Using NASA Data to predict if an Asteroid is dangerous to Earth

The recent launch of Perseverance and Ingenuity to Mars has renewed the interest of the public in the space exploration program. While there is more to be explored, this report looks at data collected already, about asteroids. Depending on how close the orbits of Asteroids and Comets could be to our planet, they can be grouped as Near-Earth Objects (NEOs). Currently there are more than 20,000 NEOs been monitored.

Possible effects of asteroid impact are well known, including the possibility of a large amount of dust thrown up into the atmosphere which could block out the sun for a while. Depending on the information collected, NEOs can be classified as been 'potentially hazardous to Earth' or not. This classification can be automated through the use of Data Science, with robust classifiers.

Experimentation has been carried out using two classifiers 'Support Vector Machines' (SVM) and 'Random Forest' (RF), to see which one could form the basis of this classification process. The data is collected from NASA(http://neo.jpl.nasa.gov/). Different combinations of the original data with two methods, Principal Component Analysis (PCA) and Synthetic minority oversampling Technique (SMOTE) are considered. Each combination along with the original data by itself is tested with both the classifiers (SVM, RF) to pick the best classifier-repo combination.

The best results of the experiments were found with the RF classifier and 'Original with SMOTE' combination as shown in Figure 1.

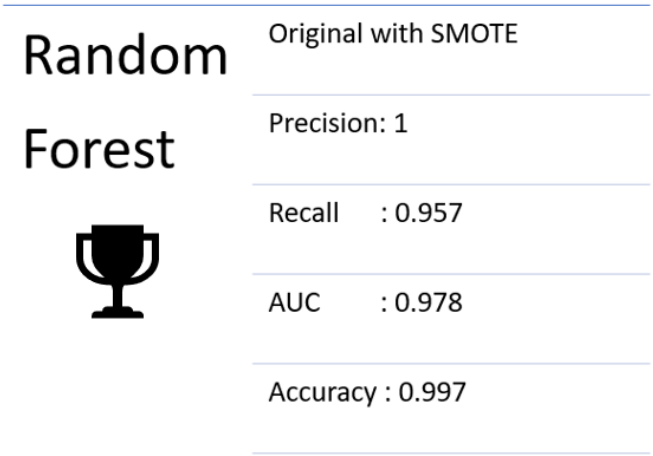

Figure 1: The best Classifier- Repo combination with metrics

The perfect classifier would have achieved a '1' for all metrics (Precision, Recall, AUC and Accuracy). And it can be seen that the chosen classifier has come quite close to achieving this. Further experimentation can be done with this combination to possibly achieve even better results and are detailed in the technical report.