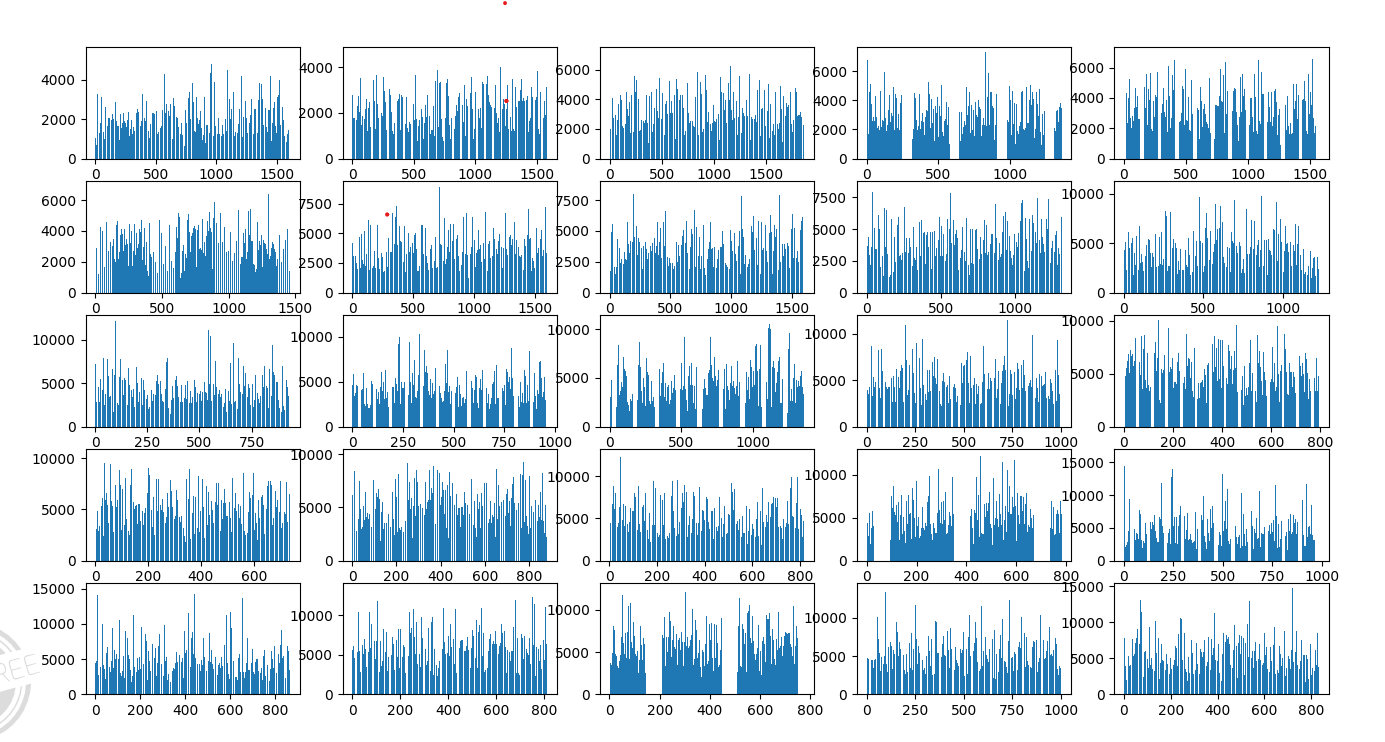
The subject of this project is classification rockets

\* First I known the data like:

Display amount of routes for each type

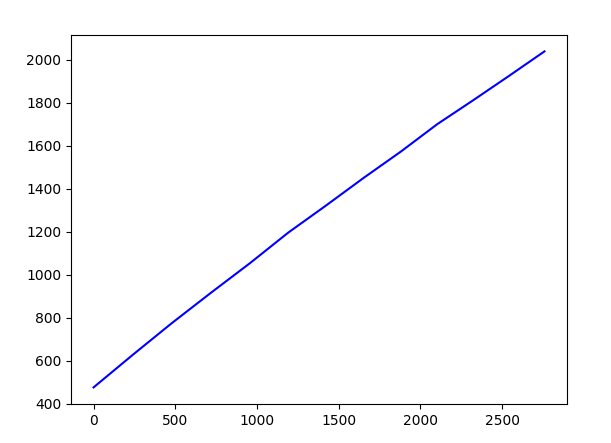
Display histograms of length of routes rockets for each type. Like this:



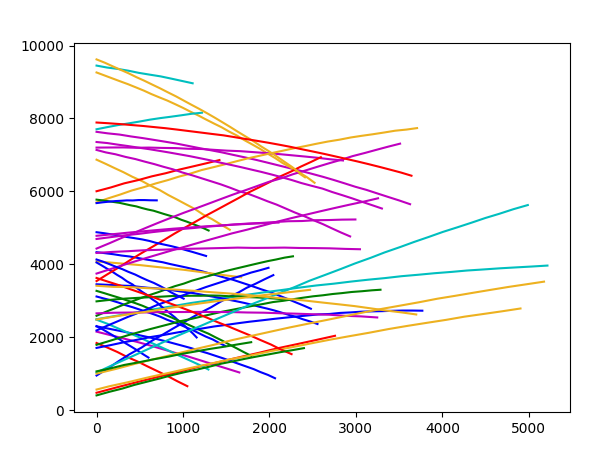
\* Drawing

I want to see different between the type 1 and 16, so I draw gragh:

- draw the first route:

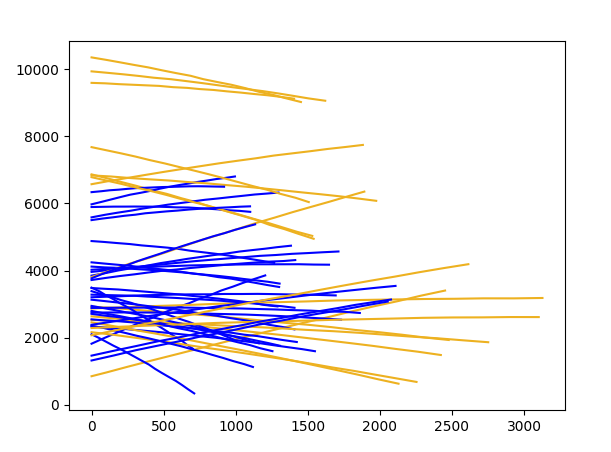


- draw 50 first routes with type 1, 2, 3, 4, 5, 6:



- draw 50 first routes with type 1 and 6 with length of routes 15 second. In order to know length how many second have to routes I add to data new column with name "Length" this column with type int that say the length of route in seconds. Now to draw 50 first routes with type 1 and 6 with length of routes 15 second I check for every routes if "Length" is 15.

See in drawing type 1 with color blue and type 16 with color yellow :



in this drawing you can see the different between type 1 and type 16, the length is different and type 16 is higher. I built the intuitive rules according this.

Making data \*

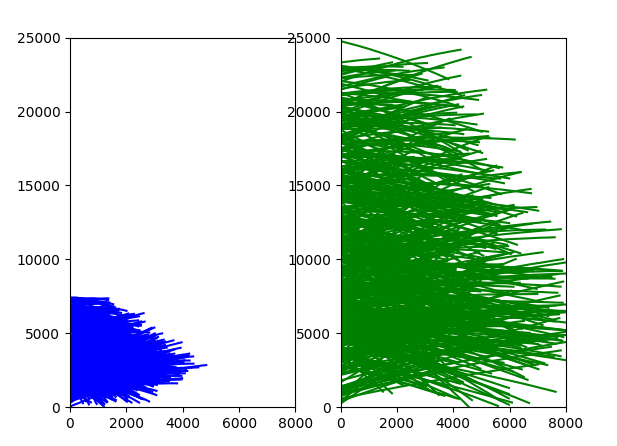
I create two new tables: split the data to two parts. One is 20% from data and second is 80% from data and create from it new tables, the tables only types 1 and 16

The table with 80% is train table in order to train the model with data for recognize if type is 1 or 16. And table with 20% is for checking the model, so I delete the column of types from check set.

\*classification with intuitive rules

First I draw all rockets with type 1 and 16 to see the different:

Type1 Type 16



Now after I have draws and two table one train set and second check set I built intuitive rules to recognize if type is 1 or 16

The intuitive rules :

- like you can see in last drawing, length of routes of rocket with type small than 5000, and high of rocket small than 7500, else type is 16.

I calculate confusion matrix in order to see percents of success:

[[331 3],

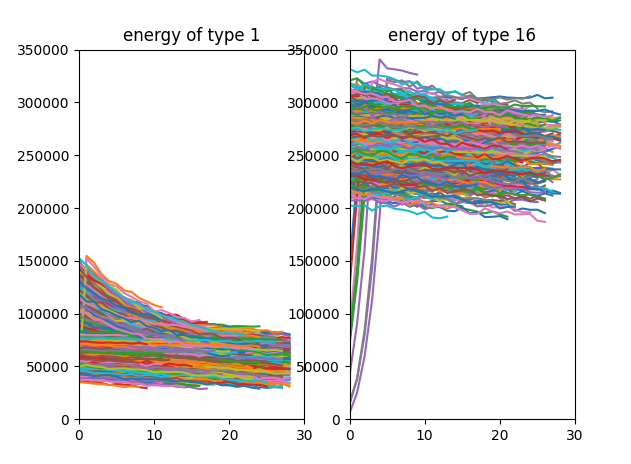
[55 79]]

And calculate f1 score:

0.9195

The result is not very good Its only 91%

Now in order to improve the result I calculate energy for all routes with type 1 and 16:



I add to intuitive rules validation of energy, I calculate energy for end of routes and check if it small than 155000 then type is 1 else type is 16.

I calculate the confusion matrix:

[[317 0]

[0 151]]

And calculate f1 score:

1.0

Now you can see that result is 100% !!!!

\*machine learning

I try RandomForestClassifier:

The confusion matrix:

[ 0 0 0 0 0 2 3 311]]

[ 0 0 0 0 0 0 0 0]

[ 0 0 0 0 0 0 0 0]

[ 0 0 0 0 0 0 0 0]

[ 0 0 0 0 0 0 0 0]

[ 0 0 0 0 0 0 0 0]

[ 0 0 0 0 0 0 0 0]

[ [ 143 4 1 4 1 0 0 0]

Average Error: 0.0700 degrees

And accuracy: 0.968017

Result is 96%

I try LogisticRegression

The confusion matrix:

[[316 0]

[0 153]]

Average Error: 0.0000 degrees

And accuracy: 1.0

Result is 100%

This is very good to classification rockets with types 1 and 16!!

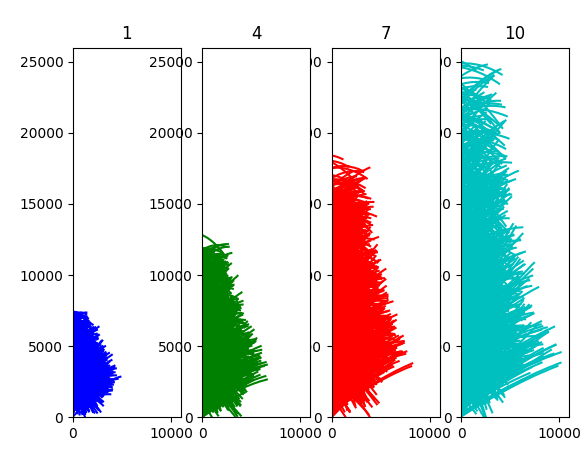
\* classification rockets with type 1, 4, 7, 10

I create two new tables: split the data to two parts. One is 20% from data and second is 80% from data and create from it new tables, the tables only with types 1, 4, 7, 10

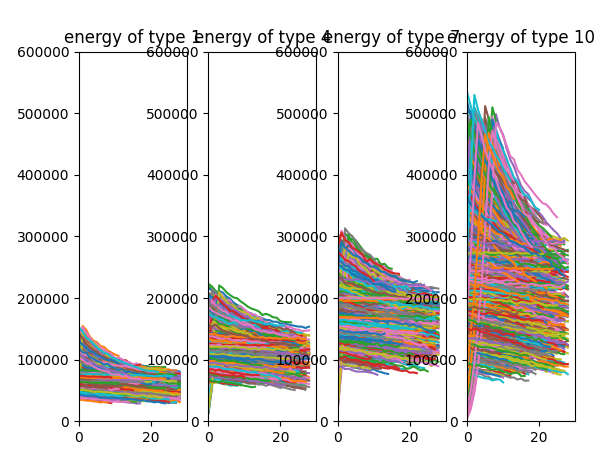
The table with 80% is train table in order to train the model with data for recognize if type is 1, 4, 7, 10. And table with 20% is for checking the model, so I delete the column of types from check set.

\*classification with intuitive rules

I draw the routes of rockets with types 1, 4, 7, 10:



I draw the calculate energy for each type:



And according the drawings I decide which type it is:

intuitive rules is validation of energy, I calculate energy for end of routes and check if it bigger than 260000 or distance between first second for last second of routes bigger than 18160 then type is 10.

Else if energy of end second bigger than 162500 or distance between first second for last second of routes bigger than 12176 then type is 7.

Else if energy of end second bigger than 106000 or distance between first second for last second of routes bigger than 7365 then type is 4.

Else type is 1.

I calculate confusion matrix:

[ 0 0 0 334]]

[ 0 3 128 128]

[0 161 143 15 ]

[[ 56 119 49 23]

And f1 score: 0.5858498705780846

Result is 58.6% is not good

\*machine learning

I try LogisticRegression

The confusion matrix:

[ 0 0 0 317]]

[ 8 71 179 1 ]

[48 274 25 1 ]

[[ 202 19 15 0 ]

Average Error: 0.5483 degrees

And accuracy: 0.8379310344827586

Result is 83.8%

It is good but not excellent

I try RandomForestClassifier

The confusion matrix:

[ 0 0 0 317]]

[ 0 2 257 0 ]

[16 329 3 0 ]

[[ 204 29 3 0 ]

Average Error: 0.1448 degrees

Accuracy: 0.9543103448275863

Result is 95.4% it is very good

I try RandomForestClassifier with default values:

Average Error: 0.1526 degrees

Accuracy: 0.9517241379310345

Result is 95.2% it is not better than last

I try RandomForestClassifier with default values but I check on train set instead test set:

Average Error: 0.0000 degrees

Accuracy: 1.0

And result is 100%

I try to change values of number of tree and depth of tree in RandomForestClassifier

|  |  |  |  |
| --- | --- | --- | --- |
| Accuracy of train set | Accuracy of test set | Depth of tree | Number of trees |
| 1.0 | 0.95172 | 200 | 50 |
| 1.0 | 0.95344 | 50 | 200 |
| 1.0 | 0.95876 | 50 | 1000 |
| 0.99655 | 0.93362 | 50 | 10 |
| 0.99676 | 0.93362 | 100 | 10 |
| 1.0 | 0.94913 | 100 | 100 |

You can see that 1000 tree and 50 depth is best: 95.87%

I see that need number of tree bigger than depth of tree.