<u>NASA Planetoid Stratification Using Logistic Regression, DecisionTree,XGBoosting Algorithms of Machine Learning</u>

Literature Survey :

There are many approaches has presented by many researchers.The former approaches which were doing using Logistic Regression Algorithm are not exactly accurate and that's the reason why astronomers are still using ground-space-based telescopes to detect the threat.so,to increase the accuracy we want to train a model with a collection of machine learning algorithms like Decision Tree,Random Forest,XGBoosting etc. In this process we want to use the dataset collected by NEOWS(Near Earth Object Web Service) consisting of 4687 data instances(rows) and 40(columns) which gives the detailed nature of every asteroid identified by the space research organizations.
The main features include:

1.***Neo Reference ID':*** This feature denotes the reference ID assigned to an asteroid.
*2.'**Name**':* This feature denotes the name given to an asteroid.
3.*'**Absolute Magnitude**':* This feature denotes the **absolute magnitude** of an asteroid. An **asteroid's absolute magnitude** is the visual **magnitude** an observer would record if the **asteroid** were placed 1 Astronomical Unit (AU) away, and 1 AU from the Sun and at a zero phase angle.
4.*'**Est Dia in KM(min)**':* This feature denotes the estimated diameter of the asteroid in kilometres (KM).
5.***'Est Dia in M(min)**':* This feature denotes the estimated diameter of the asteroid in meters(M).
6.*'**Relative Velocity km per sec**':* This feature denotes the relative velocity of the asteroid in kilometre per second.
7.*'**Relative Velocity km per hr**':* This feature denotes the relative velocity of the asteroid in kilometre per hour.
8.*'**Orbiting Body**':* This feature denotes the planet around which the asteroid is revolving.
9.*'**Jupiter Tisserand Invariant**':* This feature denotes the Tisserand's parameter for the asteroid. **Tisserand's parameter** (or **Tisserand's invariant**) is a value calculated from several orbital

elements(semi-major axis, orbital eccentricity, and inclination) of a relatively small object and a more substantial‘ perturbing body’. It is used to distinguish different kinds of orbits.

10. ***‘Eccentricity’:*** This feature denotes the value of eccentricity of the asteroid’s orbit. Just like many other bodies in the solar system, the realms made by **asteroids** are not perfect circles, but ellipses. The axis marked **eccentricity** is a measure of how far from circular each orbit is: the smaller the **eccentricity** number, the more circular the realm.

11. *‘Semi Major Axis’:* This feature denotes the value of the **Semi Major Axis** of the asteroid’s orbit. As discussed above, the realm of an asteroid is elliptical rather than circular. Hence, the Semi Major Axis exists.

12. ***‘Orbital Period’:*** This feature denotes the value of the orbital period of the asteroid. **Orbital period** refers to the time taken by the asteroid to make one full revolution around its orbiting body.

13. ***‘Perihelion Distance’:*** This feature denotes the value of the Perihelion distance of the asteroid. For a body orbiting the Sun, the point of least distance is the **perihelion.**

14. ***‘Aphelion Dist’:*** This feature denotes the value of Aphelion distance of the asteroid. For a body orbiting the Sun, the point of greatest distance is the **aphelion.**

15. ***‘Hazardous’:*** This feature denotes whether the asteroid is hazardous or not.

After the detailed study it has been noticed that the time taken for threat detection was a major issue in previous studies and because of that lead time decreases,a threat to earth increases.Hence there is a need for improvement in the speed of the detection and more focus on reliability of the model.

In this paper[1],the authors tried to make some predictions over the combinations of orbital parameters for yet undiscovered and potentially hazardous asteroids by machine learning techniques. For this reason, they have used the Support Vector Machine algorithm with the kernel RBF. By this approach, the boundaries of the potentially hazardous asteroid groups in 2-D and 3-D can be easily understood. By this algorithm, they have achieved the accuracy over 90%.

In this paper[2] , authors has provided classification of asteroids, which are observed by VISTA-VHS survey. They have used some statistical methods to classify the 18,265 asteroids.

They used a probabilistic method, KNN method and a statistical method to classify those. Later, they compared the algorithms' accuracy. They have classified the asteroids into V, S and A types. They did it on the 18265 asteroids and test set consists of over 6400 asteroids.

In this Paper[3] , they have classified the asteroids which are near to the earth and can easily pass through the orbital of Mars. For this purpose they, have used a perceptron-type network. And their output class are three, they are Apollo, Amor and Aten. They have also found that semi-major axis and focal distance are the two important major features, for which the asteroids can be separable linearly.

In this paper[4] , they identified the asteroid family by the help of the supervised hierarchical clustering method. For this method, they have found the distance between an asteroid from any reference objects. Later they compared the results with traditional hierarchical clustering method and found that their result gives 89.5% accuracy, which is better. By this approach, they found 6 new families and 13 clumps.