Kampus Merdeka
INDONESIA JAYA

Speech Processing dan Music Information Retrieval

M Octaviano Pratama, S.Kom., M.Kom

Director BISA AI Academy

Outline

- **Course Introduction**
- Speech Processing
- Music Information Retrieval
- Klasifikasi Voice Gender pada Feature low-Level Suara dengan dan Deep Neural Networks

Silabus

- Speech Processing & MIR
- Basic Feature Extraction
- ASR
- Pengolahan Sinyal Digital
- Machine Learning
- Mini Project 1
- UTS
- Research 1
- Research 2
- State-of-the-art Speech
- Mini Project 2
- Mini Project 3
- Music & Speech Apps
- UAS

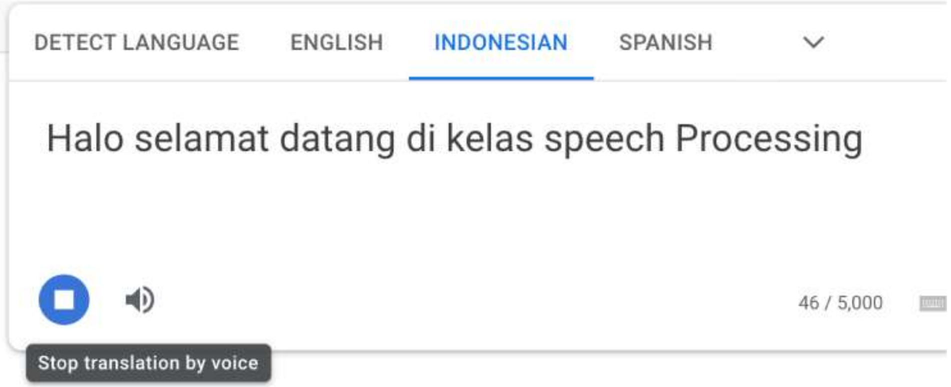
Mini Project

- Mini Project 1: Klasifikasi low-level Audio dataset dengan Algoritma Machine Learning
- Mini Project 2: Building Music Dataset from collection audio files
- Mini Project 3: Music Information Retrieval

Outline

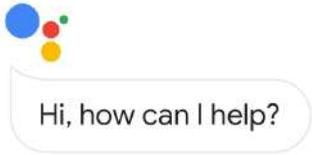
- Course Introduction
- **Speech Processing**
- Music Information Retrieval
- Klasifikasi Voice Gender pada Feature low-Level Suara dengan dan Deep Neural Networks

Spoken Language Processing Apps



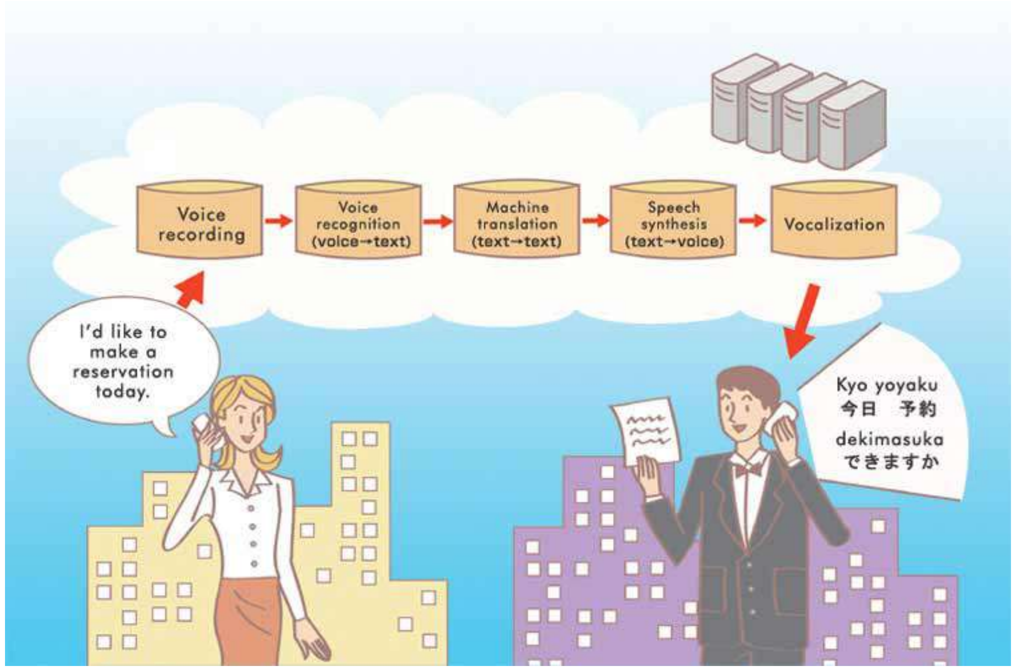
Halo selamat datang di kelas speech Processing

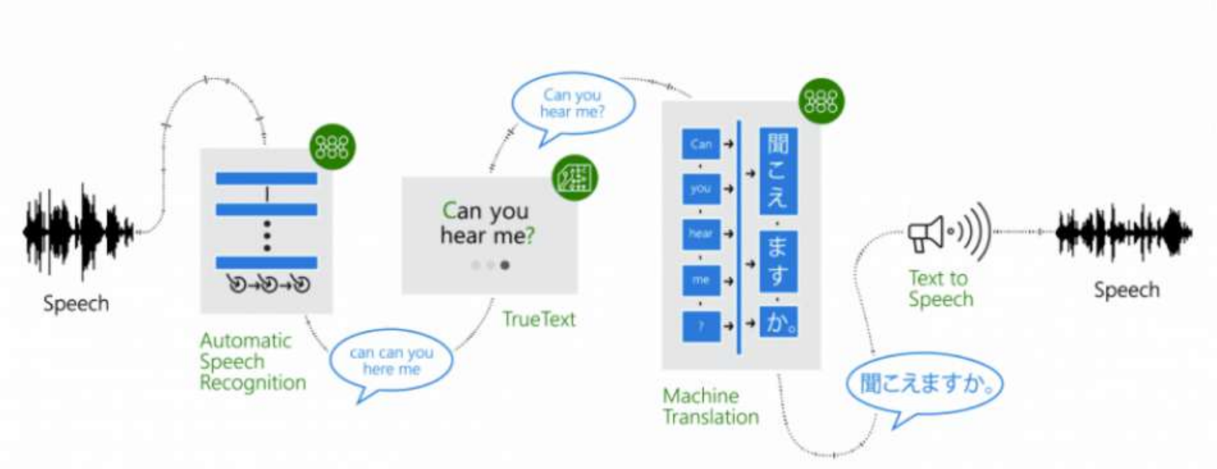
46 / 5,000



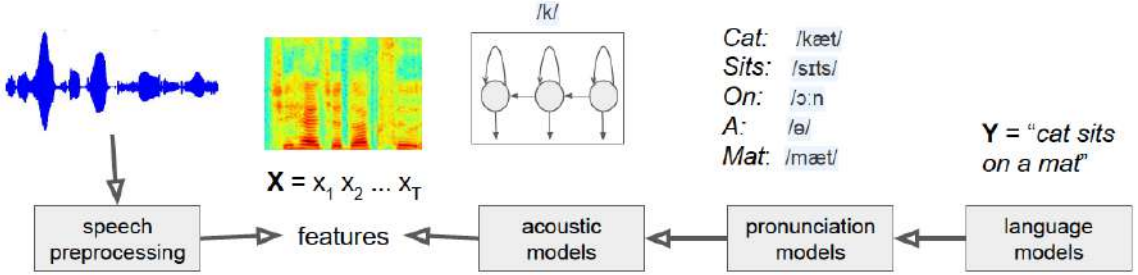
Hi, how can I help?

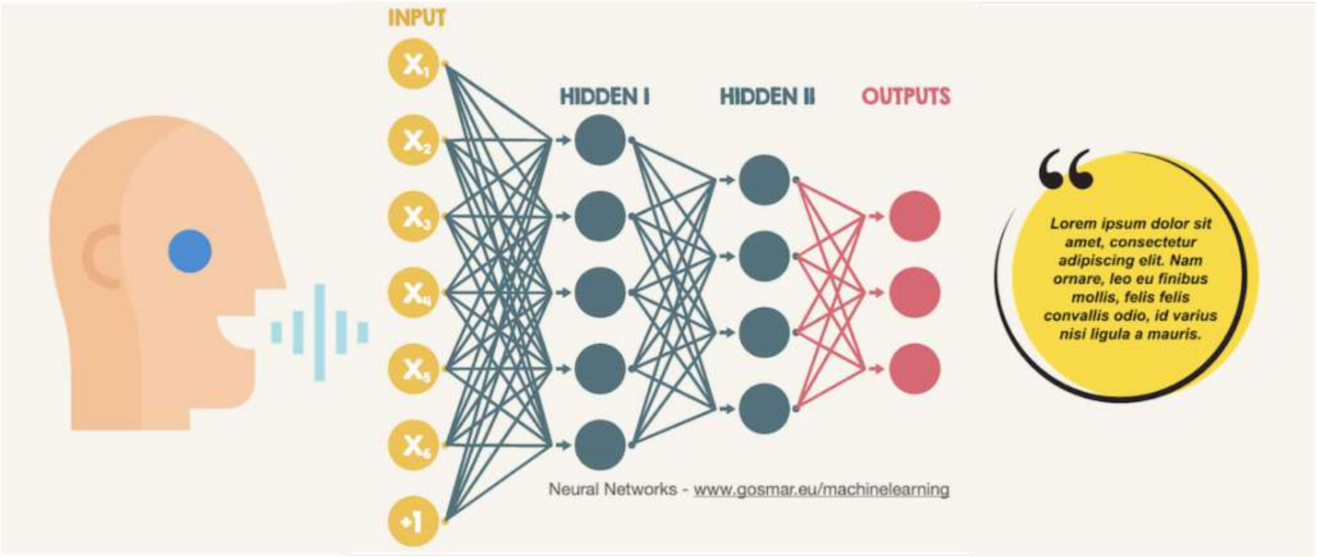
Speech Translation


Voice recording
Voice recognition (voice→text)
Machine translation (text→text)
Speech synthesis (text→voice)
Vocalization
I'd like to make a reservation today.
Kyo yoyaku 今日 予約 dekimasuka できますか

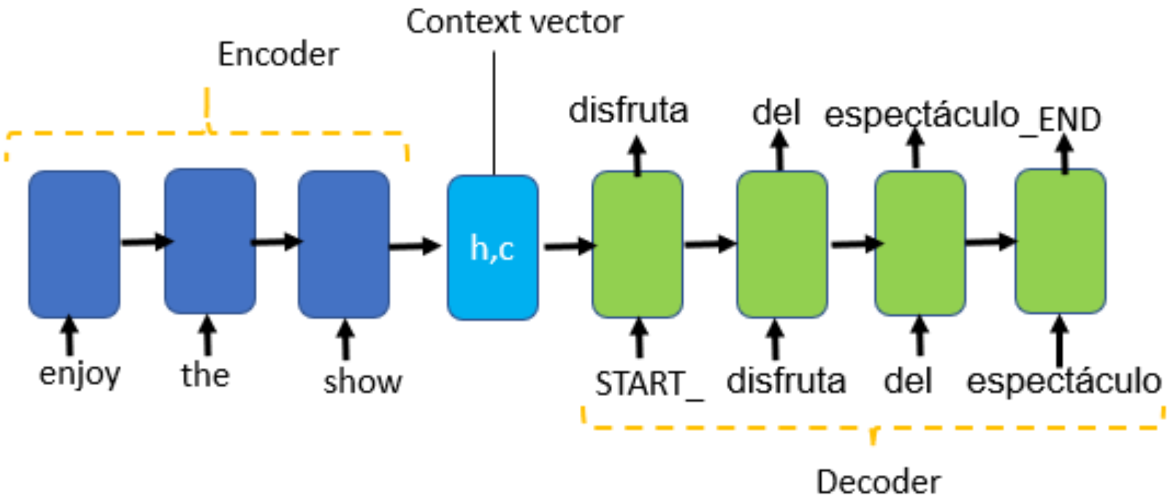


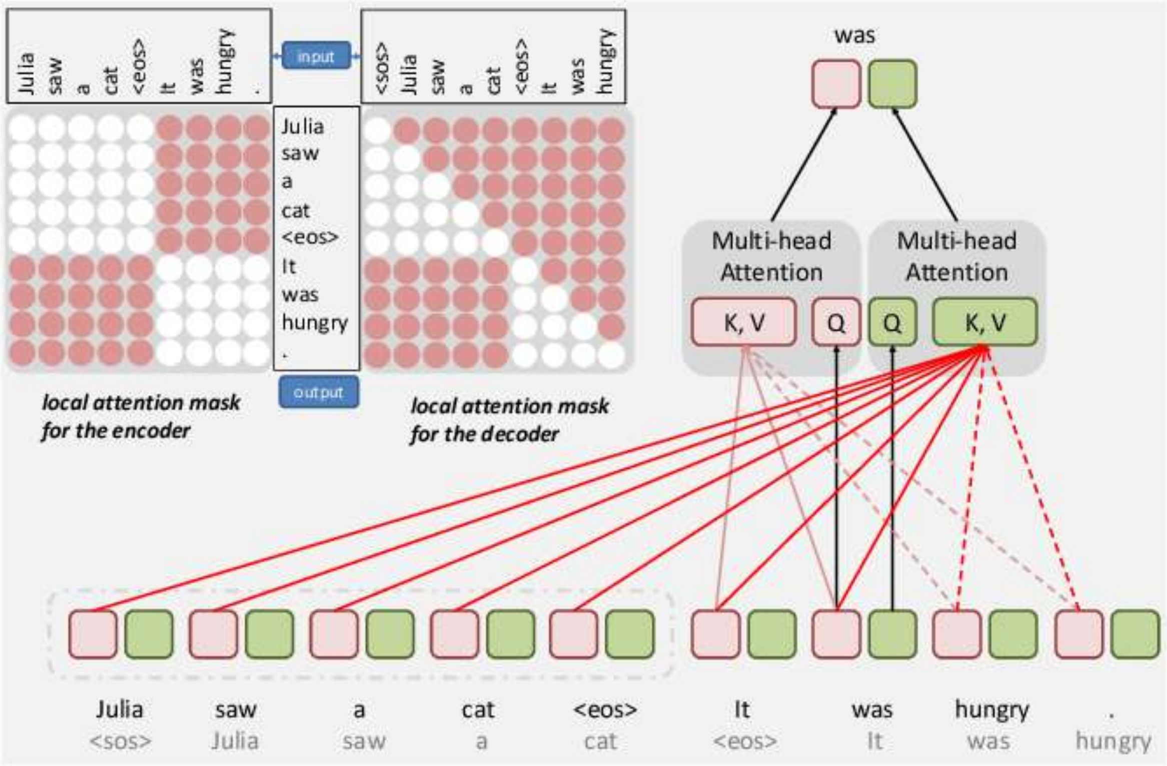
ASR


/k/
Cat: /kæt/
Sits: /sɪts/
On: /ɔːn/
A: /ə/
Mat: /mæt/
Y = "cat sits on a mat"
X = x1 x2 ... xT
speech preprocessing
features
acoustic models
pronunciation models
language models

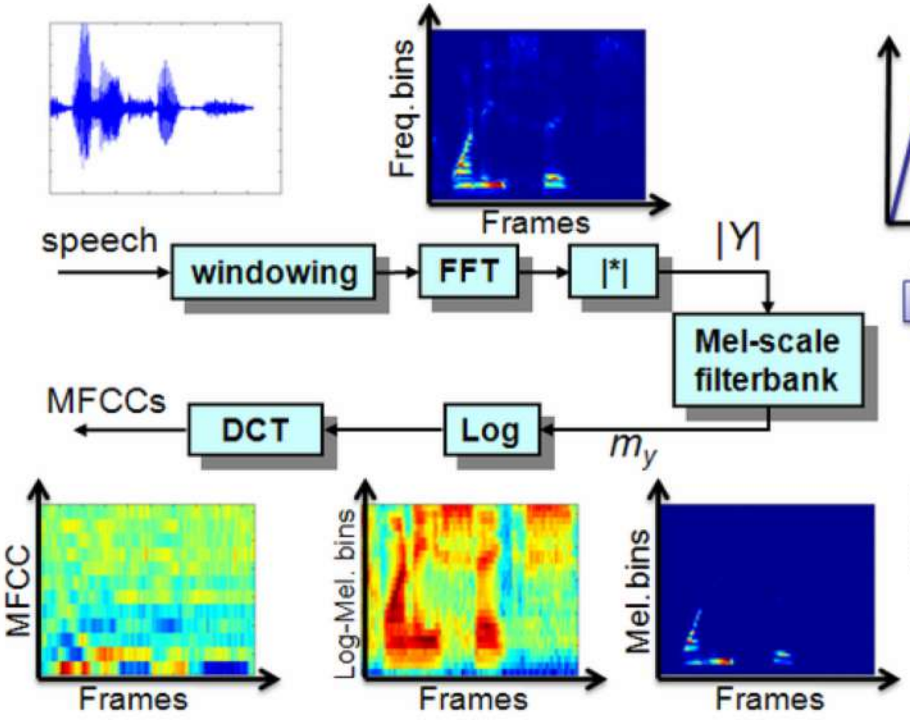

INPUT
X1
X2
X3
X4
X5
X6
+1
HIDDEN I
HIDDEN II
OUTPUTS
Lorem ipsum dolor sit amet, consectetur adipiscing elit. Nam ornare, leo eu finibus mollis, felis felis convallis odio, id varius nisi ligula a mauris.
Neural Networks - www.gosmar.eu/machinelearning

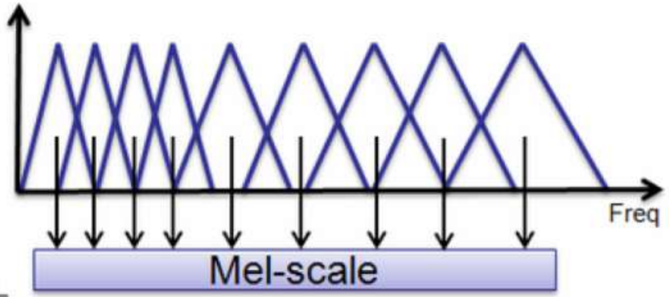
Machine Translation

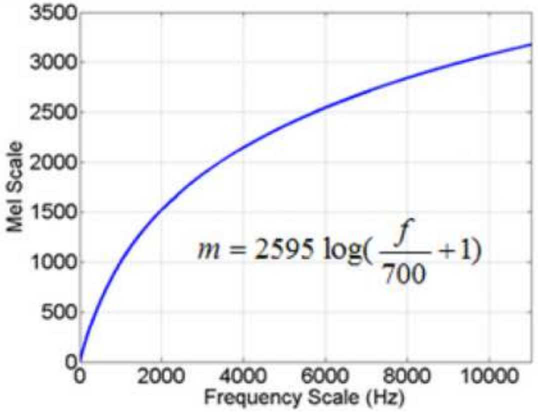




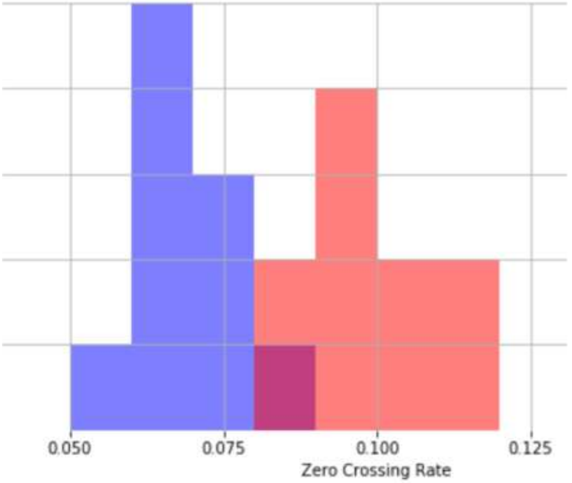
We Need Features

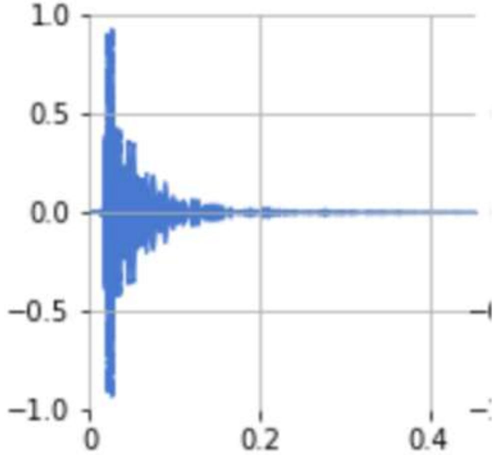






$$m = 2595 \log(\frac{f}{700} + 1)$$

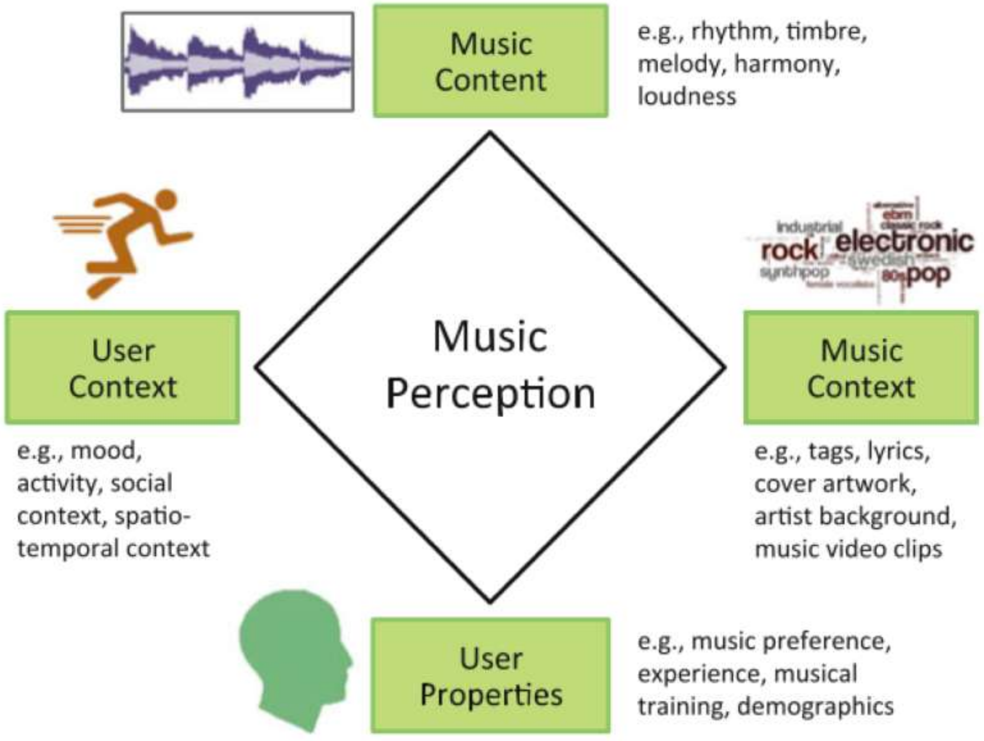




Outline

- Course Introduction
- Speech Processing
- **Music Information Retrieval**
- Klasifikasi Voice Gender pada Feature low-Level Suara dengan dan Deep Neural Networks

Music Retrieval



Music Content

e.g., rhythm, timbre, melody, harmony, loudness

User Context

e.g., mood, activity, social context, spatio-temporal context

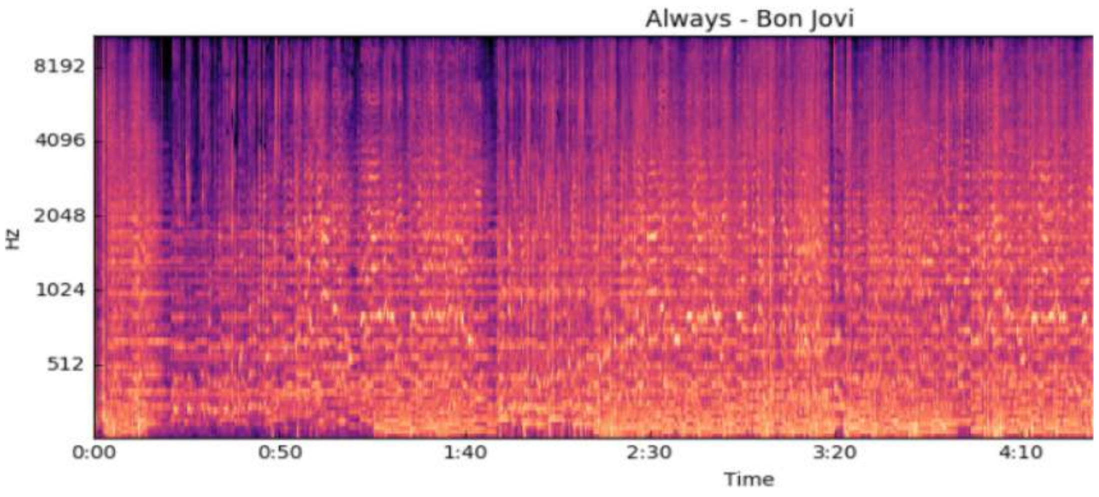
Music Perception

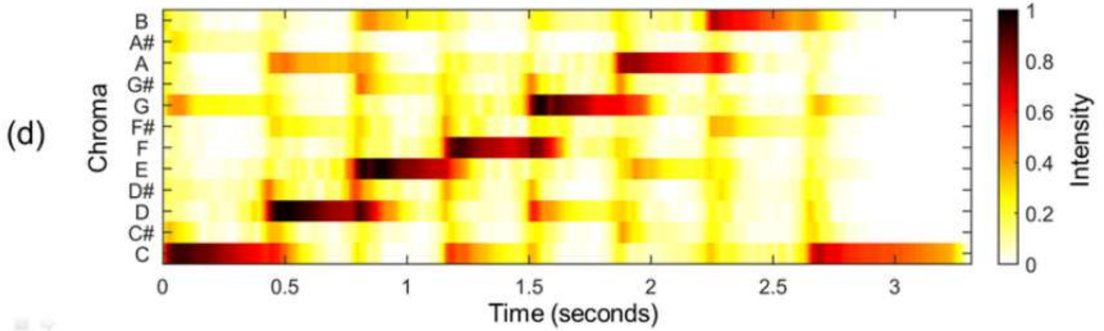
Music Context

e.g., tags, lyrics, cover artwork, artist background, music video clips

User Properties

e.g., music preference, experience, musical training, demographics

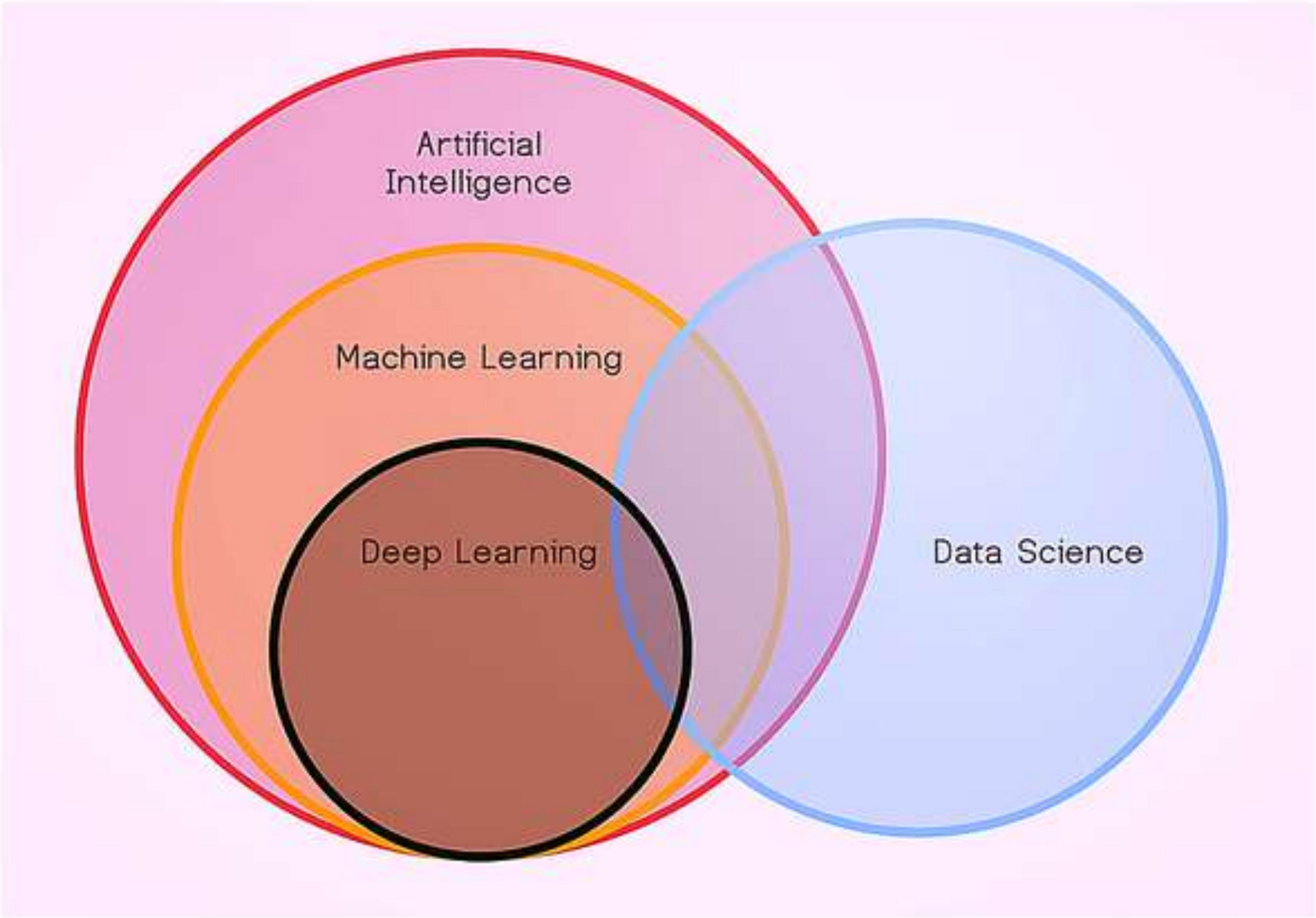




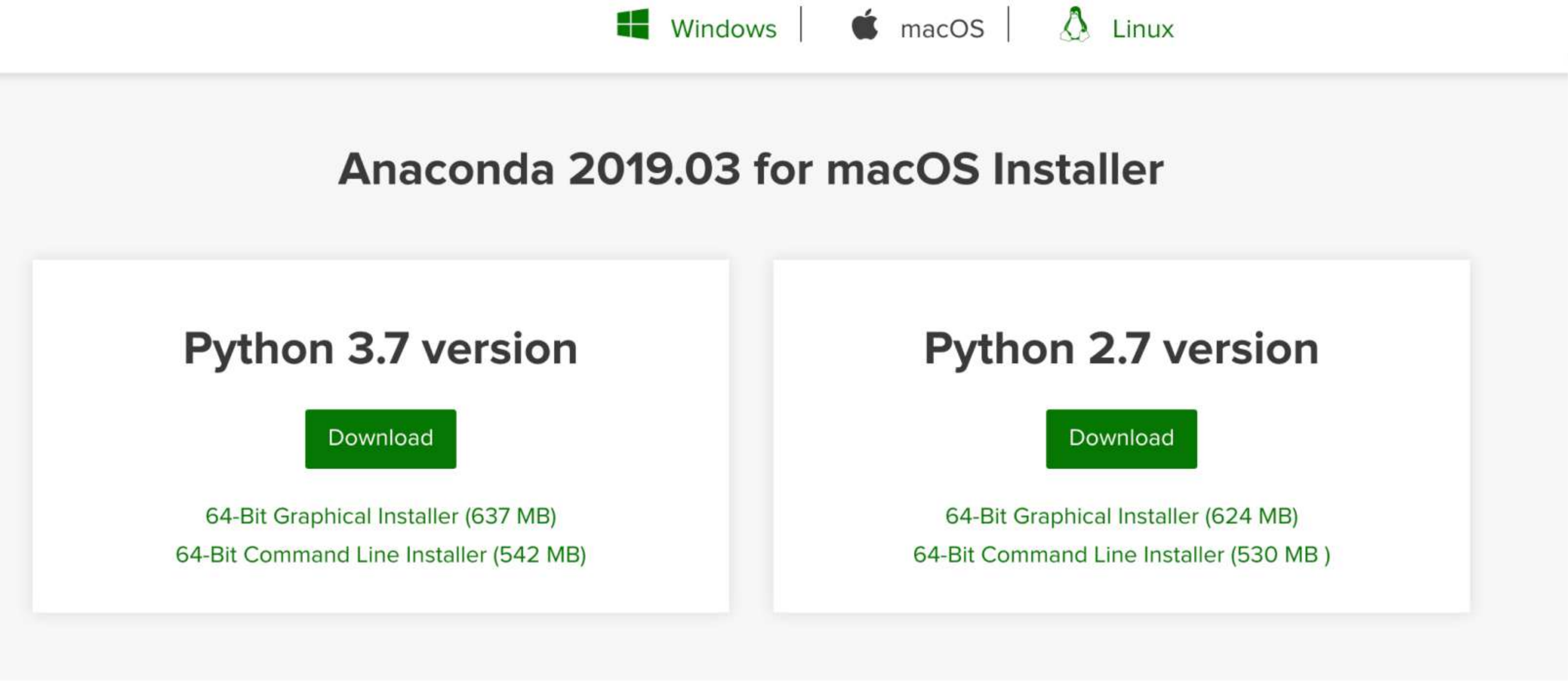
Outline

- Course Introduction
- Speech Processing
- Music Information Retrieval
- **Klasifikasi Voice Gender pada Feature low-Level Suara dengan dan Deep Neural Networks**

Artificial Intelligence



Python Install



Anaconda 2019.03 for macOS Installer

Python 3.7 version

64-Bit Graphical Installer (637 MB)

64-Bit Command Line Installer (542 MB)

Python 2.7 version

64-Bit Graphical Installer (624 MB)

64-Bit Command Line Installer (530 MB)

https://www.anaconda.com/distribution/

We Need Data!

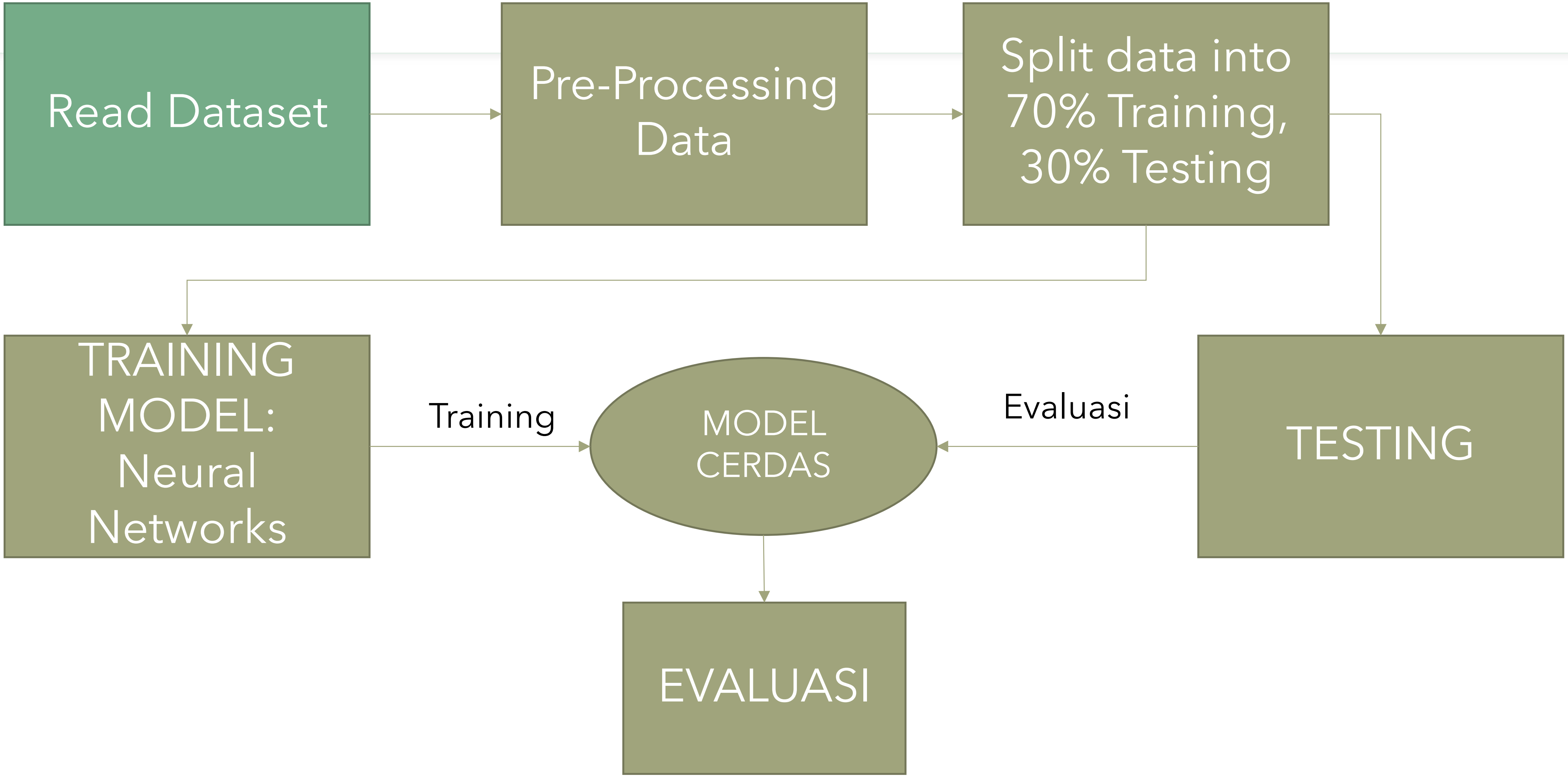
- https://archive.ics.uci.edu/ml/index.php
- https://www.kaggle.com/datasets
- https://data.go.id/

- https://www.kaggle.com/ronitf/heart-disease-uci
- http://faculty.neu.edu.cn/yunhyan/NEU_surface_defect_database.html

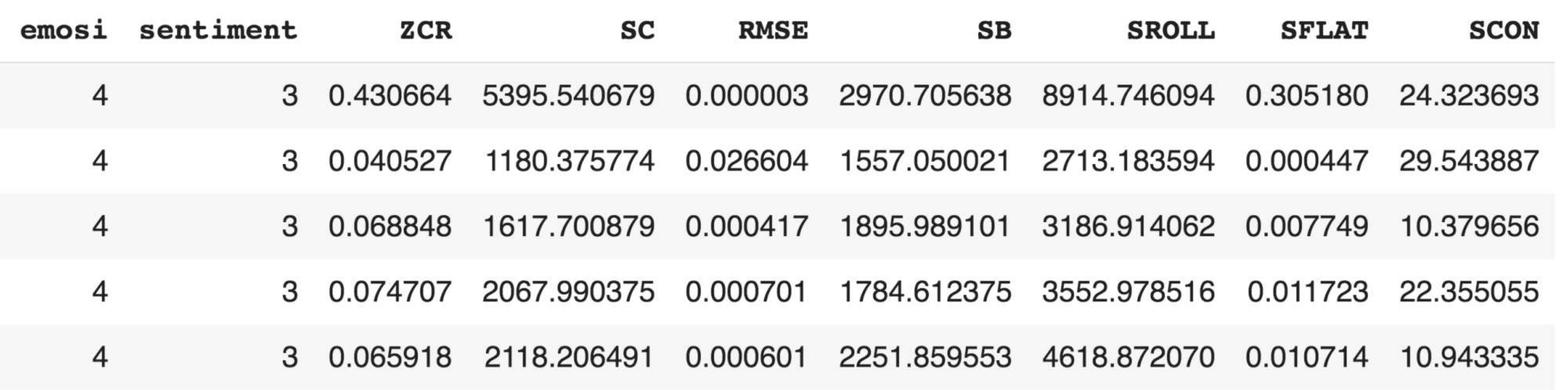
Programming

SciPy
django
NumPy
matplotlib
python
IP[y]:
IPython
Interactive Computing
pandas
$y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$
scikit
learn
OpenCV

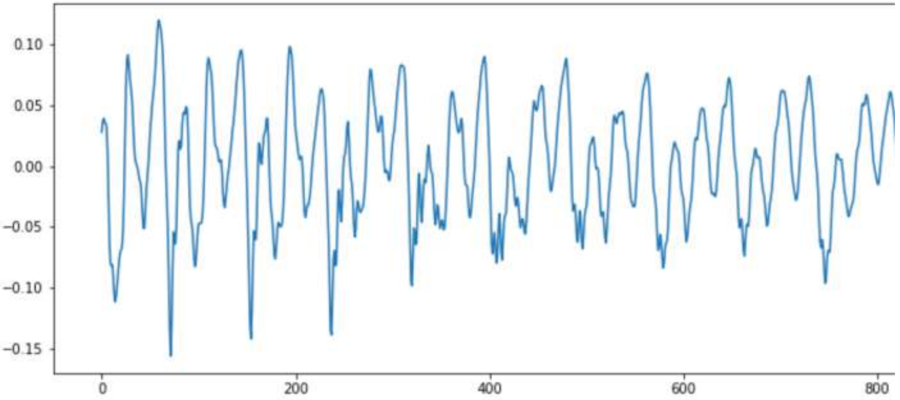
Flow Classification: Voice Gender

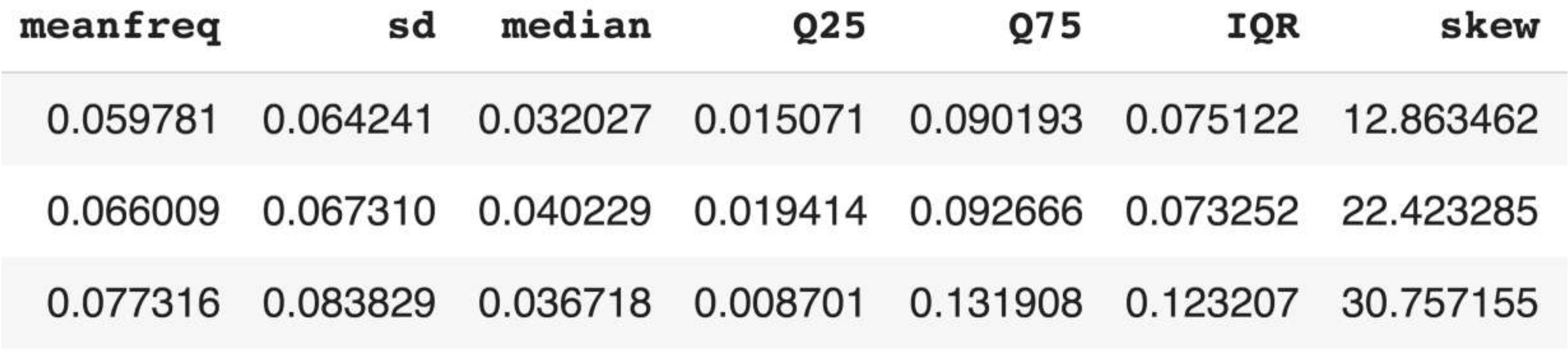


Flow Classification: Voice Gender Recognition



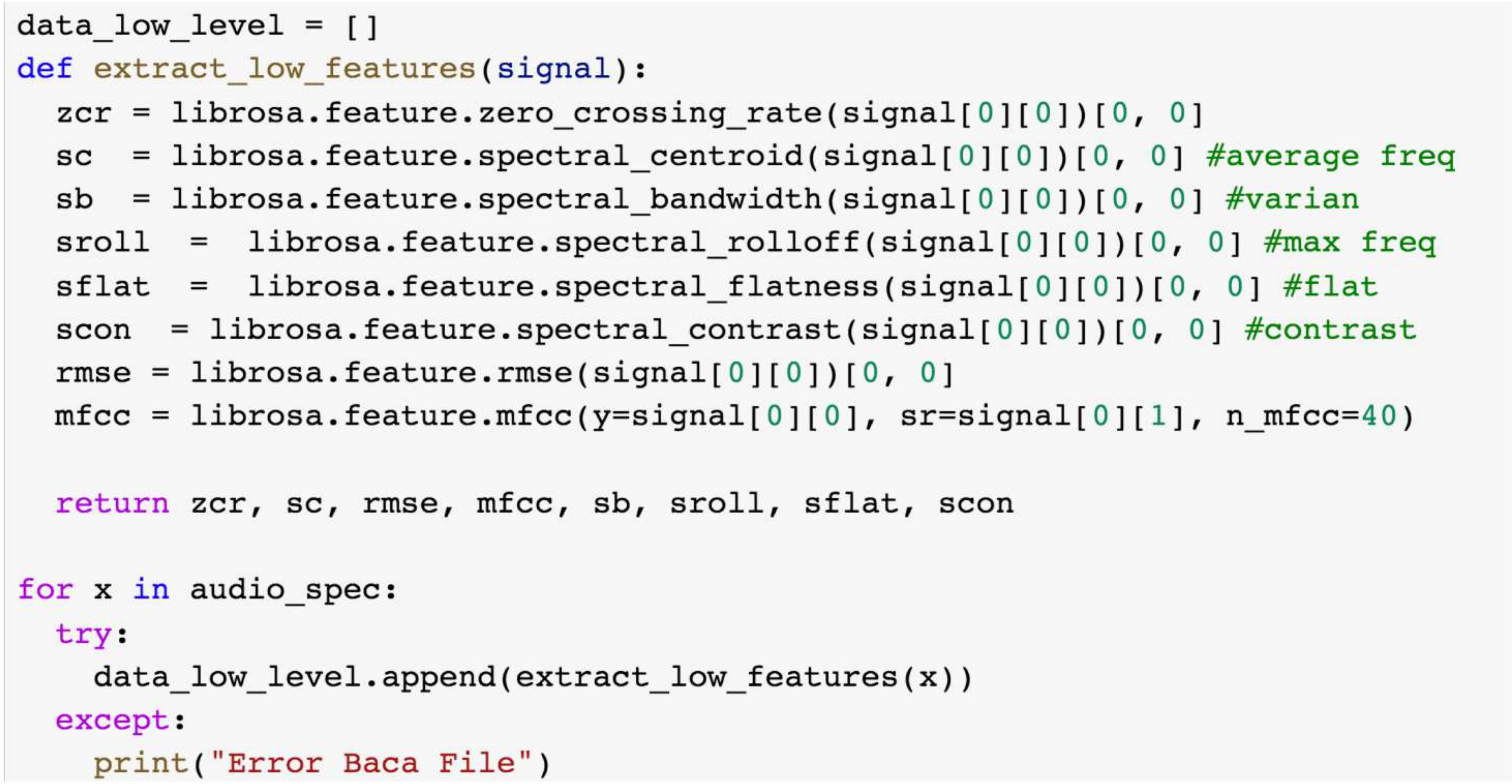
emosi	sentiment	ZCR	SC	RMSE	SB	SROLL	SFLAT	SCON
4	3	0.430664	5395.540679	0.000003	2970.705638	8914.746094	0.305180	24.323693
4	3	0.040527	1180.375774	0.026604	1557.050021	2713.183594	0.000447	29.543887
4	3	0.068848	1617.700879	0.000417	1895.989101	3186.914062	0.007749	10.379656
4	3	0.074707	2067.990375	0.000701	1784.612375	3552.978516	0.011723	22.355055
4	3	0.065918	2118.206491	0.000601	2251.859553	4618.872070	0.010714	10.943335





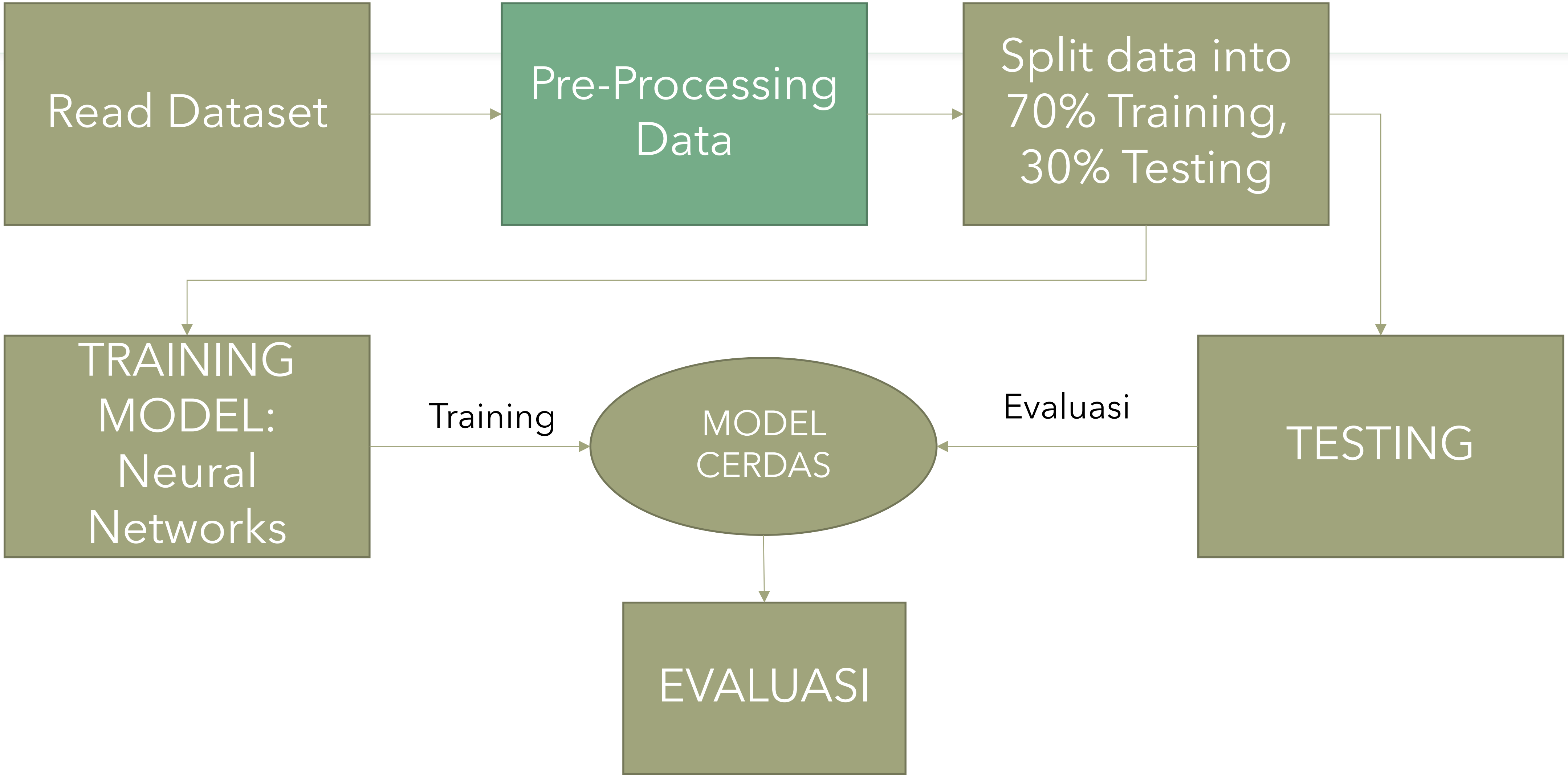
meanfreq	sd	median	Q25	Q75	IQR	skew
0.059781	0.064241	0.032027	0.015071	0.090193	0.075122	12.863462
0.066009	0.067310	0.040229	0.019414	0.092666	0.073252	22.423285
0.077316	0.083829	0.036718	0.008701	0.131908	0.123207	30.757155

Flow Classification: Voice Gender Recognition

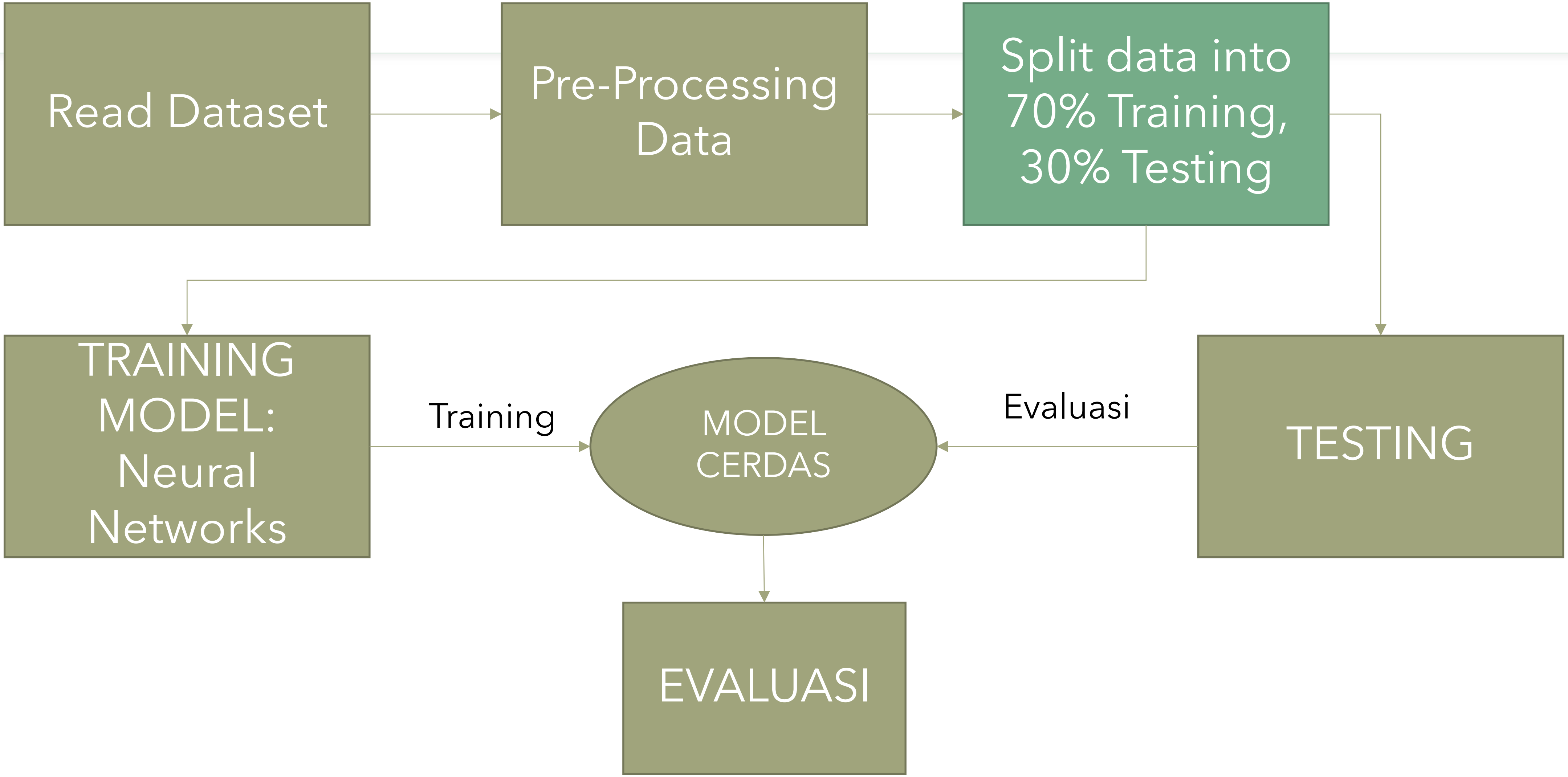


```
data_low_level = []
def extract_low_features(signal):
  zcr = librosa.feature.zero_crossing_rate(signal[0][0])[0, 0]
  sc  = librosa.feature.spectral_centroid(signal[0][0])[0, 0] #average freq
  sb  = librosa.feature.spectral_bandwidth(signal[0][0])[0, 0] #varian
  sroll  =  librosa.feature.spectral_rolloff(signal[0][0])[0, 0] #max freq
  sflat  =  librosa.feature.spectral_flatness(signal[0][0])[0, 0] #flat
  scon  = librosa.feature.spectral_contrast(signal[0][0])[0, 0] #contrast
  rmse = librosa.feature.rmse(signal[0][0])[0, 0]
  mfcc = librosa.feature.mfcc(y=signal[0][0], sr=signal[0][1], n_mfcc=40)

  return zcr, sc, rmse, mfcc, sb, sroll, sflat, scon

for x in audio_spec:
  try:
    data_low_level.append(extract_low_features(x))
  except:
    print("Error Baca File")
```

Flow Classification: Voice Gender



Flow Classification: Voice Gender



Read Dataset

Pre-Processing Data

Split data into 70% Training, 30% Testing

TRAINING MODEL: Neural Networks

Training

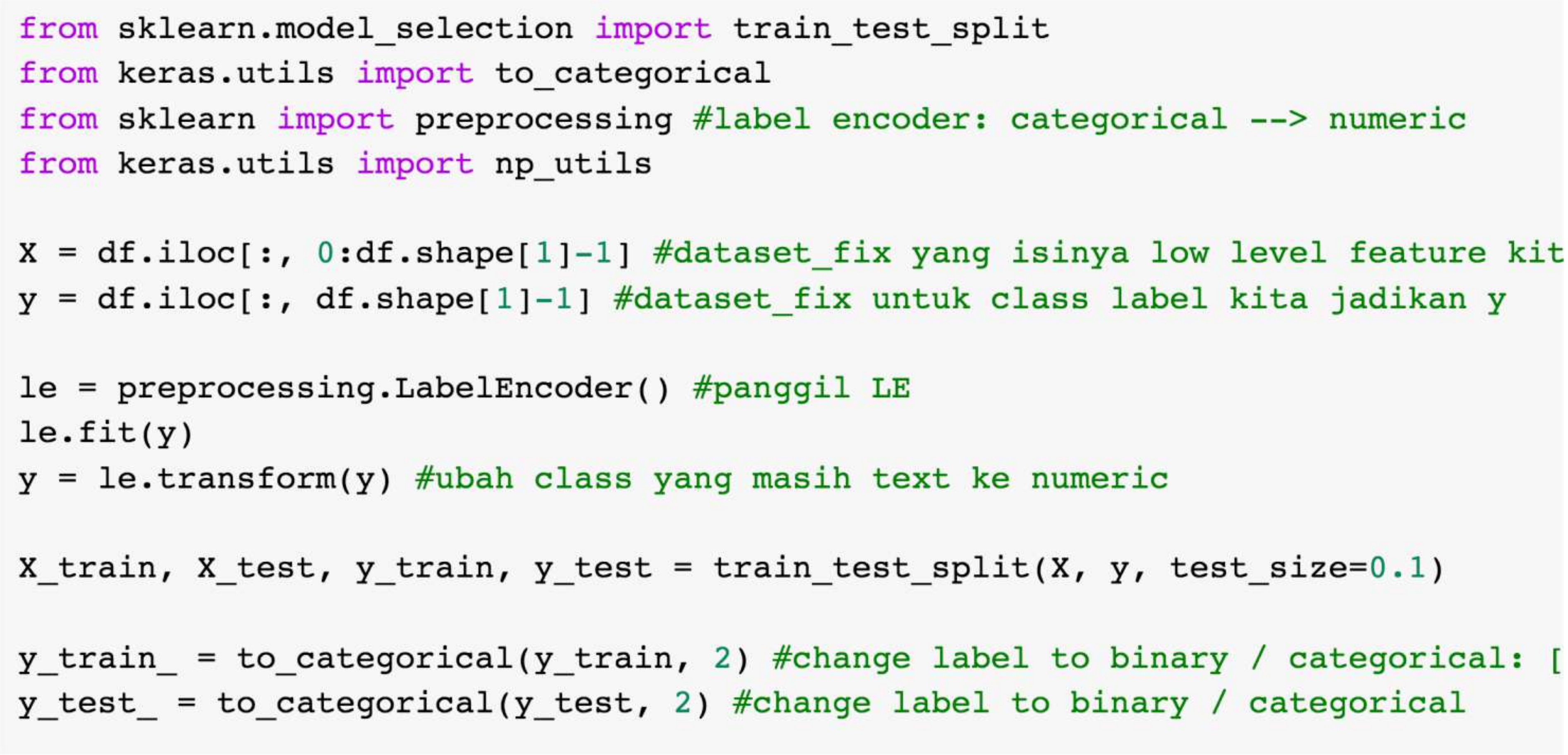
MODEL CERDAS

Evaluasi

TESTING

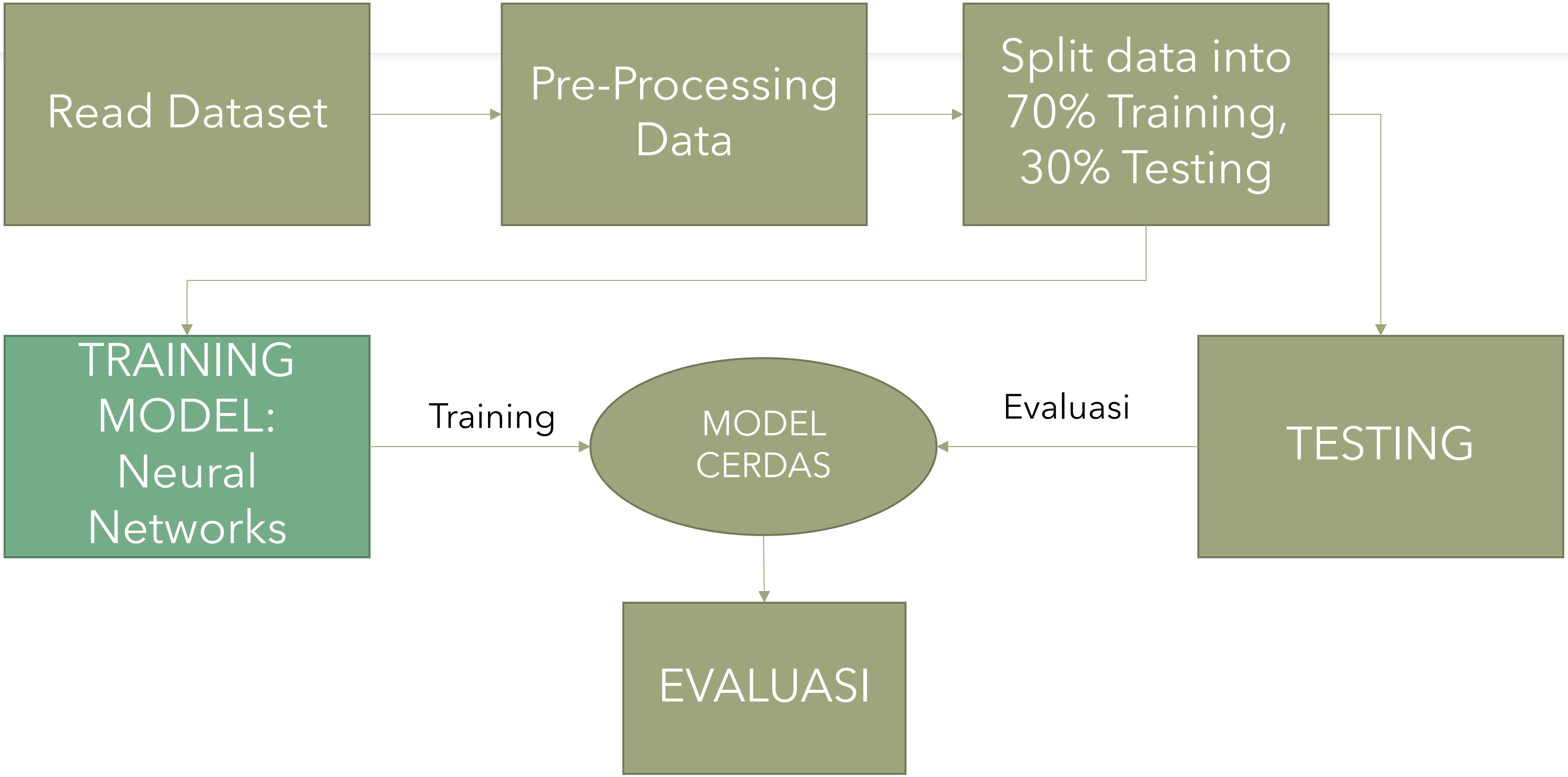
EVALUASI

Flow Classification: Voice Gender

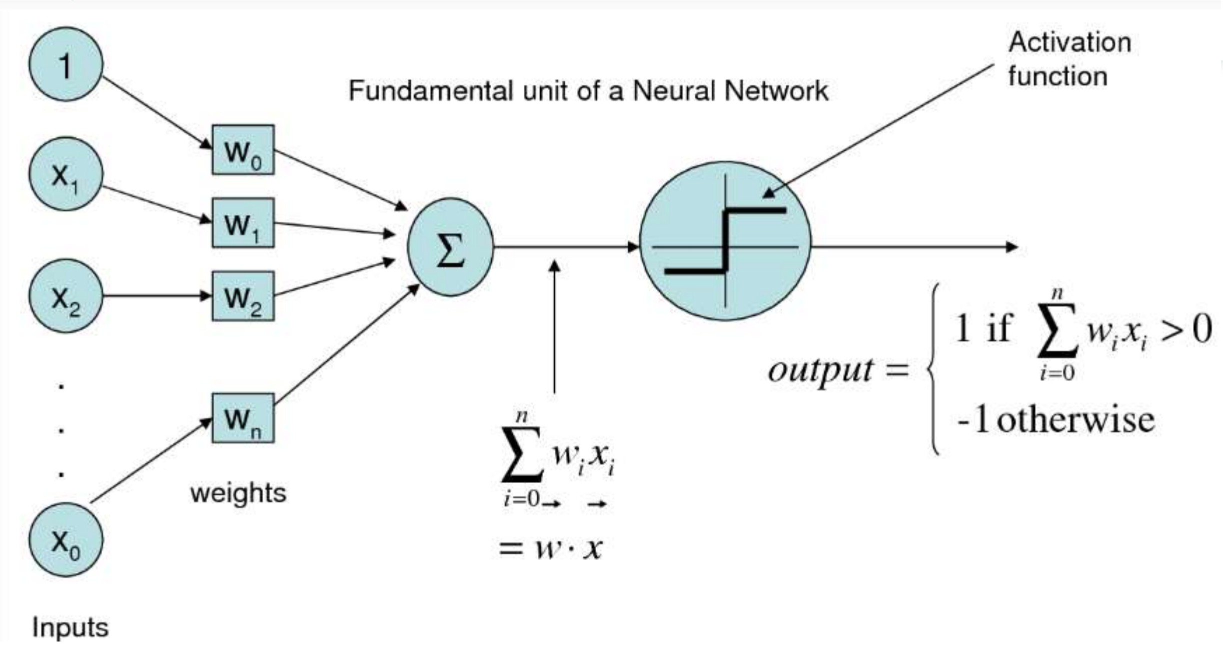


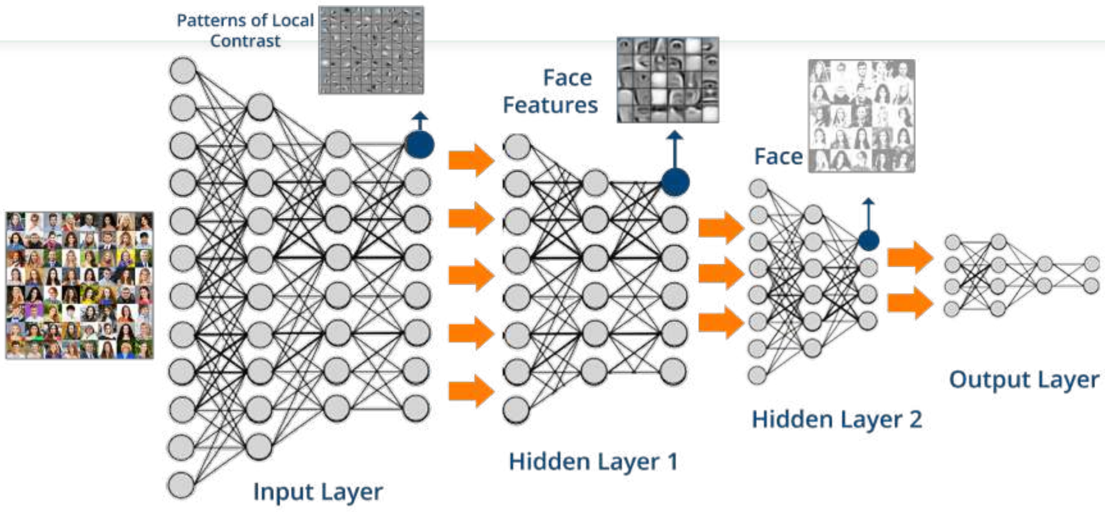
```
from sklearn.model_selection import train_test_split
from keras.utils import to_categorical
from sklearn import preprocessing #label encoder: categorical --> numeric
from keras.utils import np_utils

X = df.iloc[:, 0:df.shape[1]-1] #dataset_fix yang isinya low level feature kit
y = df.iloc[:, df.shape[1]-1] #dataset_fix untuk class label kita jadikan y

le = preprocessing.LabelEncoder() #panggil LE
le.fit(y)
y = le.transform(y) #ubah class yang masih text ke numeric

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.1)

y_train_ = to_categorical(y_train, 2) #change label to binary / categorical: [
y_test_ = to_categorical(y_test, 2) #change label to binary / categorical
```

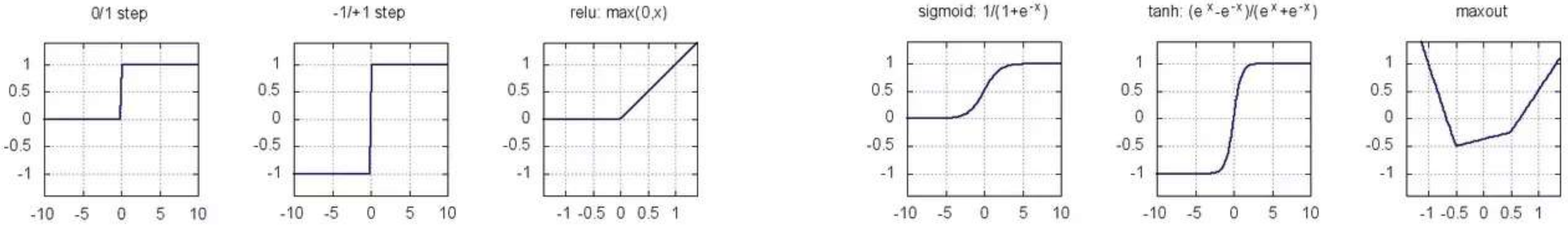
Flow Classification: Contoh Klasifikasi



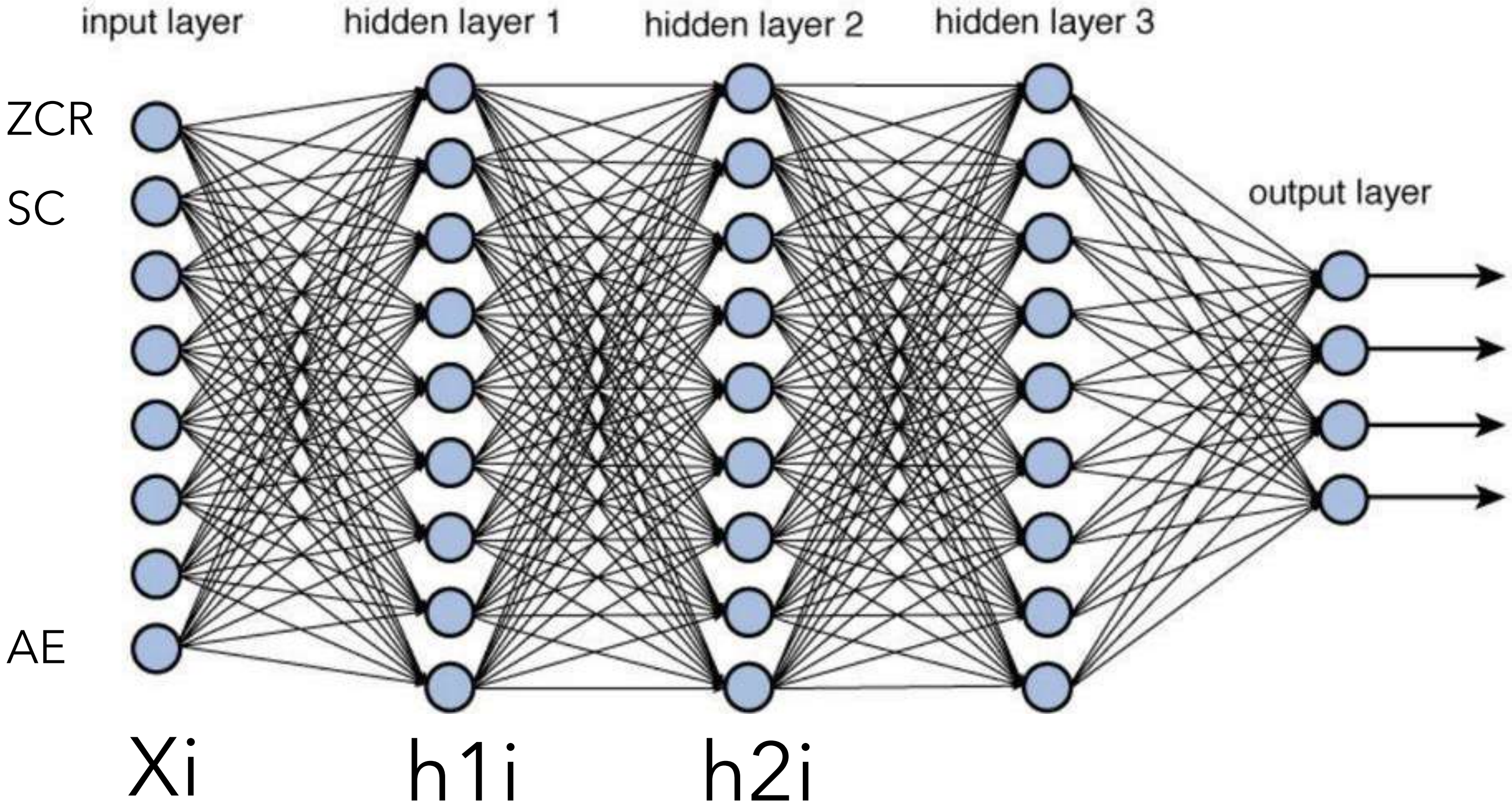
Neural Networks


Fundamental unit of a Neural Network
Activation function
1
x_1
x_2
x_0
w_0
w_1
w_2
w_n
weights
Inputs
Σ
$\sum_{i=0}^{n} w_i x_i = \vec{w} \cdot \vec{x}$
$output = \begin{cases} 1 \text{ if } \sum_{i=0}^{n} w_i x_i > 0 \\ -1 \text{ otherwise} \end{cases}$





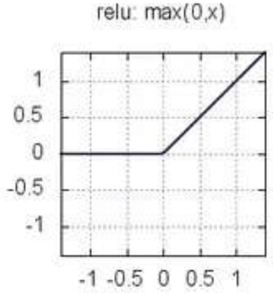
Flow Classification: Machine Learning Model



1 = male
0 = female

$$S(x, W) = \sum_{i=1}^{n} \sum_{j=1}^{m} x_{ij} W_{(i-m, j-n)}$$

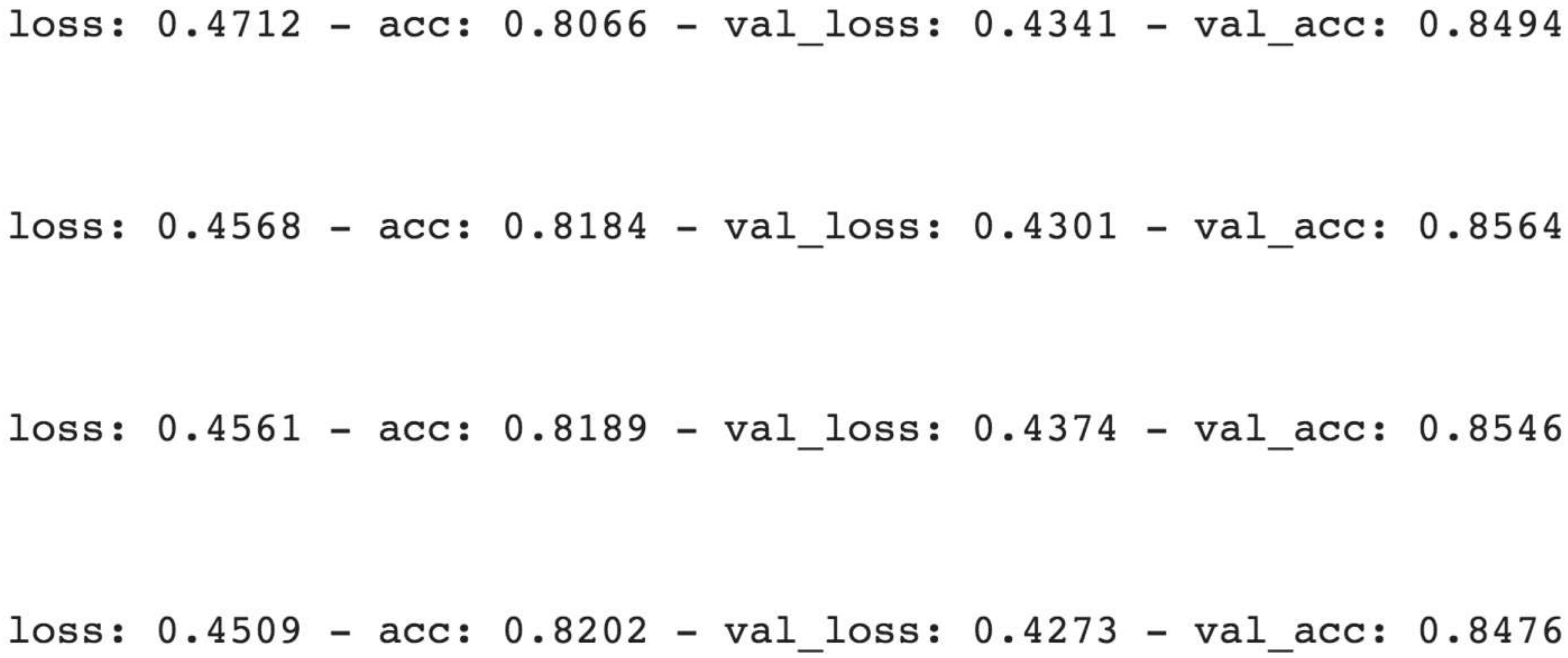
$$Z(S, U) = \sum_{i=1}^{n} \sum_{j=1}^{m} S_{ij} U_{(i-m, j-n)}$$



$$ReLU = \begin{cases} x, if\ x > 0 \\ 0, otherwise \end{cases}$$

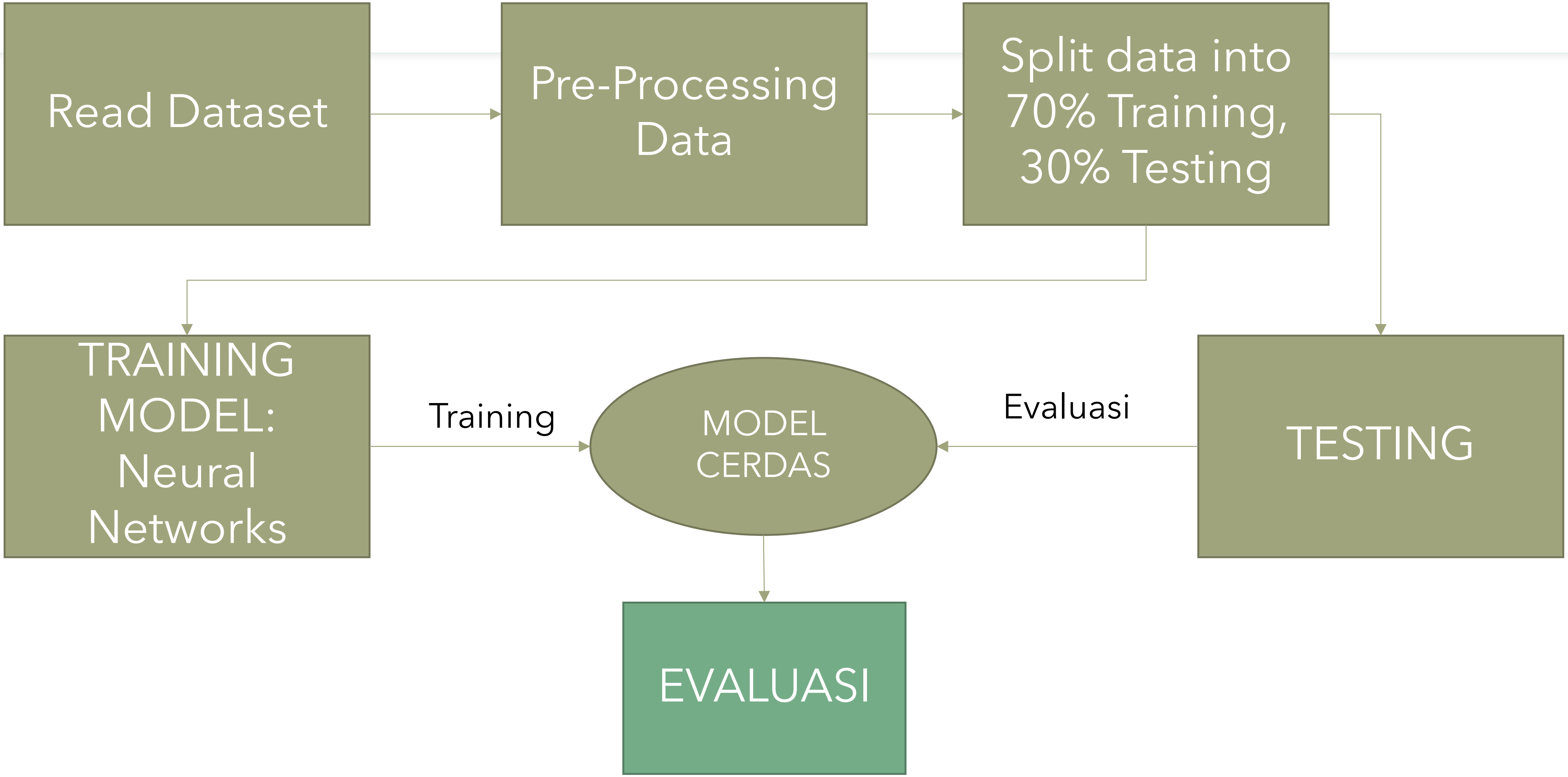
$$Softmax(z_i) = \frac{\exp(Dl(B_{ij} + h_i, W))}{\sum_{i=1}^{n} \exp(Dl(B_{ij} + h_i, W))}$$

Training Process



```
loss: 0.4712 - acc: 0.8066 - val_loss: 0.4341 - val_acc: 0.8494

loss: 0.4568 - acc: 0.8184 - val_loss: 0.4301 - val_acc: 0.8564

loss: 0.4561 - acc: 0.8189 - val_loss: 0.4374 - val_acc: 0.8546

loss: 0.4509 - acc: 0.8202 - val_loss: 0.4273 - val_acc: 0.8476
```

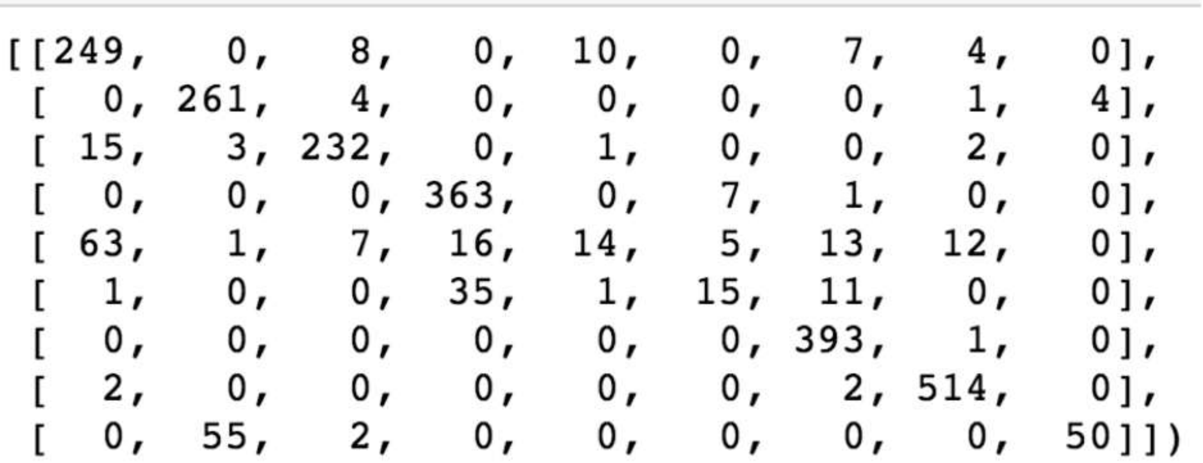
Flow Classification: Contoh Klasifikasi



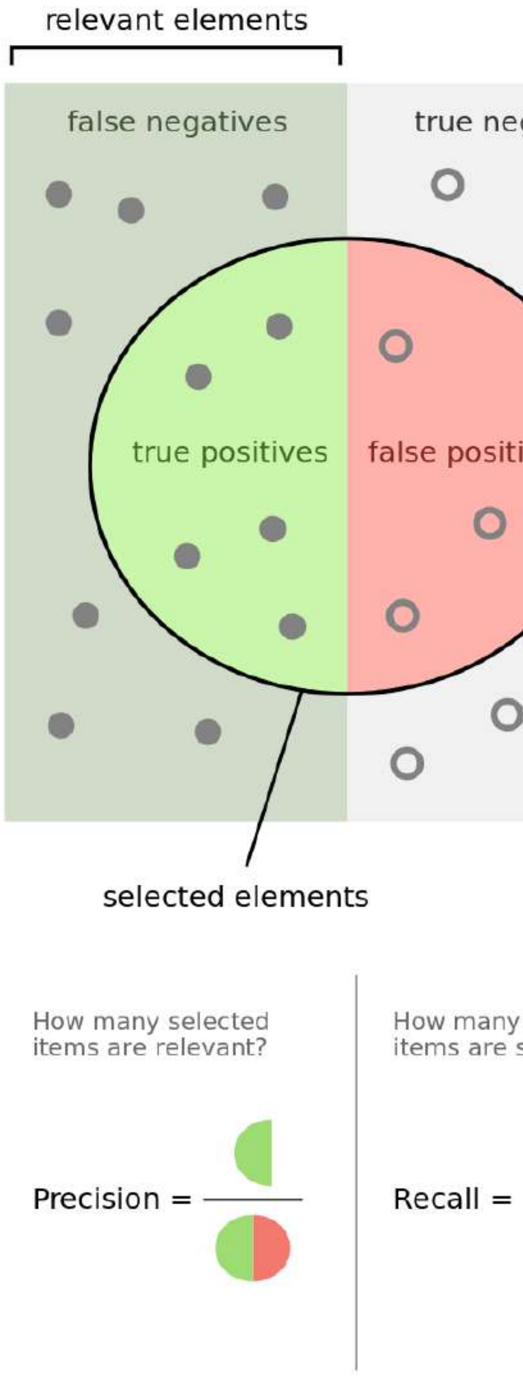
Precision Recall + Confusion Matrix

$$\text{Precision} = \frac{tp}{tp + fp} \qquad \text{Accuracy} = \frac{tp + tn}{tp + tn + fp + fn}$$

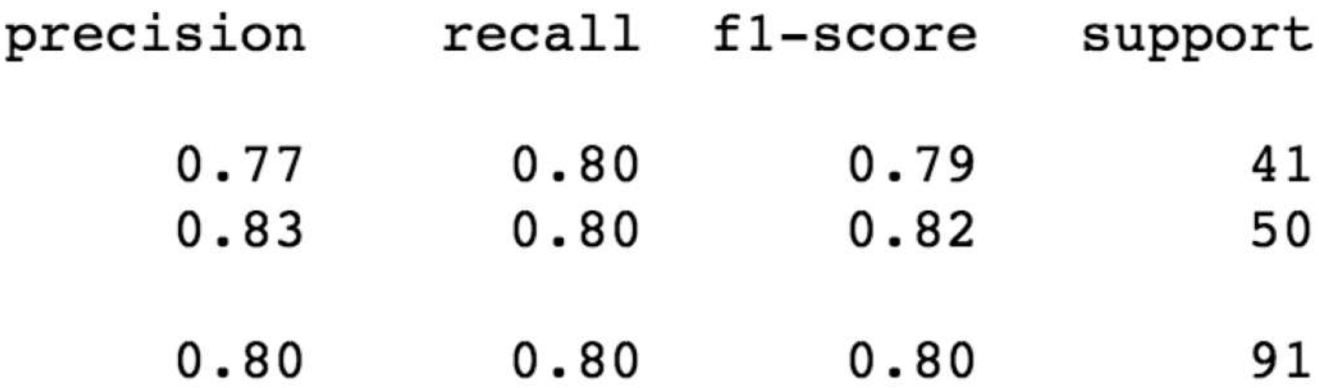
$$\text{Recall} = \frac{tp}{tp + fn} \qquad F = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$$



```
[[249,   0,   8,   0,  10,   0,   7,   4,   0],
 [  0, 261,   4,   0,   0,   0,   0,   1,   4],
 [ 15,   3, 232,   0,   1,   0,   0,   2,   0],
 [  0,   0,   0, 363,   0,   7,   1,   0,   0],
 [ 63,   1,   7,  16,  14,   5,  13,  12,   0],
 [  1,   0,   0,  35,   1,  15,  11,   0,   0],
 [  0,   0,   0,   0,   0,   0, 393,   1,   0],
 [  2,   0,   0,   0,   0,   0,   2, 514,   0],
 [  0,  55,   2,   0,   0,   0,   0,   0,  50]])
```



Flow Classification: Evaluasi



```
precision    recall  f1-score   support

     0.77      0.80      0.79        41
     0.83      0.80      0.82        50

     0.80      0.80      0.80        91
```