**Initial Project Plan**

**Student:**Srivishnu Maganti

**Date Submitted:** *09/23/2020*

**Task Selected:** Estimate the total response time of the ems.

**Task Description:** Here task is to predict the response time of an ambulance dispatch to a location in NYC, here we are going to predict the total response time (INCIDENT\_RESPONSE\_SECONDS\_QY), by using few attributes such as zip code, incident call type, city counsel district, etc. Here this task can help to improve the response time of ems services.

**Final Predictive Value:** When we clearly observe the dataset, we can see a column which has the total time of dispatch (INCIDENT\_RESPONSE\_SECONDS\_QY), this value is given in seconds. So, by considering this as a target variable we are going to predict the total response time, where we are considering few selected features for predicting. Here with the help of the independent variables that I have considered I will try to estimate a relationship using regression. Train the classifier with few features to predict the response time.

**Factors Considering:**

I am going to predict response time using few features

1. INITIAL\_CALL\_TYPE \*
2. INITIAL\_SEVERITY\_LEVEL\_CODE
3. BOROUGH
4. ATOM
5. INCIDENT\_DISPATCH\_AREA
6. ZIPCODE
7. POLICEPRECINCT
8. CITYCOUNCILDISTRICT
9. COMMUNITYDISTRICT
10. COMMUNITYSCHOOLDISTRICT
11. HELD\_INDICATOR
12. SPECIAL\_EVENT\_INDICATOR

These are the factors that I think that they are the one which will be responsible for predicting response time.

So, first I am going to analyze these factors and then firstly, clean the data because there are a lot of null values that are present in the dataset. Secondly, find the correlation matrix between each attribute and “INCIDENT\_RESPONSE\_SECONDS\_QY”. By this we understand the relation between the factors and their importance’s respectively. Finally, I am going to try regression to train the data for prediction.

**Data Sources Considering:**

Right now, I am not considering any other data set other than the given Ems data set. But first I will try with my analyses that I am thinking (regression-based approach) and if that does not produce satisfactory result, (working with test data—explained below). I will try with another analytical approach.

Test the model- I am going to split the data in to train and test and then train the model with regressor and then work with unlabeled test data to give results, and compare those results with the original and find the models confusion matrix to estimate accuracy, f1 score etc.

**Next Steps:**

1. Explore each attribute with target variable.
2. Find the importance of each attribute.
3. Find the relationship between attributes
4. Work with the regressor using regression techniques
5. Calculate the models f1 score to get the results of regressor.
6. Improvements based on those prediction metrices.