**Introduction to Statistical Learning**

**Project Proposal**

**Siva Sameer Krishna Yarlagadda**

**16271011**

**Sy9bd@mail.umkc.edu**

**Title: Nasa Asteroid Classification**

**Dataset:** <https://www.kaggle.com/shrutimehta/nasa-asteroids-classification>

**Objective:**

1. Predicting whether an asteroid is hazardous or not.
2. Features responsible for claiming an asteroid to be hazardous.

**Introduction:**

Near-Earth Objects (NEOs) are comets and asteroids that have been nudged by the gravitational attraction of nearby planets into orbits that allow them to enter the Earth’s neighborhood.

**Description:**

The data is about Asteroids - NeoWs. NeoWs (Near Earth Object Web Service) is a RESTful web service for near earth Asteroid information. With NeoWs a user can: search for Asteroids based on their closest approach date to Earth, lookup a specific Asteroid with its NASA JPL small body id, as well as browse the overall data-set.

Data set contains *40 Attributes*(columns) with *4688 Records.*

Attributes are of different types including int, string, object etc.

**Implementation:**

This is a **Classification problem** where we have to predict the label of Hazardous class which is already known. **Supervised learning** methods for classification such as Logistic Regression, Decision trees, Bayesian Classifier, K nearest neighbor , Support Vector Machine, etc. can be used to implement this. It further includes dividing the data into Training and Test sets where training data is used to train the model and test data is used to validate the trained model. Evaluation of the model can be done based on parameters like Accuracy, Precision, Recall, AUC(Area under curve). Finally visualizing the results using appropriate plots.