

Author: Tzu-Ching Wen

Date: 2021.03.04

Master Lab Course - Big Data Machine Learning

Autor:

Tzu-Ching Wen

Date: 2021.03.04

Ansprechpartner:

Philipp Wagner

Tobias Nagel



Universität Stuttgart



Fraunhofer

IPA

```
In [4]: import tensorflow as tf
from tensorflow import keras
from tensorflow.keras import layers
from keras.layers import *
from tensorflow.keras.layers.experimental import preprocessing
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import numpy as np
from scipy.stats import norm
from sklearn.preprocessing import StandardScaler
import scipy.stats as stats
import sklearn.linear_model as linear_model
from sklearn.model_selection import KFold
from IPython.display import HTML, display
from sklearn.manifold import TSNE
from sklearn.cluster import KMeans
from sklearn.decomposition import PCA
from sklearn.preprocessing import StandardScaler
from scipy import stats
import warnings
warnings.filterwarnings('ignore')
%matplotlib inline
import os
import sys
import IPython
from six.moves import urllib
import datetime
from glob import glob
```

Using TensorFlow backend.

Step1 (Aufgabe 1) Analyse the data--> get feeling on the data

- Building System with only one Folder Data(Bearing1_4)

```
In [2]: #1# MeasurementData_Bearing1_4_acc_00001.csv
df = pd.read_csv('C:/TCW/01_Uni-Stuttgart/Big Data Labs/bdml_projekt_zip/measurement
#dataset = dataset[0].str.split('; ', expand = True)
df.columns = ['Stunde', 'Minute', 'Sekunde', 'Mikrosekunde', 'Horiz Besch1', 'Vert B
print(df)
print(df.shape)
```

	Stunde	Minute	Sekunde	Mikrosekunde	Horiz Besch1	Vert Besch1
0	8	8	0	425040.0	0.065	-0.058
1	8	8	0	425080.0	0.438	0.179
2	8	8	0	425120.0	-0.079	0.646
3	8	8	0	425160.0	-0.523	-0.411
4	8	8	0	425200.0	-0.146	-0.387
...
2555	8	8	0	524840.0	-0.102	0.438
2556	8	8	0	524880.0	-0.556	0.386
2557	8	8	0	524920.0	-0.762	0.371
2558	8	8	0	524960.0	0.015	0.136
2559	8	8	0	525000.0	0.580	0.265

```
[2560 rows x 6 columns]
(2560, 6)
```

Step2 (Aufgabe 1)Plot: Mikrosekunde - Horizontal Beschleunigung

```
In [3]: x = df['Mikrosekunde']
print(x)

y = df['Horiz Besch1']
print(y)

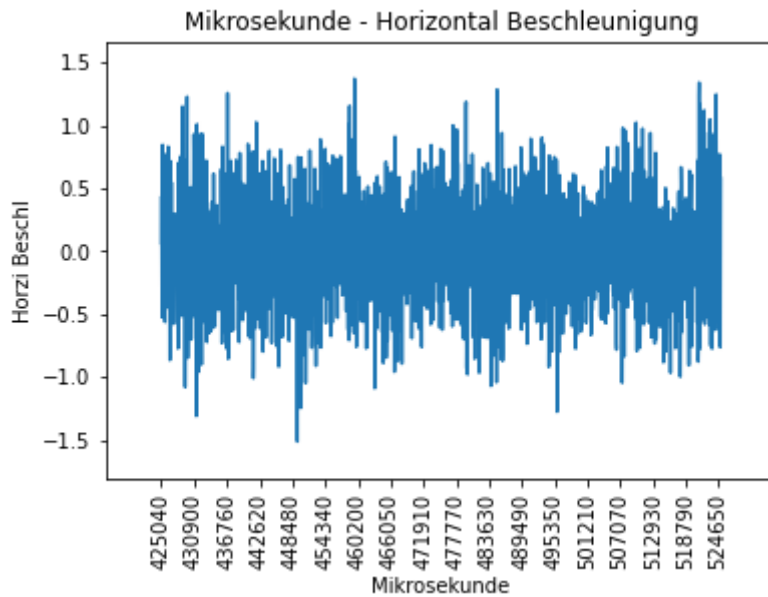
#title & labels
plt.title('Mikrosekunde - Horizontal Beschleunigung')
plt.xlabel('Mikrosekunde')
plt.ylabel('Horzi Besch1')

#print only very 150
plt.plot(x,y)
plt.xticks(x[::150], rotation='vertical')
plt.margins(0.1)
plt.show()
```

```
0      425040.0
1      425080.0
2      425120.0
3      425160.0
4      425200.0
...
2555    524840.0
2556    524880.0
2557    524920.0
2558    524960.0
2559    525000.0
Name: Mikrosekunde, Length: 2560, dtype: float64
0      0.065
1      0.438
2     -0.079
3     -0.523
4     -0.146
...
2555   -0.102
2556   -0.556
```

```
2557    -0.762
2558     0.015
2559     0.580
```

Name: Horiz Beschl, Length: 2560, dtype: float64



Step3 (Aufgabe 1)Plot: Mikrosekunde - Vertical Beschleunigung

```
In [4]: x = df['Mikrosekunde']
        print(x)

        y = df['Vert Beschl']
        print(y)

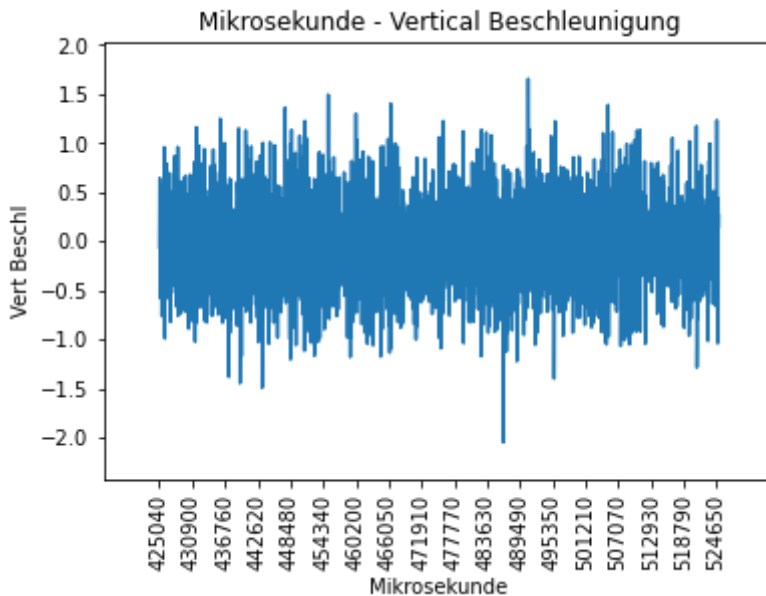
        #title & labels
        plt.title('Mikrosekunde - Vertical Beschleunigung')
        plt.xlabel('Mikrosekunde')
        plt.ylabel('Vert Beschl')

        #print only very 150
        plt.plot(x,y)
        plt.xticks(x[::150], rotation='vertical')
        plt.margins(0.1)
        plt.show()
```

```
0      425040.0
1      425080.0
2      425120.0
3      425160.0
4      425200.0
...
2555    524840.0
2556    524880.0
2557    524920.0
2558    524960.0
2559    525000.0
Name: Mikrosekunde, Length: 2560, dtype: float64
0      -0.058
1       0.179
2       0.646
3      -0.411
4      -0.387
```

```
...
2555    0.438
2556    0.386
2557    0.371
2558    0.136
2559    0.265
```

Name: Vert Beschl, Length: 2560, dtype: float64



Step4

Create func1 --> to generate 'state_df'

```
In [5]: ### create function: get each csv's label + filename ###
def get_state_filename(key, file_path_str):
    label_folder_files = sorted(glob(file_path_str))
    label_folder_files

    #create dict
    state_dict = {}
    state_list = []

    #for i in range(len(label_folder_files)):
    for i in range(len(label_folder_files)):
        filename = label_folder_files[i]
        df_ = pd.read_csv(filename, delimiter=',', header=None)
        df_ = df_.drop(df_.columns[0], axis=1)
        df_ = df_.drop([0])
        df_.columns = ['file', 'state']
        #df_filename = df_['file']
        #df_label = df_['state']
        state_list.append(df_)
        #replace 1 with 0/ replace 2 with 1(only 2 labels)

        #state_dict['{}'.format(i)] = [df_filename, df_label]
        #state_dict[key[i]] = [df_filename, df_label]

    return state_list#state_dict
```

```
In [6]: folder_files = sorted(glob('C:/TCW/01_Uni-Stuttgart/Big Data Labs/bdml_projekt_7z/bdml_projekt_7z/'))
        folder_files
```

```
Out[6]: ['C:/TCW/01_Uni-Stuttgart/Big Data Labs/bdml_projekt_7z/bdml_projekt_7z/measurement_
data/measurement_data/Bearing1_4\\acc_00001.csv',
'C:/TCW/01_Uni-Stuttgart/Big Data Labs/bdml_projekt_7z/bdml_projekt_7z/measurement_
data/measurement_data/Bearing1_4\\acc_00002.csv',
```

5/103

6/103

localhost:8888/nbconvert/html/FinalProject/MasterLab Big Data Final Project Tzu-Ching Wen .ipynb?download=false

localhost:8888/nbconvert/html/FinalProject/MasterLab Big Data Final Project Tzu-Ching Wen .ipynb?download=false

localhost:8888/nbconvert/html/FinalProject/MasterLab Big Data Final Project Tzu-Ching Wen .ipynb?download=false

localhost:8888/nbconvert/html/FinalProject/MasterLab Big Data Final Project Tzu-Ching Wen .ipynb?download=false

11/103

12/103

localhost:8888/nbconvert/html/FinalProject/MasterLab Big Data Final Project Tzu-Ching Wen .ipynb?download=false

14/103

15/103

16/103

17/103

18/103

19/103

20/103

21/103

22/103

23/103

localhost:8888/nbconvert/html/FinalProject/MasterLab Big Data Final Project Tzu-Ching Wen .ipynb?download=false

25/103

26/103

27/103

28/103

29/103

localhost:8888/nbconvert/html/FinalProject/MasterLab Big Data Final Project Tzu-Ching Wen .ipynb?download=false

31/103

32/103

stacking)

- to (Aufgabe 6): in this step will prepare extra features like: 'absolute mean value horizontal' & 'absolute mean value horizontal'. These features will be used to generate different datasets(df_features_0, df_features_1, df_features_2 and df_features_3) for training(see Step13).

```
In [7]: def calculate_each_csv(file_path_str):
        folder_files = sorted(glob(file_path_str))

        #create empty lists
        abs_mean_h_list = []
        abs_mean_v_list = []
        mean_h_list = []
        mean_v_list = []
        stand_h_list = []
        stand_v_list = []
        max_h_list = []
        min_h_list = []
        max_v_list = []
        min_v_list = []
        time_deviation_list = []

        for i in range(len(folder_files)):
            filename = folder_files[i]
            #df = pd.read_csv(filename, delimiter=',', header=None)
            df = pd.read_csv(filename, sep=';|,')

            #add columns head
            df.columns = ['Stunde', 'Minute', 'Sekunde', 'Mikrosekunde', 'Horiz Beschl']

            #calculate abs_mean_h, abs_mean_v, mean_h, mean_v, stand_h, stand_v, time_de
            abs_mean_h = df['Horiz Beschl'].abs().mean().round(5)
            abs_mean_v = df['Vert Beschl'].abs().mean().round(5)
            mean_h = df['Horiz Beschl'].mean().round(5)
            mean_v = df['Vert Beschl'].mean().round(5)
            stand_h = df['Horiz Beschl'].std().round(5)
            stand_v = df['Vert Beschl'].std().round(5)
            max_h = df['Horiz Beschl'].max().round(5)
            min_h = df['Horiz Beschl'].min().round(5)
            max_v = df['Vert Beschl'].max().round(5)
            min_v = df['Vert Beschl'].min().round(5)
            time_deviation = int(df['Mikrosekunde'].max() - df['Mikrosekunde'].min())

            #append to list
            abs_mean_h_list.append(abs_mean_h)
            abs_mean_v_list.append(abs_mean_v)
            mean_h_list.append(mean_h)
            mean_v_list.append(mean_v)
            stand_h_list.append(stand_h)
            stand_v_list.append(stand_v)
            max_h_list.append(max_h)
            min_h_list.append(min_h)
            max_v_list.append(max_v)
            min_v_list.append(min_v)
            time_deviation_list.append(time_deviation)

        #Create new dataframe
        df_new = pd.DataFrame({'abs_mean_h': abs_mean_h_list,
                               'abs_mean_v': abs_mean_v_list,
                               'mean_h': mean_h_list,
                               'mean_v': mean_v_list,
```

```

        'stand_h'      :      stand_h_list,
        'stand_v'      :      stand_v_list,
        'max_h'        :      max_h_list,
        'min_h'        :      min_h_list,
        'max_v'        :      max_v_list,
        'min_v'        :      min_v_list,
        'time_deviation': time_deviation_list
    })

    return df_new

```

Step6 (Aufgabe 2) Only load bearing1_4 folder csv file

Goal is to plot 4 plots: (x axis, y axis) --> to analyse the tendency against time

- Plot_1: (time, Mean_h)
- Plot_2: (time, Mean_v)
- Plot_3: (time, StandardDeviation_h)
- Plot_4: (time, StandardDeviation_v)

```

In [8]: #use function
key = ['Bearing1_4', 'Bearing1_5', 'Bearing1_6',
       'Bearing1_7', 'Bearing2_4', 'Bearing2_5',
       'Bearing2_6', 'Bearing2_7', 'Bearing3_3']

```

```

In [9]: #1# Load only Bearing1_4_health_state csv files
state_list_B1_4 = get_state_filename(key, 'C:/TCW/01_Uni-Stuttgart/Big Data Labs/bdm

#2#concatenate a list of dataframes together#
state_df_B1_4 = pd.concat(state_list_B1_4)
state_df_B1_4.reset_index(drop=True, inplace=True)

print('state_df_B1_4')
print(state_df_B1_4)
print(state_df_B1_4.shape)

#3# reduce from 3 to 2 classes
state_df_B1_4['state'] = state_df_B1_4['state'].replace({'0': 0, '1': 0, '2': 1})
state_df_B1_4['state']

print()
print('state_df_B1_4_reduced')
print(state_df_B1_4)

#calculat frequency
state_df_B1_4['state'].value_counts()

#4# Load only Bearing1_4 folder csv files --> generate dataframe
Bearing1_4_df = calculate_each_csv('C:/TCW/01_Uni-Stuttgart/Big Data Labs/bdml_proje
print('Bearing1_4_df')
print(Bearing1_4_df)

#5# concate 'features df' & 'label df'
B1_4_df = pd.concat([state_df_B1_4, Bearing1_4_df], axis=1)
print(B1_4_df)

```

```

print(B1_4_df.shape)
print(B1_4_df.columns.values)
B1_4_df['state'].value_counts()

#6# drop columns 'file' & 'time_deviation'
B1_4_df = B1_4_df.drop(columns=['file', 'time_deviation'], axis=1)
B1_4_df

```

state_df_B1_4

	file	state
0	acc_00001.csv	0
1	acc_00002.csv	0
2	acc_00003.csv	0
3	acc_00004.csv	0
4	acc_00005.csv	0
...
1423	acc_01424.csv	2
1424	acc_01425.csv	2
1425	acc_01426.csv	2
1426	acc_01427.csv	2
1427	acc_01428.csv	2

[1428 rows x 2 columns]
(1428, 2)

state_df_B1_4_reduced

	file	state
0	acc_00001.csv	0
1	acc_00002.csv	0
2	acc_00003.csv	0
3	acc_00004.csv	0
4	acc_00005.csv	0
...
1423	acc_01424.csv	1
1424	acc_01425.csv	1
1425	acc_01426.csv	1
1426	acc_01427.csv	1
1427	acc_01428.csv	1

[1428 rows x 2 columns]

Bearing1_4_df

	abs_mean_h	abs_mean_v	mean_h	mean_v	stand_h	stand_v	max_h	\
0	0.32318	0.35930	0.00636	0.00167	0.40337	0.45502	1.373	
1	0.31214	0.36394	-0.00900	0.00669	0.39068	0.45886	1.299	
2	0.31035	0.38840	-0.00622	-0.00830	0.39190	0.49150	1.313	
3	0.33253	0.38023	-0.00582	-0.00175	0.41586	0.47481	1.508	
4	0.31086	0.40921	-0.00202	0.00663	0.38677	0.51172	1.334	
...
1423	7.86718	11.04806	0.04985	-0.32804	10.54333	14.54749	48.128	
1424	7.91047	10.61781	-0.11316	-0.03484	10.55686	14.07207	48.128	
1425	8.29157	10.74766	-0.16524	0.06929	11.10476	14.19700	48.128	
1426	8.14881	11.08156	-0.14566	0.26825	10.89777	14.59728	48.128	
1427	7.18050	8.18296	0.17245	0.76412	9.33412	10.48393	48.128	

	min_h	max_v	min_v	time_deviation
0	-1.511	1.658	-2.045	99920
1	-1.446	1.537	-1.685	99920
2	-1.505	2.161	-1.872	99920
3	-1.476	1.637	-2.033	99920
4	-1.225	1.967	-1.690	99920
...
1423	-41.133	47.849	-47.843	99920
1424	-39.357	47.849	-47.843	99920
1425	-46.942	47.849	-47.843	99920
1426	-48.148	47.849	-47.843	99920
1427	-41.573	47.849	-41.680	99920

[1428 rows x 11 columns]

```

      file state abs_mean_h abs_mean_v mean_h mean_v \
0 acc_00001.csv 0 0.32318 0.35930 0.00636 0.00167
1 acc_00002.csv 0 0.31214 0.36394 -0.00900 0.00669
2 acc_00003.csv 0 0.31035 0.38840 -0.00622 -0.00830
3 acc_00004.csv 0 0.33253 0.38023 -0.00582 -0.00175
4 acc_00005.csv 0 0.31086 0.40921 -0.00202 0.00663
...
1423 acc_01424.csv 1 7.86718 11.04806 0.04985 -0.32804
1424 acc_01425.csv 1 7.91047 10.61781 -0.11316 -0.03484
1425 acc_01426.csv 1 8.29157 10.74766 -0.16524 0.06929
1426 acc_01427.csv 1 8.14881 11.08156 -0.14566 0.26825
1427 acc_01428.csv 1 7.18050 8.18296 0.17245 0.76412

      stand_h stand_v max_h min_h max_v min_v time_deviation
0 0.40337 0.45502 1.373 -1.511 1.658 -2.045 99920
1 0.39068 0.45886 1.299 -1.446 1.537 -1.685 99920
2 0.39190 0.49150 1.313 -1.505 2.161 -1.872 99920
3 0.41586 0.47481 1.508 -1.476 1.637 -2.033 99920
4 0.38677 0.51172 1.334 -1.225 1.967 -1.690 99920
...
1423 10.54333 14.54749 48.128 -41.133 47.849 -47.843 99920
1424 10.55686 14.07207 48.128 -39.357 47.849 -47.843 99920
1425 11.10476 14.19700 48.128 -46.942 47.849 -47.843 99920
1426 10.89777 14.59728 48.128 -48.148 47.849 -47.843 99920
1427 9.33412 10.48393 48.128 -41.573 47.849 -41.680 99920

```

[1428 rows x 13 columns]

(1428, 13)

```

['file' 'state' 'abs_mean_h' 'abs_mean_v' 'mean_h' 'mean_v' 'stand_h'
 'stand_v' 'max_h' 'min_h' 'max_v' 'min_v' 'time_deviation']

```

Out[9]:

	state	abs_mean_h	abs_mean_v	mean_h	mean_v	stand_h	stand_v	max_h	min_h	max_v
0	0	0.32318	0.35930	0.00636	0.00167	0.40337	0.45502	1.373	-1.511	1.65
1	0	0.31214	0.36394	-0.00900	0.00669	0.39068	0.45886	1.299	-1.446	1.53
2	0	0.31035	0.38840	-0.00622	-0.00830	0.39190	0.49150	1.313	-1.505	2.16
3	0	0.33253	0.38023	-0.00582	-0.00175	0.41586	0.47481	1.508	-1.476	1.63
4	0	0.31086	0.40921	-0.00202	0.00663	0.38677	0.51172	1.334	-1.225	1.96
...
1423	1	7.86718	11.04806	0.04985	-0.32804	10.54333	14.54749	48.128	-41.133	47.84
1424	1	7.91047	10.61781	-0.11316	-0.03484	10.55686	14.07207	48.128	-39.357	47.84
1425	1	8.29157	10.74766	-0.16524	0.06929	11.10476	14.19700	48.128	-46.942	47.84
1426	1	8.14881	11.08156	-0.14566	0.26825	10.89777	14.59728	48.128	-48.148	47.84
1427	1	7.18050	8.18296	0.17245	0.76412	9.33412	10.48393	48.128	-41.573	47.84

1428 rows x 11 columns

(Aufgabe 2)Plotting

- Plot_0: (time, state)
- Plot_1: (time, Abs_Mean_h)
- Plot_2: (time, Abs_Mean_v)
- Plot_3: (time, Mean_h)
- Plot_4: (time, Mean_v)

```
In [10]: # here use index as Zeitliche Verlauf
# Plot_0: (time, state)
B1_4_df.reset_index().plot(x='index', y='state')
print(B1_4_df)
plt.show()

# Plot_1: (time, Abs_Mean_h)
B1_4_df.reset_index().plot(x='index', y='abs_mean_h')
plt.show()

# Plot_2: (time, Abs_Mean_v)
B1_4_df.reset_index().plot(x='index', y='abs_mean_v')
plt.show()

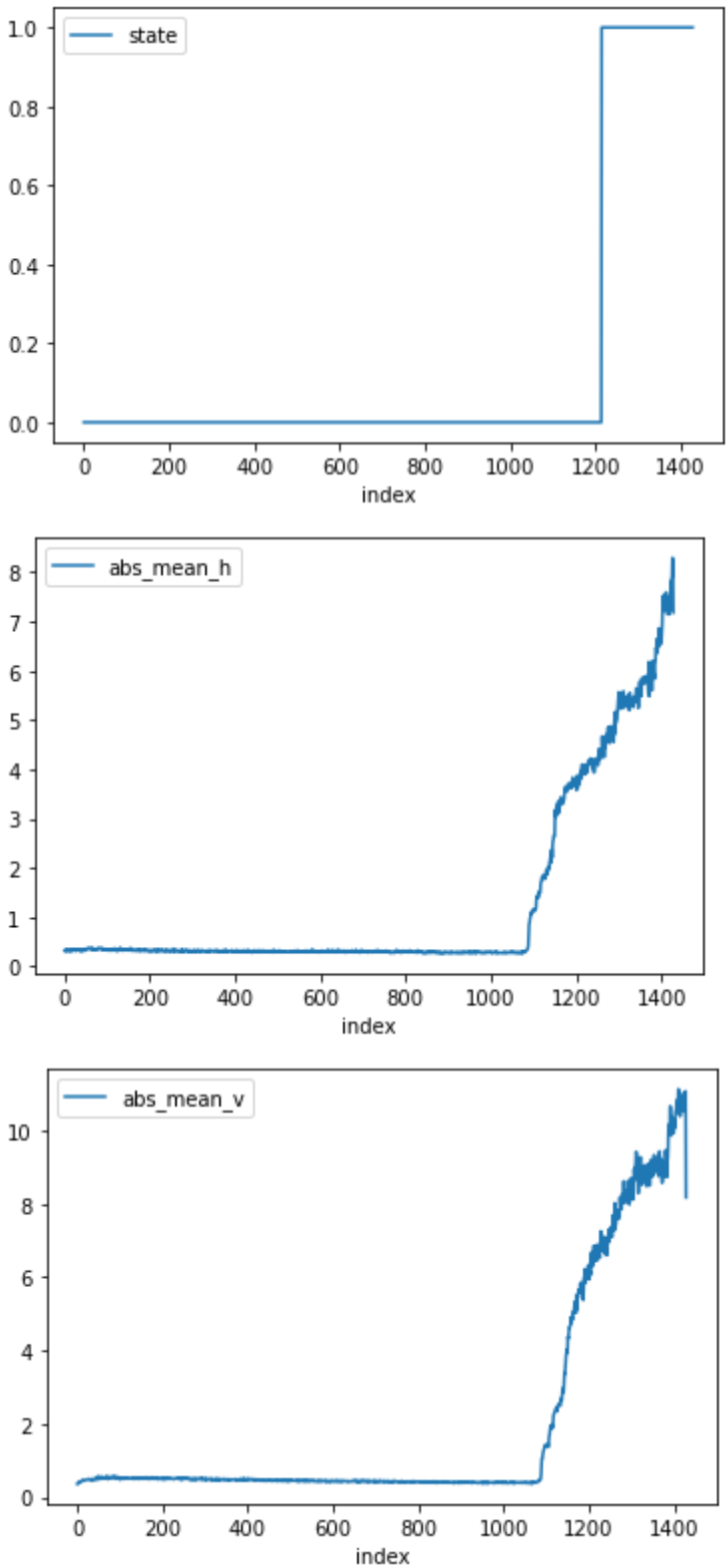
# Plot_3: (time, Mean_h)
B1_4_df.reset_index().plot(x='index', y='mean_h')
plt.show()

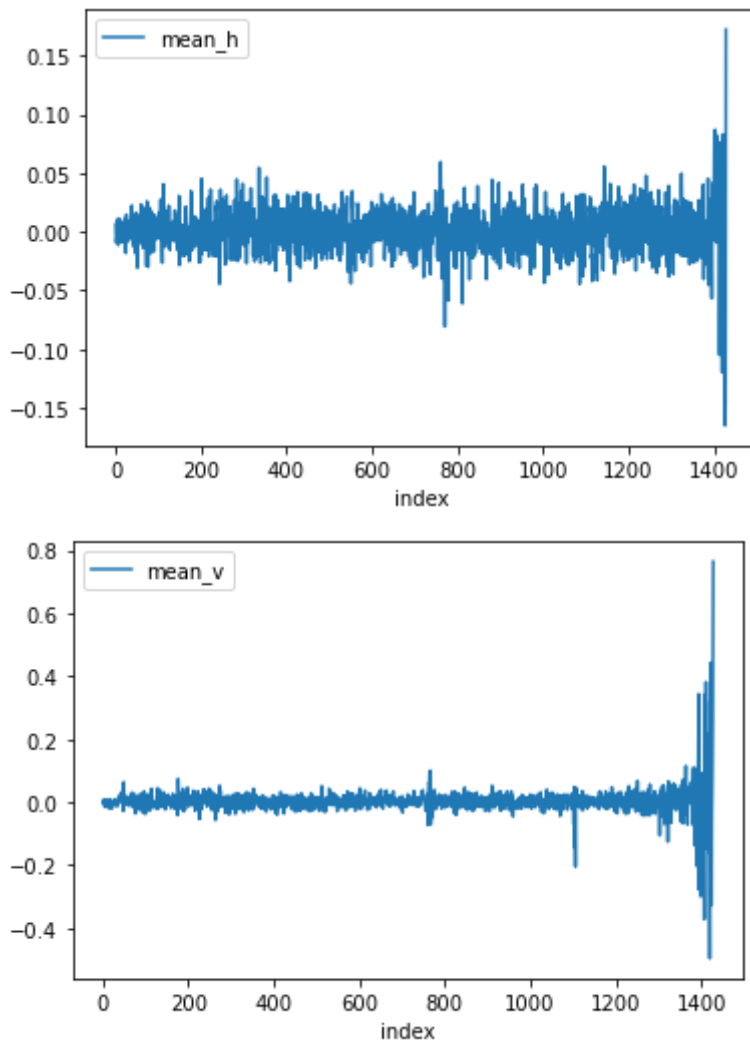
# Plot_4: (time, Mean_v)
B1_4_df.reset_index().plot(x='index', y='mean_v')
plt.show()
```

	state	abs_mean_h	abs_mean_v	mean_h	mean_v	stand_h	stand_v	\
0	0	0.32318	0.35930	0.00636	0.00167	0.40337	0.45502	
1	0	0.31214	0.36394	-0.00900	0.00669	0.39068	0.45886	
2	0	0.31035	0.38840	-0.00622	-0.00830	0.39190	0.49150	
3	0	0.33253	0.38023	-0.00582	-0.00175	0.41586	0.47481	
4	0	0.31086	0.40921	-0.00202	0.00663	0.38677	0.51172	
...	
1423	1	7.86718	11.04806	0.04985	-0.32804	10.54333	14.54749	
1424	1	7.91047	10.61781	-0.11316	-0.03484	10.55686	14.07207	
1425	1	8.29157	10.74766	-0.16524	0.06929	11.10476	14.19700	
1426	1	8.14881	11.08156	-0.14566	0.26825	10.89777	14.59728	
1427	1	7.18050	8.18296	0.17245	0.76412	9.33412	10.48393	

	max_h	min_h	max_v	min_v
0	1.373	-1.511	1.658	-2.045
1	1.299	-1.446	1.537	-1.685
2	1.313	-1.505	2.161	-1.872
3	1.508	-1.476	1.637	-2.033
4	1.334	-1.225	1.967	-1.690
...
1423	48.128	-41.133	47.849	-47.843
1424	48.128	-39.357	47.849	-47.843
1425	48.128	-46.942	47.849	-47.843
1426	48.128	-48.148	47.849	-47.843
1427	48.128	-41.573	47.849	-41.680

[1428 rows x 11 columns]





(Aufgabe 2)Plotting

- Plot_5: (time, stand_h)
- Plot_6: (time, stand_v)
- Plot_7: (time, max_h)
- Plot_8: (time, min_h)
- Plot_9: (time, max_v)
- Plot_10: (time, min_v)

```
In [11]: # here use index as Zeitliche Verlauf
# Plot_5: (time, stand_h)
B1_4_df.reset_index().plot(x='index', y='stand_h')
print(B1_4_df)
plt.show()

# Plot_6: (time, stand_v)
B1_4_df.reset_index().plot(x='index', y='stand_v')
plt.show()

# Plot_7: (time, max_h)
B1_4_df.reset_index().plot(x='index', y='max_h')
plt.show()

# Plot_8: (time, min_h)
B1_4_df.reset_index().plot(x='index', y='min_h')
plt.show()

# Plot_9: (time, max_v)
```



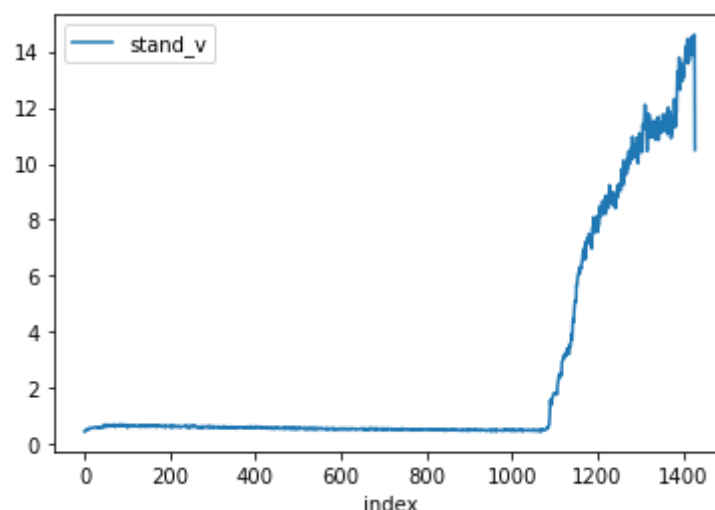
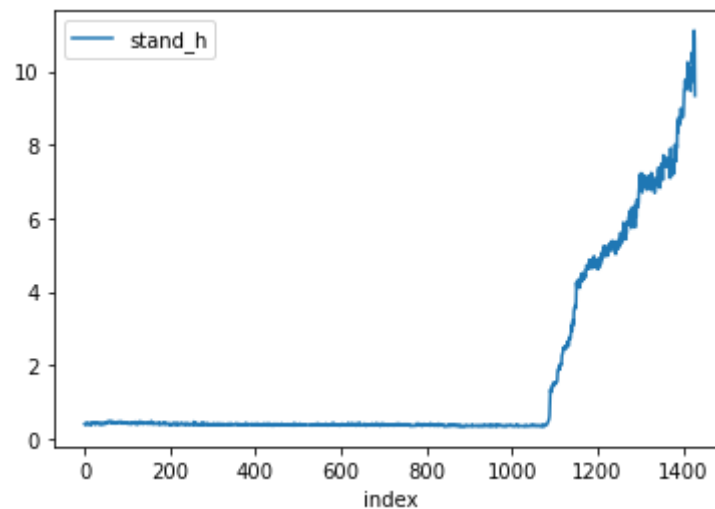
```
B1_4_df.reset_index().plot(x='index', y='max_v')
plt.show()
```

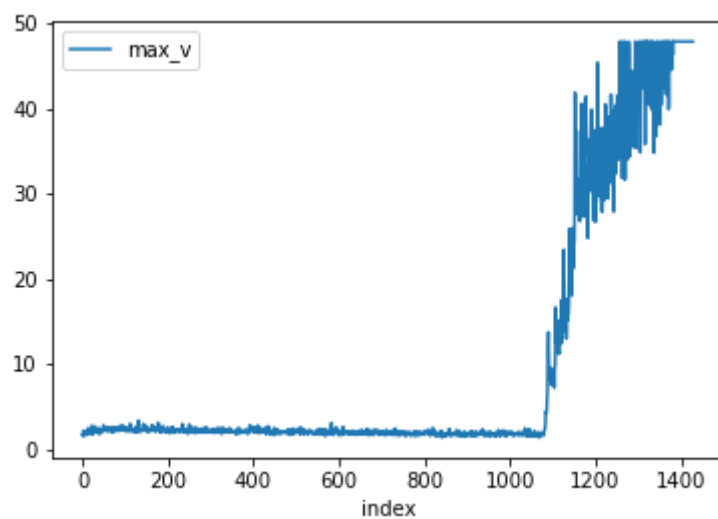
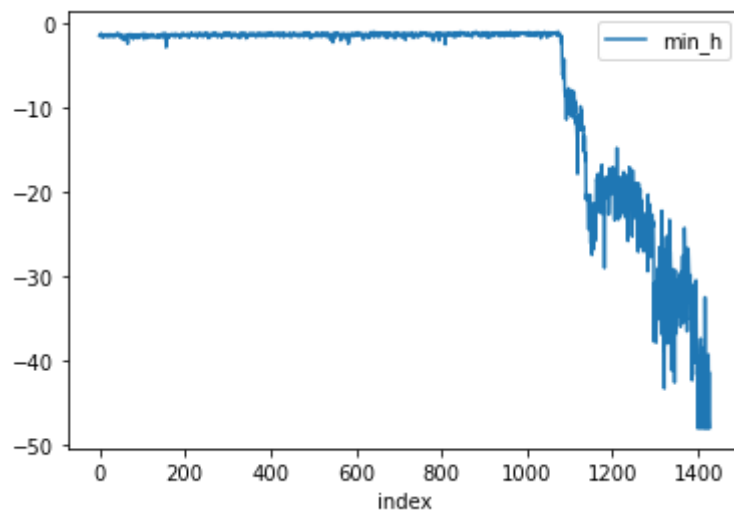
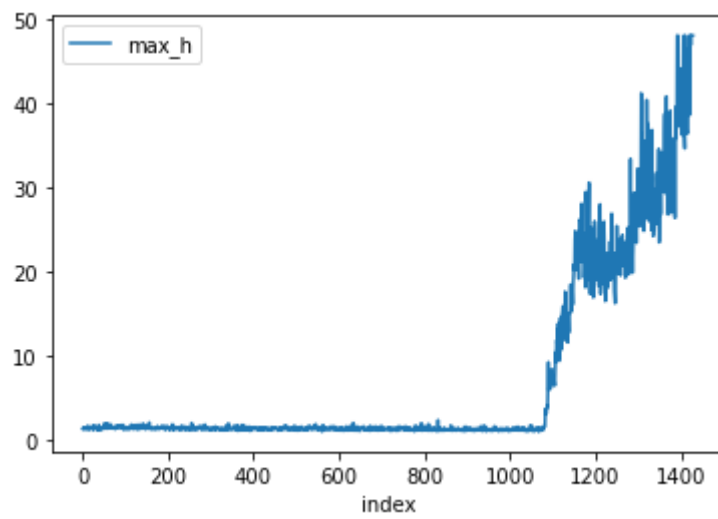
```
# Plot_10: (time, min_v)
B1_4_df.reset_index().plot(x='index', y='min_v')
plt.show()
```

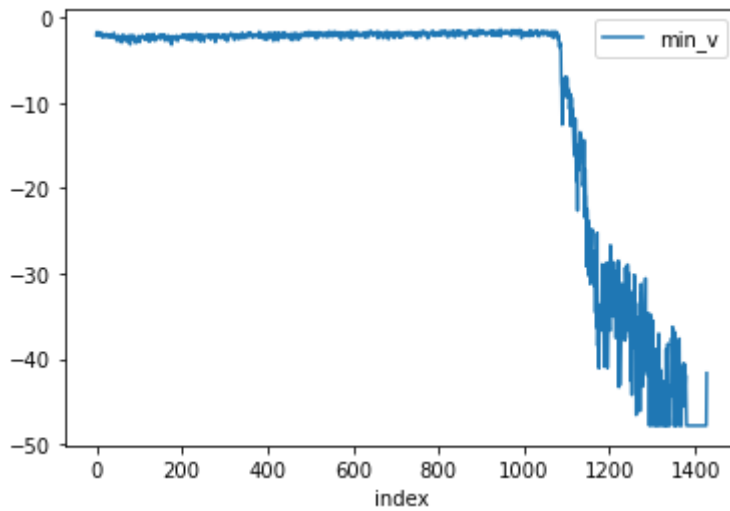
	state	abs_mean_h	abs_mean_v	mean_h	mean_v	stand_h	stand_v	\
0	0	0.32318	0.35930	0.00636	0.00167	0.40337	0.45502	
1	0	0.31214	0.36394	-0.00900	0.00669	0.39068	0.45886	
2	0	0.31035	0.38840	-0.00622	-0.00830	0.39190	0.49150	
3	0	0.33253	0.38023	-0.00582	-0.00175	0.41586	0.47481	
4	0	0.31086	0.40921	-0.00202	0.00663	0.38677	0.51172	
...	
1423	1	7.86718	11.04806	0.04985	-0.32804	10.54333	14.54749	
1424	1	7.91047	10.61781	-0.11316	-0.03484	10.55686	14.07207	
1425	1	8.29157	10.74766	-0.16524	0.06929	11.10476	14.19700	
1426	1	8.14881	11.08156	-0.14566	0.26825	10.89777	14.59728	
1427	1	7.18050	8.18296	0.17245	0.76412	9.33412	10.48393	

	max_h	min_h	max_v	min_v
0	1.373	-1.511	1.658	-2.045
1	1.299	-1.446	1.537	-1.685
2	1.313	-1.505	2.161	-1.872
3	1.508	-1.476	1.637	-2.033
4	1.334	-1.225	1.967	-1.690
...
1423	48.128	-41.133	47.849	-47.843
1424	48.128	-39.357	47.849	-47.843
1425	48.128	-46.942	47.849	-47.843
1426	48.128	-48.148	47.849	-47.843
1427	48.128	-41.573	47.849	-41.680

[1428 rows x 11 columns]







Analysis Summary on the 11 plotting above:

- all the features('state', 'abs_mean_h', 'abs_mean_v', 'mean_h', 'mean_v', 'stand_h', 'stand_v', 'max_h', 'min_h', 'max_v', 'min_v') increase their amplitude, as the time increase.

Step7 (Aufgabe 3) Generate 'State_df'

```
In [12]: #use function
key = ['Bearing1_4', 'Bearing1_5', 'Bearing1_6',
       'Bearing1_7', 'Bearing2_4', 'Bearing2_5',
       'Bearing2_6', 'Bearing2_7', 'Bearing3_3']

#1#
state_list = get_state_filename(key, 'C:/TCW/01_Uni-Stuttgart/Big Data Labs/bdml_pro

#2#concatenate a list of dataframes together#
state_df = pd.concat(state_list)
state_df.reset_index(drop=True, inplace=True)

print('state_df')
print(state_df)
print(state_df.shape)
```

```
state_df
      file state
0  acc_00001.csv  0
1  acc_00002.csv  0
2  acc_00003.csv  0
3  acc_00004.csv  0
4  acc_00005.csv  0
...
13020 acc_00430.csv  2
13021 acc_00431.csv  2
13022 acc_00432.csv  2
13023 acc_00433.csv  2
13024 acc_00434.csv  2
```

```
[13025 rows x 2 columns]
(13025, 2)
```

Step8 (Aufgabe 3 & Aufgabe 4) Reduce the class Number from 3 --> 2

```
In [13]: state_df['state'] = state_df['state'].replace({'0': 0, '1': 0, '2': 1})
state_df['state']
```

```
#calculat frequency
```

```
G,HW = state_df['state'].value_counts()
print(('Good: {}'.format(G),'HeavyWorn: {}'.format(HW)))
print('TotalNumber of csv files: ', G+HW)
```

```
#check value
```

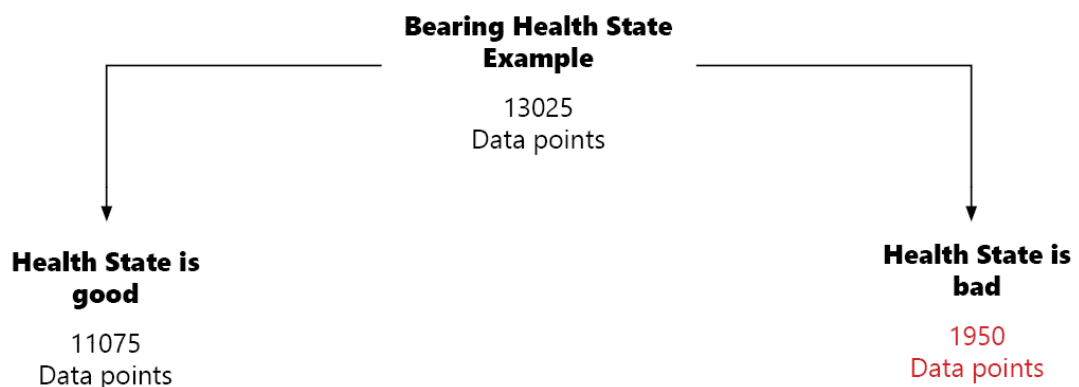
```
state_df_new = state_df
print(state_df_new)
```

```
('Good: 11075', 'HeavyWorn: 1950')
```

```
TotalNumber of csv files: 13025
```

```
file state
0 acc_00001.csv 0
1 acc_00002.csv 0
2 acc_00003.csv 0
3 acc_00004.csv 0
4 acc_00005.csv 0
...
13020 acc_00430.csv 1
13021 acc_00431.csv 1
13022 acc_00432.csv 1
13023 acc_00433.csv 1
13024 acc_00434.csv 1
```

```
[13025 rows x 2 columns]
```



Current analysis of the dataset: The dataset is imbalance(skewness). If we train a model on the dataset, since the number of data points of bearing with good health state is far more than bearing with bad health state, the model will be biased towards the target(health state is good), which has more data points

Step9

Import all the csv files from each folders

```
In [14]: #generate Bearing1_4, 1_5 .....,3_3 df
Bearing1_4_df = calculate_each_csv('C:/TCW/01_Uni-Stuttgart/Big Data Labs/bdml_proje
print('Bearing1_4_df')
print(Bearing1_4_df)

Bearing1_5_df = calculate_each_csv('C:/TCW/01_Uni-Stuttgart/Big Data Labs/bdml_proje
print('Bearing1_5_df')
print(Bearing1_5_df)

Bearing1_6_df = calculate_each_csv('C:/TCW/01_Uni-Stuttgart/Big Data Labs/bdml_proje
print('Bearing1_6_df')
```

```

print(Bearing1_6_df)

Bearing1_7_df = calculate_each_csv('C:/TCW/01_Uni-Stuttgart/Big Data Labs/bdml_proje
print('Bearing1_7_df')
print(Bearing1_7_df)

Bearing2_4_df = calculate_each_csv('C:/TCW/01_Uni-Stuttgart/Big Data Labs/bdml_proje
print('Bearing2_4_df')
print(Bearing2_4_df)

Bearing2_5_df = calculate_each_csv('C:/TCW/01_Uni-Stuttgart/Big Data Labs/bdml_proje
print('Bearing2_5_df')
print(Bearing2_5_df)

Bearing2_6_df = calculate_each_csv('C:/TCW/01_Uni-Stuttgart/Big Data Labs/bdml_proje
print('Bearing2_6_df')
print(Bearing2_6_df)

Bearing2_7_df = calculate_each_csv('C:/TCW/01_Uni-Stuttgart/Big Data Labs/bdml_proje
print('Bearing2_7_df')
print(Bearing2_7_df)

Bearing3_3_df = calculate_each_csv('C:/TCW/01_Uni-Stuttgart/Big Data Labs/bdml_proje
print('Bearing3_3_df')
print(Bearing3_3_df)

```

```

Bearing1_4_df
      abs_mean_h  abs_mean_v  mean_h  mean_v  stand_h  stand_v  max_h  \
0      0.32318    0.35930  0.00636  0.00167  0.40337  0.45502  1.373
1      0.31214    0.36394 -0.00900  0.00669  0.39068  0.45886  1.299
2      0.31035    0.38840 -0.00622 -0.00830  0.39190  0.49150  1.313
3      0.33253    0.38023 -0.00582 -0.00175  0.41586  0.47481  1.508
4      0.31086    0.40921 -0.00202  0.00663  0.38677  0.51172  1.334
...      ...      ...      ...      ...      ...      ...      ...
1423    7.86718   11.04806  0.04985 -0.32804  10.54333  14.54749  48.128
1424    7.91047   10.61781 -0.11316 -0.03484  10.55686  14.07207  48.128
1425    8.29157   10.74766 -0.16524  0.06929  11.10476  14.19700  48.128
1426    8.14881   11.08156 -0.14566  0.26825  10.89777  14.59728  48.128
1427    7.18050    8.18296  0.17245  0.76412  9.33412  10.48393  48.128

```

```

      min_h  max_v  min_v  time_deviation
0    -1.511  1.658  -2.045             99920
1    -1.446  1.537  -1.685             99920
2    -1.505  2.161  -1.872             99920
3    -1.476  1.637  -2.033             99920
4    -1.225  1.967  -1.690             99920
...      ...      ...      ...
1423 -41.133  47.849 -47.843             99920
1424 -39.357  47.849 -47.843             99920
1425 -46.942  47.849 -47.843             99920
1426 -48.148  47.849 -47.843             99920
1427 -41.573  47.849 -41.680             99920

```

[1428 rows x 11 columns]

```

Bearing1_5_df
      abs_mean_h  abs_mean_v  mean_h  mean_v  stand_h  stand_v  max_h  \
0      0.31656    0.28592  0.00337  0.00247  0.40077  0.35730  1.298
1      0.29740    0.28811  0.00306 -0.00153  0.37572  0.35934  1.304
2      0.31841    0.30212  0.00345 -0.00091  0.40191  0.37774  1.269
3      0.32935    0.29942 -0.00147  0.00045  0.41238  0.37716  1.597
4      0.31798    0.31392 -0.00231  0.00296  0.40293  0.39106  1.446
...      ...      ...      ...      ...      ...      ...      ...
2458    0.78944    1.30445  0.03493  0.00482  1.29052  1.73803  12.811
2459    0.84398    1.45504  0.01168 -0.00972  1.36539  1.92841  14.053
2460    0.81835    1.38848 -0.00267  0.01294  1.33530  1.84301  12.554
2461    0.85351    1.46709  0.00895  0.01178  1.33762  1.93461  11.680
2462    0.87393    1.42787 -0.00619 -0.00733  1.39689  1.87227  12.408

```

	min_h	max_v	min_v	time_deviation
0	-1.453	1.199	-1.267	99920
1	-1.497	1.200	-1.761	99920
2	-1.526	1.134	-1.548	99920
3	-1.476	1.362	-1.428	99920
4	-1.519	1.503	-1.280	99920
...
2458	-11.738	10.180	-9.822	99920
2459	-10.802	12.965	-11.127	99920
2460	-9.542	12.036	-9.387	99920
2461	-9.651	9.312	-9.242	99920
2462	-10.270	8.475	-10.039	99920

[2463 rows x 11 columns]

Bearing1_6_df

	abs_mean_h	abs_mean_v	mean_h	mean_v	stand_h	stand_v	max_h	\
0	0.37257	0.38361	-0.06606	-0.01485	0.47298	0.47783	2.624	
1	0.31632	0.36760	0.01297	-0.00680	0.40038	0.46393	1.427	
2	0.35554	0.38350	0.01390	0.00123	0.46514	0.47830	2.608	
3	0.34022	0.39356	0.00497	0.02892	0.43168	0.49310	2.019	
4	0.33125	0.37575	-0.00039	0.00013	0.42314	0.47254	1.429	
...	
2443	0.87490	1.38792	0.00620	0.02769	1.21665	1.79190	9.150	
2444	0.89719	1.44833	-0.00266	0.03269	1.24330	1.86343	8.220	
2445	0.87763	1.47405	0.00005	0.02033	1.22909	1.88969	6.779	
2446	0.90034	1.49086	-0.01635	-0.02096	1.25398	1.93002	8.926	
2447	0.92454	1.51083	0.00252	-0.00651	1.29322	1.95317	9.142	

	min_h	max_v	min_v	time_deviation
0	-1.999	1.496	-1.670	99920
1	-1.446	1.538	-1.798	99920
2	-2.430	2.065	-1.585	99920
3	-1.677	1.900	-1.587	99920
4	-1.835	1.724	-1.651	99920
...
2443	-9.473	7.928	-8.752	99920
2444	-7.802	9.422	-7.495	99920
2445	-7.761	9.861	-9.143	99920
2446	-6.655	6.896	-7.758	99920
2447	-6.718	7.531	-9.531	99920

[2448 rows x 11 columns]

Bearing1_7_df

	abs_mean_h	abs_mean_v	mean_h	mean_v	stand_h	stand_v	max_h	\
0	0.34329	0.31273	0.00293	0.00237	0.43816	0.39171	1.810	
1	0.33470	0.32509	0.00376	0.00263	0.42173	0.40342	1.360	
2	0.35733	0.33705	0.00337	0.00143	0.45234	0.41746	1.720	
3	0.34796	0.33305	0.00390	0.00331	0.44031	0.41773	1.565	
4	0.35848	0.34185	-0.00090	0.00256	0.45139	0.42716	1.707	
...	
2254	1.34397	1.45088	-0.00368	-0.04778	1.92288	1.89814	11.379	
2255	1.40360	1.47648	-0.01567	-0.05049	2.01051	1.99316	12.165	
2256	1.52379	1.62926	0.00516	0.02603	2.24517	2.17954	16.516	
2257	1.56360	1.67294	0.01408	0.00494	2.39805	2.29778	19.515	
2258	1.59041	1.71820	0.02106	0.00458	2.36344	2.39131	21.336	

	min_h	max_v	min_v	time_deviation
0	-1.749	1.497	-1.290	999961
1	-1.660	1.371	-1.403	999961
2	-1.467	1.394	-1.474	999961
3	-1.637	1.544	-1.352	999961
4	-1.587	1.740	-1.457	999961
...
2254	-12.913	8.733	-8.005	999961
2255	-13.152	12.401	-9.853	999961
2256	-15.311	10.585	-13.120	999961
2257	-16.872	17.698	-10.561	999961
2258	-16.490	21.531	-12.010	999961

[2259 rows x 11 columns]

Bearing2_4_df

	abs_mean_h	abs_mean_v	mean_h	mean_v	stand_h	stand_v	max_h	min_h	\
0	0.27088	0.19117	0.00659	0.00232	0.34133	0.23891	1.130	-1.142	
1	0.25441	0.19617	0.00147	0.00225	0.31824	0.24588	1.080	-1.098	
2	0.25801	0.20131	0.00314	0.00414	0.32686	0.25265	1.247	-1.086	
3	0.28052	0.20080	0.00313	-0.00293	0.34730	0.25432	1.147	-1.032	
4	0.27698	0.20619	0.00426	0.00260	0.35136	0.26060	1.288	-1.572	
..	
746	0.82048	0.74865	0.00660	-0.00958	1.08543	1.00420	4.100	-4.979	
747	0.89540	0.81136	-0.01702	0.01387	1.17412	1.06485	4.297	-4.621	
748	0.89061	0.89055	0.02332	-0.01391	1.19781	1.17173	4.999	-5.890	
749	1.12198	1.12363	-0.00320	0.00656	1.50681	1.53442	4.890	-7.899	
750	1.17333	1.18761	0.02301	0.00977	1.56361	1.59059	6.072	-8.510	
max_v	min_v	time_deviation							
0	0.775	-0.781	99920						
1	0.832	-0.784	99920						
2	0.949	-0.824	99920						
3	0.887	-1.054	99920						
4	1.029	-1.415	99920						
..						
746	6.962	-3.994	99920						
747	5.219	-4.402	99920						
748	6.456	-5.084	99920						
749	8.818	-6.483	99920						
750	8.523	-6.190	99920						

[751 rows x 11 columns]

Bearing2_5_df

	abs_mean_h	abs_mean_v	mean_h	mean_v	stand_h	stand_v	max_h	\
0	0.40556	0.17984	0.03310	0.00685	0.52770	0.22588	2.585	
1	0.51514	0.18593	-0.07174	0.00546	0.67926	0.23376	3.150	
2	0.47952	0.18504	0.02326	0.00641	0.61162	0.23121	3.631	
3	1.31519	0.18643	0.15630	-0.00824	1.49801	0.23518	4.940	
4	0.75188	0.19091	-0.01070	0.00356	0.87201	0.24063	2.273	
...	
2306	0.25763	0.38221	-0.00420	0.01240	0.49188	1.09998	6.151	
2307	0.27392	0.41429	-0.02538	-0.01595	0.53793	1.18650	4.569	
2308	0.27781	0.39508	0.01833	0.01070	0.48899	1.03347	3.422	
2309	0.32881	0.48996	0.00408	0.00602	0.68726	1.48907	8.163	
2310	0.33551	0.51870	-0.00161	-0.00247	0.73462	1.66807	9.135	
min_h	max_v	min_v	time_deviation					
0	-1.516	0.846	-0.743	99920				
1	-1.861	0.763	-0.816	99920				
2	-1.628	0.799	-0.805	99920				
3	-3.208	0.869	-0.907	99920				
4	-2.068	1.008	-0.714	99920				
...				
2306	-7.401	16.095	-26.881	99920				
2307	-9.218	14.997	-22.875	99920				
2308	-5.759	11.250	-18.679	99920				
2309	-8.879	21.732	-18.025	99920				
2310	-10.854	22.562	-19.973	99920				

[2311 rows x 11 columns]

Bearing2_6_df

	abs_mean_h	abs_mean_v	mean_h	mean_v	stand_h	stand_v	max_h	min_h	\
0	0.27303	0.26293	-0.00311	-0.00041	0.34436	0.33272	1.009	-1.248	
1	0.26749	0.27778	0.12070	0.06403	0.31395	0.34069	1.100	-0.935	
2	0.24945	0.27290	0.00198	-0.01052	0.31355	0.34022	1.153	-1.289	
3	0.24972	0.27219	0.00347	0.00293	0.31452	0.34328	1.106	-1.105	
4	0.24085	0.27095	-0.00548	0.00415	0.30021	0.33738	1.147	-0.912	
..	
696	0.85694	1.41641	0.04461	-0.00007	1.17250	1.88860	5.647	-5.538	
697	1.03927	1.75602	-0.00456	0.00442	1.38989	2.28945	5.436	-6.358	
698	1.03024	1.78191	-0.03349	0.00768	1.37195	2.31026	5.141	-6.090	
699	1.06220	1.74686	0.00827	0.00126	1.42372	2.30703	5.560	-6.053	

700	1.12649	1.74578	0.01531	-0.00417	1.49550	2.23694	6.465	-6.656
-----	---------	---------	---------	----------	---------	---------	-------	--------

	max_v	min_v	time_deviation
0	1.198	-1.184	99920
1	1.107	-1.147	99920
2	1.725	-1.231	99920
3	1.224	-1.240	99920
4	1.092	-1.196	99920
..
696	7.808	-8.333	99920
697	10.126	-10.669	99920
698	11.456	-8.147	99920
699	11.116	-9.630	99920
700	9.848	-10.082	99920

[701 rows x 11 columns]

Bearing2_7_df

	abs_mean_h	abs_mean_v	mean_h	mean_v	stand_h	stand_v	max_h	\
0	0.33031	0.28376	0.00186	0.00517	0.41297	0.35981	1.374	
1	0.31915	0.28325	-0.00699	0.00482	0.40751	0.35601	1.430	
2	0.32236	0.29040	-0.00427	0.00158	0.40778	0.37068	1.449	
3	0.31078	0.30227	-0.01309	-0.00556	0.39480	0.38382	1.561	
4	0.32549	0.29751	-0.01214	-0.00158	0.41139	0.37797	1.316	
..	
225	4.78263	1.61105	-0.00918	0.00801	6.28481	2.58437	19.870	
226	4.63704	1.26401	-0.04638	-0.00077	6.11701	2.07538	20.306	
227	4.36090	1.23944	0.05411	-0.00833	5.76569	2.02487	18.135	
228	3.90999	1.16633	0.05733	0.00468	4.90667	1.98418	16.283	
229	4.15159	1.28780	0.02473	-0.00505	5.40245	2.00472	16.804	

	min_h	max_v	min_v	time_deviation
0	-1.354	1.563	-1.316	99920
1	-1.852	1.244	-1.166	99920
2	-1.528	1.265	-1.426	99920
3	-2.283	1.293	-1.512	99920
4	-2.130	1.381	-1.344	99920
..
225	-18.254	38.889	-22.193	99920
226	-18.371	15.370	-14.625	99920
227	-17.548	16.997	-16.404	99920
228	-16.769	18.024	-23.281	99920
229	-16.751	17.547	-15.438	99920

[230 rows x 11 columns]

Bearing3_3_df

	abs_mean_h	abs_mean_v	mean_h	mean_v	stand_h	stand_v	max_h	min_h	\
0	0.22836	0.23635	0.00221	-0.00628	0.28805	0.29704	1.277	-1.131	
1	0.24678	0.24859	-0.01742	0.00210	0.31110	0.31197	0.953	-1.261	
2	0.23765	0.23651	-0.02088	-0.00637	0.29370	0.29739	0.986	-1.055	
3	0.24305	0.23576	0.05725	0.00103	0.29863	0.29555	1.047	-0.879	
4	0.25260	0.25107	-0.00430	0.00638	0.31708	0.31662	1.113	-1.039	
..	
429	0.90140	1.32143	-0.03920	0.00316	1.21123	1.79916	4.004	-6.815	
430	0.88235	1.33355	-0.02996	-0.00206	1.17919	1.81695	3.762	-6.387	
431	0.90315	1.40123	-0.05648	0.05088	1.20484	1.88076	4.365	-6.003	
432	0.95303	1.50619	0.01310	0.02866	1.25497	2.05971	4.022	-6.062	
433	0.99460	1.65159	0.04925	0.00473	1.29547	2.22718	4.578	-5.862	

	max_v	min_v	time_deviation
0	1.160	-0.900	99920
1	1.571	-1.284	99920
2	1.131	-1.134	99920
3	1.120	-1.154	99920
4	1.098	-1.253	99920
..
429	12.851	-6.985	99920
430	12.570	-8.043	99920
431	11.900	-7.761	99920
432	14.616	-8.082	99920

433 15.982 -9.992

99920

[434 rows x 11 columns]

Step10 (Aufgabe3) Combine all the Bearing df into one df: stack_bearing_df

```
In [15]: stack_bearing_df = pd.concat([Bearing1_4_df, Bearing1_5_df, Bearing1_6_df,
                                     Bearing1_7_df, Bearing2_4_df, Bearing2_5_df,
                                     Bearing2_6_df, Bearing2_7_df, Bearing3_3_df], ignore_i

stack_bearing_df.reset_index(drop=True, inplace=True)
print(stack_bearing_df)
print(stack_bearing_df.shape)
```

	abs_mean_h	abs_mean_v	mean_h	mean_v	stand_h	stand_v	max_h \
0	0.32318	0.35930	0.00636	0.00167	0.40337	0.45502	1.373
1	0.31214	0.36394	-0.00900	0.00669	0.39068	0.45886	1.299
2	0.31035	0.38840	-0.00622	-0.00830	0.39190	0.49150	1.313
3	0.33253	0.38023	-0.00582	-0.00175	0.41586	0.47481	1.508
4	0.31086	0.40921	-0.00202	0.00663	0.38677	0.51172	1.334
...
13020	0.90140	1.32143	-0.03920	0.00316	1.21123	1.79916	4.004
13021	0.88235	1.33355	-0.02996	-0.00206	1.17919	1.81695	3.762
13022	0.90315	1.40123	-0.05648	0.05088	1.20484	1.88076	4.365
13023	0.95303	1.50619	0.01310	0.02866	1.25497	2.05971	4.022
13024	0.99460	1.65159	0.04925	0.00473	1.29547	2.22718	4.578

	min_h	max_v	min_v	time_deviation
0	-1.511	1.658	-2.045	99920
1	-1.446	1.537	-1.685	99920
2	-1.505	2.161	-1.872	99920
3	-1.476	1.637	-2.033	99920
4	-1.225	1.967	-1.690	99920
...
13020	-6.815	12.851	-6.985	99920
13021	-6.387	12.570	-8.043	99920
13022	-6.003	11.900	-7.761	99920
13023	-6.062	14.616	-8.082	99920
13024	-5.862	15.982	-9.992	99920

[13025 rows x 11 columns]
(13025, 11)

Step11 (Aufgabe 3) Merge 'stack_bearing_df' with 'state_df_new' horizontally into 'training_df'

```
In [16]: training_df = pd.concat([state_df_new, stack_bearing_df], axis=1)
print(training_df)
print(training_df.shape)
print(training_df.columns.values)
training_df['state'].value_counts()
```

	file	state	abs_mean_h	abs_mean_v	mean_h	mean_v	\
0	acc_00001.csv	0	0.32318	0.35930	0.00636	0.00167	
1	acc_00002.csv	0	0.31214	0.36394	-0.00900	0.00669	
2	acc_00003.csv	0	0.31035	0.38840	-0.00622	-0.00830	
3	acc_00004.csv	0	0.33253	0.38023	-0.00582	-0.00175	
4	acc_00005.csv	0	0.31086	0.40921	-0.00202	0.00663	
...	
13020	acc_00430.csv	1	0.90140	1.32143	-0.03920	0.00316	
13021	acc_00431.csv	1	0.88235	1.33355	-0.02996	-0.00206	

```

13022 acc_00432.csv      1      0.90315      1.40123 -0.05648  0.05088
13023 acc_00433.csv      1      0.95303      1.50619  0.01310  0.02866
13024 acc_00434.csv      1      0.99460      1.65159  0.04925  0.00473

```

```

      stand_h  stand_v  max_h  min_h  max_v  min_v  time_deviation
0      0.40337  0.45502  1.373 -1.511  1.658 -2.045          99920
1      0.39068  0.45886  1.299 -1.446  1.537 -1.685          99920
2      0.39190  0.49150  1.313 -1.505  2.161 -1.872          99920
3      0.41586  0.47481  1.508 -1.476  1.637 -2.033          99920
4      0.38677  0.51172  1.334 -1.225  1.967 -1.690          99920
...      ...      ...      ...      ...      ...      ...      ...
13020  1.21123  1.79916  4.004 -6.815  12.851 -6.985          99920
13021  1.17919  1.81695  3.762 -6.387  12.570 -8.043          99920
13022  1.20484  1.88076  4.365 -6.003  11.900 -7.761          99920
13023  1.25497  2.05971  4.022 -6.062  14.616 -8.082          99920
13024  1.29547  2.22718  4.578 -5.862  15.982 -9.992          99920

```

```

[13025 rows x 13 columns]
(13025, 13)
['file' 'state' 'abs_mean_h' 'abs_mean_v' 'mean_h' 'mean_v' 'stand_h'
 'stand_v' 'max_h' 'min_h' 'max_v' 'min_v' 'time_deviation']

```

```

Out[16]: 0      11075
         1      1950
         Name: state, dtype: int64

```

Step12 (Aufgabe 3) Normalize dataframe

```

In [17]: from sklearn import preprocessing
# all columns head
# 'abs_mean_h', 'abs_mean_v', 'mean_h', 'mean_v', 'stand_h', 'stand_v', 'max_h', 'min

#drop column with string content 'file' & 'time_deviation'
df_features = training_df.drop(columns=['file', 'time_deviation'], axis=1)
df_label = training_df['state']

x = df_features.values #returns a numpy array
min_max_scaler = preprocessing.MinMaxScaler()
x_scaled = min_max_scaler.fit_transform(x)
df_features_n = pd.DataFrame(x_scaled)
df_features_n.columns = ['state', 'abs_mean_h', 'abs_mean_v', 'mean_h', 'mean_v', 's

#normalized df
df_features_n

```

```

Out[17]:
      state  abs_mean_h  abs_mean_v  mean_h  mean_v  stand_h  stand_v  max_h  min_h
0      0.0    0.020570    0.016372  0.651261  0.395370  0.019164  0.015944  0.017174  0.980737
1      0.0    0.019213    0.016796  0.634832  0.399351  0.018001  0.016211  0.015618  0.982104
2      0.0    0.018993    0.019027  0.637806  0.387464  0.018113  0.018483  0.015913  0.980863
3      0.0    0.021719    0.018282  0.638233  0.392658  0.020309  0.017321  0.020012  0.981473
4      0.0    0.019055    0.020926  0.642298  0.399304  0.017643  0.019890  0.016354  0.986752
...      ...      ...      ...      ...      ...      ...      ...      ...      ...
13020  1.0    0.091641    0.104148  0.602531  0.396552  0.093209  0.109473  0.072480  0.869199
13021  1.0    0.089300    0.105254  0.612414  0.392412  0.090272  0.110711  0.067393  0.878199
13022  1.0    0.091856    0.111428  0.584048  0.434394  0.092623  0.115151  0.080068  0.886274

```

	state	abs_mean_h	abs_mean_v	mean_h	mean_v	stand_h	stand_v	max_h	min_h
13023	1.0	0.097987	0.121004	0.658470	0.416774	0.097218	0.127603	0.072858	0.885034
13024	1.0	0.103097	0.134269	0.697136	0.397797	0.100930	0.139256	0.084546	0.889239

13025 rows × 11 columns

In []:

Step13 (Aufgabe 3) Prepare multiple datasets with different features

```
In [18]: # all columns head
# 'abs_mean_h', 'abs_mean_v', 'mean_h', 'mean_v', 'stand_h', 'stand_v', 'max_h', 'min_h'

#####
# based on the Correlation Matrix --> features['abs_mean_v', 'stand_v', 'max_h', 'min_h']
df_features_0 = training_df.drop(columns=['file', 'state', 'abs_mean_h', 'abs_mean_v'])
df_features_1 = training_df.drop(columns=['file', 'state', 'mean_h', 'mean_v', 'max_h', 'min_h'])
df_features_2 = training_df.drop(columns=['file', 'state', 'abs_mean_h', 'abs_mean_v'])
df_label      = training_df['state']
#####

print('df_features_0= ', df_features_0.columns)
print('df_features_1= ', df_features_1.columns)
print('df_features_2= ', df_features_2.columns)
print('df_features_shape = ', df_features_1.shape)
print('df_label_shape   = ', df_label.shape)

df_features_0= Index(['mean_h', 'mean_v', 'stand_h', 'stand_v'], dtype='object')
df_features_1= Index(['abs_mean_h', 'abs_mean_v', 'stand_h', 'stand_v'], dtype='object')
df_features_2= Index(['mean_h', 'mean_v', 'stand_h', 'stand_v', 'max_h', 'min_h', 'max_v', 'min_v'], dtype='object')
df_features_shape = (13025, 4)
df_label_shape   = (13025,)
```

Definition:

- Precision: Positive Predictive Value --> of all Bearings classified as having worn heavily, how many of them actually had worn heavily (Interpretation)
 - $\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$
- Recall --> what percentage of actual bad health state predictions were correctly classified by classifier
 - $\text{Recall} = \frac{\text{TP}}{\text{FN} + \text{TP}}$

	Predicted 0	Predicted 1
Actual 0	TN	FP
Actual 1	FN	TP

Step14 (Aufgabe 5) Training with various Algorithms:

- 1_SVM_0 (with features: ['mean_h', 'mean_v', 'stand_h', 'stand_v'])
- 1_SVM_1 (with features: ['abs_mean_h', 'abs_mean_v', 'stand_h', 'stand_v'])
- 1_SVM_2 (with features: ['mean_h', 'mean_v', 'stand_h', 'stand_v', 'max_h', 'min_h', 'max_v', 'min_v'])
- 2_RandomForest_0 (with features: ['mean_h', 'mean_v', 'stand_h', 'stand_v'])
- 2_RandomForest_1 (with features: ['abs_mean_h', 'abs_mean_v', 'stand_h', 'stand_v'])
- 3_GradientBoosting_0 (with features: ['mean_h', 'mean_v', 'stand_h', 'stand_v'])
- 3_GradientBoosting_1 (with features: ['abs_mean_h', 'abs_mean_v', 'stand_h', 'stand_v'])
- 4_K-NearestNeighbors_0 (with features: ['mean_h', 'mean_v', 'stand_h', 'stand_v'])
- 4_K-NearestNeighbors_1 (with features: ['abs_mean_h', 'abs_mean_v', 'stand_h', 'stand_v'])
- 5_CNN_0 (with features: [mean_h, mean_v, stand_h, stand_v]) (2 hidden layers, epochs=50)
- 5_CNN_1 (with features: [mean_h, mean_v, stand_h, stand_v]) (4 hidden layers, epochs=50)
- 5_CNN_2 (with features: [abs_mean_h, abs_mean_v, stand_h, stand_v]) (2 hidden layers, epochs=50)
- 5_CNN_3 (with features: [abs_mean_h, abs_mean_v, stand_h, stand_v]) (4 hidden layers, epochs=100)
- 5_CNN_4 (with features: [abs_mean_h, abs_mean_v, stand_h, stand_v]) (8 hidden layers, epochs=100)
- 5_CNN_5 (with features: ['mean_h', 'mean_v', 'stand_h', 'stand_v', 'max_h', 'min_h', 'max_v', 'min_v']) (2 hidden layers, epochs=50)
- 5_CNN_6 (with features: ['mean_h', 'mean_v', 'stand_h', 'stand_v', 'max_h', 'min_h', 'max_v', 'min_v']) (8 hidden layers, epochs=100)

- 5_CNN_7 (with features: ['abs_mean_v', 'stand_v', 'max_h', 'min_h', 'max_v', 'min_v']) (8 hidden layers, epochs=100)

```
In [19]: #create list for storing the result of each classifier
training_title_list      = []
train_accuracy_f1_result_list = []
test_accuracy_f1_result_list = []
train_test_difference_result_list = []
precision_result_list    = []
recall_result_list       = []
```

1_SVM_0 (with features: ['mean_h', 'mean_v', 'stand_h', 'stand_v'])

```
In [20]: from sklearn.model_selection import train_test_split
from sklearn.metrics import confusion_matrix
from sklearn.metrics import plot_confusion_matrix
from sklearn.svm import SVC
from sklearn.metrics import recall_score, precision_score
from sklearn.metrics import f1_score
from sklearn.metrics import classification_report

#1# split into 80:20
X=df_features_0
y=df_label

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_stat

clf = SVC(C=10)
clf.fit(X_train, y_train)

#train result
train_accuracy=clf.score(X_train, y_train).round(3)*100
#test result
acc_SVM      = clf.score(X_test, y_test).round(3)*100
prec_SVM     = precision_score(y_test, clf.predict(X_test)).round(3)*100
recall_SVM   = recall_score(y_test, clf.predict(X_test)).round(3)*100
Train_Test_Difference = abs(acc_SVM-train_accuracy).round(3)

print('Train_Accuracy_SVM_0 = ', train_accuracy)
print('Test_Accuracy_SVM_0 = ', acc_SVM)
print('Train-Test Difference = ', Train_Test_Difference)
print()
print('Precision_SVM = ', prec_SVM)
print('Recall_SVM = ', recall_SVM)
print()

#confusion matrix(y_true, y_pred)
tn, fp, fn, tp = confusion_matrix(y_test, clf.predict(X_test)).ravel()
print(('TN:{}'.format(tn), 'FP:{}'.format(fp), 'FN:{}'.format(fn), 'TP:{}'.format(tp)

plot_confusion_matrix(clf, X_test, y_test)
plt.show()
```

```
#append result to 'precision_result_list' & 'recall_result_list'
training_title_list.append('1_SVM_0')
train_accuracy_f1_result_list.append(train_accuracy)
test_accuracy_f1_result_list.append(acc_SVM)
train_test_difference_result_list.append(Train_Test_Difference)
precision_result_list.append(prec_SVM)
recall_result_list.append(recall_SVM)

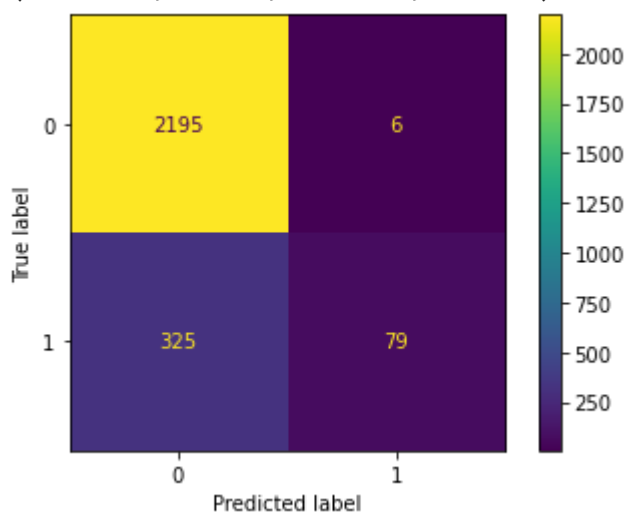
#Classification report
print()
target_names = ['state: good', 'state: bad']
print(classification_report(y_test, clf.predict(X_test), target_names=target_names))

#f1_score
f1_score=f1_score(y_test, clf.predict(X_test), average='micro')
print('f1_score= {}'.format(f1_score))
```

Train_Accuracy_SVM_0 = 87.8
 Test_Accuracy_SVM_0 = 87.3
 Train-Test Difference = 0.5

Precision_SVM = 92.9
 Recall_SVM = 19.6

('TN:2195', 'FP:6', 'FN:325', 'TP:79')



	precision	recall	f1-score	support
state: good	0.87	1.00	0.93	2201
state: bad	0.93	0.20	0.32	404
accuracy			0.87	2605
macro avg	0.90	0.60	0.63	2605
weighted avg	0.88	0.87	0.84	2605

f1_score= 0.872936660268714

```
In [21]: print(training_title_list)
print(train_accuracy_f1_result_list)
print(test_accuracy_f1_result_list)
print(train_test_difference_result_list)
print(precision_result_list)
print(recall_result_list)
```

```
['1_SVM_0']
[87.8]
[87.3]
```

```
[0.5]
[92.9]
[19.6]
```

1_SVM_1 (with features: ['abs_mean_h', 'abs_mean_v', 'stand_h', 'stand_v'])

```
In [22]: from sklearn.model_selection import train_test_split
from sklearn.metrics import confusion_matrix
from sklearn.metrics import plot_confusion_matrix
from sklearn.svm import SVC
from sklearn.metrics import recall_score, precision_score
from sklearn.metrics import f1_score
from sklearn.metrics import classification_report

#1# split into 80:20
X=df_features_1
y=df_label

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_stat

clf_1 = SVC(C = 10)
clf_1.fit(X_train, y_train)

#train result
train_accuracy=clf_1.score(X_train, y_train).round(3)*100
#test result
acc_SVM_1 = clf_1.score(X_test, y_test).round(3)*100
prec_SVM_1 = precision_score(y_test, clf_1.predict(X_test)).round(3)*100
recall_SVM_1 = recall_score(y_test, clf_1.predict(X_test)).round(3)*100
Train_Test_Difference = abs(acc_SVM_1-train_accuracy).round(3)

print('Train_Accuracy_SVM_1 =', train_accuracy.round(3))
print('Test_Accuracy_SVM_1 = ', acc_SVM_1.round(3))
print('Train-Test Difference = ', Train_Test_Difference)
print()
print('Precision_SVM = ', prec_SVM_1.round(3))
print('Recall_SVM = ', recall_SVM_1.round(3))
print()

#confusion matrix(y_true, y_pred)
tn_1, fp_1, fn_1, tp_1 = confusion_matrix(y_test, clf_1.predict(X_test)).ravel()
print(('TN:{}').format(tn_1), 'FP:{}'.format(fp_1), 'FN:{}'.format(fn_1), 'TP:{}'.for

plot_confusion_matrix(clf_1, X_test, y_test)
plt.show()

#append result to 'precision_result_list' & 'recall_result_list'
training_title_list.append('1_SVM_1')
train_accuracy_f1_result_list.append(train_accuracy)
test_accuracy_f1_result_list.append(acc_SVM_1)
train_test_difference_result_list.append(Train_Test_Difference)
precision_result_list.append(prec_SVM_1)
recall_result_list.append(recall_SVM_1)

#Classification report
print()
```

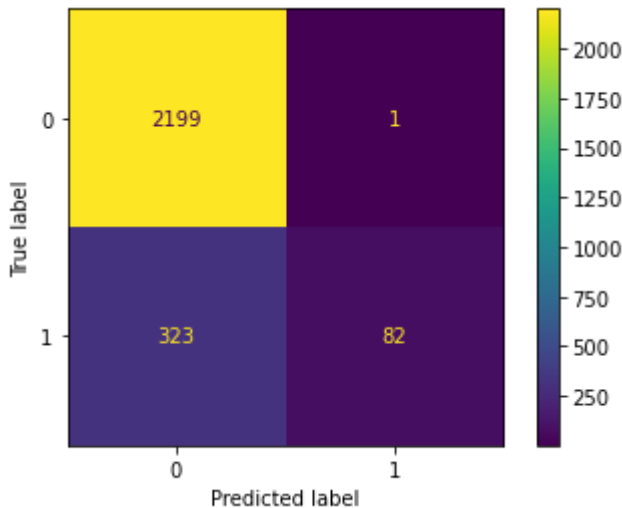
```
target_names = ['state: good', 'state: bad']
print(classification_report(y_test, clf_1.predict(X_test), target_names=target_names))
```

```
#f1_score
f1_score=f1_score(y_test, clf_1.predict(X_test), average='micro')
print('f1_score= {}'.format(f1_score))
```

Train_Accuracy_SVM_1 = 87.8
 Test_Accuracy_SVM_1 = 87.6
 Train-Test Difference = 0.2

Precision_SVM = 98.8
 Recall_SVM = 20.2

('TN:2199', 'FP:1', 'FN:323', 'TP:82')



	precision	recall	f1-score	support
state: good	0.87	1.00	0.93	2200
state: bad	0.99	0.20	0.34	405
accuracy			0.88	2605
macro avg	0.93	0.60	0.63	2605
weighted avg	0.89	0.88	0.84	2605

f1_score= 0.8756238003838771

1_SVM_2 (with features: ['mean_h', 'mean_v', 'stand_h', 'stand_v', 'max_h', 'min_h', 'max_v', 'min_v'])

```
In [23]: from sklearn.model_selection import train_test_split
from sklearn.metrics import confusion_matrix
from sklearn.metrics import plot_confusion_matrix
from sklearn.svm import SVC
from sklearn.metrics import recall_score, precision_score
from sklearn.metrics import f1_score
from sklearn.metrics import classification_report

#1# split into 80:20
X=df_features_2
y=df_label

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_stat

clf_2 = SVC(C = 10000)
```



```

clf_2.fit(X_train, y_train)

#train result
train_accuracy=clf_2.score(X_train, y_train).round(3)*100
#test result
acc_SVM_2 = clf_2.score(X_test, y_test).round(3)*100
prec_SVM_2 = precision_score(y_test, clf_2.predict(X_test)).round(3)*100
recall_SVM_2 = recall_score(y_test, clf_2.predict(X_test)).round(3)*100
Train_Test_Difference = abs(acc_SVM_2-train_accuracy).round(3)

print('Train_Accuracy_SVM =', train_accuracy.round(3))
print('Test_Accuracy_SVM = ', acc_SVM_2.round(3))
print('Train-Test Difference = ', Train_Test_Difference)
print()
print('Precision_SVM = ', prec_SVM_2.round(3))
print('Recall_SVM = ', recall_SVM_2.round(3))
print()

#confusion matrix(y_true, y_pred)
tn_2, fp_2, fn_2, tp_2 = confusion_matrix(y_test, clf_2.predict(X_test)).ravel()
print(('TN:{}'.format(tn_2), 'FP:{}'.format(fp_2), 'FN:{}'.format(fn_2), 'TP:{}'.for

plot_confusion_matrix(clf_2, X_test, y_test)
plt.show()

#train result
train_accuracy=clf.fit(X_train, y_train)
#test result
accuracy = (tp+tn)/(tp+tn+fp+fn)
precision = tp/(tp+fp)
recall = tp/(tp+fn)

#append result to 'precision_result_list' & 'recall_result_list'
training_title_list.append('1_SVM_2')
train_accuracy_f1_result_list.append(train_accuracy)
test_accuracy_f1_result_list.append(acc_SVM_2)
train_test_difference_result_list.append(Train_Test_Difference)
precision_result_list.append(prec_SVM_2)
recall_result_list.append(recall_SVM_2)

#Classification report
print()
target_names = ['state: good', 'state: bad']
print(classification_report(y_test, clf_2.predict(X_test), target_names=target_names)

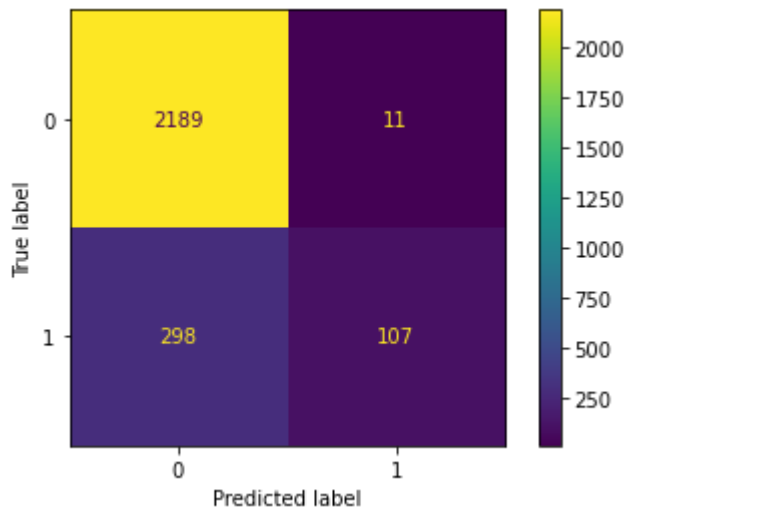
#f1_score
f1_score=f1_score(y_test, clf_2.predict(X_test), average='micro')
print('f1_score= {}'.format(f1_score))

Train_Accuracy_SVM = 88.7
Test_Accuracy_SVM = 88.1
Train-Test Difference = 0.6

Precision_SVM = 90.7
Recall_SVM = 26.4

('TN:2189', 'FP:11', 'FN:298', 'TP:107')

```



	precision	recall	f1-score	support
state: good	0.88	0.99	0.93	2200
state: bad	0.91	0.26	0.41	405
accuracy			0.88	2605
macro avg	0.89	0.63	0.67	2605
weighted avg	0.88	0.88	0.85	2605

f1_score= 0.8813819577735125

2_RandomForest_0

```
In [24]: from sklearn.preprocessing import LabelBinarizer
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import recall_score, precision_score
from sklearn.metrics import f1_score
from sklearn.metrics import classification_report

#1# split into 80:20
X=df_features_0
y=df_label

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_stat

clf = RandomForestClassifier(n_estimators=50, max_depth=5)
clf.fit(X_train, y_train)

#train result
train_accuracy=clf.score(X_train, y_train).round(3)*100
#test result
acc_RF      = clf.score(X_test, y_test).round(3)*100
prec_RF     = precision_score(y_test, clf.predict(X_test)).round(3)*100
recall_RF   = recall_score(y_test, clf.predict(X_test)).round(3)*100
Train_Test_Difference = abs(acc_RF-train_accuracy).round(3)

print('Train_Accuracy_RF =', train_accuracy.round(3))
print('Test_Accuracy_RF = ', acc_RF.round(3))
print('Train-Test Difference = ', Train_Test_Difference)
print()
print('Precision_RF = ', prec_RF.round(3))
print('Recall_RF      = ', recall_RF.round(3))
print(clf.predict(X_test))
print()
```

```

#confusion matrix(y_true, y_pred)
tn, fp, fn, tp = confusion_matrix(y_test, clf.predict(X_test)).ravel()
print(('TN:{}'.format(tn), 'FP:{}'.format(fp), 'FN:{}'.format(fn), 'TP:{}'.format(tp)

plot_confusion_matrix(clf, X_test, y_test)
plt.show()

#train result
train_accuracy=clf.fit(X_train, y_train)
#test result
accuracy = (tp+tn)/(tp+tn+fp+fn)
precision = tp/(tp+fp)
recall = tp/(tp+fn)

#append result to 'precision_result_list' & 'recall_result_list'
training_title_list.append('2_RandomForest_0')
train_accuracy_f1_result_list.append(train_accuracy)
test_accuracy_f1_result_list.append(acc_RF)
train_test_difference_result_list.append(Train_Test_Difference)
precision_result_list.append(prec_RF)
recall_result_list.append(recall_RF)

#Classification report
print()
target_names = ['state: good', 'state: bad']
print(classification_report(y_test, clf.predict(X_test), target_names=target_names))

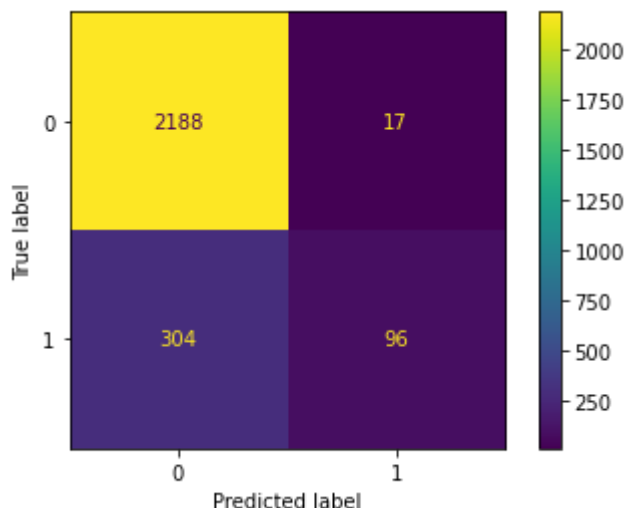
#f1_score
f1_score=f1_score(y_test, clf.predict(X_test), average='micro')
print('f1_score= {}'.format(f1_score))

```

Train_Accuracy_RF = 88.2
 Test_Accuracy_RF = 87.7
 Train-Test Difference = 0.5

Precision_RF = 85.0
 Recall_RF = 24.0
 [0 1 0 ... 0 0 0]

('TN:2188', 'FP:17', 'FN:304', 'TP:96')



	precision	recall	f1-score	support
state: good	0.88	0.99	0.93	2205
state: bad	0.86	0.24	0.38	400
accuracy			0.88	2605
macro avg	0.87	0.62	0.66	2605
weighted avg	0.88	0.88	0.85	2605

f1_score= 0.8775431861804224

2_RandomForest_1

```
In [25]: from sklearn.preprocessing import LabelBinarizer
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import recall_score, precision_score
from sklearn.metrics import f1_score
from sklearn.metrics import classification_report

#1# split into 80:20
X=df_features_1
y=df_label

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_stat

clf = RandomForestClassifier(n_estimators=50, max_depth=5)
clf.fit(X_train, y_train)

#train result
train_accuracy=clf.score(X_train, y_train).round(3)*100
#test result
acc_RF      = clf.score(X_test, y_test).round(3)*100
prec_RF     = precision_score(y_test, clf.predict(X_test)).round(3)*100
recall_RF  = recall_score(y_test, clf.predict(X_test)).round(3)*100
Train_Test_Difference = abs(acc_RF-train_accuracy).round(3)

print('Train_Accuracy_RF_1 =', train_accuracy.round(3))
print('Test_Accuracy_RF_1  = ', acc_RF.round(3))
print('Train-Test Difference = ', Train_Test_Difference)
print()
print('Precision_RF_1 = ', prec_RF.round(3))
print('Recall_RF_1    = ', recall_RF.round(3))
print(clf.predict(X_test))
print()

#confusion matrix(y_true, y_pred)
tn, fp, fn, tp = confusion_matrix(y_test, clf.predict(X_test)).ravel()
print(('TN:{}'.format(tn), 'FP:{}'.format(fp), 'FN:{}'.format(fn), 'TP:{}'.format(tp)

plot_confusion_matrix(clf, X_test, y_test)
plt.show()

#train result
train_accuracy=clf.fit(X_train, y_train)
#test result
accuracy = (tp+tn)/(tp+tn+fp+fn)
precision = tp/(tp+fp)
recall = tp/(tp+fn)
```

```
#append result to 'precision_result_list' & 'recall_result_list'
training_title_list.append('2_RandomForest_1')
train_accuracy_f1_result_list.append(train_accuracy)
test_accuracy_f1_result_list.append(acc_RF)
train_test_difference_result_list.append(Train_Test_Difference)
precision_result_list.append(prec_RF)
recall_result_list.append(recall_RF)

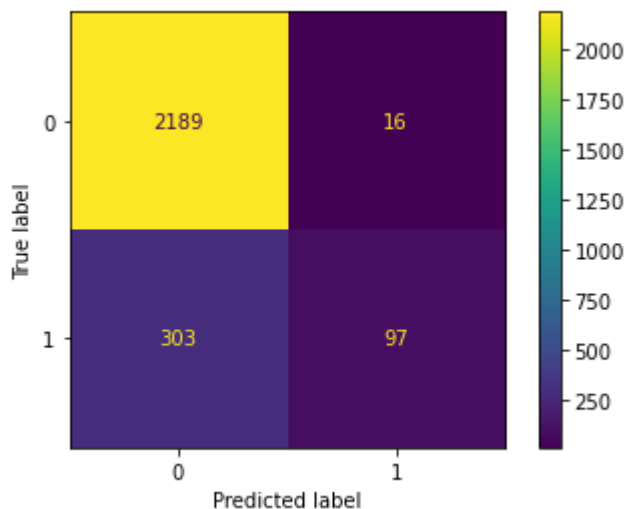
#Classification report
print()
target_names = ['state: good', 'state: bad']
print(classification_report(y_test, clf.predict(X_test), target_names=target_names))

#f1_score
f1_score=f1_score(y_test, clf.predict(X_test), average='micro')
print('f1_score= {}'.format(f1_score))
```

Train_Accuracy_RF_1 = 88.5
 Test_Accuracy_RF_1 = 87.8
 Train-Test Difference = 0.7

Precision_RF_1 = 85.8
 Recall_RF_1 = 24.2
 [0 1 0 ... 0 0 0]

('TN:2189', 'FP:16', 'FN:303', 'TP:97')



	precision	recall	f1-score	support
state: good	0.88	0.99	0.93	2205
state: bad	0.86	0.24	0.38	400
accuracy			0.88	2605
macro avg	0.87	0.62	0.65	2605
weighted avg	0.87	0.88	0.85	2605

f1_score= 0.8771593090211133

3_GradientBoosting_0

```
In [26]: from sklearn.datasets import make_hastie_10_2
from sklearn.ensemble import GradientBoostingClassifier
from sklearn.metrics import recall_score, precision_score
from sklearn.metrics import f1_score
from sklearn.metrics import classification_report
```

```

#1# split into 80:20
X=df_features_0
y=df_label

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_stat

clf_GB = GradientBoostingClassifier(n_estimators=1000, learning_rate=1, max_depth=1,

#train result
train_accuracy=clf_GB.score(X_train, y_train).round(3)*100
#test result
acc_GB      = clf_GB.score(X_test, y_test).round(3)*100
prec_GB     = precision_score(y_test, clf_GB.predict(X_test)).round(3)*100
recall_GB  = recall_score(y_test, clf_GB.predict(X_test)).round(3)*100
Train_Test_Difference = abs(acc_GB-train_accuracy).round(3)

print('Train_Accuracy_GB_0 =', train_accuracy.round(3))
print('Test_Accuracy_GB_0 = ', acc_GB.round(3))
print('Train-Test Difference = ', Train_Test_Difference)
print()
print('Precision_GB_0 = ', prec_GB.round(3))
print('Recall_GB_0     = ', recall_GB.round(3))
print()

#confusion matrix(y_true, y_pred)
tn_GB, fp_GB, fn_GB, tp_GB = confusion_matrix(y_test, clf_GB.predict(X_test)).ravel()
print(('TN:{}'.format(tn), 'FP:{}'.format(fp), 'FN:{}'.format(fn), 'TP:{}'.format(tp

plot_confusion_matrix(clf_GB, X_test, y_test)
plt.show()

#train result
train_accuracy=clf.fit(X_train, y_train)
#test result
accuracy = (tp_GB+tn_GB)/(tp_GB+tn_GB+fp_GB+fn_GB)
precision = tp_GB/(tp_GB+fp_GB)
recall = tp_GB/(tp_GB+fn_GB)

#append result to 'precision_result_list' & 'recall_result_list'
training_title_list.append('3_GradientBoosting_0')
train_accuracy_f1_result_list.append(train_accuracy)
test_accuracy_f1_result_list.append(acc_GB)
train_test_difference_result_list.append(Train_Test_Difference)
precision_result_list.append(prec_GB)
recall_result_list.append(recall_GB)

#Classification report
print()
target_names = ['state: good', 'state: bad']
print(classification_report(y_test, clf_GB.predict(X_test), target_names=target_name

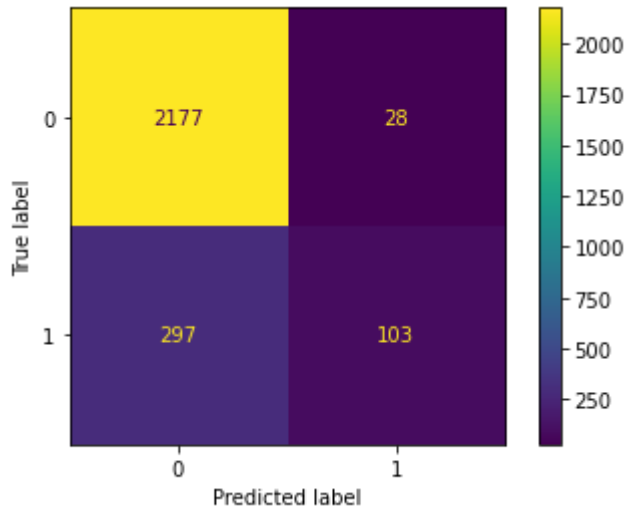
#f1_score
f1_score=f1_score(y_test, clf_GB.predict(X_test), average='micro')
print('f1_score= {}'.format(f1_score))

```

Train_Accuracy_GB_0 = 88.8
 Test_Accuracy_GB_0 = 87.5
 Train-Test Difference = 1.3

Precision_GB_0 = 78.6
 Recall_GB_0 = 25.8

('TN:2189', 'FP:16', 'FN:303', 'TP:97')



	precision	recall	f1-score	support
state: good	0.88	0.99	0.93	2205
state: bad	0.79	0.26	0.39	400
accuracy			0.88	2605
macro avg	0.83	0.62	0.66	2605
weighted avg	0.87	0.88	0.85	2605

f1_score= 0.8752399232245681

3_GradientBoosting_1

```
In [27]: from sklearn.datasets import make_hastie_10_2
from sklearn.ensemble import GradientBoostingClassifier
from sklearn.metrics import recall_score, precision_score
from sklearn.metrics import f1_score
from sklearn.metrics import classification_report

#1# split into 80:20
X=df_features_1
y=df_label

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_stat

clf_GB = GradientBoostingClassifier(n_estimators=1000, learning_rate=1, max_depth=1,

#train result
train_accuracy=clf_GB.score(X_train, y_train).round(3)*100
#test result
acc_GB = clf_GB.score(X_test, y_test).round(3)*100
prec_GB = precision_score(y_test, clf_GB.predict(X_test)).round(3)*100
recall_GB = recall_score(y_test, clf_GB.predict(X_test)).round(3)*100
Train_Test_Difference = abs(acc_GB-train_accuracy).round(3)

print('Train_Accuracy_GB_1 =', train_accuracy.round(3))
print('Test_Accuracy_GB_1 =', acc_GB.round(3))
```

```

print('Train-Test Difference = ', Train_Test_Difference)
print()
print('Precision_GB_1 = ', prec_GB.round(3))
print('Recall_GB_1 = ', recall_GB.round(3))
print()

#confusion matrix(y_true, y_pred)
tn_GB, fp_GB, fn_GB, tp_GB = confusion_matrix(y_test, clf_GB.predict(X_test)).ravel()
print(('TN:{}'.format(tn), 'FP:{}'.format(fp), 'FN:{}'.format(fn), 'TP:{}'.format(tp)

plot_confusion_matrix(clf_GB, X_test, y_test)
plt.show()

#train result
train_accuracy=clf.fit(X_train, y_train)
#test result
accuracy = (tp_GB+tn_GB)/(tp_GB+tn_GB+fp_GB+fn_GB)
precision = tp_GB/(tp_GB+fp_GB)
recall = tp_GB/(tp_GB+fn_GB)

#append result to 'precision_result_list' & 'recall_result_list'
training_title_list.append('3_GradientBoosting_1')
train_accuracy_f1_result_list.append(train_accuracy)
test_accuracy_f1_result_list.append(acc_GB)
train_test_difference_result_list.append(Train_Test_Difference)
precision_result_list.append(prec_GB)
recall_result_list.append(recall_GB)

#Classification report
print()
target_names = ['state: good', 'state: bad']
print(classification_report(y_test, clf_GB.predict(X_test), target_names=target_name

#f1_score
f1_score=f1_score(y_test, clf_GB.predict(X_test), average='micro')
print('f1_score= {}'.format(f1_score))

```

```

Train_Accuracy_GB_1 = 89.3
Test_Accuracy_GB_1 = 87.6
Train-Test Difference = 1.7

```

```

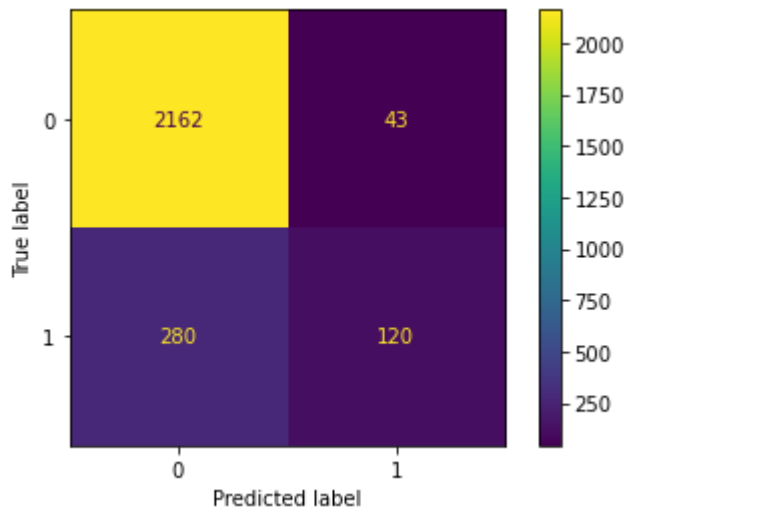
Precision_GB_1 = 73.6
Recall_GB_1 = 30.0

```

```

('TN:2189', 'FP:16', 'FN:303', 'TP:97')

```

	precision	recall	f1-score	support
state: good	0.89	0.98	0.93	2205
state: bad	0.74	0.30	0.43	400
accuracy			0.88	2605
macro avg	0.81	0.64	0.68	2605
weighted avg	0.86	0.88	0.85	2605

f1_score= 0.8760076775431862

4_K-NearestNeighbors_0

```
In [28]: from sklearn.neighbors import KNeighborsClassifier
from sklearn.model_selection import train_test_split
from sklearn.metrics import confusion_matrix
from sklearn.metrics import plot_confusion_matrix
from sklearn.svm import SVC
from sklearn.metrics import recall_score, precision_score
from sklearn.metrics import f1_score
from sklearn.metrics import classification_report

#1# split into 80:20
X=df_features_0
y=df_label

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_stat

clf_k = KNeighborsClassifier(n_neighbors=10)
clf_k.fit(X_train, y_train)

#train result
train_accuracy=clf_k.score(X_train, y_train).round(3)*100
#test result
acc_k = clf_k.score(X_test, y_test).round(3)*100
prec_k = precision_score(y_test, clf_k.predict(X_test)).round(3)*100
recall_k = recall_score(y_test, clf_k.predict(X_test)).round(3)*100
Train_Test_Difference = abs(acc_k-train_accuracy).round(3)

print('Train_Accuracy_k_0 =', train_accuracy.round(3))
print('Test_Accuracy_k_0 = ', acc_k.round(3))
print('Train-Test Difference = ', Train_Test_Difference)
print()
print('Precision_k_0 = ', prec_k.round(3))
print('Recall_k_0 = ', recall_k.round(3))
```

```

print()

#confusion matrix(y_true, y_pred)
tn, fp, fn, tp = confusion_matrix(y_test, clf_k.predict(X_test)).ravel()
print(('TN:{}'.format(tn), 'FP:{}'.format(fp), 'FN:{}'.format(fn), 'TP:{}'.format(tp)

plot_confusion_matrix(clf_k, X_test, y_test)
plt.show()

#train result
train_accuracy=clf.fit(X_train, y_train)
#test result
accuracy = (tp+tn)/(tp+tn+fp+fn)
precision = tp/(tp+fp)
recall = tp/(tp+fn)

#append result to 'precision_result_list' & 'recall_result_list'
training_title_list.append('4_K-NearestNeighbors_0')
train_accuracy_f1_result_list.append(train_accuracy)
test_accuracy_f1_result_list.append(acc_k)
train_test_difference_result_list.append(Train_Test_Difference)
precision_result_list.append(prec_k)
recall_result_list.append(recall_k)

#Classification report
print()
target_names = ['state: good', 'state: bad']
print(classification_report(y_test, clf_k.predict(X_test), target_names=target_names)

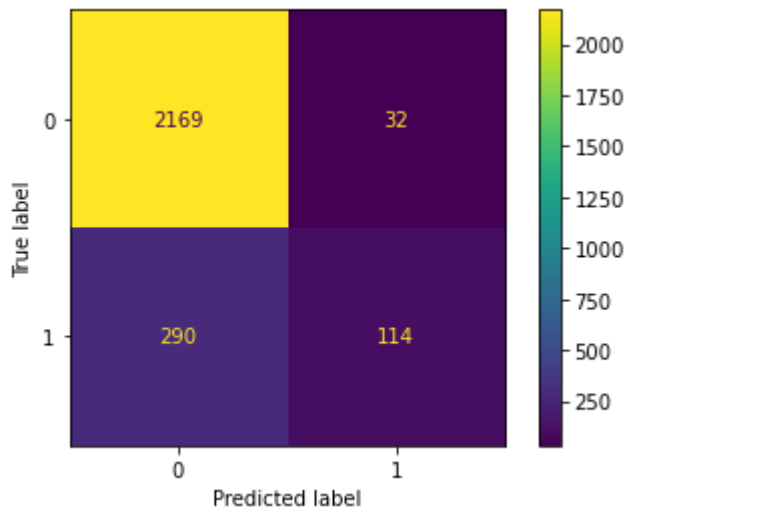
#f1_score
f1_score=f1_score(y_test, clf_k.predict(X_test), average='micro')
print('f1_score= {}'.format(f1_score))

Train_Accuracy_k_0 = 89.2
Test_Accuracy_k_0 = 87.6
Train-Test Difference = 1.6

Precision_k_0 = 78.1
Recall_k_0 = 28.2

('TN:2169', 'FP:32', 'FN:290', 'TP:114')

```



	precision	recall	f1-score	support
state: good	0.88	0.99	0.93	2201
state: bad	0.78	0.28	0.41	404
accuracy			0.88	2605
macro avg	0.83	0.63	0.67	2605
weighted avg	0.87	0.88	0.85	2605

f1_score= 0.8763915547024952

4_K-NearestNeighbors_1

```
In [29]: from sklearn.neighbors import KNeighborsClassifier
from sklearn.model_selection import train_test_split
from sklearn.metrics import confusion_matrix
from sklearn.metrics import plot_confusion_matrix
from sklearn.svm import SVC
from sklearn.metrics import recall_score, precision_score
from sklearn.metrics import f1_score
from sklearn.metrics import classification_report

#1# split into 80:20
X=df_features_1
y=df_label

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_stat

clf_k = KNeighborsClassifier(n_neighbors=10)
clf_k.fit(X_train, y_train)

#train result
train_accuracy=clf_k.score(X_train, y_train).round(3)*100
#test result
acc_k = clf_k.score(X_test, y_test).round(3)*100
prec_k = precision_score(y_test, clf_k.predict(X_test)).round(3)*100
recall_k = recall_score(y_test, clf_k.predict(X_test)).round(3)*100
Train_Test_Difference = abs(acc_k-train_accuracy).round(3)

print('Train_Accuracy_k_1 =', train_accuracy.round(3))
print('Test_Accuracy_k_1 = ', acc_k.round(3))
print('Train-Test Difference = ', Train_Test_Difference)
print()
print('Precision_k_1 = ', prec_k.round(3))
print('Recall_k_1 = ', recall_k.round(3))
```

```

print()

#confusion matrix(y_true, y_pred)
tn, fp, fn, tp = confusion_matrix(y_test, clf_k.predict(X_test)).ravel()
print(('TN:{}'.format(tn), 'FP:{}'.format(fp), 'FN:{}'.format(fn), 'TP:{}'.format(tp)

plot_confusion_matrix(clf_k, X_test, y_test)
plt.show()

#train result
train_accuracy=clf.fit(X_train, y_train)
#test result
accuracy = (tp+tn)/(tp+tn+fp+fn)
precision = tp/(tp+fp)
recall = tp/(tp+fn)

#append result to 'precision_result_list' & 'recall_result_list'
training_title_list.append('4_K-NearestNeighbors_1')
train_accuracy_f1_result_list.append(train_accuracy)
test_accuracy_f1_result_list.append(acc_k)
train_test_difference_result_list.append(Train_Test_Difference)
precision_result_list.append(prec_k)
recall_result_list.append(recall_k)

#Classification report
print()
target_names = ['state: good', 'state: bad']
print(classification_report(y_test, clf_k.predict(X_test), target_names=target_names)

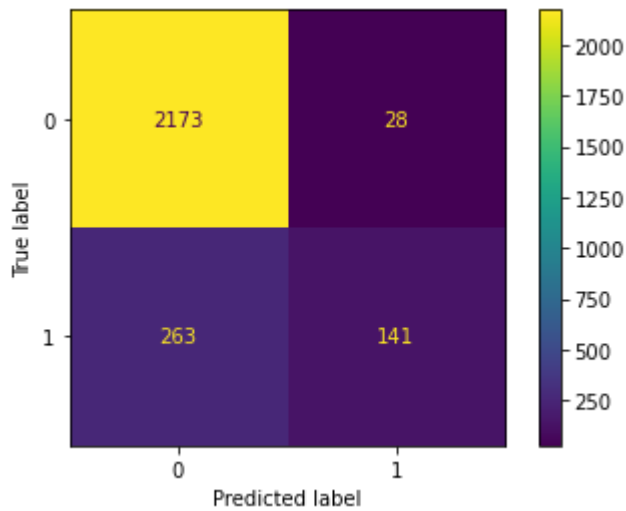
#f1_score
f1_score=f1_score(y_test, clf_k.predict(X_test), average='micro')
print('f1_score= {}'.format(f1_score))

Train_Accuracy_k_1 = 90.2
Test_Accuracy_k_1 = 88.8
Train-Test Difference = 1.4

Precision_k_1 = 83.4
Recall_k_1 = 34.9

('TN:2173', 'FP:28', 'FN:263', 'TP:141')

```



	precision	recall	f1-score	support
state: good	0.89	0.99	0.94	2201
state: bad	0.83	0.35	0.49	404
accuracy			0.89	2605
macro avg	0.86	0.67	0.71	2605
weighted avg	0.88	0.89	0.87	2605

f1_score= 0.8882917466410749

5_CNN_0 (with features: [mean_h, mean_v, stand_h, stand_v])

```
In [30]: from sklearn.metrics import classification_report
from pandas import read_csv
from keras.models import Sequential
from keras.layers import Dense
from keras.wrappers.scikit_learn import KerasClassifier
from sklearn.model_selection import cross_val_score
from sklearn.preprocessing import LabelEncoder
from sklearn.model_selection import StratifiedKFold
from sklearn.pipeline import Pipeline
from sklearn.metrics import f1_score
from sklearn.metrics import confusion_matrix
```

```
In [31]: #1# split into 80:20
X=df_features_0
y=df_label

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_stat

#2# build the model
clf_cnn = keras.Sequential([
    keras.layers.Flatten(input_shape=(4,)),
    keras.layers.Dense(16, activation=tf.nn.relu),
    keras.layers.Dense(16, activation=tf.nn.relu),
    keras.layers.Dense(1, activation=tf.nn.sigmoid),
])

#3# compile model
clf_cnn.compile(optimizer='adam',
                loss='binary_crossentropy',
                metrics=['accuracy'])
```

```
#4# training
clf_cnn.fit(X_train, y_train, epochs=50, batch_size=1)

test_loss, test_acc = clf_cnn.evaluate(X_test, y_test)
print('Test accuracy:', test_acc)
```

Train on 10420 samples

```
Epoch 1/50
10420/10420 [=====] - 18s 2ms/sample - loss: 0.3992 - accur
acy: 0.8605
Epoch 2/50
10420/10420 [=====] - 18s 2ms/sample - loss: 0.3848 - accur
acy: 0.8604
Epoch 3/50
10420/10420 [=====] - 17s 2ms/sample - loss: 0.3821 - accur
acy: 0.8630
Epoch 4/50
10420/10420 [=====] - 17s 2ms/sample - loss: 0.3800 - accur
acy: 0.8667
Epoch 5/50
10420/10420 [=====] - 17s 2ms/sample - loss: 0.3777 - accur
acy: 0.8702
Epoch 6/50
10420/10420 [=====] - 17s 2ms/sample - loss: 0.3752 - accur
acy: 0.8713
Epoch 7/50
10420/10420 [=====] - 17s 2ms/sample - loss: 0.3734 - accur
acy: 0.8713
Epoch 8/50
10420/10420 [=====] - 18s 2ms/sample - loss: 0.3703 - accur
acy: 0.8711
Epoch 9/50
10420/10420 [=====] - 17s 2ms/sample - loss: 0.3700 - accur
acy: 0.8720
Epoch 10/50
10420/10420 [=====] - 17s 2ms/sample - loss: 0.3675 - accur
acy: 0.8721
Epoch 11/50
10420/10420 [=====] - 17s 2ms/sample - loss: 0.3664 - accur
acy: 0.8730
Epoch 12/50
10420/10420 [=====] - 17s 2ms/sample - loss: 0.3646 - accur
acy: 0.8738
Epoch 13/50
10420/10420 [=====] - 17s 2ms/sample - loss: 0.3642 - accur
acy: 0.8741
Epoch 14/50
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3619 - accur
acy: 0.8743
Epoch 15/50
10420/10420 [=====] - 19s 2ms/sample - loss: 0.3613 - accur
acy: 0.8750
Epoch 16/50
10420/10420 [=====] - 18s 2ms/sample - loss: 0.3601 - accur
acy: 0.8755
Epoch 17/50
10420/10420 [=====] - 19s 2ms/sample - loss: 0.3599 - accur
acy: 0.8756
Epoch 18/50
10420/10420 [=====] - 19s 2ms/sample - loss: 0.3586 - accur
acy: 0.8758
Epoch 19/50
10420/10420 [=====] - 24s 2ms/sample - loss: 0.3576 - accur
acy: 0.8764
Epoch 20/50
10420/10420 [=====] - 25s 2ms/sample - loss: 0.3584 - accur
acy: 0.8763
Epoch 21/50
10420/10420 [=====] - 23s 2ms/sample - loss: 0.3567 - accur
```

```
acy: 0.8771
Epoch 22/50
10420/10420 [=====] - 20s 2ms/sample - loss: 0.3557 - accur
acy: 0.8768
Epoch 23/50
10420/10420 [=====] - 22s 2ms/sample - loss: 0.3556 - accur
acy: 0.8761
Epoch 24/50
10420/10420 [=====] - 23s 2ms/sample - loss: 0.3545 - accur
acy: 0.8772
Epoch 25/50
10420/10420 [=====] - 23s 2ms/sample - loss: 0.3548 - accur
acy: 0.8765
Epoch 26/50
10420/10420 [=====] - 22s 2ms/sample - loss: 0.3533 - accur
acy: 0.8771
Epoch 27/50
10420/10420 [=====] - 22s 2ms/sample - loss: 0.3531 - accur
acy: 0.8771
Epoch 28/50
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3522 - accur
acy: 0.8766
Epoch 29/50
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3539 - accur
acy: 0.8769
Epoch 30/50
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3520 - accur
acy: 0.8760
Epoch 31/50
10420/10420 [=====] - 20s 2ms/sample - loss: 0.3513 - accur
acy: 0.8771
Epoch 32/50
10420/10420 [=====] - 20s 2ms/sample - loss: 0.3512 - accur
acy: 0.8773
Epoch 33/50
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3511 - accur
acy: 0.8774
Epoch 34/50
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3515 - accur
acy: 0.8771
Epoch 35/50
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3508 - accur
acy: 0.8769
Epoch 36/50
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3514 - accur
acy: 0.8774
Epoch 37/50
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3512 - accur
acy: 0.8774
Epoch 38/50
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3503 - accur
acy: 0.8774
Epoch 39/50
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3510 - accur
acy: 0.8770
Epoch 40/50
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3500 - accur
acy: 0.8767
Epoch 41/50
10420/10420 [=====] - 23s 2ms/sample - loss: 0.3495 - accur
acy: 0.8770
Epoch 42/50
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3490 - accur
acy: 0.8770
Epoch 43/50
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3499 - accur
acy: 0.8771
Epoch 44/50
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3489 - accur
```

```

acy: 0.8776
Epoch 45/50
10420/10420 [=====] - 20s 2ms/sample - loss: 0.3491 - accur
acy: 0.8777
Epoch 46/50
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3495 - accur
acy: 0.8776
Epoch 47/50
10420/10420 [=====] - 20s 2ms/sample - loss: 0.3481 - accur
acy: 0.8781
Epoch 48/50
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3471 - accur
acy: 0.8774
Epoch 49/50
10420/10420 [=====] - 20s 2ms/sample - loss: 0.3482 - accur
acy: 0.8777
Epoch 50/50
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3470 - accur
acy: 0.8781
2605/2605 [=====] - 0s 73us/sample - loss: 0.3482 - accurac
y: 0.8775
Test accuracy: 0.8775432

```

```

In [32]: Y_pred = clf_cnn.predict(X_test)
         #print(Y_pred)
         y_pred = np.where(Y_pred>0.5, 1, 0)

         unique_y_pred, counts_y_pred = np.unique(y_pred, return_counts=True)
         y_pred_static = dict(zip(unique_y_pred, counts_y_pred))
         #print(y_pred_static)

         #confusion matrix(y_true, y_pred)
         tn_cnn, fp_cnn, fn_cnn, tp_cnn = confusion_matrix(y_test, y_pred).ravel()

         #train result
         train_accuracy= clf_cnn.evaluate(X_train, y_train)[1].round(3)*100
         #test result
         accuracy = (tp_cnn+tn_cnn)/(tp_cnn+tn_cnn+fp_cnn+fn_cnn).round(3)*100
         precision = tp_cnn/(tp_cnn+fp_cnn).round(3)*100
         recall = tp_cnn/(tp_cnn+fn_cnn).round(3)*100
         Train_Test_Difference = abs(accuracy-train_accuracy).round(3)

         print()
         print('Train_Accuracy_cnn_0 = ', train_accuracy.round(3))
         print('Test_Accuracy_cnn_0 = ', accuracy.round(3))
         print('Train-Test Difference = ', Train_Test_Difference)
         print()
         print('Precision_cnn_0 = ', precision.round(3))
         print('Recall_cnn_0 = ', recall.round(3))
         print()

         #append result to 'precision_result_list' & 'recall_result_list'
         training_title_list.append('5_CNN_0')
         train_accuracy_f1_result_list.append(train_accuracy)
         test_accuracy_f1_result_list.append(accuracy)
         train_test_difference_result_list.append(Train_Test_Difference)
         precision_result_list.append(precision)
         recall_result_list.append(recall)

         #Classification report

```



```
print()
target_names = ['state: good', 'state: bad']
print(classification_report(y_test, y_pred, target_names=target_names))
```

```
10420/10420 [=====] - 0s 43us/sample - loss: 0.3435 - accur
acy: 0.8782
```

```
Train_Accuracy_cnn_0 = 87.8
Test_Accuracy_cnn_0 = 87.754
Train-Test Difference = 0.046
```

```
Precision_cnn_0 = 94.792
Recall_cnn_0 = 22.469
```

	precision	recall	f1-score	support
state: good	0.87	1.00	0.93	2200
state: bad	0.95	0.22	0.36	405
accuracy			0.88	2605
macro avg	0.91	0.61	0.65	2605
weighted avg	0.89	0.88	0.84	2605

5_CNN_1 (with features: [mean_h, mean_v, stand_h, stand_v])

```
In [33]: from sklearn.metrics import classification_report
from pandas import read_csv
from keras.models import Sequential
from keras.layers import Dense
from keras.wrappers.scikit_learn import KerasClassifier
from sklearn.model_selection import cross_val_score
from sklearn.preprocessing import LabelEncoder
from sklearn.model_selection import StratifiedKFold
from sklearn.pipeline import Pipeline
from sklearn.metrics import f1_score
```

```
In [34]: #1# split into 80:20
X=df_features_0
y=df_label

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_stat

#2# build the model
clf_cnn_1 = keras.Sequential([
    keras.layers.Flatten(input_shape=(4,)),
    keras.layers.Dense(16, activation=tf.nn.relu),
    keras.layers.Dense(16, activation=tf.nn.relu),
    keras.layers.Dense(16, activation=tf.nn.relu),
    keras.layers.Dense(16, activation=tf.nn.relu),
    keras.layers.Dense(1, activation=tf.nn.sigmoid),
])

#3# compile model
clf_cnn_1.compile(optimizer='adam',
                  loss='binary_crossentropy',
                  metrics=['accuracy'])

#4# training
clf_cnn_1.fit(X_train, y_train, epochs=50, batch_size=1)
```

```
test_loss, test_acc = clf_cnn_1.evaluate(X_test, y_test)
print('Test accuracy:', test_acc)
```

Train on 10420 samples

```
Epoch 1/50
10420/10420 [=====] - 25s 2ms/sample - loss: 0.3964 - accur
acy: 0.8615
Epoch 2/50
10420/10420 [=====] - 24s 2ms/sample - loss: 0.3819 - accur
acy: 0.8686
Epoch 3/50
10420/10420 [=====] - 24s 2ms/sample - loss: 0.3773 - accur
acy: 0.8705
Epoch 4/50
10420/10420 [=====] - 24s 2ms/sample - loss: 0.3723 - accur
acy: 0.8716
Epoch 5/50
10420/10420 [=====] - 25s 2ms/sample - loss: 0.3687 - accur
acy: 0.8735
Epoch 6/50
10420/10420 [=====] - 24s 2ms/sample - loss: 0.3643 - accur
acy: 0.8745
Epoch 7/50
10420/10420 [=====] - 24s 2ms/sample - loss: 0.3617 - accur
acy: 0.8763
Epoch 8/50
10420/10420 [=====] - 23s 2ms/sample - loss: 0.3592 - accur
acy: 0.8765
Epoch 9/50
10420/10420 [=====] - 24s 2ms/sample - loss: 0.3557 - accur
acy: 0.8784
Epoch 10/50
10420/10420 [=====] - 24s 2ms/sample - loss: 0.3540 - accur
acy: 0.8783
Epoch 11/50
10420/10420 [=====] - 24s 2ms/sample - loss: 0.3522 - accur
acy: 0.8792
Epoch 12/50
10420/10420 [=====] - 24s 2ms/sample - loss: 0.3502 - accur
acy: 0.8789
Epoch 13/50
10420/10420 [=====] - 24s 2ms/sample - loss: 0.3483 - accur
acy: 0.8784
Epoch 14/50
10420/10420 [=====] - 26s 2ms/sample - loss: 0.3485 - accur
acy: 0.8790
Epoch 15/50
10420/10420 [=====] - 25s 2ms/sample - loss: 0.3490 - accur
acy: 0.8794
Epoch 16/50
10420/10420 [=====] - 24s 2ms/sample - loss: 0.3460 - accur
acy: 0.8792
Epoch 17/50
10420/10420 [=====] - 24s 2ms/sample - loss: 0.3470 - accur
acy: 0.8795
Epoch 18/50
10420/10420 [=====] - 24s 2ms/sample - loss: 0.3457 - accur
acy: 0.8790
Epoch 19/50
10420/10420 [=====] - 24s 2ms/sample - loss: 0.3442 - accur
acy: 0.8802
Epoch 20/50
10420/10420 [=====] - 26s 3ms/sample - loss: 0.3439 - accur
acy: 0.8790
Epoch 21/50
10420/10420 [=====] - 25s 2ms/sample - loss: 0.3458 - accur
acy: 0.8796
Epoch 22/50
10420/10420 [=====] - 26s 3ms/sample - loss: 0.3443 - accur
```

```
acy: 0.8799
Epoch 23/50
10420/10420 [=====] - 25s 2ms/sample - loss: 0.3451 - accur
acy: 0.8803
Epoch 24/50
10420/10420 [=====] - 25s 2ms/sample - loss: 0.3456 - accur
acy: 0.8783
Epoch 25/50
10420/10420 [=====] - 25s 2ms/sample - loss: 0.3429 - accur
acy: 0.8794
Epoch 26/50
10420/10420 [=====] - 29s 3ms/sample - loss: 0.3434 - accur
acy: 0.8787
Epoch 27/50
10420/10420 [=====] - 28s 3ms/sample - loss: 0.3426 - accur
acy: 0.8802
Epoch 28/50
10420/10420 [=====] - 27s 3ms/sample - loss: 0.3414 - accur
acy: 0.8810
Epoch 29/50
10420/10420 [=====] - 26s 3ms/sample - loss: 0.3418 - accur
acy: 0.8801
Epoch 30/50
10420/10420 [=====] - 26s 2ms/sample - loss: 0.3413 - accur
acy: 0.8802
Epoch 31/50
10420/10420 [=====] - 26s 3ms/sample - loss: 0.3404 - accur
acy: 0.8808
Epoch 32/50
10420/10420 [=====] - 25s 2ms/sample - loss: 0.3417 - accur
acy: 0.8808
Epoch 33/50
10420/10420 [=====] - 25s 2ms/sample - loss: 0.3405 - accur
acy: 0.8810
Epoch 34/50
10420/10420 [=====] - 26s 2ms/sample - loss: 0.3394 - accur
acy: 0.8809
Epoch 35/50
10420/10420 [=====] - 26s 2ms/sample - loss: 0.3395 - accur
acy: 0.8802
Epoch 36/50
10420/10420 [=====] - 25s 2ms/sample - loss: 0.3385 - accur
acy: 0.8803
Epoch 37/50
10420/10420 [=====] - 26s 3ms/sample - loss: 0.3375 - accur
acy: 0.8807
Epoch 38/50
10420/10420 [=====] - 26s 3ms/sample - loss: 0.3380 - accur
acy: 0.8811
Epoch 39/50
10420/10420 [=====] - 26s 3ms/sample - loss: 0.3379 - accur
acy: 0.8801
Epoch 40/50
10420/10420 [=====] - 26s 2ms/sample - loss: 0.3394 - accur
acy: 0.8799
Epoch 41/50
10420/10420 [=====] - 26s 3ms/sample - loss: 0.3371 - accur
acy: 0.8812
Epoch 42/50
10420/10420 [=====] - 27s 3ms/sample - loss: 0.3380 - accur
acy: 0.8807
Epoch 43/50
10420/10420 [=====] - 27s 3ms/sample - loss: 0.3383 - accur
acy: 0.8810
Epoch 44/50
10420/10420 [=====] - 25s 2ms/sample - loss: 0.3365 - accur
acy: 0.8812
Epoch 45/50
10420/10420 [=====] - 26s 2ms/sample - loss: 0.3366 - accur
```

```

acy: 0.8809
Epoch 46/50
10420/10420 [=====] - 25s 2ms/sample - loss: 0.3361 - accur
acy: 0.8814
Epoch 47/50
10420/10420 [=====] - 24s 2ms/sample - loss: 0.3345 - accur
acy: 0.8816
Epoch 48/50
10420/10420 [=====] - 24s 2ms/sample - loss: 0.3373 - accur
acy: 0.8807
Epoch 49/50
10420/10420 [=====] - 27s 3ms/sample - loss: 0.3364 - accur
acy: 0.8812
Epoch 50/50
10420/10420 [=====] - 25s 2ms/sample - loss: 0.3367 - accur
acy: 0.8814
2605/2605 [=====] - 0s 86us/sample - loss: 0.3321 - accurac
y: 0.8818
Test accuracy: 0.88176584

```

```

In [35]: Y_pred = clf_cnn_1.predict(X_test)
         #print(Y_pred)
         y_pred = np.where(Y_pred>0.5, 1, 0)

         unique_y_pred, counts_y_pred = np.unique(y_pred, return_counts=True)
         y_pred_static = dict(zip(unique_y_pred, counts_y_pred))
         #print(y_pred_static)

         #confusion matrix(y_true, y_pred)
         tn_cnn, fp_cnn, fn_cnn, tp_cnn = confusion_matrix(y_test, y_pred).ravel()

         #train result
         train_accuracy= clf_cnn_1.evaluate(X_train, y_train)[1].round(3)*100
         #test result
         accuracy = (tp_cnn+tn_cnn)/(tp_cnn+tn_cnn+fp_cnn+fn_cnn).round(3)*100
         precision = tp_cnn/(tp_cnn+fp_cnn).round(3)*100
         recall = tp_cnn/(tp_cnn+fn_cnn).round(3)*100
         Train_Test_Difference = abs(accuracy-train_accuracy).round(3)

         print('Train_Accuracy_cnn_1 = ', train_accuracy.round(3))
         print('Test_Accuracy_cnn_1 = ', accuracy.round(3))
         print('Train-Test Difference = ', Train_Test_Difference)
         print()
         print('Precision_cnn_1 = ', precision.round(3))
         print('Recall_cnn_1 = ', recall.round(3))
         print()

         #append result to 'precision_result_list' & 'recall_result_list'
         training_title_list.append('5_CNN_1')
         train_accuracy_f1_result_list.append(train_accuracy)
         test_accuracy_f1_result_list.append(accuracy)
         train_test_difference_result_list.append(Train_Test_Difference)
         precision_result_list.append(precision)
         recall_result_list.append(recall)

         #Classification report
         print()
         target_names = ['state: good', 'state: bad']
         print(classification_report(y_test, y_pred, target_names=target_names))

```

```

10420/10420 [=====] - 0s 47us/sample - loss: 0.3265 - accur
acy: 0.8823
Train_Accuracy_cnn_1 = 88.2
Test_Accuracy_cnn_1 = 88.177
Train-Test Difference = 0.023

Precision_cnn_1 = 97.087
Recall_cnn_1 = 24.691

```

	precision	recall	f1-score	support
state: good	0.88	1.00	0.93	2200
state: bad	0.97	0.25	0.39	405
accuracy			0.88	2605
macro avg	0.92	0.62	0.66	2605
weighted avg	0.89	0.88	0.85	2605

5_CNN_2 (with features: [abs_mean_h, abs_mean_v, stand_h, stand_v])

```

In [36]: from sklearn.metrics import classification_report
from pandas import read_csv
from keras.models import Sequential
from keras.layers import Dense
from keras.wrappers.scikit_learn import KerasClassifier
from sklearn.model_selection import cross_val_score
from sklearn.preprocessing import LabelEncoder
from sklearn.model_selection import StratifiedKFold
from sklearn.pipeline import Pipeline
from sklearn.metrics import f1_score

```

```

In [37]: #1# split into 80:20
X=df_features_1
y=df_label

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_stat

#2# build the model
clf_cnn_1 = keras.Sequential([
    keras.layers.Flatten(input_shape=(4,)),
    keras.layers.Dense(16, activation=tf.nn.relu),
    keras.layers.Dense(16, activation=tf.nn.relu),
    keras.layers.Dense(1, activation=tf.nn.sigmoid),
])

#3# compile model
clf_cnn_1.compile(optimizer='adam',
                  loss='binary_crossentropy',
                  metrics=['accuracy'])

#4# training
clf_cnn_1.fit(X_train, y_train, epochs=50, batch_size=1)

test_loss, test_acc = clf_cnn_1.evaluate(X_test, y_test)
print('Test accuracy:', test_acc)

```

```

Train on 10420 samples
Epoch 1/50
10420/10420 [=====] - 22s 2ms/sample - loss: 0.3998 - accur
acy: 0.8607

```

```
Epoch 2/50
10420/10420 [=====] - 23s 2ms/sample - loss: 0.3870 - accur
acy: 0.8615
Epoch 3/50
10420/10420 [=====] - 26s 2ms/sample - loss: 0.3816 - accur
acy: 0.8614
Epoch 4/50
10420/10420 [=====] - 26s 2ms/sample - loss: 0.3769 - accur
acy: 0.8622
Epoch 5/50
10420/10420 [=====] - 25s 2ms/sample - loss: 0.3723 - accur
acy: 0.8673
Epoch 6/50
10420/10420 [=====] - 24s 2ms/sample - loss: 0.3695 - accur
acy: 0.8708
Epoch 7/50
10420/10420 [=====] - 23s 2ms/sample - loss: 0.3653 - accur
acy: 0.8725
Epoch 8/50
10420/10420 [=====] - 23s 2ms/sample - loss: 0.3623 - accur
acy: 0.8726
Epoch 9/50
10420/10420 [=====] - 24s 2ms/sample - loss: 0.3595 - accur
acy: 0.8743
Epoch 10/50
10420/10420 [=====] - 24s 2ms/sample - loss: 0.3588 - accur
acy: 0.8754
Epoch 11/50
10420/10420 [=====] - 24s 2ms/sample - loss: 0.3574 - accur
acy: 0.8744
Epoch 12/50
10420/10420 [=====] - 23s 2ms/sample - loss: 0.3556 - accur
acy: 0.8770
Epoch 13/50
10420/10420 [=====] - 22s 2ms/sample - loss: 0.3547 - accur
acy: 0.8762
Epoch 14/50
10420/10420 [=====] - 22s 2ms/sample - loss: 0.3529 - accur
acy: 0.8768
Epoch 15/50
10420/10420 [=====] - 23s 2ms/sample - loss: 0.3532 - accur
acy: 0.8776
Epoch 16/50
10420/10420 [=====] - 22s 2ms/sample - loss: 0.3514 - accur
acy: 0.8770
Epoch 17/50
10420/10420 [=====] - 22s 2ms/sample - loss: 0.3507 - accur
acy: 0.8767
Epoch 18/50
10420/10420 [=====] - 23s 2ms/sample - loss: 0.3514 - accur
acy: 0.8764
Epoch 19/50
10420/10420 [=====] - 24s 2ms/sample - loss: 0.3487 - accur
acy: 0.8780
Epoch 20/50
10420/10420 [=====] - 22s 2ms/sample - loss: 0.3482 - accur
acy: 0.8769
Epoch 21/50
10420/10420 [=====] - 23s 2ms/sample - loss: 0.3462 - accur
acy: 0.8785
Epoch 22/50
10420/10420 [=====] - 22s 2ms/sample - loss: 0.3462 - accur
acy: 0.8770
Epoch 23/50
10420/10420 [=====] - 22s 2ms/sample - loss: 0.3455 - accur
acy: 0.8780
Epoch 24/50
10420/10420 [=====] - 22s 2ms/sample - loss: 0.3432 - accur
acy: 0.8781
```

```
Epoch 25/50
10420/10420 [=====] - 22s 2ms/sample - loss: 0.3419 - accur
acy: 0.8789
Epoch 26/50
10420/10420 [=====] - 22s 2ms/sample - loss: 0.3411 - accur
acy: 0.8785
Epoch 27/50
10420/10420 [=====] - 23s 2ms/sample - loss: 0.3401 - accur
acy: 0.8789
Epoch 28/50
10420/10420 [=====] - 24s 2ms/sample - loss: 0.3402 - accur
acy: 0.8792
Epoch 29/50
10420/10420 [=====] - 23s 2ms/sample - loss: 0.3371 - accur
acy: 0.8792
Epoch 30/50
10420/10420 [=====] - 23s 2ms/sample - loss: 0.3389 - accur
acy: 0.8788
Epoch 31/50
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3363 - accur
acy: 0.8794
Epoch 32/50
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3336 - accur
acy: 0.8790
Epoch 33/50
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3353 - accur
acy: 0.8790
Epoch 34/50
10420/10420 [=====] - 18s 2ms/sample - loss: 0.3330 - accur
acy: 0.8788
Epoch 35/50
10420/10420 [=====] - 18s 2ms/sample - loss: 0.3308 - accur
acy: 0.8798
Epoch 36/50
10420/10420 [=====] - 18s 2ms/sample - loss: 0.3297 - accur
acy: 0.8797
Epoch 37/50
10420/10420 [=====] - 19s 2ms/sample - loss: 0.3293 - accur
acy: 0.8788
Epoch 38/50
10420/10420 [=====] - 18s 2ms/sample - loss: 0.3295 - accur
acy: 0.8792
Epoch 39/50
10420/10420 [=====] - 18s 2ms/sample - loss: 0.3288 - accur
acy: 0.8801
Epoch 40/50
10420/10420 [=====] - 18s 2ms/sample - loss: 0.3272 - accur
acy: 0.8808
Epoch 41/50
10420/10420 [=====] - 18s 2ms/sample - loss: 0.3251 - accur
acy: 0.8805
Epoch 42/50
10420/10420 [=====] - 19s 2ms/sample - loss: 0.3272 - accur
acy: 0.8803
Epoch 43/50
10420/10420 [=====] - 18s 2ms/sample - loss: 0.3284 - accur
acy: 0.8805
Epoch 44/50
10420/10420 [=====] - 18s 2ms/sample - loss: 0.3252 - accur
acy: 0.8802
Epoch 45/50
10420/10420 [=====] - 19s 2ms/sample - loss: 0.3244 - accur
acy: 0.8809
Epoch 46/50
10420/10420 [=====] - 18s 2ms/sample - loss: 0.3244 - accur
acy: 0.8806
Epoch 47/50
10420/10420 [=====] - 18s 2ms/sample - loss: 0.3231 - accur
acy: 0.8813
```

```
Epoch 48/50
10420/10420 [=====] - 18s 2ms/sample - loss: 0.3210 - accur
acy: 0.8815
Epoch 49/50
10420/10420 [=====] - 18s 2ms/sample - loss: 0.3234 - accur
acy: 0.8808
Epoch 50/50
10420/10420 [=====] - 18s 2ms/sample - loss: 0.3210 - accur
acy: 0.8812
2605/2605 [=====] - 0s 76us/sample - loss: 0.3362 - accurac
y: 0.8810
Test accuracy: 0.8809981
```

```
In [38]: Y_pred = clf_cnn_1.predict(X_test)
# print(Y_pred)
y_pred = np.where(Y_pred>0.5, 1, 0)

unique_y_pred, counts_y_pred = np.unique(y_pred, return_counts=True)
y_pred_static = dict(zip(unique_y_pred, counts_y_pred))
# print(y_pred_static)

# confusion matrix(y_true, y_pred)
tn_cnn, fp_cnn, fn_cnn, tp_cnn = confusion_matrix(y_test, y_pred).ravel()

# train result
train_accuracy = clf_cnn_1.evaluate(X_train, y_train)[1].round(3)*100
# test result
accuracy = (tp_cnn+tn_cnn)/(tp_cnn+tn_cnn+fp_cnn+fn_cnn).round(3)*100
precision = tp_cnn/(tp_cnn+fp_cnn).round(3)*100
recall = tp_cnn/(tp_cnn+fn_cnn).round(3)*100
Train_Test_Difference = abs(accuracy-train_accuracy).round(3)

print('Train_Accuracy_cnn_2 = ', train_accuracy.round(3))
print('Test_Accuracy_cnn_2 = ', accuracy.round(3))
print('Train-Test Difference = ', Train_Test_Difference)
print()
print('Precision_cnn_2 = ', precision.round(3))
print('Recall_cnn_2 = ', recall.round(3))
print()

# append result to 'precision_result_list' & 'recall_result_list'
training_title_list.append('5_CNN_2')
train_accuracy_f1_result_list.append(train_accuracy)
test_accuracy_f1_result_list.append(accuracy)
train_test_difference_result_list.append(Train_Test_Difference)
precision_result_list.append(precision)
recall_result_list.append(recall)

# Classification report
print()
target_names = ['state: good', 'state: bad']
print(classification_report(y_test, y_pred, target_names=target_names))

10420/10420 [=====] - 0s 43us/sample - loss: 0.3363 - accur
acy: 0.8827
Train_Accuracy_cnn_2 = 88.3
Test_Accuracy_cnn_2 = 88.1
Train-Test Difference = 0.2

Precision_cnn_2 = 92.793
Recall_cnn_2 = 25.432
```


	precision	recall	f1-score	support
state: good	0.88	1.00	0.93	2200
state: bad	0.93	0.25	0.40	405
accuracy			0.88	2605
macro avg	0.90	0.63	0.67	2605
weighted avg	0.89	0.88	0.85	2605

5_CNN_3 (with features: [abs_mean_h, abs_mean_v, stand_h, stand_v])

```
In [39]: from sklearn.metrics import classification_report
from pandas import read_csv
from keras.models import Sequential
from keras.layers import Dense
from keras.wrappers.scikit_learn import KerasClassifier
from sklearn.model_selection import cross_val_score
from sklearn.preprocessing import LabelEncoder
from sklearn.model_selection import StratifiedKFold
from sklearn.pipeline import Pipeline
from sklearn.metrics import f1_score
```

```
In [40]: #1# split into 80:20
X=df_features_1
y=df_label

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_stat

#2# build the model
clf_cnn_1 = keras.Sequential([
    keras.layers.Flatten(input_shape=(4,)),
    keras.layers.Dense(16, activation=tf.nn.relu),
    keras.layers.Dense(16, activation=tf.nn.relu),
    keras.layers.Dense(16, activation=tf.nn.relu),
    keras.layers.Dense(16, activation=tf.nn.relu),
    keras.layers.Dense(1, activation=tf.nn.sigmoid),
])

#3# compile model
clf_cnn_1.compile(optimizer='adam',
                  loss='binary_crossentropy',
                  metrics=['accuracy'])

#4# training
clf_cnn_1.fit(X_train, y_train, epochs=100, batch_size=1)

test_loss, test_acc = clf_cnn_1.evaluate(X_test, y_test)
print('Test accuracy:', test_acc)
```

Train on 10420 samples

Epoch 1/100

10420/10420 [=====] - 22s 2ms/sample - loss: 0.3960 - accuracy: 0.8614

Epoch 2/100

10420/10420 [=====] - 20s 2ms/sample - loss: 0.3822 - accuracy: 0.8671

Epoch 3/100

10420/10420 [=====] - 21s 2ms/sample - loss: 0.3791 - accuracy: 0.8708

```
Epoch 4/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3743 - accur
acy: 0.8726
Epoch 5/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3686 - accur
acy: 0.8749
Epoch 6/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3648 - accur
acy: 0.8770
Epoch 7/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3607 - accur
acy: 0.8781
Epoch 8/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3576 - accur
acy: 0.8790
Epoch 9/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3533 - accur
acy: 0.8789
Epoch 10/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3527 - accur
acy: 0.8788
Epoch 11/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3508 - accur
acy: 0.8789
Epoch 12/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3476 - accur
acy: 0.8807
Epoch 13/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3459 - accur
acy: 0.8797
Epoch 14/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3448 - accur
acy: 0.8805
Epoch 15/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3434 - accur
acy: 0.8806
Epoch 16/100
10420/10420 [=====] - 20s 2ms/sample - loss: 0.3401 - accur
acy: 0.8821
Epoch 17/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3395 - accur
acy: 0.8803
Epoch 18/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3419 - accur
acy: 0.8812
Epoch 19/100
10420/10420 [=====] - 22s 2ms/sample - loss: 0.3378 - accur
acy: 0.8810
Epoch 20/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3366 - accur
acy: 0.8810
Epoch 21/100
10420/10420 [=====] - 20s 2ms/sample - loss: 0.3356 - accur
acy: 0.8803
Epoch 22/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3335 - accur
acy: 0.8821
Epoch 23/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3335 - accur
acy: 0.8813
Epoch 24/100
10420/10420 [=====] - 22s 2ms/sample - loss: 0.3326 - accur
acy: 0.8817
Epoch 25/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3293 - accur
acy: 0.8824
Epoch 26/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3323 - accur
acy: 0.8819
```

```
Epoch 27/100
10420/10420 [=====] - 22s 2ms/sample - loss: 0.3309 - accur
acy: 0.8821
Epoch 28/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3280 - accur
acy: 0.8829
Epoch 29/100
10420/10420 [=====] - 22s 2ms/sample - loss: 0.3301 - accur
acy: 0.8820
Epoch 30/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3269 - accur
acy: 0.8826
Epoch 31/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3282 - accur
acy: 0.8829
Epoch 32/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3304 - accur
acy: 0.8819
Epoch 33/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3261 - accur
acy: 0.8820
Epoch 34/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3280 - accur
acy: 0.8824
Epoch 35/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3268 - accur
acy: 0.8825
Epoch 36/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3275 - accur
acy: 0.8826
Epoch 37/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3256 - accur
acy: 0.8835
Epoch 38/100
10420/10420 [=====] - 22s 2ms/sample - loss: 0.3217 - accur
acy: 0.8827
Epoch 39/100
10420/10420 [=====] - 22s 2ms/sample - loss: 0.3262 - accur
acy: 0.8825
Epoch 40/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3235 - accur
acy: 0.8824
Epoch 41/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3246 - accur
acy: 0.8822
Epoch 42/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3243 - accur
acy: 0.8830
Epoch 43/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3243 - accur
acy: 0.8828
Epoch 44/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3231 - accur
acy: 0.8834
Epoch 45/100
10420/10420 [=====] - 22s 2ms/sample - loss: 0.3212 - accur
acy: 0.8829
Epoch 46/100
10420/10420 [=====] - 22s 2ms/sample - loss: 0.3227 - accur
acy: 0.8831
Epoch 47/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3224 - accur
acy: 0.8832
Epoch 48/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3219 - accur
acy: 0.8833
Epoch 49/100
10420/10420 [=====] - 20s 2ms/sample - loss: 0.3207 - accur
acy: 0.8833
```

```
Epoch 50/100
10420/10420 [=====] - 22s 2ms/sample - loss: 0.3212 - accur
acy: 0.8828
Epoch 51/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3222 - accur
acy: 0.8830
Epoch 52/100
10420/10420 [=====] - 22s 2ms/sample - loss: 0.3187 - accur
acy: 0.8841
Epoch 53/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3212 - accur
acy: 0.8826
Epoch 54/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3202 - accur
acy: 0.8838
Epoch 55/100
10420/10420 [=====] - 23s 2ms/sample - loss: 0.3175 - accur
acy: 0.8833
Epoch 56/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3196 - accur
acy: 0.8833
Epoch 57/100
10420/10420 [=====] - 22s 2ms/sample - loss: 0.3161 - accur
acy: 0.8836
Epoch 58/100
10420/10420 [=====] - 22s 2ms/sample - loss: 0.3197 - accur
acy: 0.8827
Epoch 59/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3215 - accur
acy: 0.8824
Epoch 60/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3172 - accur
acy: 0.8833
Epoch 61/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3184 - accur
acy: 0.8834
Epoch 62/100
10420/10420 [=====] - 22s 2ms/sample - loss: 0.3157 - accur
acy: 0.8837
Epoch 63/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3184 - accur
acy: 0.8837
Epoch 64/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3191 - accur
acy: 0.8819
Epoch 65/100
10420/10420 [=====] - 22s 2ms/sample - loss: 0.3194 - accur
acy: 0.8835
Epoch 66/100
10420/10420 [=====] - 22s 2ms/sample - loss: 0.3197 - accur
acy: 0.8833
Epoch 67/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3170 - accur
acy: 0.8832
Epoch 68/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3144 - accur
acy: 0.8836
Epoch 69/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3175 - accur
acy: 0.8830
Epoch 70/100
10420/10420 [=====] - 22s 2ms/sample - loss: 0.3171 - accur
acy: 0.8823
Epoch 71/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3164 - accur
acy: 0.8830
Epoch 72/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3165 - accur
acy: 0.8833
```

```
Epoch 73/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3181 - accur
acy: 0.8829
Epoch 74/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3153 - accur
acy: 0.8839
Epoch 75/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3165 - accur
acy: 0.8837
Epoch 76/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3154 - accur
acy: 0.8832
Epoch 77/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3160 - accur
acy: 0.8832
Epoch 78/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3124 - accur
acy: 0.8835
Epoch 79/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3133 - accur
acy: 0.8837
Epoch 80/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3133 - accur
acy: 0.8845
Epoch 81/100
10420/10420 [=====] - 22s 2ms/sample - loss: 0.3142 - accur
acy: 0.8837
Epoch 82/100
10420/10420 [=====] - 22s 2ms/sample - loss: 0.3140 - accur
acy: 0.8846
Epoch 83/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3132 - accur
acy: 0.8824
Epoch 84/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3123 - accur
acy: 0.8839
Epoch 85/100
10420/10420 [=====] - 22s 2ms/sample - loss: 0.3142 - accur
acy: 0.8831
Epoch 86/100
10420/10420 [=====] - 22s 2ms/sample - loss: 0.3132 - accur
acy: 0.8839
Epoch 87/100
10420/10420 [=====] - 22s 2ms/sample - loss: 0.3148 - accur
acy: 0.8829
Epoch 88/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3126 - accur
acy: 0.8831
Epoch 89/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3126 - accur
acy: 0.8843
Epoch 90/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3153 - accur
acy: 0.8831
Epoch 91/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3125 - accur
acy: 0.8843
Epoch 92/100
10420/10420 [=====] - 22s 2ms/sample - loss: 0.3125 - accur
acy: 0.8831
Epoch 93/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3120 - accur
acy: 0.8840
Epoch 94/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3158 - accur
acy: 0.8838
Epoch 95/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3124 - accur
acy: 0.8829
```

```

Epoch 96/100
10420/10420 [=====] - 22s 2ms/sample - loss: 0.3107 - accur
acy: 0.8839
Epoch 97/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3134 - accur
acy: 0.8839
Epoch 98/100
10420/10420 [=====] - 20s 2ms/sample - loss: 0.3095 - accur
acy: 0.8838
Epoch 99/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3125 - accur
acy: 0.8836
Epoch 100/100
10420/10420 [=====] - 20s 2ms/sample - loss: 0.3124 - accur
acy: 0.8841
2605/2605 [=====] - 0s 65us/sample - loss: 0.3213 - accurac
y: 0.8841
Test accuracy: 0.8840691

```

```

In [41]: Y_pred = clf_cnn_1.predict(X_test)
          #print(Y_pred)
          y_pred = np.where(Y_pred>0.5, 1, 0)

          unique_y_pred, counts_y_pred = np.unique(y_pred, return_counts=True)
          y_pred_static = dict(zip(unique_y_pred, counts_y_pred))
          #print(y_pred_static)

          #confusion matrix(y_true, y_pred)
          tn_cnn, fp_cnn, fn_cnn, tp_cnn = confusion_matrix(y_test, y_pred).ravel()

          #train result
          train_accuracy= clf_cnn_1.evaluate(X_train, y_train)[1].round(3)*100
          #test result
          accuracy = (tp_cnn+tn_cnn)/(tp_cnn+tn_cnn+fp_cnn+fn_cnn).round(3)*100
          precision = tp_cnn/(tp_cnn+fp_cnn).round(3)*100
          recall = tp_cnn/(tp_cnn+fn_cnn).round(3)*100
          Train_Test_Difference = abs(accuracy-train_accuracy).round(3)

          print('Train_Accuracy_cnn_3 = ', train_accuracy.round(3))
          print('Test_Accuracy_cnn_3 = ', accuracy.round(3))
          print('Train-Test Difference = ', Train_Test_Difference)
          print()
          print('Precision_cnn_3 = ', precision.round(3))
          print('Recall_cnn_3 = ', recall.round(3))
          print()

          #append result to 'precision_result_list' & 'recall_result_list'
          training_title_list.append('5_CNN_3')
          train_accuracy_f1_result_list.append(train_accuracy)
          test_accuracy_f1_result_list.append(accuracy)
          train_test_difference_result_list.append(Train_Test_Difference)
          precision_result_list.append(precision)
          recall_result_list.append(recall)

          #Classification report
          print()
          target_names = ['state: good', 'state: bad']
          print(classification_report(y_test, y_pred, target_names=target_names))

          10420/10420 [=====] - 0s 44us/sample - loss: 0.3305 - accur
          acy: 0.8859

```

Train_Accuracy_cnn_3 = 88.6
 Test_Accuracy_cnn_3 = 88.407
 Train-Test Difference = 0.193

Precision_cnn_3 = 97.248
 Recall_cnn_3 = 26.173

	precision	recall	f1-score	support
state: good	0.88	1.00	0.94	2200
state: bad	0.97	0.26	0.41	405
accuracy			0.88	2605
macro avg	0.93	0.63	0.67	2605
weighted avg	0.89	0.88	0.85	2605

5_CNN_4 (with features: [abs_mean_h, abs_mean_v, stand_h, stand_v])

```
In [42]: from sklearn.metrics import classification_report
from pandas import read_csv
from keras.models import Sequential
from keras.layers import Dense
from keras.wrappers.scikit_learn import KerasClassifier
from sklearn.model_selection import cross_val_score
from sklearn.preprocessing import LabelEncoder
from sklearn.model_selection import StratifiedKFold
from sklearn.pipeline import Pipeline
from sklearn.metrics import f1_score
```

```
In [43]: #1# split into 80:20
X=df_features_1
y=df_label

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_stat

#2# build the model
clf_cnn_1 = keras.Sequential([
    keras.layers.Flatten(input_shape=(4,)),
    keras.layers.Dense(16, activation=tf.nn.relu),
    keras.layers.Dense(16, activation=tf.nn.relu),
    keras.layers.Dense(16, activation=tf.nn.relu),
    keras.layers.Dense(16, activation=tf.nn.relu),
    keras.layers.Dense(1, activation=tf.nn.sigmoid),
])

#3# compile model
clf_cnn_1.compile(optimizer='adam',
                  loss='binary_crossentropy',
                  metrics=['accuracy'])

#4# training
clf_cnn_1.fit(X_train, y_train, epochs=100, batch_size=1)

test_loss, test_acc = clf_cnn_1.evaluate(X_test, y_test)
print('Test accuracy:', test_acc)
```

Train on 10420 samples

Epoch 1/100

10420/10420 [=====] - 21s 2ms/sample - loss: 0.3911 - accuracy: 0.8629

```
Epoch 2/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3803 - accur
acy: 0.8697
Epoch 3/100
10420/10420 [=====] - 20s 2ms/sample - loss: 0.3767 - accur
acy: 0.8669
Epoch 4/100
10420/10420 [=====] - 20s 2ms/sample - loss: 0.3732 - accur
acy: 0.8696
Epoch 5/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3707 - accur
acy: 0.8715
Epoch 6/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3692 - accur
acy: 0.8702
Epoch 7/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3650 - accur
acy: 0.8717
Epoch 8/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3618 - accur
acy: 0.8736
Epoch 9/100
10420/10420 [=====] - 20s 2ms/sample - loss: 0.3605 - accur
acy: 0.8740
Epoch 10/100
10420/10420 [=====] - 20s 2ms/sample - loss: 0.3564 - accur
acy: 0.8760
Epoch 11/100
10420/10420 [=====] - 20s 2ms/sample - loss: 0.3535 - accur
acy: 0.8755
Epoch 12/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3499 - accur
acy: 0.8772
Epoch 13/100
10420/10420 [=====] - 22s 2ms/sample - loss: 0.3469 - accur
acy: 0.8779
Epoch 14/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3426 - accur
acy: 0.8776
Epoch 15/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3380 - accur
acy: 0.8794
Epoch 16/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3405 - accur
acy: 0.8784
Epoch 17/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3373 - accur
acy: 0.8793
Epoch 18/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3356 - accur
acy: 0.8808
Epoch 19/100
10420/10420 [=====] - 20s 2ms/sample - loss: 0.3346 - accur
acy: 0.8805
Epoch 20/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3332 - accur
acy: 0.8804
Epoch 21/100
10420/10420 [=====] - 20s 2ms/sample - loss: 0.3320 - accur
acy: 0.8798
Epoch 22/100
10420/10420 [=====] - 20s 2ms/sample - loss: 0.3302 - accur
acy: 0.8810
Epoch 23/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3297 - accur
acy: 0.8815
Epoch 24/100
10420/10420 [=====] - 23s 2ms/sample - loss: 0.3307 - accur
acy: 0.8802
```



```
Epoch 25/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3303 - accur
acy: 0.8807
Epoch 26/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3290 - accur
acy: 0.8801
Epoch 27/100
10420/10420 [=====] - 22s 2ms/sample - loss: 0.3300 - accur
acy: 0.8813
Epoch 28/100
10420/10420 [=====] - 22s 2ms/sample - loss: 0.3318 - accur
acy: 0.8805
Epoch 29/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3304 - accur
acy: 0.8801
Epoch 30/100
10420/10420 [=====] - 20s 2ms/sample - loss: 0.3292 - accur
acy: 0.8812
Epoch 31/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3262 - accur
acy: 0.8818
Epoch 32/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3257 - accur
acy: 0.8816
Epoch 33/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3246 - accur
acy: 0.8825
Epoch 34/100
10420/10420 [=====] - 20s 2ms/sample - loss: 0.3260 - accur
acy: 0.8806
Epoch 35/100
10420/10420 [=====] - 20s 2ms/sample - loss: 0.3251 - accur
acy: 0.8816
Epoch 36/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3241 - accur
acy: 0.8820
Epoch 37/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3247 - accur
acy: 0.8813
Epoch 38/100
10420/10420 [=====] - 20s 2ms/sample - loss: 0.3238 - accur
acy: 0.8815
Epoch 39/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3223 - accur
acy: 0.8814
Epoch 40/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3247 - accur
acy: 0.8809
Epoch 41/100
10420/10420 [=====] - 22s 2ms/sample - loss: 0.3244 - accur
acy: 0.8806
Epoch 42/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3227 - accur
acy: 0.8821
Epoch 43/100
10420/10420 [=====] - 20s 2ms/sample - loss: 0.3217 - accur
acy: 0.8834
Epoch 44/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3228 - accur
acy: 0.8815
Epoch 45/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3204 - accur
acy: 0.8831
Epoch 46/100
10420/10420 [=====] - 20s 2ms/sample - loss: 0.3200 - accur
acy: 0.8818
Epoch 47/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3228 - accur
acy: 0.8817
```

```
Epoch 48/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3201 - accur
acy: 0.8818
Epoch 49/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3196 - accur
acy: 0.8820
Epoch 50/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3189 - accur
acy: 0.8842
Epoch 51/100
10420/10420 [=====] - 20s 2ms/sample - loss: 0.3200 - accur
acy: 0.8822
Epoch 52/100
10420/10420 [=====] - 20s 2ms/sample - loss: 0.3190 - accur
acy: 0.8828
Epoch 53/100
10420/10420 [=====] - 20s 2ms/sample - loss: 0.3206 - accur
acy: 0.8834
Epoch 54/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3203 - accur
acy: 0.8829
Epoch 55/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3205 - accur
acy: 0.8829
Epoch 56/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3203 - accur
acy: 0.8821
Epoch 57/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3185 - accur
acy: 0.8831
Epoch 58/100
10420/10420 [=====] - 20s 2ms/sample - loss: 0.3215 - accur
acy: 0.8845
Epoch 59/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3170 - accur
acy: 0.8827
Epoch 60/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3175 - accur
acy: 0.8845
Epoch 61/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3183 - accur
acy: 0.8827
Epoch 62/100
10420/10420 [=====] - 20s 2ms/sample - loss: 0.3173 - accur
acy: 0.8841
Epoch 63/100
10420/10420 [=====] - 20s 2ms/sample - loss: 0.3203 - accur
acy: 0.8832
Epoch 64/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3177 - accur
acy: 0.8840
Epoch 65/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3206 - accur
acy: 0.8834
Epoch 66/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3173 - accur
acy: 0.8852
Epoch 67/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3158 - accur
acy: 0.8846
Epoch 68/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3186 - accur
acy: 0.8829
Epoch 69/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3145 - accur
acy: 0.8832
Epoch 70/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3192 - accur
acy: 0.8837
```

```
Epoch 71/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3227 - accur
acy: 0.8841
Epoch 72/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3167 - accur
acy: 0.8849
Epoch 73/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3168 - accur
acy: 0.8845
Epoch 74/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3143 - accur
acy: 0.8838
Epoch 75/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3171 - accur
acy: 0.8849
Epoch 76/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3177 - accur
acy: 0.8846
Epoch 77/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3169 - accur
acy: 0.8846
Epoch 78/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3148 - accur
acy: 0.8845
Epoch 79/100
10420/10420 [=====] - 20s 2ms/sample - loss: 0.3151 - accur
acy: 0.8841
Epoch 80/100
10420/10420 [=====] - 20s 2ms/sample - loss: 0.3145 - accur
acy: 0.8845
Epoch 81/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3145 - accur
acy: 0.8843
Epoch 82/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3143 - accur
acy: 0.8840
Epoch 83/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3206 - accur
acy: 0.8842
Epoch 84/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3124 - accur
acy: 0.8860
Epoch 85/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3159 - accur
acy: 0.8835
Epoch 86/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3134 - accur
acy: 0.8845
Epoch 87/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3158 - accur
acy: 0.8849
Epoch 88/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3146 - accur
acy: 0.8834
Epoch 89/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3121 - accur
acy: 0.8851
Epoch 90/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3157 - accur
acy: 0.8837
Epoch 91/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3142 - accur
acy: 0.8841
Epoch 92/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3143 - accur
acy: 0.8835
Epoch 93/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3139 - accur
acy: 0.8839
```

```

Epoch 94/100
10420/10420 [=====] - 20s 2ms/sample - loss: 0.3123 - accur
acy: 0.8857
Epoch 95/100
10420/10420 [=====] - 20s 2ms/sample - loss: 0.3135 - accur
acy: 0.8841
Epoch 96/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3133 - accur
acy: 0.8844
Epoch 97/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3115 - accur
acy: 0.8845
Epoch 98/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3126 - accur
acy: 0.8843
Epoch 99/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3114 - accur
acy: 0.8849
Epoch 100/100
10420/10420 [=====] - 21s 2ms/sample - loss: 0.3127 - accur
acy: 0.8846
2605/2605 [=====] - 0s 67us/sample - loss: 0.2949 - accurac
y: 0.8887
Test accuracy: 0.88867563

```

```

In [44]: Y_pred = clf_cnn_1.predict(X_test)
          #print(Y_pred)
          y_pred = np.where(Y_pred>0.5, 1, 0)

          unique_y_pred, counts_y_pred = np.unique(y_pred, return_counts=True)
          y_pred_static = dict(zip(unique_y_pred, counts_y_pred))
          #print(y_pred_static)

          #confusion matrix(y_true, y_pred)
          tn_cnn, fp_cnn, fn_cnn, tp_cnn = confusion_matrix(y_test, y_pred).ravel()

          #train result
          train_accuracy= clf_cnn_1.evaluate(X_train, y_train)[1].round(3)*100
          #test result
          accuracy = (tp_cnn+tn_cnn)/(tp_cnn+tn_cnn+fp_cnn+fn_cnn).round(3)*100
          precision = tp_cnn/(tp_cnn+fp_cnn).round(3)*100
          recall = tp_cnn/(tp_cnn+fn_cnn).round(3)*100
          Train_Test_Difference = abs(accuracy-train_accuracy).round(3)

          print('Train_Accuracy_cnn_4 = ', train_accuracy.round(3))
          print('Test_Accuracy_cnn_4 = ', accuracy.round(3))
          print('Train-Test Difference = ', Train_Test_Difference)
          print()
          print('Precision_cnn_4 = ', precision.round(3))
          print('Recall_cnn_4 = ', recall.round(3))
          print()

          #append result to 'precision_result_list' & 'recall_result_list'
          training_title_list.append('5_CNN_4')
          train_accuracy_f1_result_list.append(train_accuracy)
          test_accuracy_f1_result_list.append(accuracy)
          train_test_difference_result_list.append(Train_Test_Difference)
          precision_result_list.append(precision)
          recall_result_list.append(recall)

          #Classification report

```

```
print()
target_names = ['state: good', 'state: bad']
print(classification_report(y_test, y_pred, target_names=target_names))
```

```
10420/10420 [=====] - 0s 42us/sample - loss: 0.3012 - accur
acy: 0.8877
```

```
Train_Accuracy_cnn_4 = 88.8
Test_Accuracy_cnn_4 = 88.868
Train-Test Difference = 0.068
```

```
Precision_cnn_4 = 91.971
Recall_cnn_4 = 31.111
```

	precision	recall	f1-score	support
state: good	0.89	0.99	0.94	2200
state: bad	0.92	0.31	0.46	405
accuracy			0.89	2605
macro avg	0.90	0.65	0.70	2605
weighted avg	0.89	0.89	0.86	2605

5_CNN_5 (with features: [mean_h, mean_v, stand_h, stand_v, max_h, min_h, max_v, min_v])

```
In [45]: from sklearn.metrics import classification_report
from pandas import read_csv
from keras.models import Sequential
from keras.layers import Dense
from keras.wrappers.scikit_learn import KerasClassifier
from sklearn.model_selection import cross_val_score
from sklearn.preprocessing import LabelEncoder
from sklearn.model_selection import StratifiedKFold
from sklearn.pipeline import Pipeline
from sklearn.metrics import f1_score
```

```
In [49]: #1# split into 80:20
X=df_features_2
y=df_label

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_stat

#2# build the model
clf_cnn_1 = keras.Sequential([
    keras.layers.Flatten(input_shape=(8,)),
    keras.layers.Dense(32, activation=tf.nn.relu),
    keras.layers.Dense(32, activation=tf.nn.relu),
    keras.layers.Dense(1, activation=tf.nn.sigmoid),
])

#3# compile model
clf_cnn_1.compile(optimizer='adam',
                  loss='binary_crossentropy',
                  metrics=['accuracy'])

#4# training
clf_cnn_1.fit(X_train, y_train, epochs=50, batch_size=1)

test_loss, test_acc = clf_cnn_1.evaluate(X_test, y_test)
print('Test accuracy:', test_acc)
```

Train on 10420 samples

Epoch 1/50

10420/10420 [=====] - 18s 2ms/sample - loss: 0.3992 - accuracy: 0.8607

Epoch 2/50

10420/10420 [=====] - 18s 2ms/sample - loss: 0.3779 - accuracy: 0.8635

Epoch 3/50

10420/10420 [=====] - 18s 2ms/sample - loss: 0.3718 - accuracy: 0.8661

Epoch 4/50

10420/10420 [=====] - 18s 2ms/sample - loss: 0.3651 - accuracy: 0.8676

Epoch 5/50

10420/10420 [=====] - 17s 2ms/sample - loss: 0.3583 - accuracy: 0.8669

Epoch 6/50

10420/10420 [=====] - 17s 2ms/sample - loss: 0.3531 - accuracy: 0.8670

Epoch 7/50

10420/10420 [=====] - 17s 2ms/sample - loss: 0.3483 - accuracy: 0.8699

Epoch 8/50

10420/10420 [=====] - 17s 2ms/sample - loss: 0.3459 - accuracy: 0.8697

Epoch 9/50

10420/10420 [=====] - 18s 2ms/sample - loss: 0.3422 - accuracy: 0.8707

Epoch 10/50

10420/10420 [=====] - 18s 2ms/sample - loss: 0.3424 - accuracy: 0.8710

Epoch 11/50

10420/10420 [=====] - 18s 2ms/sample - loss: 0.3373 - accuracy: 0.8713

Epoch 12/50

10420/10420 [=====] - 17s 2ms/sample - loss: 0.3377 - accuracy: 0.8726

Epoch 13/50

10420/10420 [=====] - 17s 2ms/sample - loss: 0.3369 - accuracy: 0.8727

Epoch 14/50

10420/10420 [=====] - 18s 2ms/sample - loss: 0.3348 - accuracy: 0.8731

Epoch 15/50

10420/10420 [=====] - 18s 2ms/sample - loss: 0.3365 - accuracy: 0.8751

Epoch 16/50

10420/10420 [=====] - 19s 2ms/sample - loss: 0.3342 - accuracy: 0.8753

Epoch 17/50

10420/10420 [=====] - 20s 2ms/sample - loss: 0.3321 - accuracy: 0.8775

Epoch 18/50

10420/10420 [=====] - 17s 2ms/sample - loss: 0.3348 - accuracy: 0.8764

Epoch 19/50

10420/10420 [=====] - 17s 2ms/sample - loss: 0.3328 - accuracy: 0.8755

Epoch 20/50

10420/10420 [=====] - 18s 2ms/sample - loss: 0.3327 - accuracy: 0.8771

Epoch 21/50

10420/10420 [=====] - 17s 2ms/sample - loss: 0.3314 - accuracy: 0.8767

Epoch 22/50

10420/10420 [=====] - 17s 2ms/sample - loss: 0.3339 - accuracy: 0.8750

Epoch 23/50

10420/10420 [=====] - 17s 2ms/sample - loss: 0.3297 - accuracy: 0.8750

```
acy: 0.8773
Epoch 24/50
10420/10420 [=====] - 17s 2ms/sample - loss: 0.3310 - accur
acy: 0.8778
Epoch 25/50
10420/10420 [=====] - 17s 2ms/sample - loss: 0.3315 - accur
acy: 0.8772
Epoch 26/50
10420/10420 [=====] - 17s 2ms/sample - loss: 0.3337 - accur
acy: 0.8766
Epoch 27/50
10420/10420 [=====] - 17s 2ms/sample - loss: 0.3315 - accur
acy: 0.8760
Epoch 28/50
10420/10420 [=====] - 16s 2ms/sample - loss: 0.3308 - accur
acy: 0.8778
Epoch 29/50
10420/10420 [=====] - 17s 2ms/sample - loss: 0.3296 - accur
acy: 0.8775
Epoch 30/50
10420/10420 [=====] - 17s 2ms/sample - loss: 0.3308 - accur
acy: 0.8771
Epoch 31/50
10420/10420 [=====] - 17s 2ms/sample - loss: 0.3277 - accur
acy: 0.8789
Epoch 32/50
10420/10420 [=====] - 18s 2ms/sample - loss: 0.3274 - accur
acy: 0.8782
Epoch 33/50
10420/10420 [=====] - 17s 2ms/sample - loss: 0.3267 - accur
acy: 0.8770
Epoch 34/50
10420/10420 [=====] - 17s 2ms/sample - loss: 0.3258 - accur
acy: 0.8774
Epoch 35/50
10420/10420 [=====] - 18s 2ms/sample - loss: 0.3289 - accur
acy: 0.8781
Epoch 36/50
10420/10420 [=====] - 18s 2ms/sample - loss: 0.3263 - accur
acy: 0.8783
Epoch 37/50
10420/10420 [=====] - 17s 2ms/sample - loss: 0.3276 - accur
acy: 0.8774
Epoch 38/50
10420/10420 [=====] - 19s 2ms/sample - loss: 0.3253 - accur
acy: 0.8778
Epoch 39/50
10420/10420 [=====] - 18s 2ms/sample - loss: 0.3258 - accur
acy: 0.8779
Epoch 40/50
10420/10420 [=====] - 17s 2ms/sample - loss: 0.3283 - accur
acy: 0.8772
Epoch 41/50
10420/10420 [=====] - 16s 2ms/sample - loss: 0.3240 - accur
acy: 0.8781
Epoch 42/50
10420/10420 [=====] - 17s 2ms/sample - loss: 0.3266 - accur
acy: 0.8784
Epoch 43/50
10420/10420 [=====] - 17s 2ms/sample - loss: 0.3273 - accur
acy: 0.8774
Epoch 44/50
10420/10420 [=====] - 17s 2ms/sample - loss: 0.3267 - accur
acy: 0.8779
Epoch 45/50
10420/10420 [=====] - 18s 2ms/sample - loss: 0.3248 - accur
acy: 0.8786
Epoch 46/50
10420/10420 [=====] - 19s 2ms/sample - loss: 0.3219 - accur
```

```

acy: 0.8781
Epoch 47/50
10420/10420 [=====] - 17s 2ms/sample - loss: 0.3254 - accur
acy: 0.8793
Epoch 48/50
10420/10420 [=====] - 18s 2ms/sample - loss: 0.3245 - accur
acy: 0.8771
Epoch 49/50
10420/10420 [=====] - 17s 2ms/sample - loss: 0.3232 - accur
acy: 0.8794
Epoch 50/50
10420/10420 [=====] - 17s 2ms/sample - loss: 0.3268 - accur
acy: 0.8788
2605/2605 [=====] - 0s 60us/sample - loss: 0.3257 - accurac
y: 0.8752
Test accuracy: 0.8752399

```

```

In [50]: Y_pred = clf_cnn_1.predict(X_test)
         #print(Y_pred)
         y_pred = np.where(Y_pred>0.5, 1, 0)

         unique_y_pred, counts_y_pred = np.unique(y_pred, return_counts=True)
         y_pred_static = dict(zip(unique_y_pred, counts_y_pred))
         #print(y_pred_static)

         #confusion matrix(y_true, y_pred)
         tn_cnn, fp_cnn, fn_cnn, tp_cnn = confusion_matrix(y_test, y_pred).ravel()

         #train result
         train_accuracy= clf_cnn_1.evaluate(X_train, y_train)[1].round(3)*100
         #test result
         accuracy = (tp_cnn+tn_cnn)/(tp_cnn+tn_cnn+fp_cnn+fn_cnn).round(3)*100
         precision = tp_cnn/(tp_cnn+fp_cnn).round(3)*100
         recall = tp_cnn/(tp_cnn+fn_cnn).round(3)*100
         Train_Test_Difference = abs(accuracy-train_accuracy).round(3)

         print('Train_Accuracy_cnn_5 = ', train_accuracy.round(3))
         print('Test_Accuracy_cnn_5 = ', accuracy.round(3))
         print('Train-Test Difference = ', Train_Test_Difference)
         print()
         print('Precision_cnn_5 = ', precision.round(3))
         print('Recall_cnn_5 = ', recall.round(3))
         print()

         #append result to 'precision_result_list' & 'recall_result_list'
         training_title_list.append('5_CNN_5')
         train_accuracy_f1_result_list.append(train_accuracy)
         test_accuracy_f1_result_list.append(accuracy)
         train_test_difference_result_list.append(Train_Test_Difference)
         precision_result_list.append(precision)
         recall_result_list.append(recall)

         #Classification report
         print()
         target_names = ['state: good', 'state: bad']
         print(classification_report(y_test, y_pred, target_names=target_names))

10420/10420 [=====] - 0s 35us/sample - loss: 0.3169 - accur
acy: 0.8793
Train_Accuracy_cnn_5 = 87.9
Test_Accuracy_cnn_5 = 87.524

```


Train-Test Difference = 0.376

Precision_cnn_5 = 78.571

Recall_cnn_5 = 27.16

	precision	recall	f1-score	support
state: good	0.88	0.99	0.93	2200
state: bad	0.79	0.27	0.40	405
accuracy			0.88	2605
macro avg	0.83	0.63	0.67	2605
weighted avg	0.87	0.88	0.85	2605

5_CNN_6 (with features: [mean_h, mean_v, stand_h, stand_v, max_h, min_h, max_v, min_v])

```
In [ ]: from sklearn.metrics import classification_report
from pandas import read_csv
from keras.models import Sequential
from keras.layers import Dense
from keras.wrappers.scikit_learn import KerasClassifier
from sklearn.model_selection import cross_val_score
from sklearn.preprocessing import LabelEncoder
from sklearn.model_selection import StratifiedKFold
from sklearn.pipeline import Pipeline
from sklearn.metrics import f1_score
```

```
In [ ]: #1# split into 80:20
X=df_features_2
y=df_label

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_stat

#2# build the model
clf_cnn_1 = keras.Sequential([
    keras.layers.Flatten(input_shape=(8,)),
    keras.layers.Dense(32, activation=tf.nn.relu),
    keras.layers.Dense(32, activation=tf.nn.relu),
    keras.layers.Dense(32, activation=tf.nn.relu),
    keras.layers.Dense(32, activation=tf.nn.relu),
    keras.layers.Dense(32, activation=tf.nn.relu),
    keras.layers.Dense(32, activation=tf.nn.relu),
    keras.layers.Dense(32, activation=tf.nn.relu),
    keras.layers.Dense(1, activation=tf.nn.sigmoid),
])

#3# compile model
clf_cnn_1.compile(optimizer='adam',
                  loss='binary_crossentropy',
                  metrics=['accuracy'])

#4# training
clf_cnn_1.fit(X_train, y_train, epochs=100, batch_size=1)

test_loss, test_acc = clf_cnn_1.evaluate(X_test, y_test)
print('Test accuracy:', test_acc)
```

```

In [ ]: Y_pred = clf_cnn_1.predict(X_test)
        #print(Y_pred)
        y_pred = np.where(Y_pred>0.5, 1, 0)

        unique_y_pred, counts_y_pred = np.unique(y_pred, return_counts=True)
        y_pred_static = dict(zip(unique_y_pred, counts_y_pred))
        #print(y_pred_static)

        #confusion matrix(y_true, y_pred)
        tn_cnn, fp_cnn, fn_cnn, tp_cnn = confusion_matrix(y_test, y_pred).ravel()

        #train result
        train_accuracy= clf_cnn_1.evaluate(X_train, y_train)[1].round(3)*100
        #test result
        accuracy = (tp_cnn+tn_cnn)/(tp_cnn+tn_cnn+fp_cnn+fn_cnn).round(3)*100
        precision = tp_cnn/(tp_cnn+fp_cnn).round(3)*100
        recall = tp_cnn/(tp_cnn+fn_cnn).round(3)*100
        Train_Test_Difference = abs(accuracy-train_accuracy).round(3)

        print('Train_Accuracy_cnn_7 = ', train_accuracy.round(3))
        print('Test_Accuracy_cnn_7 = ', accuracy.round(3))
        print('Train-Test Difference = ', Train_Test_Difference)
        print()
        print('Precision_cnn_7 = ', precision.round(3))
        print('Recall_cnn_7 = ', recall.round(3))
        print()

        #append result to 'precision_result_list' & 'recall_result_list'
        training_title_list.append('5_CNN_7')
        train_accuracy_f1_result_list.append(train_accuracy)
        test_accuracy_f1_result_list.append(accuracy)
        train_test_difference_result_list.append(Train_Test_Difference)
        precision_result_list.append(precision)
        recall_result_list.append(recall)

        #Classification report
        print()
        target_names = ['state: good', 'state: bad']
        print(classification_report(y_test, y_pred, target_names=target_names))

```

Step15_0

Run Correlation Matrix --> evaluate the relationships btw features & state

```

In [ ]: corrmatrix = training_df.corr()
        sns.heatmap(corrmatrix, annot=True, annot_kws= {'size':7}, square=True)

```

```

In [ ]: # index in matrix
        corrdat = df_features_n.corr()
        corrdat
        corrdat.index

```

Step15_1

Create a func --> to find which features have most related relationship with 'state'

```
In [ ]: def getCorrelatedFeature(Corrdata, threshold):
        feature = []
        value = []

        for i, index in enumerate(Corrdata.index):
            if abs(Corrdata[index])>threshold:
                feature.append(index)
                print(index)
                value.append(Corrdata[index])
        df = pd.DataFrame(data = value, index = feature, columns = ['corr value'])
        return df
```

```
In [ ]: threshold = 0.29
        corr_value = getCorrelatedFeature(corrdat['state'],threshold)
```

- According to the Correlation Matrix Map, above are the features most correlated with 'state'
- Will keep above features in the dataset, and filter out the rest features in the following step

```
In [ ]: # check correlation value of each feature
        corr_value
```

Step15_2

Generate a new dataset(df_features_3) with most important features based on the Correlation Matrix Result

```
In [ ]: #Generate new dataset based on CorrelationMatrix
        df_features_3 = training_df.drop(columns=['file', 'state', 'abs_mean_h', 'mean_h',
        df_label      = training_df['state']
        #####

        print('df_features_3= ', df_features_3.columns)
```

Step16

5_CNN_7 (with features: ['abs_mean_v', 'stand_v', 'max_h', 'min_h', 'max_v', 'min_v'])

- here with use new dataset: df_features_3

```
In [ ]: from sklearn.metrics import classification_report
        from pandas import read_csv
        from keras.models import Sequential
        from keras.layers import Dense
        from keras.wrappers.scikit_learn import KerasClassifier
        from sklearn.model_selection import cross_val_score
        from sklearn.preprocessing import LabelEncoder
        from sklearn.model_selection import StratifiedKFold
        from sklearn.pipeline import Pipeline
        from sklearn.metrics import f1_score
```

```
In [ ]: #1# split into 80:20
        X=df_features_3
        y=df_label

        X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_stat
```

```

#2# build the model
clf_cnn_1 = keras.Sequential([
    keras.layers.Flatten(input_shape=(6,)),
    keras.layers.Dense(32, activation=tf.nn.relu),
    keras.layers.Dense(32, activation=tf.nn.relu),
    keras.layers.Dense(32, activation=tf.nn.relu),
    keras.layers.Dense(32, activation=tf.nn.relu),
    keras.layers.Dense(32, activation=tf.nn.relu),
    keras.layers.Dense(32, activation=tf.nn.relu),
    keras.layers.Dense(32, activation=tf.nn.relu),
    keras.layers.Dense(1, activation=tf.nn.sigmoid),
])

#3# compile model
clf_cnn_1.compile(optimizer='adam',
                  loss='binary_crossentropy',
                  metrics=['accuracy'])

#4# training
clf_cnn_1.fit(X_train, y_train, epochs=100, batch_size=1)

test_loss, test_acc = clf_cnn_1.evaluate(X_test, y_test)
print('Test accuracy:', test_acc)

```

```

In [ ]: Y_pred = clf_cnn_1.predict(X_test)
        #print(Y_pred)
        y_pred = np.where(Y_pred>0.5, 1, 0)

        unique_y_pred, counts_y_pred = np.unique(y_pred, return_counts=True)
        y_pred_static = dict(zip(unique_y_pred, counts_y_pred))
        #print(y_pred_static)

        #confusion matrix(y_true, y_pred)
        tn_cnn, fp_cnn, fn_cnn, tp_cnn = confusion_matrix(y_test, y_pred).ravel()

        #train result
        train_accuracy= clf_cnn_1.evaluate(X_train, y_train)[1].round(3)*100
        #test result
        accuracy = (tp_cnn+tn_cnn)/(tp_cnn+tn_cnn+fp_cnn+fn_cnn).round(3)*100
        precision = tp_cnn/(tp_cnn+fp_cnn).round(3)*100
        recall = tp_cnn/(tp_cnn+fn_cnn).round(3)*100
        Train_Test_Difference = abs(accuracy-train_accuracy).round(3)

        print('Train_Accuracy_cnn_8 = ', train_accuracy.round(3))
        print('Test_Accuracy_cnn_8 = ', accuracy.round(3))
        print('Train-Test Difference = ', Train_Test_Difference)
        print()
        print('Precision_cnn_8 = ', precision.round(3))
        print('Recall_cnn_8 = ', recall.round(3))
        print()

        #append result to 'precision_result_list' & 'recall_result_list'
        training_title_list.append('5_CNN_8')
        train_accuracy_f1_result_list.append(train_accuracy)
        test_accuracy_f1_result_list.append(accuracy)
        train_test_difference_result_list.append(Train_Test_Difference)
        precision_result_list.append(precision)
        recall_result_list.append(recall)

```

```
#Classification report
print()
target_names = ['state: good', 'state: bad']
print(classification_report(y_test, y_pred, target_names=target_names))
```

In []:

Step17 Display Results

```
In [ ]: #result list
print(training_title_list)
print(train_accuracy_f1_result_list)
print(test_accuracy_f1_result_list)
print(train_test_difference_result_list)
print(precision_result_list)
print(recall_result_list)
```

Result Dataframe

Comparisions

- 1_SVM_0 (with features: ['mean_h', 'mean_v', 'stand_h', 'stand_v'])
- 1_SVM_1 (with features: ['abs_mean_h', 'abs_mean_v', 'stand_h', 'stand_v'])
- 1_SVM_2 (with features: ['mean_h', 'mean_v', 'stand_h', 'stand_v', 'max_h', 'min_h', 'max_v', 'min_v'])
- 2_RandomForest_0 (with features: ['mean_h', 'mean_v', 'stand_h', 'stand_v'])
- 2_RandomForest_1 (with features: ['abs_mean_h', 'abs_mean_v', 'stand_h', 'stand_v'])
- 3_GradientBoosting_0 (with features: ['mean_h', 'mean_v', 'stand_h', 'stand_v'])
- 3_GradientBoosting_1 (with features: ['abs_mean_h', 'abs_mean_v', 'stand_h', 'stand_v'])
- 4_K-NearestNeighbors_0 (with features: ['mean_h', 'mean_v', 'stand_h', 'stand_v'])
- 4_K-NearestNeighbors_1 (with features: ['abs_mean_h', 'abs_mean_v', 'stand_h', 'stand_v'])
- 5_CNN_0 (with features: [mean_h, mean_v, stand_h, stand_v]) (2 hidden layers, epochs=50)
- 5_CNN_1 (with features: [mean_h, mean_v, stand_h, stand_v]) (4 hidden layers, epochs=50)
- 5_CNN_2 (with features: [abs_mean_h, abs_mean_v, stand_h, stand_v]) (2 hidden layers, epochs=50)
- 5_CNN_3 (with features: [abs_mean_h, abs_mean_v, stand_h, stand_v]) (4 hidden layers, epochs=100)
- 5_CNN_4 (with features: [abs_mean_h, abs_mean_v, stand_h, stand_v]) (8 hidden layers, epochs=100)
- 5_CNN_5 (with features: ['mean_h', 'mean_v', 'stand_h', 'stand_v', 'max_h', 'min_h', 'max_v', 'min_v']) (2 hidden layers, epochs=50)

- 5_CNN_6 (with features: ['mean_h', 'mean_v', 'stand_h', 'stand_v', 'max_h', 'min_h', 'max_v', 'min_v']) (8 hidden layers, epochs=100)
- 5_CNN_7 (with features: ['abs_mean_v', 'stand_v', 'max_h', 'min_h', 'max_v', 'min_v']) (8 hidden layers, epochs=100)

Step18 Summary

The following results are gathered from each result of each algorithm.

```
In [5]: result = [(87.8, 87.3, 0.5, 92.9, 19.6),
                  (87.8, 87.6, 0.2, 98.8, 20.2),
                  (88.7, 88.1, 0.6, 90.7, 26.4),
                  (88.2, 87.7, 0.5, 85.0, 24.0),
                  (88.5, 87.8, 0.7, 85.8, 24.2),
                  (88.8, 87.5, 1.3, 78.6, 25.8),
                  (89.3, 87.6, 1.7, 73.6, 30.0),
                  (89.2, 87.6, 1.6, 78.1, 28.2),
                  (90.2, 88.8, 1.4, 83.4, 34.9),
                  (87.8, 87.754, 0.046, 94.79, 22.46),
                  (88.2, 88.177, 0.023, 97.08, 24.69),
                  (88.3, 88.100, 0.2, 92.79, 25.43),
                  (88.6, 88.407, 0.193, 97.25, 26.17),
                  (88.8, 88.868, 0.068, 91.97, 31.11),
                  (87.9, 87.524, 0.376, 78.57, 27.16)]

columns = ['train_acc(%)', 'test_acc(%)', 'train-test diff(%)', 'precision(%)', 'rec

indexs = ['1_SMM_0',
          '1_SVM_1',
          '1_SVM_2',
          '2_RF_0',
          '2_RF_1',
          '3_GB_0',
          '3_GB_1',
          '4_k_0',
          '4_k_1',
          '5_CNN_0',
          '5_CNN_1',
          '5_CNN_2',
          '5_CNN_3',
          '5_CNN_4',
          '5_CNN_5']
```

```
In [10]: result_df = pd.DataFrame(result,
                                   columns=['train_acc(%)',
                                           'test_acc(%)',
                                           'train-test diff(%)',
                                           'precision(%)',
                                           'recall(%)'],
                                   index = ['1_SMM_0',
                                           '1_SVM_1',
                                           '1_SVM_2',
                                           '2_RF_0',
                                           '2_RF_1',
                                           '3_GB_0',
                                           '3_GB_1',
                                           '4_k_0',
```

```
'4_k_1',
'5_CNN_0',
'5_CNN_1',
'5_CNN_2',
'5_CNN_3',
'5_CNN_4',
'5_CNN_5']])
```

```
In [11]: print(result_df)
```

	train_acc(%)	test_acc(%)	train-test diff(%)	precision(%)	\
1_SMM_0	87.8	87.300	0.500	92.90	
1_SVM_1	87.8	87.600	0.200	98.80	
1_SVM_2	88.7	88.100	0.600	90.70	
2_RF_0	88.2	87.700	0.500	85.00	
2_RF_1	88.5	87.800	0.700	85.80	
3_GB_0	88.8	87.500	1.300	78.60	
3_GB_1	89.3	87.600	1.700	73.60	
4_k_0	89.2	87.600	1.600	78.10	
4_k_1	90.2	88.800	1.400	83.40	
5_CNN_0	87.8	87.754	0.046	94.79	
5_CNN_1	88.2	88.177	0.023	97.08	
5_CNN_2	88.3	88.100	0.200	92.79	
5_CNN_3	88.6	88.407	0.193	97.25	
5_CNN_4	88.8	88.868	0.068	91.97	
5_CNN_5	87.9	87.524	0.376	78.57	

	recall(%)
1_SMM_0	19.60
1_SVM_1	20.20
1_SVM_2	26.40
2_RF_0	24.00
2_RF_1	24.20
3_GB_0	25.80
3_GB_1	30.00
4_k_0	28.20
4_k_1	34.90
5_CNN_0	22.46
5_CNN_1	24.69
5_CNN_2	25.43
5_CNN_3	26.17
5_CNN_4	31.11
5_CNN_5	27.16

Summary

- From the results, most of the trained classifier have no overfitting issue. And most of the difference of 'Train-Test Accuracy' are less than 1%. Only the 'Train-Test Accuracy' from Gradient Boosting and K-nearest Neighbors are more than 1%.
- CNN_5 is trained with 8 features. Compared to previous Model, the features are increased from 4 to 8. With only 2 hidden layers and 50 training epoch. We can see the test accuracy is decreasing a bit. To improve the accuracy of the model, maybe can try to increase the hidden layers and the node in each layer and also increase the training epoch.
- To this case, CNN are slightly better than the other algorithm. I believe, by increase the hidden layer and training epoch, CNN can achieve better result.
- Because the training time is too long, so i have not run through CNN_6 and CNN_7.

```
In [ ]:
```