**Student Name: Sarmad Tanwir**

**Project Ideas**

**Title: Used Car Prices: Advanced Regression Techniques**

**Description:**

This project requires scrapping data about car features from a number of websites e.g. cargurus.com autos.yahoo.com etc. Then, I’ll use this data to predict the listed price of a particular car.

I got this idea from Kaggle competitions. Kaggle has a dataset of residential homes in Ames, Iowa that has 79 explanatory variables that attempt to describe every aspect of these homes. According to Kaggle, the dataset lends itself to advanced regression techniques like random forests and gradient boosting with popular XGBoost library. The challenge is to predict the final price of each home from this information. I’ve changed the data set to online cars data to have a more real-world data situation especially w.r.t to data collection and cleaning.

**Title: Identify question pairs with the same intent in the Quora questions pair corpus**

**Description**: Currently, Quora uses a Random Forest model to identify duplicate questions. In this competition, Kagglers are challenged to tackle this natural language processing problem by applying advanced techniques to classify whether question pairs are duplicates or not. Doing so will make it easier to find high quality answers to questions resulting in an improved experience for Quora writers, seekers, and readers.

**Title: Study the Complex Contagion of Campaign Donations.**

**Description:**

In this project, I will study how campaign donations diffuse through a network of more than 50000 elites and

examine how connectivity among previous donors reinforces contagion. This analysis will help give useful suggestions to political campaign managers regarding fund raising.

The idea is taken from the following paper:-

Traag, Vincent A. "Complex Contagion of Campaign Donations." *PloS one* 11.4 (2016): e0153539. paper attached.

I will try to reproduce the findings of the paper, create useful visualizations and/or find some new insights.