

Blend Tree Models

Import

```
In [1]: import pandas as pd
from pycaret.classification import *
import time
from DB_scripts import rnd_bln_split_CSV as shf
```

Settings

```
In [2]: # set constants
target_label = 'tuple'
learning_model = ['rf', 'et', 'lightgbm', 'xgboost']
num_features = ['min_packet_size', 'min_fpkt', 'min_bpkt']
file_name = "all_features_"
path = target_label + "_dataset/"
model = []
```

```
In [3]: # function for making model-prediction over the data set and measure the run time
def timed_prediction(in_data,in_model):
    t = time.process_time()
    predicted = predict_model(in_model, data=in_data)
    elapsed_time = time.process_time() - t
    print("prediction took: " + str(elapsed_time))
    return predicted
```

```
In [4]: # function for checkign the correction of the model-prediction over the data
def check_correction(in_predicted):
    count=0
    index = in_predicted.index
    number_of_rows = len(index)
    for i in range(0,number_of_rows):
        if str(int(in_predicted.iloc[i][target_label])) != str(int(in_predicted.iloc[i]['Label'])):
            #print("prediction not matched in Line " + str(i) + " as " + str(in_predicted.iloc[i]['app']) + "!=" + str(in_predicted.iloc[i]['Label']))
            count=count+1
    print("number of error: " + str(count) + " from " + str(number_of_rows) +
" test samples \n which is " + str(count/number_of_rows) + " precent of erro
r.")
```

```
In [5]: # compare answers and labeled test
def compare_prediction_with_answers(in_predicted, in_answers):
    count=0
    index = in_predicted.index
    number_of_rows = len(index)
    for i in range(0,number_of_rows):
        if str(in_answers.iloc[i]) != str(int(in_predicted.iloc[i]['Label'])):
            count=count+1
            # print the unmatched answers
            #print("answer os and test Label are not matched in Line " + str(i) +
            " as " + str(answers.iloc[i]['os']) + "!=" + str(predict_test.iloc[i]['Label']))
            print("number of error: " + str(count) + " from " + str(number_of_rows) +
            " test samples \n which is " + str(count/number_of_rows) + " precent of erro
r.")
```

```
In [6]: # activating balanced random data shuffling
shf.split_CSV_randomly_balanced(target_label,file_name)
```

Read Data

the data set "../app_dataset/all_features_app.csv" is the main resource and has been split to train and test.

the prepared "all_features_app_train" is a random split of the "all_features_app" data set and got 70% of the data. the data set CSV format separate columns by TAB ('\t'), and the [1] line is the columns index line, so skipped.

```
In [7]: data = pd.read_csv(path+file_name+target_label+'_train.csv',
                      sep='\t',
                      skiprows=[1])
```

In [8]: `data.info()`

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 14442 entries, 0 to 14441
Data columns (total 70 columns):
 #   Column           Non-Null Count Dtype  
--- 
 0   fSSL_session_id_len    14442 non-null float64
 1   fSSL_num_extensions   14442 non-null float64
 2   fSSL_num_compression_methods 14442 non-null float64
 3   SYN_tcp_scale        14442 non-null float64
 4   SYN_MSS              14442 non-null float64
 5   SYN_tcp_winsize      14442 non-null float64
 6   fcipher_suites_1     14442 non-null float64
 7   fcipher_suites_2     14442 non-null float64
 8   fcipher_suites_3     14442 non-null float64
 9   fSSLv_1               14442 non-null float64
 10  fSSLv_2               14442 non-null float64
 11  fSSLv_3               14442 non-null float64
 12  fSSLv_4               14442 non-null float64
 13  size_histogram_1     14442 non-null float64
 14  size_histogram_2     14442 non-null float64
 15  size_histogram_3     14442 non-null float64
 16  size_histogram_4     14442 non-null float64
 17  size_histogram_5     14442 non-null float64
 18  size_histogram_6     14442 non-null float64
 19  size_histogram_7     14442 non-null float64
 20  size_histogram_8     14442 non-null float64
 21  size_histogram_9     14442 non-null float64
 22  size_histogram_10    14442 non-null float64
 23  fpeak_features_1    14442 non-null float64
 24  fpeak_features_2    14442 non-null float64
 25  fpeak_features_3    14442 non-null float64
 26  fpeak_features_4    14442 non-null float64
 27  fpeak_features_5    14442 non-null float64
 28  fpeak_features_6    14442 non-null float64
 29  fpeak_features_7    14442 non-null float64
 30  fpeak_features_8    14442 non-null float64
 31  fpeak_features_9    14442 non-null float64
 32  bpeak_features_1    14442 non-null float64
 33  bpeak_features_2    14442 non-null float64
 34  bpeak_features_3    14442 non-null float64
 35  bpeak_features_4    14442 non-null float64
 36  bpeak_features_5    14442 non-null float64
 37  bpeak_features_6    14442 non-null float64
 38  bpeak_features_7    14442 non-null float64
 39  bpeak_features_8    14442 non-null float64
 40  bpeak_features_9    14442 non-null float64
 41  packet_count        14442 non-null float64
 42  min_packet_size     14442 non-null float64
 43  max_packet_size     14442 non-null float64
 44  mean_packet_size    14442 non-null float64
 45  sizevar             14442 non-null float64
 46  std_fiat            14442 non-null float64
 47  fpackets            14442 non-null float64
 48  bpackets            14442 non-null float64
 49  fbytes              14442 non-null float64
 50  bbytes              14442 non-null float64
 51  min_fiat            14442 non-null float64

```

```

et_unseen_test
52 min_biat           14442 non-null float64
53 max_fiat           14442 non-null float64
54 max_biat           14442 non-null float64
55 std_biat           14442 non-null float64
56 mean_fiat          14442 non-null float64
57 mean_biat          14442 non-null float64
58 min_fpkt           14442 non-null float64
59 min_bpkt           14442 non-null float64
60 max_fpkt           14442 non-null float64
61 max_bpkt           14442 non-null float64
62 std_fpkt           14442 non-null float64
63 std_bpkt           14442 non-null float64
64 mean_fpkt          14442 non-null float64
65 mean_bpkt          14442 non-null float64
66 mean_fttl_1         14442 non-null float64
67 mean_fttl_2         14442 non-null float64
68 num_keep_alive      14442 non-null float64
69 tuple               14442 non-null int64
dtypes: float64(69), int64(1)
memory usage: 7.7 MB

```

In [9]: `data[target_label].unique()`

Out[9]: `array([16102, 14602, 15602, 18102, 13102, 11602, 16201, 13201, 18201,
 16101, 13101, 18101, 16202, 18202, 13202, 16302, 18302, 13302,
 13403, 18403, 17403, 14601, 17101, 12101, 17201, 12201, 16403,
 16103, 13103, 18103], dtype=int64)`

Setup Classifier

which splits the data into train and test for model building

```
In [10]: setup(data=data,
             target=target_label,
             numeric_features=num_features,
             silent=True)
# train_size=0.99
```

Setup Successfully Completed!

	Description	Value
0	session_id	5730
1	Target Type	Multiclass
2	Label Encoded	None
3	Original Data	(14442, 70)
4	Missing Values	False
5	Numeric Features	61
6	Categorical Features	8
7	Ordinal Features	False
8	High Cardinality Features	False
9	High Cardinality Method	None
10	Sampled Data	(14442, 70)
11	Transformed Train Set	(10109, 77)
12	Transformed Test Set	(4333, 77)
13	Numeric Imputer	mean
14	Categorical Imputer	constant
15	Normalize	False
16	Normalize Method	None
17	Transformation	False
18	Transformation Method	None
19	PCA	False
20	PCA Method	None
21	PCA Components	None
22	Ignore Low Variance	False
23	Combine Rare Levels	False
24	Rare Level Threshold	None
25	Numeric Binning	False
26	Remove Outliers	False
27	Outliers Threshold	None
28	Remove Multicollinearity	False
29	Multicollinearity Threshold	None
30	Clustering	False
31	Clustering Iteration	None
32	Polynomial Features	False
33	Polynomial Degree	None
34	Trigonometry Features	False

	Description	Value
35	Polynomial Threshold	None
36	Group Features	False
37	Feature Selection	False
38	Features Selection Threshold	None
39	Feature Interaction	False
40	Feature Ratio	False
41	Interaction Threshold	None
42	Fix Imbalance	False
43	Fix Imbalance Method	SMOTE

```
Out[10]: (   fSSL_num_extensions  fSSL_num_compression_methods  SYN_tcp_scale \
0               13.0                      1.0                8.0
1               11.0                      1.0                8.0
2               11.0                      1.0                8.0
3               13.0                      1.0                8.0
4               13.0                      1.0                8.0
...
14437            ...                      ...                ...
14438            ...                      ...                ...
14439            ...                      ...                ...
14440            ...                      ...                ...
14441            ...                      ...                ...

      SYN_MSS  SYN_tcp_winsize  fSSLv_1  fSSLv_3  size_histogram_1 \
0       1460.0          8192.0    0.0    0.0        67.0
1       1460.0          8192.0    0.0    0.0       270.0
2       1460.0          8192.0    0.0    0.0       199.0
3       1460.0          8192.0    0.0    0.0        43.0
4       1460.0          8192.0    0.0    0.0        34.0
...
14437            ...            ...    ...    ...
14438            ...            ...    ...    ...
14439            ...            ...    ...    ...
14440            ...            ...    ...    ...
14441            ...            ...    ...    ...

      size_histogram_2  size_histogram_3  ...  fcipher_suites_3_0.0 \
0           5.0          3.0    ...        1.0
1           4.0          11.0   ...       0.0
2           4.0           8.0   ...       0.0
3           4.0           4.0   ...       1.0
4           4.0           1.0   ...       1.0
...
14437            ...            ...    ...    ...
14438            ...            ...    ...    ...
14439            ...            ...    ...    ...
14440            ...            ...    ...    ...
14441            ...            ...    ...    ...

      fcipher_suites_3_1.0  fSSLv_2_0.0  fSSLv_2_1.0  fSSLv_4_0.0 \
0             0.0          0.0        1.0        1.0
1             1.0          0.0        1.0        1.0
2             1.0          0.0        1.0        1.0
3             0.0          0.0        1.0        1.0
4             0.0          0.0        1.0        1.0
...
14437            ...            ...    ...    ...
14438            ...            ...    ...    ...
14439            ...            ...    ...    ...
14440            ...            ...    ...    ...
14441            ...            ...    ...    ...

      fSSLv_4_1.0  mean_fttl_1_0.0  mean_fttl_1_1.0  mean_fttl_2_0.0 \
0             0.0            1.0            0.0            0.0
1             0.0            1.0            0.0            0.0
2             0.0            1.0            0.0            0.0
3             0.0            1.0            0.0            0.0
```

4	0.0	1.0	0.0	0.0
...
14437	0.0	0.0	1.0	1.0
14438	0.0	0.0	1.0	1.0
14439	0.0	0.0	1.0	1.0
14440	0.0	0.0	1.0	1.0
14441	0.0	0.0	1.0	1.0

mean_fttl_2_1.0	
0	1.0
1	1.0
2	1.0
3	1.0
4	1.0
...	...
14437	0.0
14438	0.0
14439	0.0
14440	0.0
14441	0.0

[14442 rows x 77 columns],

0	16102
1	16102
2	16102
3	16102
4	16102
...	...
14437	13103
14438	18103
14439	18103
14440	18103
14441	18103

Name: tuple, Length: 14442, dtype: int64,

	fSSL_num_extensions	fSSL_num_compression_methods	SYN_tcp_scale	\
11238	9.0		1.0	8.0
1362	7.0		1.0	8.0
12492	5.0		1.0	5.0
11586	7.0		1.0	8.0
12856	5.0		1.0	5.0
...
7768	7.0		1.0	8.0
3488	11.0		1.0	7.0
8285	7.0		1.0	8.0
8546	9.0		1.0	8.0
2292	11.0		1.0	8.0

	SYN_MSS	SYN_tcp_winsize	fSSLv_1	fSSLv_3	size_histogram_1	\
11238	1460.0	65535.0	0.0	0.0	137.0	
1362	1460.0	65535.0	0.0	0.0	7.0	
12492	1460.0	65535.0	0.0	0.0	54.0	
11586	1460.0	65535.0	0.0	0.0	137.0	
12856	1460.0	65535.0	0.0	0.0	128.0	
...	
7768	1460.0	65535.0	0.0	0.0	138.0	
3488	1460.0	29200.0	0.0	0.0	31.0	
8285	1460.0	65535.0	0.0	0.0	96.0	

et_unseen_test					
8546	1460.0	65535.0	0.0	0.0	130.0
2292	1460.0	8192.0	0.0	0.0	37.0
11238	size_histogram_2	size_histogram_3	...	fcipher_suites_3_0.0	\
1362	7.0	21.0	...		0.0
12492	3.0	0.0	...		0.0
11586	1.0	1.0	...		1.0
12856	7.0	10.0	...		0.0
...	4.0	1.0	...		1.0
7768
3488	7.0	8.0	...		0.0
8285	3.0	3.0	...		1.0
8546	8.0	21.0	...		0.0
2292	6.0	17.0	...		0.0
	4.0	1.0	...		1.0
11238	fcipher_suites_3_1.0	fSSLv_2_0.0	fSSLv_2_1.0	fSSLv_4_0.0	\
1362	1.0	1.0	0.0		0.0
12492	1.0	1.0	0.0		0.0
11586	0.0	0.0	1.0		1.0
12856	1.0	1.0	0.0		0.0
...	0.0	0.0	1.0		0.0
7768
3488	1.0	1.0	0.0		0.0
8285	0.0	0.0	1.0		1.0
8546	1.0	1.0	0.0		0.0
2292	1.0	0.0	1.0		1.0
11238	fSSLv_4_1.0	mean_fttl_1_0.0	mean_fttl_1_1.0	mean_fttl_2_0.0	\
1362	1.0	1.0	0.0		0.0
12492	1.0	1.0	0.0		0.0
11586	0.0	0.0	1.0		1.0
12856	1.0	1.0	0.0		0.0
...	0.0	0.0	1.0		1.0
7768
3488	1.0	1.0	0.0		0.0
8285	0.0	0.0	1.0		1.0
8546	1.0	1.0	0.0		0.0
2292	1.0	1.0	0.0		0.0
11238	mean_fttl_2_1.0				
1362	1.0				
12492	1.0				
11586	0.0				
12856	1.0				
...	0.0				
7768	...				
3488	1.0				
8285	0.0				
8546	1.0				
2292	1.0				
	[10109 rows x 77 columns],				
	fSSL_num_extensions	fSSL_num_compression_methods	SYN_tcp_scale	\	

2175	7.0		1.0	8.0
11800	7.0		1.0	8.0
12095	7.0		1.0	8.0
1713	7.0		1.0	8.0
8268	7.0		1.0	8.0
...
12121	7.0		1.0	8.0
1135	7.0		1.0	8.0
304	13.0		1.0	8.0
13326	4.0		1.0	5.0
1165	7.0		1.0	8.0

	SYN_MSS	SYN_tcp_winsize	fSSLv_1	fSSLv_3	size_histogram_1	\
2175	1460.0	8192.0	0.0	0.0	11.0	
11800	1460.0	65535.0	0.0	0.0	137.0	
12095	1460.0	65535.0	0.0	0.0	140.0	
1713	1460.0	8192.0	0.0	0.0	9.0	
8268	1460.0	65535.0	0.0	0.0	44.0	
...	
12121	1460.0	65535.0	0.0	0.0	67.0	
1135	1460.0	65535.0	0.0	0.0	8.0	
304	1460.0	8192.0	0.0	0.0	42.0	
13326	1460.0	65535.0	0.0	0.0	1602.0	
1165	1460.0	8192.0	0.0	0.0	51.0	

	size_histogram_2	size_histogram_3	...	fcipher_suites_3_0.0	\
2175	3.0	0.0	...	1.0	
11800	4.0	11.0	...	0.0	
12095	7.0	17.0	...	0.0	
1713	1.0	1.0	...	0.0	
8268	5.0	3.0	...	0.0	
...	
12121	4.0	8.0	...	0.0	
1135	2.0	1.0	...	0.0	
304	3.0	0.0	...	1.0	
13326	3.0	9.0	...	1.0	
1165	9.0	4.0	...	0.0	

	fcipher_suites_3_1.0	fSSLv_2_0.0	fSSLv_2_1.0	fSSLv_4_0.0	\
2175	0.0	1.0	0.0	0.0	
11800	1.0	1.0	0.0	0.0	
12095	1.0	1.0	0.0	0.0	
1713	1.0	1.0	0.0	0.0	
8268	1.0	1.0	0.0	0.0	
...	
12121	1.0	1.0	0.0	0.0	
1135	1.0	1.0	0.0	0.0	
304	0.0	0.0	1.0	1.0	
13326	0.0	0.0	1.0	1.0	
1165	1.0	1.0	0.0	0.0	

	fSSLv_4_1.0	mean_fttl_1_0.0	mean_fttl_1_1.0	mean_fttl_2_0.0	\
2175	1.0	1.0	0.0	0.0	
11800	1.0	1.0	0.0	0.0	
12095	1.0	1.0	0.0	0.0	
1713	1.0	1.0	0.0	0.0	
8268	1.0	1.0	0.0	0.0	

```

...
12121      1.0      1.0      0.0      0.0
1135       1.0      1.0      0.0      0.0
304        0.0      1.0      0.0      0.0
13326      0.0      0.0      1.0      1.0
1165       1.0      1.0      0.0      0.0

mean_fttl_2_1.0
2175       1.0
11800      1.0
12095      1.0
1713       1.0
8268       1.0
...
12121      1.0
1135       1.0
304        1.0
13326      0.0
1165       1.0

[4333 rows x 77 columns],
11238      16302
1362       14602
12492      13403
11586      16302
12856      13403
...
7768       16302
3488       13201
8285       16302
8546       16302
2292       18102
Name: tuple, Length: 10109, dtype: int64,
2175       15602
11800      16302
12095      16302
1713       14602
8268       16302
...
12121      16302
1135       14602
304        16102
13326      17403
1165       14602
Name: tuple, Length: 4333, dtype: int64,
5730,
Pipeline(memory=None,
          steps=[('dtypes',
                  DataTypes_Auto_infer(categorical_features=[],
                                        display_types=False, features_todrop=
          []),
          ml_usecase='regression',
          numerical_features=['min_packet_size',
                               'min_fpkt',
                               'min_bpkt'],
          target='tuple', time_features=[])),
          ('imputer',

```

```

et_unseen_test
Simple_Imputer(categorical_strategy='not_available',
                 numeric_strategy='mean',
                 target_variable=None))...
('group', Empty()), ('nonliner', Empty()), ('scaling', Empty
()),
('P_transform', Empty()), ('pt_target', Empty()),
('binn', Empty()), ('rem_outliers', Empty()),
('cluster_all', Empty()), ('dummy', Dummify(target='tuple
e')),
('fix_perfect', Empty()), ('clean_names', Clean_Colum_Names
()),
('feature_select', Empty()), ('fix_multi', Empty()),
('dfs', Empty()), ('pca', Empty())],
verbose=False),
[('Classification Setup Config',
  [
    ('session_id', 5730),
    ('Target Type', Multiclass),
    ('Label Encoded', None),
    ('Original Data', (14442, 70)),
    ('Missing Values', False),
    ('Numeric Features', 61),
    ('Categorical Features', 8),
    ('Ordinal Features', False),
    ('High Cardinality Features', False),
    ('High Cardinality Method', None),
    ('Sampled Data', (14442, 70)),
    ('Transformed Train Set', (10109, 77)),
    ('Transformed Test Set', (4333, 77)),
    ('Numeric Imputer', mean),
    ('Categorical Imputer', constant),
    ('Normalize', False),
    ('Normalize Method', None),
    ('Transformation', False),
    ('Transformation Method', None),
    ('PCA', False),
    ('PCA Method', None),
    ('PCA Components', None),
    ('Ignore Low Variance', False),
    ('Combine Rare Levels', False),
    ('Rare Level Threshold', None),
    ('Numeric Binning', False),
    ('Remove Outliers', False),
    ('Outliers Threshold', None),
    ('Remove Multicollinearity', False),
    ('Multicollinearity Threshold', None),
    ('Clustering', False),
    ('Clustering Iteration', None),
    ('Polynomial Features', False),
    ('Polynomial Degree', None),
    ('Trigonometry Features', False),
    ('Polynomial Threshold', None),
    ('Group Features', False),
    ('Feature Selection', False),
    ('Features Selection Threshold', None),
    ('Feature Interaction', False),
    ('Feature Ratio', False)
  ]
)

```

```

41      Interaction Threshold      None
42      Fix Imbalance        False
43      Fix Imbalance Method    SMOTE),
('X_training Set',
     fSSL_num_extensions  fSSL_num_compression_methods  SYN_tcp_scale  \
11238          9.0                  1.0                8.0
1362          7.0                  1.0                8.0
12492          5.0                  1.0                5.0
11586          7.0                  1.0                8.0
12856          5.0                  1.0                5.0
...
7768          7.0                  1.0                8.0
3488          11.0                 1.0                7.0
8285          7.0                  1.0                8.0
8546          9.0                  1.0                8.0
2292          11.0                 1.0                8.0

     SYN_MSS  SYN_tcp_winsize  fSSLv_1   fSSLv_3  size_histogram_1  \
11238  1460.0       65535.0     0.0      0.0      137.0
1362   1460.0       65535.0     0.0      0.0       7.0
12492  1460.0       65535.0     0.0      0.0      54.0
11586  1460.0       65535.0     0.0      0.0      137.0
12856  1460.0       65535.0     0.0      0.0      128.0
...
7768   1460.0       65535.0     0.0      0.0      138.0
3488   1460.0       29200.0    0.0      0.0       31.0
8285   1460.0       65535.0     0.0      0.0       96.0
8546   1460.0       65535.0     0.0      0.0      130.0
2292   1460.0       8192.0      0.0      0.0       37.0

     size_histogram_2  size_histogram_3  ...  fcipher_suites_3_0.0  \
11238          7.0          21.0  ...          0.0
1362           3.0          0.0  ...          0.0
12492          1.0          1.0  ...          1.0
11586          7.0          10.0  ...          0.0
12856          4.0          1.0  ...          1.0
...
7768           7.0          8.0  ...          0.0
3488           3.0          3.0  ...          1.0
8285           8.0          21.0  ...          0.0
8546           6.0          17.0  ...          0.0
2292           4.0          1.0  ...          1.0

     fcipher_suites_3_1.0  fSSLv_2_0.0  fSSLv_2_1.0  fSSLv_4_0.0  \
11238            1.0          1.0          0.0          0.0
1362            1.0          1.0          0.0          0.0
12492            0.0          0.0          1.0          1.0
11586            1.0          1.0          0.0          0.0
12856            0.0          0.0          1.0          1.0
...
7768            1.0          1.0          0.0          0.0
3488            0.0          0.0          1.0          1.0
8285            1.0          1.0          0.0          0.0
8546            1.0          1.0          0.0          0.0
2292            0.0          0.0          1.0          1.0

     fSSLv_4_1.0  mean_fttl_1_0.0  mean_fttl_1_1.0  mean_fttl_2_0.0  \

```

11238	1.0	1.0	0.0	0.0
1362	1.0	1.0	0.0	0.0
12492	0.0	0.0	1.0	1.0
11586	1.0	1.0	0.0	0.0
12856	0.0	0.0	1.0	1.0
...
7768	1.0	1.0	0.0	0.0
3488	0.0	0.0	1.0	1.0
8285	1.0	1.0	0.0	0.0
8546	1.0	1.0	0.0	0.0
2292	0.0	1.0	0.0	0.0

mean_fttl_2_1.0

11238	1.0
1362	1.0
12492	0.0
11586	1.0
12856	0.0
...	...
7768	1.0
3488	0.0
8285	1.0
8546	1.0
2292	1.0

[10109 rows x 77 columns]),
('y_training Set',

11238	16302
1362	14602
12492	13403
11586	16302
12856	13403
...	
7768	16302
3488	13201
8285	16302
8546	16302
2292	18102

Name: tuple, Length: 10109, dtype: int64),
('X_test Set',

	fSSL_num_extensions	fSSL_num_compression_methods	SYN_tcp_scale	\
2175	7.0	1.0	8.0	
11800	7.0	1.0	8.0	
12095	7.0	1.0	8.0	
1713	7.0	1.0	8.0	
8268	7.0	1.0	8.0	
...
12121	7.0	1.0	8.0	
1135	7.0	1.0	8.0	
304	13.0	1.0	8.0	
13326	4.0	1.0	5.0	
1165	7.0	1.0	8.0	

	SYN_MSS	SYN_tcp_winsize	fSSLv_1	fSSLv_3	size_histogram_1	\
2175	1460.0	8192.0	0.0	0.0	11.0	
11800	1460.0	65535.0	0.0	0.0	137.0	
12095	1460.0	65535.0	0.0	0.0	140.0	

1713	1460.0	8192.0	0.0	0.0	9.0
8268	1460.0	65535.0	0.0	0.0	44.0
...
12121	1460.0	65535.0	0.0	0.0	67.0
1135	1460.0	65535.0	0.0	0.0	8.0
304	1460.0	8192.0	0.0	0.0	42.0
13326	1460.0	65535.0	0.0	0.0	1602.0
1165	1460.0	8192.0	0.0	0.0	51.0
 size_histogram_2 size_histogram_3 ... fcipher_suites_3_0.0 \					
2175	3.0	0.0	...		1.0
11800	4.0	11.0	...		0.0
12095	7.0	17.0	...		0.0
1713	1.0	1.0	...		0.0
8268	5.0	3.0	...		0.0
...
12121	4.0	8.0	...		0.0
1135	2.0	1.0	...		0.0
304	3.0	0.0	...		1.0
13326	3.0	9.0	...		1.0
1165	9.0	4.0	...		0.0
 fcipher_suites_3_1.0 fSSLv_2_0.0 fSSLv_2_1.0 fSSLv_4_0.0 \					
2175	0.0	1.0	0.0		0.0
11800	1.0	1.0	0.0		0.0
12095	1.0	1.0	0.0		0.0
1713	1.0	1.0	0.0		0.0
8268	1.0	1.0	0.0		0.0
...
12121	1.0	1.0	0.0		0.0
1135	1.0	1.0	0.0		0.0
304	0.0	0.0	1.0		1.0
13326	0.0	0.0	1.0		1.0
1165	1.0	1.0	0.0		0.0
 fSSLv_4_1.0 mean_fttl_1_0.0 mean_fttl_1_1.0 mean_fttl_2_0.0 \					
2175	1.0	1.0	0.0		0.0
11800	1.0	1.0	0.0		0.0
12095	1.0	1.0	0.0		0.0
1713	1.0	1.0	0.0		0.0
8268	1.0	1.0	0.0		0.0
...
12121	1.0	1.0	0.0		0.0
1135	1.0	1.0	0.0		0.0
304	0.0	1.0	0.0		0.0
13326	0.0	0.0	1.0		1.0
1165	1.0	1.0	0.0		0.0
 mean_fttl_2_1.0					
2175	1.0				
11800	1.0				
12095	1.0				
1713	1.0				
8268	1.0				
...	...				
12121	1.0				
1135	1.0				

```

304           1.0
13326         0.0
1165          1.0

[4333 rows x 77 columns]),
('y_test Set',
 2175    15602
 11800   16302
 12095   16302
 1713    14602
 8268    16302
...
 12121   16302
 1135    14602
 304     16102
 13326   17403
 1165    14602
Name: tuple, Length: 4333, dtype: int64),
('Transformation Pipeline',
 Pipeline(memory=None,
           steps=[('dtypes',
                    DataTypes_Auto_infer(categorical_features=[],
                                         display_types=False, features_todrop
= [],
                                         ml_usecase='regression',
                                         numerical_features=['min_packet_siz
e',
                                         'min_fpkt',
                                         'min_bpkt'],
                                         target='tuple', time_features=[])),
          ('imputer',
           Simple_Imputer(categorical_strategy='not_available',
                           numeric_strategy='mean',
                           target_variable=None))...
          ('group', Empty()), ('nonliner', Empty()), ('scaling', Emp
ty()),
          ('P_transform', Empty()), ('pt_target', Empty()),
          ('binn', Empty()), ('rem_outliers', Empty()),
          ('cluster_all', Empty()), ('dummy', Dummify(target='tupl
e')),
          ('fix_perfect', Empty()), ('clean_names', Clean_Colum_Nam
e),
          ('feature_select', Empty()), ('fix_multi', Empty()),
          ('dfs', Empty()), ('pca', Empty())],
           verbose=False)],
 False,
 -1,
 True,
 [],
 [],
 [],
 [],
 'no_logging',
 False,
 False,
 '6571',
 False,

```

None,
<Logger logs (DEBUG)>)

Creating Learning Model

In [11]: `model.append(create_model(learning_model[0]))`

	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC
0	0.9713	0.0000	0.8588	0.9696	0.9691	0.9664	0.9665
1	0.9654	0.0000	0.8790	0.9657	0.9643	0.9595	0.9595
2	0.9644	0.0000	0.8264	0.9673	0.9629	0.9583	0.9584
3	0.9674	0.0000	0.7953	0.9659	0.9648	0.9618	0.9618
4	0.9654	0.0000	0.8049	0.9649	0.9627	0.9594	0.9596
5	0.9674	0.0000	0.8235	0.9662	0.9647	0.9617	0.9618
6	0.9703	0.0000	0.8281	0.9694	0.9683	0.9653	0.9654
7	0.9674	0.0000	0.8241	0.9668	0.9635	0.9618	0.9619
8	0.9693	0.0000	0.8619	0.9690	0.9676	0.9641	0.9642
9	0.9703	0.0000	0.8873	0.9706	0.9683	0.9652	0.9653
Mean	0.9679	0.0000	0.8389	0.9675	0.9656	0.9624	0.9624
SD	0.0023	0.0000	0.0294	0.0019	0.0023	0.0027	0.0027

In [12]: `model.append(create_model(learning_model[1]))`

	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC
0	0.9812	0.0000	0.9223	0.9807	0.9802	0.9780	0.9780
1	0.9773	0.0000	0.9252	0.9761	0.9762	0.9734	0.9734
2	0.9733	0.0000	0.9105	0.9739	0.9724	0.9687	0.9688
3	0.9812	0.0000	0.8975	0.9814	0.9801	0.9780	0.9780
4	0.9733	0.0000	0.8763	0.9725	0.9721	0.9687	0.9688
5	0.9782	0.0000	0.9036	0.9781	0.9773	0.9745	0.9746
6	0.9743	0.0000	0.8765	0.9729	0.9726	0.9699	0.9700
7	0.9733	0.0000	0.8376	0.9715	0.9717	0.9687	0.9688
8	0.9713	0.0000	0.8803	0.9694	0.9695	0.9664	0.9665
9	0.9762	0.0000	0.9132	0.9767	0.9757	0.9722	0.9722
Mean	0.9760	0.0000	0.8943	0.9753	0.9748	0.9719	0.9719
SD	0.0033	0.0000	0.0255	0.0038	0.0035	0.0038	0.0038

In [13]: `model.append(create_model(learning_model[2]))`

	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC
0	0.9654	0.0000	0.8362	0.9712	0.9674	0.9595	0.9596
1	0.9357	0.0000	0.8268	0.9715	0.9505	0.9250	0.9260
2	0.9674	0.0000	0.7924	0.9689	0.9669	0.9618	0.9618
3	0.9683	0.0000	0.8299	0.9712	0.9678	0.9630	0.9631
4	0.9614	0.0000	0.7850	0.9725	0.9656	0.9549	0.9550
5	0.8833	0.0000	0.7402	0.9688	0.9145	0.8641	0.8682
6	0.9505	0.0000	0.7538	0.9696	0.9593	0.9423	0.9424
7	0.9575	0.0000	0.7885	0.9698	0.9617	0.9503	0.9505
8	0.8991	0.0000	0.8182	0.9765	0.9334	0.8849	0.8903
9	0.9673	0.0000	0.8453	0.9739	0.9691	0.9618	0.9618
Mean	0.9456	0.0000	0.8016	0.9714	0.9556	0.9368	0.9379
SD	0.0290	0.0000	0.0338	0.0023	0.0172	0.0333	0.0316

In [14]: `model.append(create_model(learning_model[3]))`

	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC
0	0.9832	0.0000	0.9062	0.9817	0.9817	0.9803	0.9803
1	0.9802	0.0000	0.9275	0.9808	0.9799	0.9768	0.9769
2	0.9723	0.0000	0.8591	0.9711	0.9706	0.9676	0.9676
3	0.9802	0.0000	0.8557	0.9803	0.9789	0.9768	0.9769
4	0.9753	0.0000	0.9033	0.9751	0.9747	0.9711	0.9711
5	0.9713	0.0000	0.8818	0.9718	0.9705	0.9664	0.9665
6	0.9753	0.0000	0.8660	0.9744	0.9740	0.9711	0.9711
7	0.9773	0.0000	0.8414	0.9755	0.9751	0.9734	0.9734
8	0.9832	0.0000	0.8742	0.9800	0.9813	0.9803	0.9803
9	0.9772	0.0000	0.9055	0.9766	0.9755	0.9733	0.9734
Mean	0.9775	0.0000	0.8821	0.9767	0.9762	0.9737	0.9737
SD	0.0039	0.0000	0.0262	0.0036	0.0039	0.0046	0.0046

```
In [16]: model = blend_models(model)
```

	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC
0	0.9812	0.0000	0.9062	0.9799	0.9798	0.9780	0.9780
1	0.9822	0.0000	0.9350	0.9813	0.9813	0.9792	0.9792
2	0.9753	0.0000	0.8756	0.9758	0.9742	0.9710	0.9711
3	0.9812	0.0000	0.8640	0.9811	0.9799	0.9780	0.9780
4	0.9753	0.0000	0.8692	0.9742	0.9741	0.9711	0.9711
5	0.9773	0.0000	0.8894	0.9777	0.9762	0.9734	0.9734
6	0.9773	0.0000	0.8726	0.9765	0.9762	0.9734	0.9734
7	0.9763	0.0000	0.8349	0.9757	0.9745	0.9722	0.9723
8	0.9812	0.0000	0.8988	0.9788	0.9796	0.9780	0.9780
9	0.9743	0.0000	0.9013	0.9749	0.9731	0.9699	0.9699
Mean	0.9781	0.0000	0.8847	0.9776	0.9769	0.9744	0.9744
SD	0.0029	0.0000	0.0262	0.0024	0.0028	0.0033	0.0033

Prediction

make a prediction process over the trained data, see validation results.

```
In [17]: predicted = timed_prediction(data,model)
```

prediction took: 32.40625

```
In [18]: check_correction(predicted)
```

number of error: 73 from 14442 test samples
which is 0.00505470156488021 percent of error.

Read Test

Read the unseen test data set

```
In [19]: unseen_data = pd.read_csv(target_label+'dataset/all_features_'+target_label+
'_test.csv',
                                sep='\t',
                                skiprows=[1])
```

```
In [20]: # check for target column values (we got all wanted values...)
unseen_data[target_label].unique()
```

```
Out[20]: array([16102, 14602, 15602, 18102, 13102, 11602, 16201, 13201, 18201,
   16101, 13101, 18101, 16202, 18202, 13202, 16302, 18302, 13302,
   13403, 18403, 17403, 14601, 17101, 12101, 17201, 12201, 16403,
   16103, 13103, 18103], dtype=int64)
```

```
In [21]: # saving the target column
answers = unseen_data[target_label]
```

```
In [22]: # dropping target column from test.
unseen_data = unseen_data.drop(columns=[target_label])
```

In [23]: `unseen_data.info()`

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 6189 entries, 0 to 6188
Data columns (total 69 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   fSSL_session_id_len      6189 non-null   float64
 1   fSSL_num_extensions     6189 non-null   float64
 2   fSSL_num_compression_methods 6189 non-null   float64
 3   SYN_tcp_scale          6189 non-null   float64
 4   SYN_MSS                6189 non-null   float64
 5   SYN_tcp_winsize        6189 non-null   float64
 6   fcipher_suites_1       6189 non-null   float64
 7   fcipher_suites_2       6189 non-null   float64
 8   fcipher_suites_3       6189 non-null   float64
 9   fSSLv_1                 6189 non-null   float64
 10  fSSLv_2                 6189 non-null   float64
 11  fSSLv_3                 6189 non-null   float64
 12  fSSLv_4                 6189 non-null   float64
 13  size_histogram_1        6189 non-null   float64
 14  size_histogram_2        6189 non-null   float64
 15  size_histogram_3        6189 non-null   float64
 16  size_histogram_4        6189 non-null   float64
 17  size_histogram_5        6189 non-null   float64
 18  size_histogram_6        6189 non-null   float64
 19  size_histogram_7        6189 non-null   float64
 20  size_histogram_8        6189 non-null   float64
 21  size_histogram_9        6189 non-null   float64
 22  size_histogram_10       6189 non-null   float64
 23  fpeak_features_1        6189 non-null   float64
 24  fpeak_features_2        6189 non-null   float64
 25  fpeak_features_3        6189 non-null   float64
 26  fpeak_features_4        6189 non-null   float64
 27  fpeak_features_5        6189 non-null   float64
 28  fpeak_features_6        6189 non-null   float64
 29  fpeak_features_7        6189 non-null   float64
 30  fpeak_features_8        6189 non-null   float64
 31  fpeak_features_9        6189 non-null   float64
 32  bpeak_features_1        6189 non-null   float64
 33  bpeak_features_2        6189 non-null   float64
 34  bpeak_features_3        6189 non-null   float64
 35  bpeak_features_4        6189 non-null   float64
 36  bpeak_features_5        6189 non-null   float64
 37  bpeak_features_6        6189 non-null   float64
 38  bpeak_features_7        6189 non-null   float64
 39  bpeak_features_8        6189 non-null   float64
 40  bpeak_features_9        6189 non-null   float64
 41  packet_count            6189 non-null   float64
 42  min_packet_size         6189 non-null   float64
 43  max_packet_size         6189 non-null   float64
 44  mean_packet_size        6189 non-null   float64
 45  sizevar                6189 non-null   float64
 46  std_fiat                6189 non-null   float64
 47  fpackets               6189 non-null   float64
 48  bpackets               6189 non-null   float64
 49  fbytes                 6189 non-null   float64
 50  bbytes                 6189 non-null   float64
 51  min_fiat                6189 non-null   float64
```

```
et_unseen_test
52 min_biat          6189 non-null float64
53 max_fiat          6189 non-null float64
54 max_biat          6189 non-null float64
55 std_biat          6189 non-null float64
56 mean_fiat         6189 non-null float64
57 mean_biat         6189 non-null float64
58 min_fpkt          6189 non-null float64
59 min_bpkt          6189 non-null float64
60 max_fpkt          6189 non-null float64
61 max_bpkt          6189 non-null float64
62 std_fpkt          6189 non-null float64
63 std_bpkt          6189 non-null float64
64 mean_fpkt         6189 non-null float64
65 mean_bpkt         6189 non-null float64
66 mean_fttl_1       6189 non-null float64
67 mean_fttl_2       6189 non-null float64
68 num_keep_alive    6189 non-null float64
dtypes: float64(69)
memory usage: 3.3 MB
```

Make Unseen Test

In [24]: `predicted = predict_model(model, data=unseen_data)`

In [25]: `predicted.head()`

Out[25]:

	fSSL_session_id_len	fSSL_num_extensions	fSSL_num_compression_methods	SYN_tcp_scale
0	32.0	13.0		1.0
1	0.0	12.0		1.0
2	0.0	12.0		1.0
3	32.0	13.0		1.0
4	32.0	13.0		1.0

5 rows × 70 columns

```
In [26]: predicted.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 6189 entries, 0 to 6188
Data columns (total 70 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   fSSL_session_id_len      6189 non-null   float64
 1   fSSL_num_extensions     6189 non-null   float64
 2   fSSL_num_compression_methods 6189 non-null   float64
 3   SYN_tcp_scale          6189 non-null   float64
 4   SYN_MSS                6189 non-null   float64
 5   SYN_tcp_winsize        6189 non-null   float64
 6   fcipher_suites_1       6189 non-null   float64
 7   fcipher_suites_2       6189 non-null   float64
 8   fcipher_suites_3       6189 non-null   float64
 9   fSSLv_1                 6189 non-null   float64
 10  fSSLv_2                 6189 non-null   float64
 11  fSSLv_3                 6189 non-null   float64
 12  fSSLv_4                 6189 non-null   float64
 13  size_histogram_1        6189 non-null   float64
 14  size_histogram_2        6189 non-null   float64
 15  size_histogram_3        6189 non-null   float64
 16  size_histogram_4        6189 non-null   float64
 17  size_histogram_5        6189 non-null   float64
 18  size_histogram_6        6189 non-null   float64
 19  size_histogram_7        6189 non-null   float64
 20  size_histogram_8        6189 non-null   float64
 21  size_histogram_9        6189 non-null   float64
 22  size_histogram_10       6189 non-null   float64
 23  fpeak_features_1        6189 non-null   float64
 24  fpeak_features_2        6189 non-null   float64
 25  fpeak_features_3        6189 non-null   float64
 26  fpeak_features_4        6189 non-null   float64
 27  fpeak_features_5        6189 non-null   float64
 28  fpeak_features_6        6189 non-null   float64
 29  fpeak_features_7        6189 non-null   float64
 30  fpeak_features_8        6189 non-null   float64
 31  fpeak_features_9        6189 non-null   float64
 32  bpeak_features_1        6189 non-null   float64
 33  bpeak_features_2        6189 non-null   float64
 34  bpeak_features_3        6189 non-null   float64
 35  bpeak_features_4        6189 non-null   float64
 36  bpeak_features_5        6189 non-null   float64
 37  bpeak_features_6        6189 non-null   float64
 38  bpeak_features_7        6189 non-null   float64
 39  bpeak_features_8        6189 non-null   float64
 40  bpeak_features_9        6189 non-null   float64
 41  packet_count            6189 non-null   float64
 42  min_packet_size         6189 non-null   float64
 43  max_packet_size         6189 non-null   float64
 44  mean_packet_size        6189 non-null   float64
 45  sizevar                6189 non-null   float64
 46  std_fiat                6189 non-null   float64
 47  fpackets               6189 non-null   float64
 48  bpackets               6189 non-null   float64
 49  fbytes                 6189 non-null   float64
 50  bbytes                 6189 non-null   float64
 51  min_fiat                6189 non-null   float64
```

```

et_unseen_test
52 min_biat          6189 non-null float64
53 max_fiat          6189 non-null float64
54 max_biat          6189 non-null float64
55 std_biat          6189 non-null float64
56 mean_fiat         6189 non-null float64
57 mean_biat         6189 non-null float64
58 min_fpkt          6189 non-null float64
59 min_bpkt          6189 non-null float64
60 max_fpkt          6189 non-null float64
61 max_bpkt          6189 non-null float64
62 std_fpkt          6189 non-null float64
63 std_bpkt          6189 non-null float64
64 mean_fpkt         6189 non-null float64
65 mean_bpkt         6189 non-null float64
66 mean_fttl_1       6189 non-null float64
67 mean_fttl_2       6189 non-null float64
68 num_keep_alive    6189 non-null float64
69 Label              6189 non-null int32
dtypes: float64(69), int32(1)
memory usage: 3.3 MB

```

In [27]: `predicted['Label'].unique()`

Out[27]: `array([16102, 14602, 18102, 15602, 13102, 11602, 16201, 18201, 13201,
18101, 13101, 17201, 12201, 16101, 14601, 16202, 18202, 13202,
16302, 18302, 13302, 13403, 16403, 17403, 18403, 17101, 12101,
16103, 13103, 18103])`

check prediction correction

In [28]: `answers.unique()`

Out[28]: `array([16102, 14602, 15602, 18102, 13102, 11602, 16201, 13201, 18201,
16101, 13101, 18101, 16202, 18202, 13202, 16302, 18302, 13302,
13403, 18403, 17403, 14601, 17101, 12101, 17201, 12201, 16403,
16103, 13103, 18103], dtype=int64)`

In [29]: `compare_prediction_with_answers(predicted, answers)`

number of error: 139 from 6189 test samples
which is 0.022459201809662304 percent of error.

In []: