Project Draft/Milestone 3, Project 2

**Introduction**

Kaggle competition is a competition platform which hosts competitions in the data science realm. Kaggle has data sets and a wide variety of competition types to the data science community in order to challenge individuals regardless of where they are within their data science careers. Kaggle hosts several competition types, including featured which are what they are best known for, which feature a cash prize. The playground competition type, which the Taxi Fare prediction competition falls into, is a for fun type of competition. These competition types are typically targeted towards newcomers ("How to use Kaggle", n.d).

**Background**

What is a data pipeline

SKlearn Data Pipeline

What is Docker

What is Kubernetes

Text Preprocessing

**Data Overview**

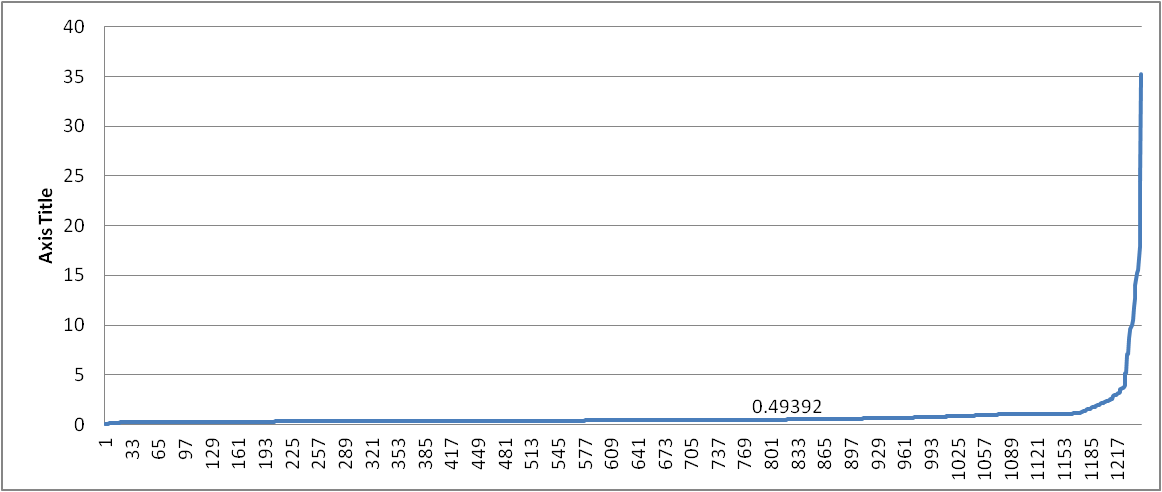
**Software**

The software packages used for this project included the anaconda distribution of the programming language Python and Jupyter Notebook. The project was written in a Jupyter Notebook. The popular version control Git was used to store various different versions. While the project was written in Python, with several imported packages were used as well. Those imported packages include Numpy, Pandas, and SkLearn.

**Data Preprocessing**

**Model Creation**

**Results**



**Results Analysis**

**Conclusions**

**Future Studies**

References