DIGANTARA

Objective:

To create a method that automatically detects maneuvers in orbital data.

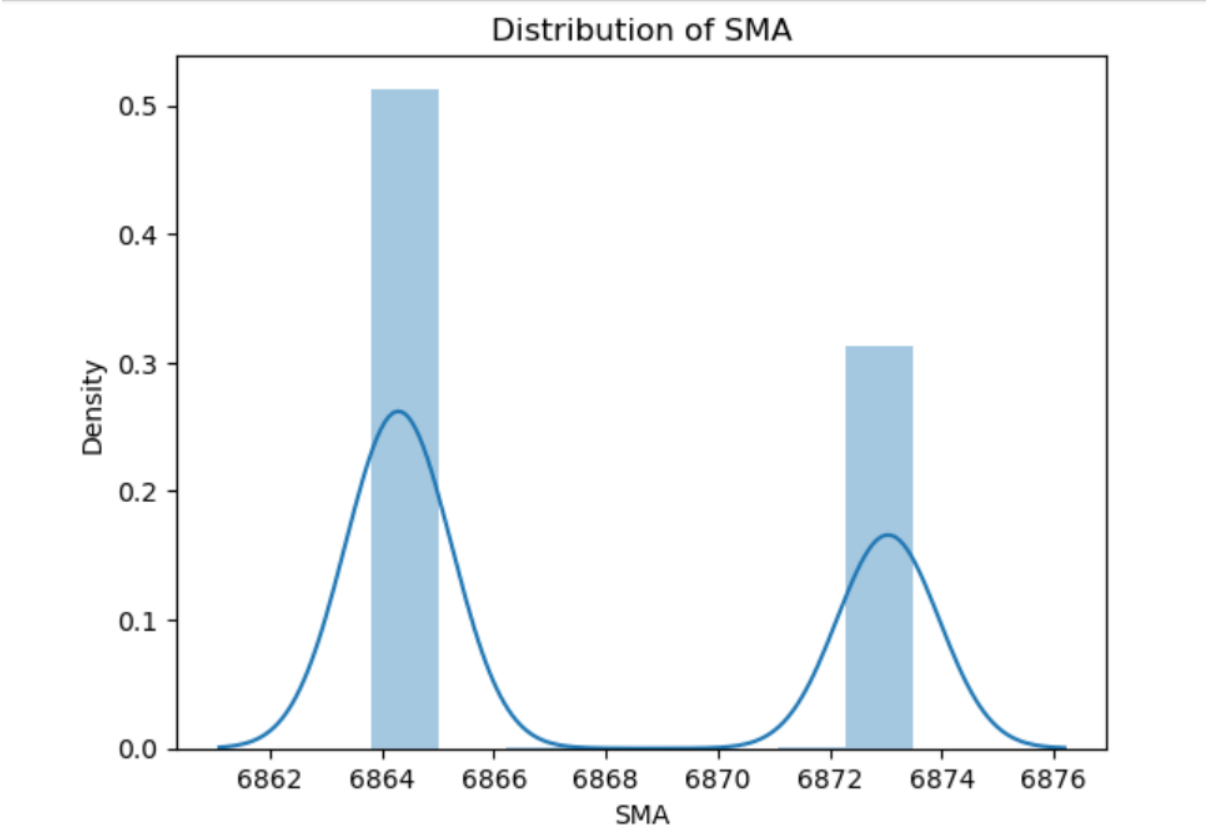
Terminology:

Maneuvers refer to carefully planned adjustments made to a satellite's orbit in space. These maneuvers allow us to achieve specific goals by altering the satellite's speed, direction, or position.

Data:

The data contains datetime column and corresponding SMA in kms. It contains 2991 records with no null values in both the columns. The max and min values of SMA are close to 6863 and 6873 respectively.

Distribution of the data is shown below:



Methodology:

Following observations are made from the data given:

- There is a sudden jump in the SMA values corresponding to a maneuver.
- The SMA values are decreasing with time with a steady slope except when maneuvers occur with a quick jump, in which case the values are again falling.

Heuristic method:

Approach-1

One way to identify maneuvers is to subtract every value from its previous values. Upon doing, the points which results in positive can be classified as maneuvers.

Here to quantify a jump to be maneuver a threshold of **value 0.1** is considered. The threshold value has been decided by observing the occurrences of maneuvers from the given maneuver data. To calculate the SMA_diff between two consecutive SMA values df.diff() of pandas has been used.

But from the given data it is clear that not every point which has positive SMA_diff is maneuver.

Kindly refer to the jupyter note book attached with the mail. For better understanding of my thought process.

Approach-2

This approach is a little modification over above approach. It is observed that not every jump is maneuver. For instance, consider the dates 2019-03-26, 2019-03-27, 2019-03-27 these dates are consecutive and have positive SMA_diff. But the maneuver occurred on 2019-03-27. Same applies for 2019-05-15, 2019-05-16, 2019-05-17. Hence a conclusion can be drawn that when there is a maneuver then the following dates has non increasing rather decreasing SMA values.

To look at whether following date as decreasing or non-increasing value of SMA_diff. SMA_diff values have been shifted by one place a head to get new column SMA_diff_shifted. The points whose SMA_diff is greater than 0.01 and SMA_diff_shifted less than 0.05 are taken as maneuvers.

The choice of choosing threshold value is merely by observation. And one might come up with certain threshold which classifies the each data point to be maneuver or not.

Approach -3 Exponential Distribution

Exponential distribution is generally used to model the time frame between two events. For example number of hours between two car intervals at a toll gate etc. In short exponential distribution is the modelling of waiting period. However it is expected that events must occur at a constant rate. And events must be independent to each other.

On same lines we can think of applying this knowledge to this study. But in this case occurrence of a maneuver depends on the previous maneuver (based on my understanding about a maneuver). And in addition to this occurrence of a maneuver not happening at a constant rate. Consider the days between two consecutive maneuver.

1. 2018-05-03
2. 2018-10-11
3. 2019-03-27
4. 2019-05-17
5. 2019-09-11
6. 2019-11-01

The number of days between consecutive occurrences of two maneuvers are 161, 167,51,117,51.

Machine Learning:

Approach -1 Supervised Learning (Classification)

Logistics Regression:

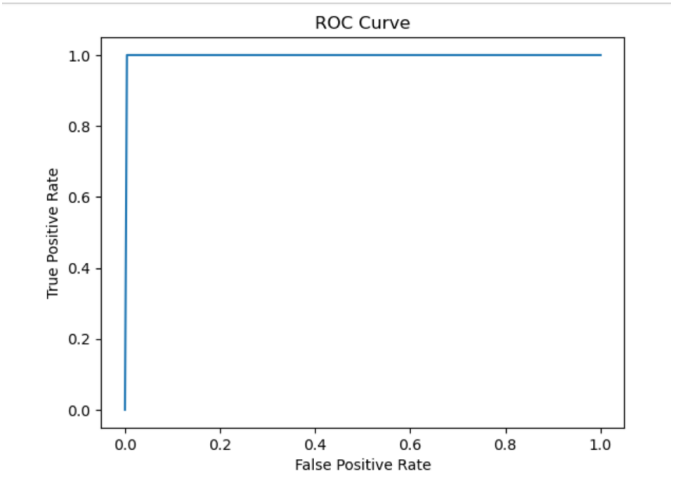
Consider maneuver detection to be binary classification problem, where we classify each datapoint to be maneuver or not. But now the question is what features are we going to consider?

Considering SMA, SMA_diff which we computed earlier as features based on heuristic methods. And based on the maneuver data given label each data point to be either maneuver or not. Clearly it forms highly imbalanced dataset with

only 6 positive classes (maneuvers) and rest to be negative (no maneuvers). Now one way to overcome this issue is to oversampling the class with less records. SMOTE was used to oversample the data. The oversampled was trained on logistic regression with test size of 0.2. Following are the results obtained.

	precision	recall	f1-score	support
0	1.00	1.00	1.00	478
1	1.00	1.00	1.00	436
accuracy			1.00	914
macro avg	1.00	1.00	1.00	914
weighted avg	1.00	1.00	1.00	914

Note: Kindly refer to the jupyter notebook attached with the mail



SUMMARY

1. Such high metrics could indicate overfitting to the training data or the model might have memorized the synthetic examples created by SMOTE. It's essential to evaluate the model on an unseen test set (not just the oversampled data) to assess its true performance.
2. Heuristic method discussed are intuitive and simple to implement can be used to identify the maneuvers.
3. Although Logistic regression giving us the better results but they might be misleading as the data is highly imbalanced