INTERIM REPORT

Project Guide: Sulekha Aloorravi

Project Leader: Sanku Lahiri

Team Members: Amol Ravatale, Devendra Bharadwaj, Kushal Thakral

Project Title: Predict Ambulation Locations to serve accident hotspots

**Problem Definition**

Road accidents contribute to approximately 1.35 million deaths [*source: WHO]* worldwide annually. A major factor in road accidents which decide mortality is post trauma care. This is dependent on how fast medical attention is provided through ambulances after accident. The accessibility of ambulance service is fragmented based on the development index of a nation. Wealthier and developed economies have very fast and reliable ambulance care services while developing and weaker economies lag drastically. Research shows that the survival rate of accident victim increases if an ambulance is in the vicinity of 3-4 miles of accident site to provide post trauma care. This means that if an ambulance could be placed beforehand in the vicinity of all such accident locations it would be able to serve victims faster. Our paper aims to develop a Machine Learning based models using two approaches to predict locations of 6 ambulances for 8 timeslots in a day (24 Hrs.) based on the history of accident locations. For our analysis we have taken data set from “Uber Nairobi Ambulation Pre-ambulation” challenge which provides accident history, locations etc. for Nairobi.

To re-iterate, Approach 1 attempts to find the locations to park ambulances based on time-series forecasting model. Approach 2 builds a predictive model using Classification learning where it attempts to predicts possibility of accident at a particular location and accordingly finds an optimum location to park ambulance(s).As we are attempting to find solutions through two approaches, the following sections covers the contents for both these approaches.

**Base Model & Architecture**

**Approach 1:**

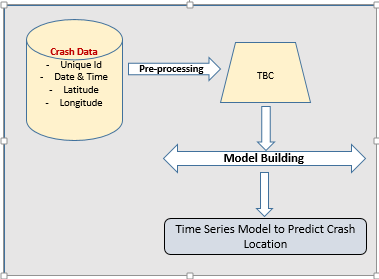


Figure 1: Solution– Time Series Forecasting Approach

**Approach 2:**

This approach is based on feature engineering and creating classifier to predict crash location. The primary data source and features which are used to build model are depicted in the below picture. The solution approach is depicted in below figure

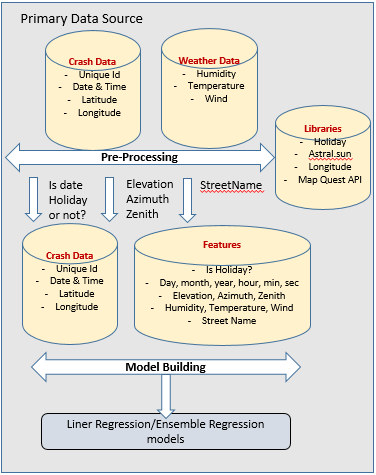


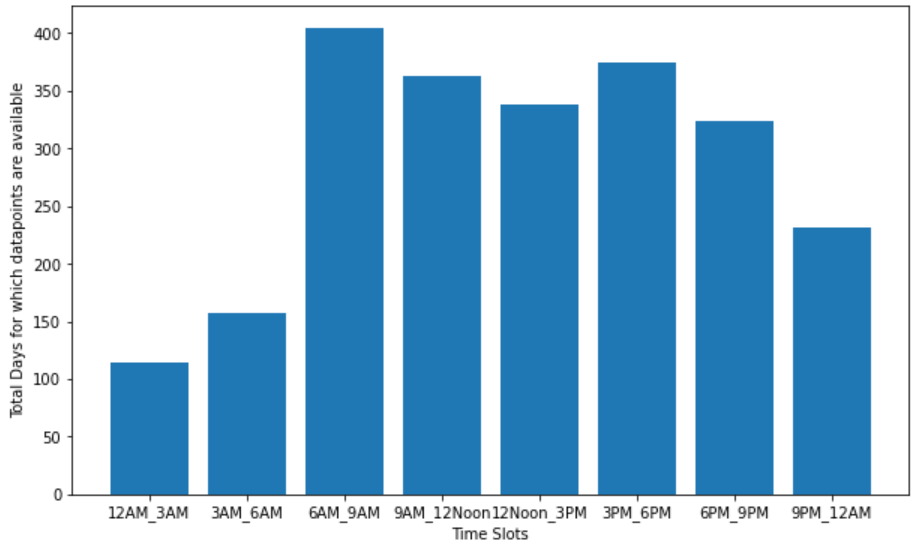
Figure 2: Solution– Classification Approach

**Exploratory Data Analysis**

**Approach 1:**

The entire training dataset is split asper 8 different time slots namely, 12AM-3AM, 3AM – 6AM, 6AM – 9AM, 9AM – 12 Noon, 12 Noon – 3PM, 3PM – 6 PM, 6PM – 9 PM and 9PM – 12 AM.

The following bar graph shows the number of days for which the data points are available for each time slot. On an average out of 545 days (Jan 01, 2018 – June 30, 2019) the data is available only for 288 days. In order to build a time series model we need to impute the missing dates and values.



*Figure 3: No of days Crash data points*

**Approach 2:**

This approach is based on feature engineering and creating classifier to predict crash location. The primary data source and features which are used to build model are depicted in the below picture

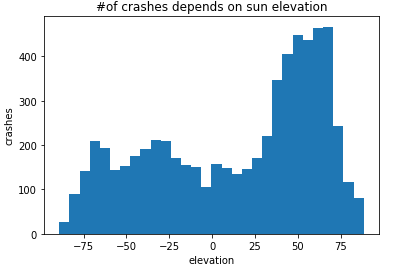


Figure 4: No of Crashes - Elevation

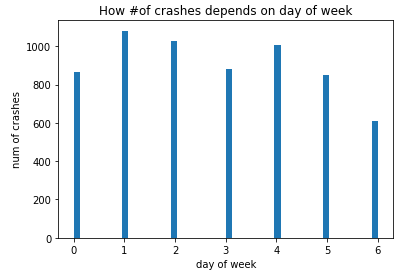
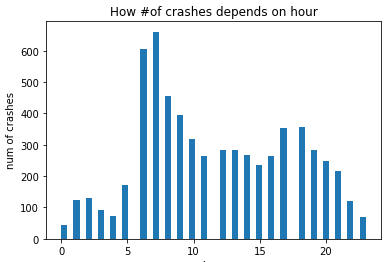
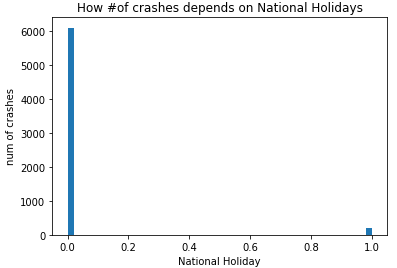


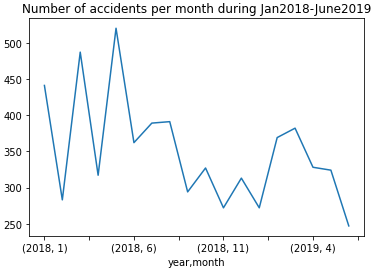
Figure 5: No of Crashes – Day of Week



*Figure 6: No of Crashes – Hour of day*



*Figure 7: No of Crashes – National Holiday or not*



*Figure 8: No of Crash – Month*

**Early Results**

**Approach 1:**

The time series model is yet to be developed.

**Approach 2:**

The classier models are yet to be built to classify if accident will happen or not at given location (latitude, longitude)

**Tentative Algorithms**

**Approach 1:**

The first approach using time series analysis is formulated in the following steps:

1. Segregate data as per respective 8 time slots for entire duration of provided data.
2. Derive a location centroid for each daily timeslot.
3. Using Bayesian Structural Time Series /Continuous Time Series Markov Model to predict the next location (centroid).
4. Validation whether the prediction locations are at optimum distance.
5. Identify clusters using DBSCAN in a particular time slot in similar period (date-time).
6. Apply RMSE or Euclidean measure to derive distance on test set.

As mentioned previously, the first approach uses time-series as the core solution approach. The accident locations provided in the data are of continuous nature and the location-based service model is the best fit for such scenario. Moreover, since this is a changing or dynamic location, we decided to attempt Continuous Time Series – Markov Model and Gaussian Mixed Model [**3**]. A brief snapshot of the time series “Location Centroid” approach is depicted in Figures below.

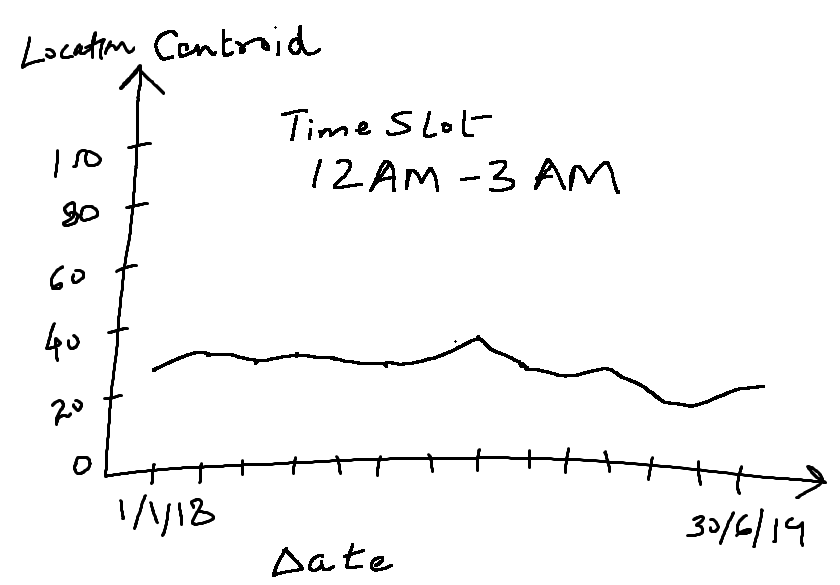


Figure 9: Accident Location Centroid – Training Data Trend

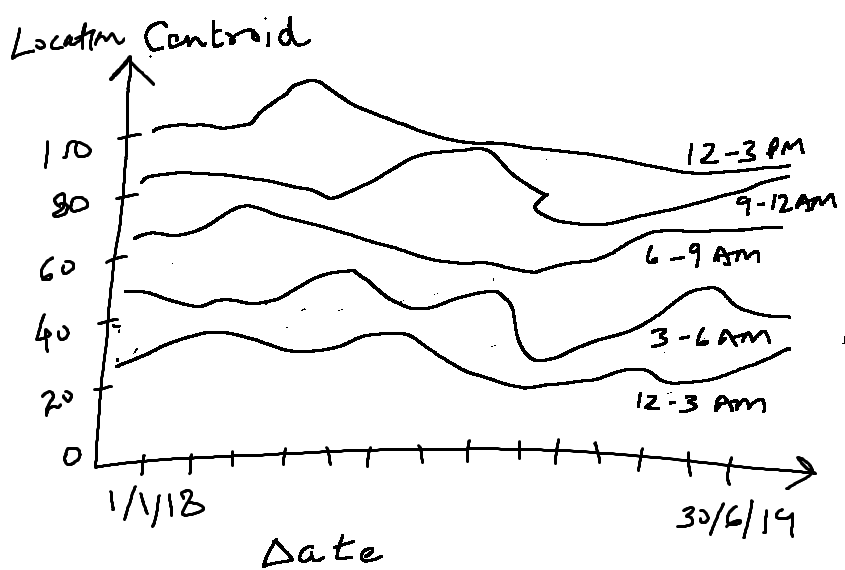


Figure 10: Accident Locations Centroid: All Timeslots

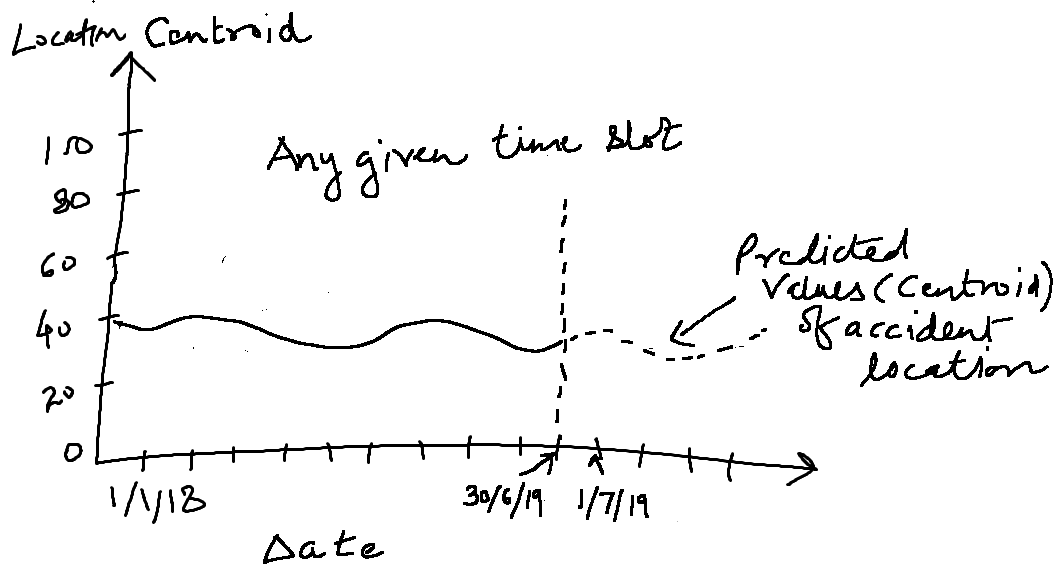


Figure 11: Predicted Accident Locations Centroids

**Approach 2:**

Libraries used for feature extraction

1) holidays – To extract is given date is holiday or not.

2) Astral - is python package for calculating the time of various aspcts of Sun and phases of the moon. We have used it to extrtact elevation (the time when the sun is at a specific elevation for either a rising or a setting sun), azimuth (The number of degrees clockwise from North at which the sun can be seen) and zenith (The angle of the sun down from directly above the observer) of Sun.

Algorithms – We are yet to work on buidling classifer models.

**References:**

1. Brodsky, H. (n.d.). Emergency medical service rescue time in fatal road accidents. Retrieved April 30, 2021, from Trb.org website: http://onlinepubs.trb.org/Onlinepubs/trr/1990/1270/1270-011.pdf
2. Coats, T. J., & Davies, G. (2002). Prehospital care for road traffic casualties. *BMJ (Clinical Research Ed.)*, *324*(7346), 1135–1138.
3. Du, Y., Wang, C., Qiao, Y., Zhao, D., & Guo, W. (2018). A geographical location prediction method based on continuous time series Markov model. *PloS One*, *13*(11), e0207063.
4. F. Sam, E., K. Blay, D., Antwi, S., Anaafi, C., & A. Adoma, J. (2019). Pre-hospital and trauma care to road traffic accident victims: Experiences of residents living along accident-prone highways in Ghana. In O. Karcioglu & M. G. Eneyli (Eds.), *Emergency Medicine and Trauma*. London, England: IntechOpen.
5. Dong, C., Shao, C., Li, J., & Xiong, Z. (2018). An improved deep learning model for traffic crash prediction. Journal of Advanced Transportation, 2018, 1–13.
6. <https://zindi.africa/competitions/uber-nairobi-ambulance-perambulation-challenge/data>
7. Yu, R., Xia, X., Liao, S., & Wang, X. (2015). A location prediction algorithm with daily routines in location-based participatory sensing systems. International Journal of Distributed Sensor Networks, 2015, 1–12.
8. Yang, Z.-Q. (2011). Marine traffic accident prediction based on particle swarm optimization-based RBF neural network. 2011 3rd International Conference on Computer Research and Development, 1, 435–438. IEEE.
9. Su, H., Zhang, W., & Li, Z. (2018). Analysis and prediction of water traffic accidents in jingtang port based on improved GM (1,1) model. 2018 37th Chinese Control Conference (CCC), 2212–2217. IEEE.
10. J. C. Yin, Z. J. Zou, F. Xu and N. N. Wang, "Online ship roll motion prediction based on grey sequential extreme learning machine", Neurocomputing, vol. 129, pp. 168-174, 2014
11. J. C. Yin and N. N. Wang, "Online grey prediction of ship roll motion using variable RBFN", Applied Artificial Intelligence, vol. 27, no. 10, pp. 941-960, 2013
12. Lv, Y., Tang, S., & Zhao, H. (2009). Real-time highway traffic accident prediction based on the k-nearest neighbor method. 2009 International Conference on Measuring Technology and Mechatronics Automation, 3, 547–550. IEEE.
13. Literature established the model of the relationship between traffic volume, lane number, turning lane number, bicycle facilities, control mode and traffic safety at intersections

Greibe Poul, "Accident prediction models for urban roads", Accident Analysis and Prevention, vol. 35, pp. 273-285, 2003.

1. Zhou, Z. (2019). Attention based stack ResNet for citywide traffic accident prediction. 2019 20th IEEE International Conference on Mobile Data Management (MDM), 369–370. IEEE.

**Glossary**

ACF – Auto Correlation Function

AIC - Akaike Information Criterion

AR – Auto Regression

ARIMA - Auto Regressive Integrated Moving Average

BIC - Bayesian Information Criterion

MA – Moving Average

MAE – Mean Average Error

PACF – Partial Correlation Function

RMSE – Root Mean Square Error

WHO – World Health Organization