**Mike Dean**

**Final Project Proposal**

**Prediction**

**10/9/14**

For my final project, I plan to complete a programming project to analyze data from Kaggle.com. The dataset I am particularly interested in is titled, “Titanic: Machine Learning from Disaster”. The dataset is a list of passengers from the Titanic along with each passenger’s class, name, sex, age, number of siblings and spouses on board, the number of parents and children on board, ticket number, passenger fare, cabin number, where they embarked from, and finally whether or not they survived.

I think this is a very interesting dataset to work on for my final project for a few reasons. First, it is a relatively simple set. Most of the values are either already numerical or can easily be quantified. Additionally, it has to do with one of the most infamous tragedies to occur in human history. It is commonly known that passengers were more likely to survive if they were in second or, ideally, first class and if they were women or children. Perhaps there were other factors such as where their cabin was located on the ship. Maybe if they had a last name that allowed them to convince the crew that they were related to someone important they could get on a lifeboat. That is why I want to look at this data, if it says something about the Titanic that no one has ever noticed before, I want to find it.

I hope to use the machine learning techniques we will be learning in the class along with my previous programming experience to analyze the data. If time allows, I would also like to explore the possibility of using genetic algorithms to analyze this data as well so I can compare those results with traditional machine learning techniques. Regardless of whether or not I find something interesting in the data, I hope this project helps me learn more about data analysis, machine learning, and the challenges of making predictions.

Link to data: http://www.kaggle.com/c/titanic-gettingStarted