datasets

- Kemudian ketik perintah berikut untuk membuat variable iris yang berisi datasets.

```
>>> iris = datasets.load_iris()
```

Figure 1.5: Buat variable iris

- Lalu ketik perintah berikut untuk membuat variable digits yang berisi datasets, dan juga untuk melihat isi data dari datasets seperti gambar 1.6

1.5 Jesron Marudut Hatuan/1164077

1.5.1 Teori

1. Definisi, sejarah, dan perkembangan kecerdasan buatan.

Kecerdasan Buatan (Artificial Intelligence atau AI) dapat didefinisikan sebagai kecerdasan yang ditunjukkan oleh suatu entitas buatan. Sistem seperti ini biasanya dianggap komputer. Kecerdasan diciptakan lalu dimasukkan ke dalam suatu mesin atau komputer supaya dapat melakukan pekerjaan-pekerjaan yang dapat dilakukan manusia.

Sebenarnya area Kecerdasan Buatan (Artificial Intelligence) atau disingkat dengan AI, dimulai dari munculnya komputer sekitar tahun 1940-an, meskipun

```

>>> digits = datasets.load_digits()
>>> print(digits.data)
[[ 0.  0.  5. ... 0.  0.  0.]
 [ 0.  0.  0. ... 10.  0.  0.]
 [ 0.  0.  0. ... 16.  9.  0.]
 ...
 [ 0.  0.  1. ... 6.  0.  0.]
 [ 0.  0.  2. ... 12.  0.  0.]
 [ 0.  0. 10. ... 12.  1.  0.]]
>>>

```

Figure 1.6: Buat variable digits

sejarah perkembangannya dapat dilacak dari zaman Mesir kuno. Pada akhir tahun 1955, Newell dan Simon mengembangkan The Logic Theorist atau program AI terdahulu. Program ini merepresentasikan masalah sebagai model pohon, lalu penyelesaiannya dengan memilih cabang yang akan menghasilkan kesimpulan terbenar. Program tersebut berdampak besar dan menjadi batu loncatan dalam mengembangkan bidang AI. Pada tahun 1956 John McCarthy dari Massachusetts Institute of Technology dianggap sebagai bapak AI, menyelenggarakan konferensi untuk menarik para ahli komputer bertemu, dengan nama kegiatan The Dartmouth Summer Research Project On AI. Konferensi Dartmouth saat itu mempertemukan para pendiri dalam AI, dan bertugas untuk meletakkan dasar bagi masa depan pengembangan dan penelitian AI. John McCarthy disaat itu mengusulkan definisi AI adalah AI merupakan cabang dari ilmu komputer yang berfokus pada pengembangan komputer agar mempunyai kemampuan dan berperilaku seperti manusia.

2. Definisi supervised learning, klasifikasi, regresi, dan unsupervised learning. Data set, training set dan testing set.

Supervised learning merupakan sebuah pendekatan dimana sudah terdapat data yang dilatih, dan terdapat variable yang ditargetkan sehingga tujuan dari pendekatan ini adalah mengelompokkan suatu data ke data yang sudah ada. Sedangkan unsupervised learning tidak memiliki data latih, sehingga dari data yang ada, kita mengelompokkan data tersebut menjadi 2 bagian atau 3 bagian dan seterusnya.

Klasifikasi adalah salah satu topik utama dalam data mining atau machine learning. Klasifikasi yaitu suatu pengelompokan data dimana data yang digunakan tersebut mempunyai kelas label atau target.

Regresi adalah Supervised learning tidak hanya mempelajari classifier, tetapi juga mempelajari fungsi yang dapat memprediksi suatu nilai numerik. Contoh,

ketika diberi foto seseorang, kita ingin memprediksi umur, tinggi, dan berat orang yang ada pada foto tersebut.

Data set adalah cabang aplikasi dari Artificial Intelligence/Kecerdasan Buatan yang fokus pada pengembangan sebuah sistem yang mampu belajar sendiri tanpa harus berulang kali di program oleh manusia.

Training set yaitu jika pasangan objek, dan kelas yang menunjuk pada objek tersebut adalah suatu contoh yang telah diberi label akan menghasilkan suatu algoritma pembelajaran.

Testing set digunakan untuk mengukur sejauh mana classifier berhasil melakukan klasifikasi dengan benar[?].

1.5.2 Instalasi

1.5.2.1 Instalasi Library Scikit dari Anaconda

1. Sediakan aplikasi Anaconda terlebih dahulu

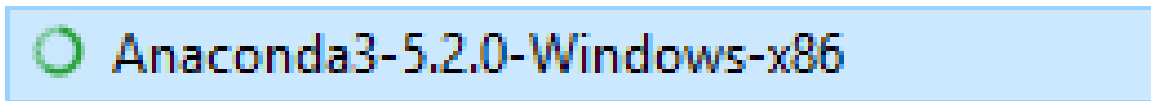


Figure 1.7: Aplikasi Anaconda.

2. Setelah di install, masukkan script dibawah ini untuk melihat versi Python dan Anacondanya

```
C:\WINDOWS\system32>conda --version
conda 4.6.7

C:\WINDOWS\system32>python --version
Python 3.6.5 :: Anaconda, Inc.

C:\WINDOWS\system32>
```

Figure 1.8: Versi Anaconda.

3. Selanjutnya masukkan perintah 'pip install -U scikit-learn'
4. Selanjutnya masukkan perintah 'conda install scikit-learn'


```

C:\WINDOWS\system32>pip install -U scikit-learn
Collecting scikit-learn
  Using cached https://files.pythonhosted.org/packages/ee/c8/c89ebdc0d7dbba6e6fd222daabd257da3c28a967dd7c352d4272b2e1cef6/scikit_learn-0.20.2-cp36-cp36m-win32.whl
Requirement not upgraded as not directly required: numpy>=1.8.2 in c:\programdata\anaconda3\lib\site-packages (from scikit-learn) (1.14.3)
Requirement not upgraded as not directly required: scipy>=0.13.3 in c:\programdata\anaconda3\lib\site-packages (from scikit-learn) (1.1.0)
Distributed 1.21.8 requires msgpack, which is not installed.
Installing collected packages: scikit-learn
  Found existing installation: scikit-learn 0.19.1
    Uninstalling scikit-learn-0.19.1:
      Successfully uninstalled scikit-learn-0.19.1
Successfully installed scikit-learn-0.20.2
You are using pip version 10.0.1, however version 19.0.3 is available.
You should consider upgrading via the 'python -m pip install --upgrade pip' command.

```

Figure 1.9: Instalasi.

```

C:\WINDOWS\system32>conda install scikit-learn
Solving environment: done

## Package Plan ##

  environment location: C:\ProgramData\Anaconda3

  added / updated specs:
    - scikit-learn

The following packages will be UPDATED:

  conda: 4.5.4-py36_0 --> 4.6.7-py36_0

Proceed ([y]/n)? y

Preparing transaction: done
Verifying transaction: done
Executing transaction: done

C:\WINDOWS\system32>

```

Figure 1.10: Langkah instalasi anaconda.

5. Selanjutnya masukkan perintah 'python' dan 'print ('jesron')

```

##### HEAD
=====

```

1.6 Teori/Puad Hamdani/1164084

1. Definisi, Sejarah dan Perkembangan Kecerdasan Buatan

- Definisi

Kecerdasan buatan adalah ilmu pengetahuan yang berhubungan dengan mesin untuk memecahkan persoalan rumit dengan cara yang mudah, dilakukan

```

C:\WINDOWS\system32\python
Python 3.6.5 |Anaconda, Inc.| (default, Mar 29 2018, 13:23:52) [MSC v.1900 32 bit (Intel)] on
win32
Type "help", "copyright", "credits" or "license()" for more information.
>> print('Jeeron')
Jeeron
>>>

```

Figure 1.11: Langkah terakhir.

dengan mengikuti kecerdasan manusia dan menerapkannya di computer sebagai algoritma

- Sejarah dan Perkembangan

AI (artificial Intelligence) di kenal sekitar tahun 1943 Teori tentang jaringan saraf tiruan (artificial neuron network, ANN) menyatakan bahwa setiap neuron dapat dimisalkan dalam keadaan biner, yaitu ON dan OFF. Dari setiap percobaan, setiap fungsi perhitungan dapat diselesaikan melalui jaringan neuron yang dimodelkan. Pada tahun 1965, Lotfi Zadeh, professor teknik elektro di University of California, memublikasikan konsepnya yang disebut dengan “fuzzy sets”. Beliau menjabarkan FL dengan pernyataan matematis dan visual yang mudah dipahami. Karena kajian ini berkaitan dengan sistem kontrol, konsep tersebut banyak dikembangkan dalam konteks pemrograman komputer hingga saat ini.

2. Definisi Supervised Learning, Unsupervised Learning, Klasifikasi, Regresi, Data Set, Training Set dan Testing Set

- Supervised Learning

Supervised learning adalah pembelajaran yang terawasi dimana jika output yang diharapkan telah diketahui sebelumnya. Biasanya pembelajaran ini dilakukan dengan menggunakan data yang telah ada

- Unsupervised Learning

Unsupervised learning adalah pembelajaran yang tidak terawasi dimana tidak memerlukan target output. Metode ini tidak dapat ditentukan hasil seperti apa yang diharapkan selama proses pembelajaran, Nilai bobot yang disusun dalam proses range tertentu tergantung pada output yang diberikan.

- Klasifikasi

Klasifikasi adalah Proses pengelompokkan berdasarkan ciri-ciri persamaan dan perbedaan

- Regresi

Regresi adalah metode analisis statistik yang digunakan untuk melihat pengaruh antara dua atau lebih variabel

- Data Set

Data set adalah objek yang merepresentasikan data dan relasinya di memory, Strukturnya hampir mirip dengan data di data base. Data set berisi koleksi dari data table dan data relation

- Training Set

Training set adalah bagian dataset yang kita latih untuk membuat prediksi atau algoritma ML lainnya sesuai tujuannya masing-masing. Kita memberikan petunjuk melalui algoritma agar mesin yang kita latih bisa mencari korelasinya sendiri. Walau demikian proses belajar harusnya proporsional. Layaknya seorang murid yang terlalu diforsir belajar, maka hasilnya pun tidak akan baik. Dalam istilah ML disebut dengan overfitting. Akan lebih mudah memahami konsep overfitting melalui praktek.

- Testing Set

Test set adalah bagian dataset yang kita tes untuk melihat keakuratannya, atau dengan kata lain melihat performanya.

3. Instalasi Scikit-Learn dari Anaconda

- Pertama install Anaconda di pc masing-masing
- Kemudian buka cmd untuk menginstall scikit-learn
- Ketikan "conda install scikit-learn" dan pilih "y"
- ketik "pip install -U scikit-learn" untuk menggabungkan anaconda dan python
- Setelah itu, kompilasi kode di dalam python dengan ketik "python", lalu "print('puad')" maka akan menghasilkan seperti gambar berikut.

4. Loading an Example Dataset

- Ketik perintah berikut "from sklearn import datasets" untuk mengimport dataset dari sklearn.

```

Administrator: Command Prompt
Microsoft Windows [Version 10.0.17134.590]
(c) 2018 Microsoft Corporation. All rights reserved.

C:\WINDOWS\system32>conda install scikit-learn
Solving environment: done

## Package Plan ##

  environment location: C:\ProgramData\Anaconda3

  added / updated specs:
    - scikit-learn

The following packages will be UPDATED:

    conda: 4.5.4-py36_0 --> 4.6.7-py36_0

Proceed ([y]/n)? y

Preparing transaction: done
Verifying transaction: done
Executing transaction: done

C:\WINDOWS\system32>

```

Figure 1.12: Proses Instalasi

```

C:\WINDOWS\system32>pip install -U scikit-learn
Collecting scikit-learn
  Downloading https://files.pythonhosted.org/packages/ee/c8/c89ebdc0d7dbba6e6fd222daabd257da3c28a967dd7c352d4272b2e1cef6/scikit_learn-0.20.2-cp36-cp36m-win32.whl (4.3MB)
    100% |#####| 4.3MB 1.2MB/s
Requirement not upgraded as not directly required: scipy>=0.13.3 in c:\programdata\anaconda3\lib\site-packages (from scikit-learn) (1.1.0)
Requirement not upgraded as not directly required: numpy>=1.8.2 in c:\programdata\anaconda3\lib\site-packages (from scikit-learn) (1.14.3)
distributed 1.21.0 requires msgpack, which is not installed.
Installing collected packages: scikit-learn
  Found existing installation: scikit-learn 0.19.1
    Uninstalling scikit-learn-0.19.1:
      Successfully uninstalled scikit-learn-0.19.1
Successfully installed scikit-learn-0.20.2
You are using pip version 10.0.1, however version 19.0.3 is available.
You should consider upgrading via the 'python -m pip install --upgrade pip' command.

C:\WINDOWS\system32>

```

Figure 1.13: Gabung Conda dan Python

- ketika perintah "iris = datasets.load iris" untuk membuat variable iris yang berisi datasets.
- ketika perintah berikut "digits = datasets.load digits" untuk membuat variable digits yang berisi datasets, dan juga "print(digits.data) " untuk melihat isi data dari datasets seperti gambar
- kemudian ketika "digits target"
- kemudian ketika "digits.images[0] "
- kemudian ketika "from sklearn import svm" dan kemudian clf = svm.SVC(gamma=0.001, C=100.) "
- kemudian ketika "clf.fit(digits.data[:-1], digits.target[:-1]) "

```

C:\WINDOWS\system32>python
Python 3.6.5 [Anaconda, Inc.] (default, Mar 29 2018, 13:23:52) [MSC v.1900 32 bit (Intel)] on win32
Type "help", "copyright", "credits" or "license()" for more information.
>>> print('puad')
puad
>>>

```

Figure 1.14: Kompilasi Kode

```

>>> from sklearn import datasets
>>> iris = datasets.load_iris()
>>> digits = datasets.load_digits()
>>> print(digits.data)
[[ 0.  0.  5. ...  0.  0.  0.]
 [ 0.  0.  0. ... 10.  0.  0.]
 [ 0.  0.  0. ... 16.  9.  0.]
 ...
 [ 0.  0.  1. ...  6.  0.  0.]
 [ 0.  0.  2. ... 12.  0.  0.]
 [ 0.  0. 10. ... 12.  1.  0.]]

```

Figure 1.15: Variable Digits

- kemudian ketik "clf.predict(digits.data[-1:])"
- kemudian ketik "from sklearn import svm"
- kemudian ketik "from sklearn import datasets"
- kemudian ketik "clf = svm.SVC(gamma='scale')"
- kemudian ketik "iris = datasets.load_and_load_iris()"
- kemudian ketik "X, y = iris.data, iris.target"
- kemudian ketik "clf.fit(X, y) "
- kemudian ketik "import pickle"
- kemudian ketik "s = pickle.dumps(clf)"
- kemudian ketik "clf2 = pickle.loads(s)"
- kemudian ketik "clf2.predict(X[0:1])"
- kemudian ketik " y[0]"
- kemudian ketik "from joblib import dump, load"
- kemudian ketik "dump(clf, 'filename.joblib')"
- conventions

```
>>> digits.target
array([0, 1, 2, ..., 8, 9, 8])
```

Figure 1.16:

```
>>> digits.images[0]
array([[ 0.,  0.,  5., 13.,  9.,  1.,  0.,  0.],
       [ 0.,  0., 13., 15., 10., 15.,  5.,  0.],
       [ 0.,  3., 15.,  2.,  0., 11.,  8.,  0.],
       [ 0.,  4., 12.,  0.,  0.,  8.,  8.,  0.],
       [ 0.,  5.,  8.,  0.,  0.,  9.,  8.,  0.],
       [ 0.,  4., 11.,  0.,  1., 12.,  7.,  0.],
       [ 0.,  2., 14.,  5., 10., 12.,  0.,  0.],
       [ 0.,  0.,  6., 13., 10.,  0.,  0.,  0.]])
```

Figure 1.17:

- ketikan "import numpy as np"
- ketikan "from sklearn import random(an-deskor)projection"
- ketikan "rng = np.random.RandomState(0)"
- ketikan "X = rng.rand(10, 2000)"
- ketikan "X = np.array(X, dtype='float32')"
- ketikan "X.dtype"
- ketikan "transformer = random projection.GaussianRandomProjection"
- ketikan "X new = transformer.fit transform(X)"
- ketikan "X new.dtype"

5. screenshoot error

6. kode error "no module named 'joblib'"

7. penanganannya instal joblib dengan mengetikan "conda instal -c anaconda joblib"

lllllll efef6eca5df5ee495d6544dd76ec774eee595c8a ===== iiiiii HEAD

1.7 Instalasi/Mhd Zulfikar Akram Nasution/1164081

1. Menjelaskan Kode dari Learning and Predicting

- Pertama import file smv dari sklearn seperti pada gambar 1.12
- Kemudian buat variabel clf seperti pada gambar 1.13
- Lalu ketik kode berikut untuk meliat array baru dari syntax python [:-1] seperti pada gambar 1.14

```
>>> clf = svm.SVC(gamma=0.001, C=100)
>>> clf.fit(digits.data[:-1], digits.target[:-1])
SVC(C=100, cache_size=200, class_weight=None, coef0=0.0,
    decision_function_shape='ovr', degree=3, gamma=0.001, kernel='rbf',
    max_iter=-1, probability=False, random_state=None, shrinking=True,
    tol=0.001, verbose=False)
>>> clf.predict(digits.data[-1:])
array([8])
```

Figure 1.18:

```
>>> clf.predict(digits.data[-1:])
array([8])
```

Figure 1.19:

- Selanjutnya ketikkan kode berikut untuk melihat penggolongan array seperti pada gambar 1.15

2. Model Persistence

- Pertama Import dulu file dari sklearn
- Kemudian buat variable classifier dengan gamma=scale
- Lalu buat variable iris dan (X,y)
- Selanjutnya kita akan melihat penyesuaian classifier
- Kemudian import pickle untuk melihat hasil array dan hasil y

3. Conventions

- Pertama import numpy menjadi np serta import random projection
- Kemudian buat variable rng dengan type random
- Lalu buat variable X, dan lihat hasil rng random yang keluar
- Setelah itu buat variable transformer dengan type random
- Berikutnya itu buat variable X new dengan type yang ada pada tranformer
- Kemudian lihat hasil dari X new

```

>>> from sklearn import svm
>>> from sklearn import datasets
>>> clf = svm.SVC(gamma='scale')
>>> iris = datasets.load_iris()
>>> X,y = iris.data, iris.target
>>> clf.fit(X, y)
SVC(C=1.0, cache_size=200, class_weight=None, coef0=0.0,
    decision_function_shape='ovr', degree=3, gamma='scale', kernel='rbf',
    max_iter=-1, probability=False, random_state=None, shrinking=True,
    tol=0.001, verbose=False)

```

Figure 1.20:

```

>>> import pickle
>>> s = pickle.dumps(clf)
>>> clf2 = pickle.loads(s)
>>> clf2.predict(X[0:1])
array([0])

```

Figure 1.21:

4. Screenshoot Error pada gambar 1.27
5. Kode yang error yaitu "joblib" karena belum ada library nya seperti pada gambar 1.28
6. Solusi dari masalah yang error seperti pada gambar 1.29

=====

1.7.1 Instalasi

1.7.1.1 Loading an Example Datasets

1. Loading an Example Dataset

- Ketik perintah berikut "from sklearn import datasets" untuk mengimport dataset dari file sklearn tadi.
- Selanjutnya ketik perintah berikut ini untuk membuat variable iris yang berisi datasets.
- Masukkan perintah ini untuk membuat variable digits yang berisi datasets, dapat juga untuk melihat isi data dari datasets tadi.

1.8 Learning and Predicting

- from sklearn import svm (pada baris berikut ini merupakan sebuah perintah untuk mengimport class svm dari package sklearn).


```
>>> y[0]
0
```

Figure 1.22:

```
>>> from joblib import dump, load
>>> dump(clf, 'filename.joblib')
['filename.joblib']
```

Figure 1.23:

- `clf = svm.SVC (gamma=0.001, C=100.)` (pada baris kedua ini `clf` sebagai estimator atau parameter, `svm.SVC` menjadi sebuah class, dan `gamma` sebagai parameter untuk menetapkan nilai secara manual)
- `clf.fit(digits.data[:-1], digits.target[:-1])` (pada baris ketiga ini `clf` sebagai estimator atau parameter, `fit` sebagai metode, `digits.data` sebagai item, `[:-1]` sebagai syntax pythonnya dan menampilkan outputannya)
- `clf.predict(digits.data[-1:])`

1.8.0.1 Model Persistence

- `from sklearn import svm`
- `from sklearn import datasets`
- `clf = svm.SVC(gamma='scale')`
- `iris = datasets.load_iris()`
- `X, y = iris.data, iris.target`
- `clf.fit(X, y)`hasil
- `import pickle`
- `s = pickle.dumps(clf)`
- `clf2 = pickle.loads(s)`
- `clf2.predict(X[0:1])`hasil
- `y[0]`hasil
- `from joblib import dump, load`
- `dump(clf, 'filename.joblib')`error
- `clf = load('filename.joblib')`error

```

>>> import numpy as np
>>> from sklearn import random_projection
>>> rng = np.random.RandomState(0)
>>> X = rng.rand(10, 2000)
>>> X = np.array(X, dtype='float32')
>>> X.dtype
dtype('float32')
>>> transformer = random_projection.GaussianRandomProjection()
>>> X_new = transformer.fit_transform(X)
>>> X_new.dtype
dtype('float64')
>>>

```

Figure 1.24:

```

>>> from joblib import dump, load
Traceback (most recent call last):
  File "<stdin>", line 1, in <module>
ModuleNotFoundError: No module named 'joblib'

```

Figure 1.25:

1.8.0.2 Conventions

1. Type Casting

- from sklearn import svm
- from sklearn import random_projection
- rng = np.random.RandomState(0)
- X = rng.rand(10, 2000)
- X = np.array(X, dtype='float32')
- X.dtype hasil
- transformer = random_projection.GaussianRandomProjection()
- X_new = transformer.fit_transform(X)
- X_new.dtype hasil
- from sklearn import datasets
- from sklearn.svm import SVC
- iris = datasets.load_iris()
- clf = SVC(gamma='scale')
- clf.fit(iris.data, iris.target)hasil
- list(clf.predict(iris.data[:3])) hasil
- clf.fit(iris.data, iris.target_names[iris.target]) hasil
- list(clf.predict(iris.data[:3])) hasil

2. Refitting and Updating Parameters

```
>>> from sklearn import svm
```

Figure 1.26: Import file svm

```
>>> clf = svm.SVC(gamma=0.001, C=100.)
```

Figure 1.27: Buat variable Classifier

- import numpy as np
- from sklearn.svm import SVC
- rng = np.random.RandomState(0)
- X = rng.rand(100, 10)
- y = rng.binomial(1, 0.5, 100)
- X_test = rng.rand(5, 10)
- clf = SVC()
- clf.set_params(kernel='linear').fit(X, y) hasil
- clf.predict(X_test) hasil
- clf.set_params(kernel='rbf', gamma='scale').fit(X, y) hasil
- clf.predict(X_test) hasil

3. Multiclass vs. Multilabel Fitting

- from sklearn.svm import SVC
- from sklearn.multiclass import OneVsRestClassifier
- from sklearn.preprocessing import LabelBinarizer
- X = [[1, 2], [2, 4], [4, 5], [3, 2], [3, 1]]
- y = [0, 0, 1, 1, 2]
- classif = OneVsRestClassifier(estimator=SVC(gamma='scale', random_state=0))
- classif.fit(X, y).predict(X) hasil
- y = LabelBinarizer().fit_transform(y)
- classif.fit(X, y).predict(X) hasil
- from sklearn.preprocessing import MultiLabelBinarizer
- y = [[0, 1], [0, 2], [1, 3], [0, 2, 3], [2, 4]]
- y = MultiLabelBinarizer().fit_transform(y)
- classif.fit(X, y).predict(X) hasil

```
>>> clf.fit(digits.data[:-1],digits.target[:-1])
SVC(C=100.0, cache_size=200, class_weight=None, coef0
=0.0,
    decision_function_shape='ovr', degree=3, gamma=0.00
1, kernel='rbf',
    max_iter=-1, probability=False, random_state=None,
shrinking=True,
    tol=0.001, verbose=False)
```

Figure 1.28: Lihat array baru dengan syntac Python

```
>>> clf.predict(digits.data[-1:])
array([8])
```

Figure 1.29: Lihat classifier array

1.9 Penanganan Error

1. Dibawah ini merupakan error yang ditemukan pada saat melakukan percobaan import.
2. Pada gambar diatas, terjadi error ketika sedang mengimport modul yang telah ditetapkan.
3. Solusinya dapat dilakukan dengan berikut ini :
Error tadi terjadi akibat Library Joblib pada PC belum terinstall. Oleh sebab itu, install terlebih dahulu.
4. Dengan membuka CMD (Admin), kemudian masukkan perintah "pip install joblib" dan tunggu sampai instalasi berhasil seperti gambar berikut.
5. Ketika sudah terinstall, maka bisa dilakukan lagi import library joblib, dan hasilnya akan tampil seperti dibawah ini

```
~~~~~ ~~~~~ f594df5d4cf2e0a7297e57a9c544d6f3fac837ab ~~~~~ 2951ffb45da2a049bd0331e
```

```
>>> from sklearn import svm
>>> from sklearn import datasets
```

Figure 1.30: Import file

```
>>> clf = svm.SVC(gamma='scale')
```

Figure 1.31: Variable classifier

```
>>> iris = datasets.load_iris()
>>> X, y = iris.data, iris.target
```

Figure 1.32: Variable iris

```
>>> clf.fit(X,y)
SVC(C=1.0, cache_size=200, class_weight=None, coef0=0.0,
    decision_function_shape='ovr', degree=3, gamma='scale',
    kernel='rbf',
    max_iter=-1, probability=False, random_state=None,
    shrinking=True,
    tol=0.001, verbose=False)
```

Figure 1.33: Penyesuaian Classifier

```
>>> import pickle
>>> s = pickle.dumps(clf)
>>> clf2 = pickle.loads(s)
>>> clf2.predict(X[0:1])
array([0])
>>> y[0]
0
>>>
```

Figure 1.34: Import Pickle

```
>>> import numpy as np
>>> from sklearn import random_projection
```

Figure 1.35: Import numpy

```
>>> rng = np.random.RandomState(0)
```

Figure 1.36: Variable rng

```
>>> X = rng.rand(10, 2000)
>>> X = np.array(X, dtype='float32')
>>> X.dtype
dtype('float32')
```

Figure 1.37: Variable X dan hasil random

```
>>> transformer = random_projection.GaussianRandomProjection()
```

Figure 1.38: Variable transformer random

```
>>> X_new = transformer.fit_transform(X)
```

Figure 1.39: Variable X new type pada transformer

```
>>> X_new.dtype
dtype('float64')
```

Figure 1.40: Hasil dari X new type pada transformer

```
>>> from joblib import dump, load
Traceback (most recent call last):
  File "<stdin>", line 1, in <module>
ModuleNotFoundError: No module named 'joblib'
>>> dump(clf, 'filename.joblib')
Traceback (most recent call last):
  File "<stdin>", line 1, in <module>
NameError: name 'dump' is not defined
>>> clf = load('filename.joblib')
Traceback (most recent call last):
  File "<stdin>", line 1, in <module>
NameError: name 'load' is not defined
>>>
```

Figure 1.41: Screenshoot Error

```
C:\Users\User>conda install -c anaconda joblib
Collecting package metadata: done
Solving environment: done

## Package Plan ##

  environment location: C:\ProgramData\Anaconda3

added / updated specs:
- joblib

The following packages will be downloaded:
-----
package                                     build
-----
ca-certificates-2018.03.07                 0          155 KB anaconda
certifi-2018.4.16                           py36_0     143 KB anaconda
conda-4.6.7                                 py36_0     1.7 MB anaconda
joblib-0.13.2                               py36_0     261 KB anaconda
openssl-1.0.2o                              h839777_0   4.2 MB anaconda
qt-5.9.5                                    vc14he4a7d60_0 90.2 MB anaconda
-----
Total: 96.7 MB

The following NEW packages will be INSTALLED:
-----
joblib          anaconda/win-32::joblib-0.13.2-py36_0

The following packages will be SUPERSEDED by a higher-priority channel:
-----
ca-certificates pkgs/main --> anaconda
certifi         pkgs/main --> anaconda
conda           pkgs/main --> anaconda
openssl         pkgs/main --> anaconda
qt              pkgs/main --> anaconda

Proceed ([y]/n)? y

Downloading and Extracting Packages
joblib-0.13.2 | 301 KB | ##### 100%
certifi-2018.4.16 | 143 KB | ##### 100%
ca-certificates-2018 | 155 KB | ##### 100%
qt-5.9.5 | 90.2 MB | ##### 100%
```

Figure 1.42: Install Joblib

```
>>> from sklearn import svm
>>> clf = svm.SVC(gamma='scale')
>>> from joblib import dump, load
>>> dump(clf, 'filename.joblib')
['filename.joblib']
```

Figure 1.43: Solusi Error

```
C:\WINDOWS\system32>python
Python 3.6.5 [Anaconda, Inc.] (default, Mar 29 2018, 13:23:52) [MSC v.1
900 32 bit (Intel)] on win32
Type 'help', 'copyright', 'credits' or 'license' for more information.
>>> from sklearn import datasets
```

Figure 1.44: Perintah sklearn import datasets

```
>>> iris = datasets.load_iris()
```

Figure 1.45: Perintah Variabel Iris

```
>>> digits = datasets.load_digits()
>>> print(digits.data)
[[ 0.  0.  5. ...  0.  0.  0.]
 [ 0.  0.  0. ... 10.  0.  0.]
 [ 0.  0.  0. ... 16.  9.  0.]
...
 [ 0.  0.  1. ...  6.  0.  0.]
 [ 0.  0.  2. ... 12.  0.  0.]
 [ 0.  0. 10. ... 12.  1.  0.]]
>>>
```

Figure 1.46: Perintah Variabel Digits

```
C:\WINDOWS\system32>python
Python 3.6.5 |Anaconda, Inc.| (default, Mar 29 2018, 13:23:52) [MSC v.1900 32 bi
t (Intel)] on win32
Type "help", "copyright", "credits" or "license" for more information.
>>> from joblib import dump, load
Traceback (most recent call last):
  File "<stdin>", line 1, in <module>
ModuleNotFoundError: No module named 'joblib'
```

Figure 1.47: Error Import

```
C:\WINDOWS\system32>pip install joblib
Collecting joblib
  Downloading https://files.pythonhosted.org/packages/cd/c1/50a753e8247561e58cb8
7305b1e90b171b8c767b15b12a1734001f41d356/joblib-0.13.2-py2.py3-none-any.whl (278
kB)
    11% |#####|
    40kB 593kB/s eta 0:00:01 18% |#####|
    51kB 721kB/s eta 0:00:00 22% |#####|
    61kB 842kB/s eta 0:00:00 25% |#####|
    71kB 969kB/s eta 0:00:00 29% |#####|
    81kB 1.1MB/s eta 0:00:00 33% |#####|
    92kB 1.2MB/s eta 0:00:00 36% |#####|
   102kB 1.3MB/s eta 0:00:00 40% |#####|
   112kB 1.4MB/s eta 0:00:00 44% |#####|
   122kB 1.5MB/s eta 0:00:00 47% |#####|
   133kB 1.6MB/s eta 0:00:00 51% |#####|
   143kB 1.7MB/s eta 0:00:00 55% |#####|
   153kB 1.8MB/s eta 0:00:00 58% |#####|
   163kB 1.9MB/s eta 0:00:00 62% |#####|
   174kB 2.0MB/s eta 0:00:00 66% |#####|
   184kB 2.1MB/s eta 0:00:00 69% |#####|
   194kB 2.2MB/s eta 0:00:00 73% |#####|
   204kB 2.3MB/s eta 0:00:00 77% |#####|
   215kB 2.4MB/s eta 0:00:00 80% |#####|
   225kB 2.5MB/s eta 0:00:00 84% |#####|
   236kB 2.6MB/s eta 0:00:00 88% |#####|
   246kB 2.7MB/s eta 0:00:00 91% |#####|
   256kB 2.8MB/s eta 0:00:00 95% |#####|
   266kB 2.9MB/s eta 0:00:00 99% |#####|
   276kB 3.0MB/s eta 0:00:00 100% |#####|
distributed 1.21.8 requires msgpack, which is not installed.
Installing collected packages: joblib
Successfully installed joblib-0.13.2
You are using pip version 10.0.1, however version 19.0.3 is available.
You should consider upgrading via the 'python -m pip install --upgrade pip' comm
and.
C:\WINDOWS\system32>
```

Figure 1.48: Install Library Joblib

```
>>> from joblib import dump, load
>>> dump(clf, 'filename.joblib')
['filename.joblib']
>>> clf = load('filename.joblib')
>>>
```

Figure 1.49: Berhasil Import Library Joblib

Chapter 2

Related Works

Your related works, and your purpose and contribution which must be different as below.

2.1 puad hamdani/1164084

2.1.1 binary classification

1. Output aktual dari banyak algoritma binary classification adalah skor prediksi. Skor menunjukkan kepastian sistem bahwa pengamatan yang diberikan adalah milik kelas positif. Untuk membuat keputusan tentang apakah pengamatan harus diklasifikasikan sebagai positif atau negatif

2.1.2 supervised learning dan unsupervised learning

1. Supervised learning adalah sebuah pendekatan dimana sudah terdapat data yang dilatih, dan terdapat variable yang di targetkan sehingga tujuan dari pendekatan ini adalah mengelompokkan suatu data ke data yang sudah ada.
2. Unsupervised learning adalah istilah yang digunakan untuk pembelajaran bahasa Ibrani, yang terkait dengan pembelajaran tanpa guru, juga dikenal sebagai organisasi mandiri dan metode pemodelan kepadatan probabilitas input. Unsupervised learning tidak memiliki data latih, sehingga dari data yang ada, kita mengelompokkan data tersebut menjadi dua bagian atau tiga bagian dan seterusnya.

2.1.3 evaluasi dan akurasi

1. Evaluasi merupakan suatu proses identifikasi untuk mengukur atau menilai apakah suatu kegiatan/program yang dilaksanakan sesuai dengan perencanaan atau tujuan yang ingin dicapai. Akurasi merupakan tingkat kedekatan pengukuran kuantitas terhadap nilai yang sebenarnya. Kepresisian dari suatu sistem pengukuran

2.1.4 bagaimana cara membuat dan membaca confusion matrix, buat confusion matrix

1. Cara membuat dan membaca confusion matrix :
 - 1) Tentukan pokok permasalahan dan atributnya
 - 2) Buat pohon keputusan
 - 3) Lalu data testingnya
 - 4) Lalu mencari nilai a, b, c, dan d. Semisal a = 5, b = 1, c = 1, dan d = 3.
 - 5) Selanjutnya mencari nilai recall, precision, accuracy, serta dan error rate.
2. Berikut adalah contoh dari confusion matrix :
 - $\text{Recall} = 3/(1+3) = 0,75$
 - $\text{Precision} = 3/(1+3) = 0,75$
 - $\text{Accuracy} = (5+3)/(5+1+1+3) = 0,8$
 - $\text{Error Rate} = (1+1)/(5+1+1+3) = 0,2$

2.1.5 bagaimana K-fold cross validation bekerja dengan gambar ilustrasi

1. Cara kerja K-fold cross validation :
 - 1) Total instance dibagi menjadi N bagian.
 - 2) Fold yang pertama adalah bagian pertama menjadi data uji (testing data) dan sisanya menjadi training data.
 - 3) Lalu hitung akurasi berdasarkan porsi data tersebut dengan menggunakan persamaan.

- 4) Fold yang ke dua adalah bagian ke dua menjadi data uji (testing data) dan sisanya training data.
- 5) Kemudian hitung akurasi berdasarkan porsi data tersebut.
- 6) Dan seterusnya hingga habis mencapai fold ke-K.
- 7) Terakhir hitung rata-rata akurasi K buah.

2.1.6 decision tree

1. Decision tree merupakan model prediksi menggunakan struktur pohon atau struktur berhirarki.

2.1.7 Information Gain

1. Information gain. Metode tersebut akan melakukan proses komputasi untuk mendapatkan atribut-atribut yang paling berpengaruh terhadap dataset

2.2 Puad Hamdani/ 1164084

2.2.1 Scikit-Learn

1. # load dataset (student mat pakenya)
import pandas as pd
sate = pd.read_csv('student-mat.csv', sep=';')
len(sate)

```
In [2]: import pandas as pd
...: sate = pd.read_csv('student-por.csv', sep=';')
...: len(sate)
Out[2]: 649
```

Figure 2.1: Load Dataset

Codingan ini akan meload (menampilkan) data pada file yang ditentukan. hasilnya berupa angka 649 .

2. # generate binary label (pass/fail) based on G1+G2+G3
(test grades, each 0-20 pts); threshold for passing is sum>=30
sate['pass'] = sate.apply(lambda row: 1 if (row['G1']+row['G2']+row['G3'])
>= 35 else 0, axis=1)
sate = sate.drop(['G1', 'G2', 'G3'], axis=1)
sate.head()

```
In [3]: sate['pass'] = sate.apply(lambda row: 1 if (row['G1']+row['G2']+row['G3']) >= 35 else 0, axis=1)
...: sate = sate.drop(['G1', 'G2', 'G3'], axis=1)
...: sate.head()
Out[3]:
```

Figure 2.2: Generate Binary Label

Codingan ini secara keseluruhan menampilkan baris G1, G2 dan G3 (berdasarkan kriterianya). Untuk lebih jelasnya, pada codingan terdapat pendefinisian pembacaan "lambda" (panjang gelombang) dari baris G1, G2 dan G3. Apabila row-row tersebut bernilai lebih dari 35 maka akan terdefiniskan angka "1" apabila tidak, maka akan terdefiniskan angka "0"

```
3. # use one-hot encoding on categorical columns
sate = pd.get_dummies(sate, columns=['sex', 'school', 'address',
'famsize',
'Pstatus', 'Mjob', 'Fjob',
'reason', 'guardian', 'schoolsup',
'famsup', 'paid', 'activities',
'nursery', 'higher', 'internet',
'romantic'])
sate.head()
```

```
In [4]: sate = pd.get_dummies(sate, columns=['sex', 'school', 'address', 'famsize', 'Pstatus', 'Mjob', 'Fjob',
...: 'reason', 'guardian', 'schoolsup', 'famsup', 'paid', 'activities',
...: 'nursery', 'higher', 'internet', 'romantic'])
...: sate.head()
Out[4]:
```

	age	Medu	Fedu	...	internet_yes	romantic_no	romantic_yes
0	18	4	4	...	0	1	0
1	17	1	1	...	1	1	0
2	15	1	1	...	1	1	0
3	15	4	2	...	1	0	1
4	16	3	3	...	0	1	0

Figure 2.3: Pemanggilan get dummies

Secara keseluruhan, codingan ini mendefinisikan pemanggilan get dummies ke dalam variabel sate.

```
4. # shuffle rows
sate = sate.sample(frac=1)
# split training and testing data
sate_train = d[:500]
sate_test = d[500:]

sate_train_att = sate_train.drop(['pass'], axis=1)
```

```

sate_train_pass = sate_train['pass']

sate_test_att = sate_test.drop(['pass'], axis=1)
sate_test_pass = sate_test['pass']

sate_att = sate.drop(['pass'], axis=1)
sate_pass = sate['pass']

# number of passing students in whole dataset:
import numpy as np
print("Passing: %d out of %d (%.2f%%)" % (np.sum(lontong_pass), len(lontong_pass),
    100*float(np.sum(sate_pass)) / len(sate_pass)))

```

```

In [5]: sate = sate.sample(frac=1)
...: # split training and testing data
...: sate_train = sate[:500]
...: sate_test = sate[500:]
...:
...: sate_train_att = sate_train.drop(['pass'], axis=1)
...: sate_train_pass = sate_train['pass']
...:
...: sate_test_att = sate_test.drop(['pass'], axis=1)
...: sate_test_pass = sate_test['pass']
...:
...: sate_att = sate.drop(['pass'], axis=1)
...: sate_pass = sate['pass']
...:
...: # number of passing students in whole dataset:
...: import numpy as np
...: print("Passing: %d out of %d (%.2f%%)" % (np.sum(sate_pass), len(sate_pass), 100*float(np.sum(sate_pass)) / len(sate_pass)))
Passing: 328 out of 649 (50.54%)

```

Figure 2.4: Mendefinisikan pembagian data

Secara keseluruhan codingan ini difungsikan untuk mendefinisikan pembagian data yang berupa training dan testing data

5. # fit a decision tree

```

from sklearn import tree
rendang = tree.DecisionTreeClassifier(criterion="entropy", max_depth=5)
rendang = rendang.fit(sate_train_att, sate_train_pass)

```

```

In [6]: from sklearn import tree
...: rendang = tree.DecisionTreeClassifier(criterion="entropy", max_depth=5)
...: rendang = rendang.fit(sate_train_att, sate_train_pass)

```

Figure 2.5: Membuktikan pengujian

Secara keseluruhan, codingan ini hanya membuktikan pengujian dari Klasifikasi Decision Tree yang ada, apakah true atau tidak dan hasilnya true

6. # visualize tree

```
import graphviz

dot_data = tree.export_graphviz(rendang, out_file=None, label="all",
                                impurity=False, proportion=True,
                                feature_names=list(sate_train_att),
                                class_names=["fail", "pass"],
                                filled=True, rounded=True)

graph = graphviz.Source(dot_data)
graph
```

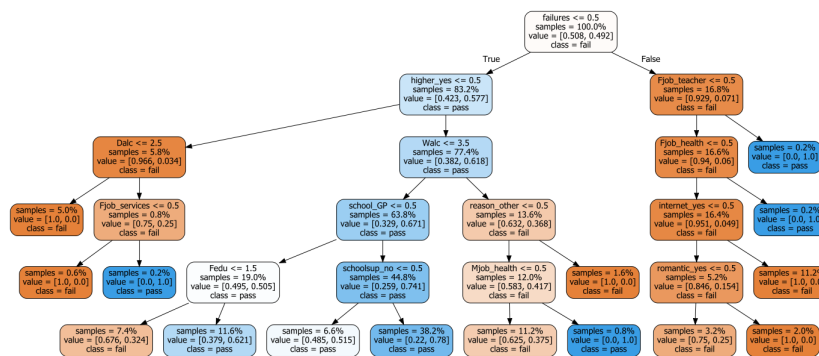


Figure 2.6: Gambaran decision tree

Codingan ini memberikan gambaran dari klasifikasi decision tree dari pengolahan parameter yang dieksekusi kedalam variabel soto. Tentunya dengan pemanfaatan library graphviz yang telah diimport dan difungsikan.

7. # save tree

```
tree.export_graphviz(rendang, out_file="student-performance.dot",
                    label="all", impurity=False,
                    proportion=True,
                    feature_names=list(sate_train_att),
                    class_names=["fail", "pass"],
                    filled=True, rounded=True)
```

```
In [9]: tree.export_graphviz(rendang, out_file="student-performance.dot", label="all", impurity=False, proportion=True,
...:                         feature_names=list(sate_train_att), class_names=["fail", "pass"],
...:                         filled=True, rounded=True)
```

Figure 2.7: Library Graphviz

Pada gambar 4.6 akan menampilkan yang terdapat pada Library Graphviz,

```
In [10]: rendang.score(sate_test_att, sate_test_pass)
Out[10]: 0.6778523489932886
```

Figure 2.8: Menampilkan hasil perhitungan 2 parameter

8. `rendang.score(sate_test_att, sate_test_pass)`

merupakan perhitungan hasil prediksi silang akan kemungkinan nilai di masa mendatang.

9. `from sklearn.model_selection import cross_val_score`
`semur = cross_val_score(rendang, sate_att, sate_pass, cv=5)`
`# show average score and +/- two standard deviations away`
 `#(covering 95% of scores)`
`print("Accuracy: %0.2f (+/- %0.2f)" % (semur.mean(), semur.std() * 2))`

```
In [11]: from sklearn.model_selection import cross_val_score
...: semur = cross_val_score(rendang, sate_att, sate_pass, cv=5)
...: # show average score and +/- two standard deviations away (covering 95% of scores)
...: print("Accuracy: %0.2f (+/- %0.2f)" % (semur.mean(), semur.std() * 2))
Accuracy: 0.69 (+/- 0.02)
```

Figure 2.9: Mendefinisikan library sklearn

Kodingan tersebut mendefinisikan library sklearn model selection dan import cross val score

10. `for max_depth in range(1, 20):`
 `rendang = tree.DecisionTreeClassifier(criterion="entropy",`
 `max_depth=max_depth)`
 `semur = cross_val_score(rendang, sate_att, sate_pass, cv=5)`
 `print("Max depth: %d, Accuracy: %0.2f (+/- %0.2f)" %`
 `(max_depth, semur.mean(), semur.std() * 2)`
 `)`

menampilkan data dari angka 1-20.

11. `depth_acc = np.empty((19,3), float)`
`i = 0`
`for max_depth in range(1, 20):`
 `soto = tree.DecisionTreeClassifier(criterion="entropy",`
 `max_depth=max_depth)`
 `semur = cross_val_score(rendang, sate_att, sate_pass, cv=5)`
 `depth_acc[i,0] = max_depth`

```

In [12]: for max_depth in range(1, 20):
...:     rendang = tree.DecisionTreeClassifier(criterion="entropy", max_depth=max_depth)
...:     semur = cross_val_score(rendang, sate_att, sate_pass, cv=5)
...:     print("Max depth: %d, Accuracy: %0.2f (+/- %0.2f)" % (max_depth, semur.mean(), semur.std() * 2))
Max depth: 1, Accuracy: 0.64 (+/- 0.05)
Max depth: 2, Accuracy: 0.69 (+/- 0.08)
Max depth: 3, Accuracy: 0.70 (+/- 0.05)
Max depth: 4, Accuracy: 0.69 (+/- 0.06)
Max depth: 5, Accuracy: 0.69 (+/- 0.01)
Max depth: 6, Accuracy: 0.66 (+/- 0.05)
Max depth: 7, Accuracy: 0.66 (+/- 0.05)
Max depth: 8, Accuracy: 0.66 (+/- 0.08)
Max depth: 9, Accuracy: 0.66 (+/- 0.07)
Max depth: 10, Accuracy: 0.64 (+/- 0.07)
Max depth: 11, Accuracy: 0.62 (+/- 0.09)
Max depth: 12, Accuracy: 0.63 (+/- 0.08)
Max depth: 13, Accuracy: 0.62 (+/- 0.04)
Max depth: 14, Accuracy: 0.63 (+/- 0.09)
Max depth: 15, Accuracy: 0.62 (+/- 0.08)
Max depth: 16, Accuracy: 0.63 (+/- 0.08)
Max depth: 17, Accuracy: 0.62 (+/- 0.07)
Max depth: 18, Accuracy: 0.63 (+/- 0.05)
Max depth: 19, Accuracy: 0.62 (+/- 0.06)

```

Figure 2.10: Menampilkan hasil fungsi max depth dan accuracy

```

depth_acc[i,1] = semur.mean()
depth_acc[i,2] = semur.std() * 2
i += 1

```

depth_acc

```

In [13]: depth_acc = np.empty((19,3), float)
...: i = 0
...: for max_depth in range(1, 20):
...:     rendang = tree.DecisionTreeClassifier(criterion="entropy", max_depth=max_depth)
...:     semur = cross_val_score(rendang, sate_att, sate_pass, cv=5)
...:     depth_acc[i,0] = max_depth
...:     depth_acc[i,1] = semur.mean()
...:     depth_acc[i,2] = semur.std() * 2
...:     i += 1
...:
...:
...: depth_acc
Out[13]:
array([[ 1.,  0.63779741,  0.0507489 ],
       [ 2.,  0.68704156,  0.07858131],
       [ 3.,  0.70247443,  0.0525174 ],
       [ 4.,  0.69185823,  0.06240303],
       [ 5.,  0.69801281,  0.02145731],
       [ 6.,  0.65641344,  0.05413215],
       [ 7.,  0.66083947,  0.06754535],
       [ 8.,  0.65177421,  0.06648372],
       [ 9.,  0.65482764,  0.06422626],
       [10.,  0.62552459,  0.09252526],
       [11.,  0.64392777,  0.08378081],
       [12.,  0.62396282,  0.05819135],
       [13.,  0.62238803,  0.10767083],
       [14.,  0.62551321,  0.08211362],
       [15.,  0.63327653,  0.06220894],
       [16.,  0.62548899,  0.07332788],
       [17.,  0.6270636 ,  0.06425702],
       [18.,  0.6193001 ,  0.07943198],
       [19.,  0.61016364,  0.05593769]])

```

Figure 2.11: Menjelaskan variable kari

Bahwa variable semur akan menampilkan atau mendefinisikan nilai dari variabel score yang mana isi dari variable score yaitu rendang

```

12. import matplotlib.pyplot as plt
fig, ax = plt.subplots()

```



```
ax.errorbar(depth_acc[:,0], depth_acc[:,1], yerr=depth_acc[:,2])
plt.show()
```

```
In [14]: import matplotlib.pyplot as plt
...: fig, ax = plt.subplots()
...: ax.errorbar(depth_acc[:,0], depth_acc[:,1], yerr=depth_acc[:,2])
...: plt.show()
```

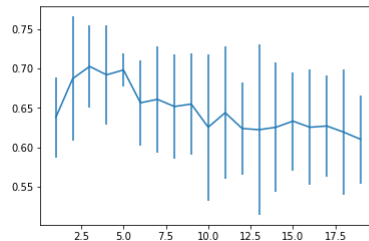


Figure 2.12: Menjelaskan dan menampilkan gambar grafik

menampilkan gambar grafik pada gambar 4.11 dari eksekusi fungsi `ax.errorbar`.

2.3 Penanganan Error

Dari percobaan yang dilakukan di atas, tidak ada eror:

2.4 Same Topics

Cite every latest journal with same topic

2.4.1 Topic 1

cite for first topic

2.4.2 Topic 2

if you have two topics you can include here to

2.5 Same Method

write and cite latest journal with same method

2.5.1 Method 1

cite and paraphrase method 1

2.5.2 Method 2

cite and paraphrase method 2 if you have more method please add new subsection.

Chapter 3

Methods

3.1 The data

Please tell where is the data come from, a little brief of company can be put here.

3.2 Method 1

Definition, steps, algorithm or equation of method 1 and how to apply into your data

3.3 Method 2

Definition, steps, algorithm or equation of method 2 and how to apply into your data

Chapter 4

Experiment and Result

brief of experiment and result.

4.1 Experiment

Please tell how the experiment conducted from method.

4.2 Result

Please provide the result of experiment

Chapter 5

Conclusion

brief of conclusion

5.1 Conclusion of Problems

Tell about solving the problem

5.2 Conclusion of Method

Tell about solving using method

5.3 Conclusion of Experiment

Tell about solving in the experiment

5.4 Conclusion of Result

tell about result for purpose of this research.

Chapter 6

Discussion

Please tell more about conclusion and how to the next work of this study.

Chapter 7

Discussion

Please tell more about conclusion and how to the next work of this study.

Chapter 8

Discussion

Please tell more about conclusion and how to the next work of this study.

Chapter 9

Discussion

Please tell more about conclusion and how to the next work of this study.

Chapter 10

Discussion

Please tell more about conclusion and how to the next work of this study.

Chapter 11

Discussion

Please tell more about conclusion and how to the next work of this study.

Chapter 12

Discussion

Please tell more about conclusion and how to the next work of this study.

Chapter 13

Discussion

Please tell more about conclusion and how to the next work of this study.

Chapter 14

Discussion

Please tell more about conclusion and how to the next work of this study.

Appendix A

Form Penilaian Jurnal

gambar A.1 dan A.2 merupakan contoh bagaimana reviewer menilai jurnal kita.

NO	UNSUR	KETERANGAN	MAKS	KETERANGAN
1	Keefektifan Judul Artikel	Maksimal 12 (dua belas) kata dalam Bahasa Indonesia atau 10 (sepuluh) kata dalam Bahasa Inggris	2	a. Tidak lugas dan tidak ringkas (0) b. Kurang lugas dan kurang ringkas (1) c. Ringkas dan lugas (2)
2	Pencantuman Nama Penulis dan Lembaga Penulis		1	a. Tidak lengkap dan tidak konsisten (0) b. Lengkap tetapi tidak konsisten (0,5) c. Lengkap dan konsisten (1)
3	Abstrak	Dalam Bahasa Indonesia dan Bahasa Inggris yang baik, jumlah 150-200 kata. Isi terdiri dari latar belakang, metode, hasil, dan kesimpulan. Isi tertuang dengan kalimat yang jelas.	2	a. Tidak dalam Bahasa Indonesia dan Bahasa Inggris (0) b. Abstrak kurang jelas dan ringkas, atau hanya dalam Bahasa Inggris, atau dalam Bahasa Indonesia saja (1) c. Abstrak yang jelas dan ringkas dalam Bahasa Indonesia dan Bahasa Inggris (2)
4	Kata Kunci	Maksimal 5 kata kunci terpenting dalam paper	1	a. Tidak ada (0) b. Ada tetapi kurang mencerminkan konsep penting dalam artikel (0,5) c. Ada dan mencerminkan konsep penting dalam artikel (1)
5	Sistematika Pembahasan	Terdiri dari pendahuluan, tinjauan pustaka, metode penelitian, hasil dan pembahasan, kesimpulan dan saran, daftar pustaka	1	a. Tidak lengkap (0) b. Lengkap tetapi tidak sesuai sistematika (0,5) c. Lengkap dan bersistem (1)
6	Pemanfaatan Instrumen Pendukung	Pemanfaatan Instrumen Pendukung seperti gambar dan tabel	1	a. Tidak dimanfaatkan (0) b. Kurang informatif atau komplementer (0,5) c. Informatif dan komplementer (1)
7	Cara Pengacuan dan Pengutipan		1	a. Tidak baku (0) b. Kurang baku (0,5) c. Baku (1)
8	Penyusunan Daftar Pustaka	Penyusunan Daftar Pustaka	1	a. Tidak baku (0) b. Kurang baku (0,5) c. Baku (1)
9	Peristilahan dan Kebahasaan		2	a. Buruk (0) b. Baik (1) c. Cukup (2)
10	Makna Sumbangan bagi Kemajuan		4	a. Tidak ada (0) b. Kurang (1) c. Sedang (2) d. Cukup (3) e. Tinggi (4)

Figure A.1: Form nilai bagian 1.

11	Dampak Ilmiah		7	a. Tidak ada (0) b. Kurang (1) c. Sedang (3) d. Cukup (5) e. Besar (7)
12	Nisbah Sumber Acuan Primer berbanding Sumber lainnya	Sumber acuan yang langsung merujuk pada bidang ilmiah tertentu, sesuai topik penelitian dan sudah teruji.	3	a. < 40% (1) b. 40-80% (2) c. > 80% (3)
13	Derajat Kemutakhiran Pustaka Acuan	Derajat Kemutakhiran Pustaka Acuan	3	a. < 40% (1) b. 40-80% (2) c. > 80% (3)
14	Analisis dan Sintesis	Analisis dan Sintesis	4	a. Sedang (2) b. Cukup (3) c. Baik (4)
15	Penyimpulan	Sangat jelas relevasinya dengan latar belakang dan pembahasan, dirumuskan dengan singkat	3	a. Kurang (1) b. Cukup (2) c. Baik (3)
16	Unsur Plagiat		0	a. Tidak mengandung plagiat (0) b. Terdapat bagian-bagian yang merupakan plagiat (-5) c. Keseluruhannya merupakan plagiat (-20)
TOTAL			36	
Catatan : Nilai minimal untuk diterima 25				

Figure A.2: form nilai bagian 2.

Appendix B

FAQ

M : Kalo Intership II atau TA harus buat aplikasi ? D : Ga harus buat aplikasi tapi harus ngoding

M : Pa saya bingung mau ngapain, saya juga bingung mau presentasi apa? D : Makanya baca de, buka jurnal topik ‘ganteng’ nah kamu baca dulu sehari 5 kali ya, 4 hari udah 20 tuh. Bingung itu tanda kurang wawasan alias kurang baca.

M : Pa saya sudah cari jurnal terindeks scopus tapi ga nemu. D : Kamu punya mata de? coba dicolok dulu. Kamu udah lakuin apa aja? tolong di list laporkan ke grup Tingkat Akhir. Tinggal buka google scholar klik dari tahun 2014, cek nama jurnalnya di scimagojr.com beres.

M : Pa saya belum dapat tempat intership, jadi ga tau mau presentasi apa? D : kamu kok ga nyambung, yang dipresentasikan itu yang kamu baca bukan yang akan kamu lakukan.

M : Pa ini jurnal harus yang terindex scopus ga bisa yang lain ? D : Index scopus menandakan artikel tersebut dalam standar semantik yang mudah dipahami dan dibaca serta bukan artikel asal jadi. Jika diluar scopus biasanya lebih sukar untuk dibaca dan dipahami karena tidak adanya proses review yang baik dan benar terhadap artikel.

M : Pa saya tidak mengerti D : Coba lihat standar alasan

M : Pa saya bingung D : Coba lihat standar alasan

M : Pa saya sibuk D : Mbahmu....

M : Pa saya ganteng D : Ndasmu....

M : Pa saya kece D : wes karepmu lah....

Biasanya anda memiliki alasan tertentu jika menghadapi kendala saat proses bimbingan, disini saya akan melakukan standar alasan agar persepsi yang diterima sama dan tidak salah kaprah. Penggunaan kata alasan tersebut antara lain :

1. Tidak Mengerti : anda boleh menggunakan alasan ini jika anda sudah melakukan tahapan membaca dan meresumekan 15 jurnal. Sudah mencoba dan mempraktekkan teorinya dengan mencari di youtube dan google minimal 6 jam sehari selama 3 hari berturut-turut.

2. Bingung : anda boleh mengatakan alasan bingung setelah maksimal dalam berusaha menyelesaikan tugas bimbingan dari dosen(sudah dilakukan semua). Anda belum bisa mengatakan alasan bingung jika anda masih belum menyelesaikan tugas bimbingan dan poin nomor 1 diatas. Setelah anda menyelesaikan tugas bimbingan secara maksimal dan tahap 1 poin diatas, tapi anda masih tetap bingung maka anda boleh memakai alasan ini.

Bibliography

- [1] Joshua Eckroth. *Python Artificial Intelligence Projects for Beginners: Get up and running with Artificial Intelligence using 8 smart and exciting AI applications*. Packt Publishing Ltd, 2018.
- [2] Stuart J Russell and Peter Norvig. *Artificial intelligence: a modern approach*. Malaysia; Pearson Education Limited,, 2016.