**INFO 5082 - Project Proposal**

**Seminar in Research and Research Methodology**

**Red Wine Quality**

**Introduction:**

**The data set is collected from the Kaggle website. The data have variety of red wine is collected and its properties have been noted. The main intension of the data set is to know which physicochemical properties. The wine samples are related to the north of Portugal. The goal is to model the wine quality based on the physicochemical test.**

**The data set consist of 12 number of attributes. The attributes are multivariant we can perform outlier detection algorithm for detecting the poor or good quality wine.**

**Dataset link:** [**https://www.kaggle.com/uciml/red-wine-quality-cortez-et-al-2009**](https://www.kaggle.com/uciml/red-wine-quality-cortez-et-al-2009)

**Analysis:**

**I would like to use different Classification models to predict the quality of the wine. And figure out the which model is best by using the optimizing techniques such as the Cross-validation score and the Random Forest classifier.**

**Classifier model:**

* **Random Forest Classifier**
* **Stochastic Gradient Descent Classifier**
* **Support Vector Classifier**