Report Date: 9/23/2022
To: ematson@purdue.edu, ahsmith@purdue.edu and lee3450@purdue.edu
From: Chasing FOX

- Daye Kim (rlaek4793@khu.ac.kr)
- Hyun Roh (yeshyun@khu.ac.kr)
- Heewon Jeong (jhjmo0719h@chungbuk.ac.kr)
- Jihu Yang (zihooy@handong.ac.kr)
- Juwon Baek (21700340@handong.ac.kr)

Summary

- Analyzed crime data considering population density
- Determined the environment of backend development
- Discussed and determined UI of the application
- Tried to prove the novelty of the research through searching papers
- Wrote the abstract of the paper and defined the structure of introduction

What Chasing FOX completed this week:

- **Data analysis**
  - For the 12 types of city facility and request datasets, the correlation between them and the crime density was analyzed to find the information that could be related to crime density.
  - Normalization of crime density was conducted in consideration of the fact that crime data is greatly affected by the population number. In order to remove effects of population from the data, normalization was performed using the formula 1-2.

$$density_{population} = number_{zipcode\ person} \div area_{zip\ code} \quad (1)$$

$$normed\ crime\ density = density_{crime} \div density_{population} \quad (2)$$

  - The mash map of the crime dataset that has 100j of the separate cell was made, and was applied to Kernel Density Estimation(KDE). In Figure 1, the value of crime density and position of the facility was plotted. Crime density is a dependent value to analyze.

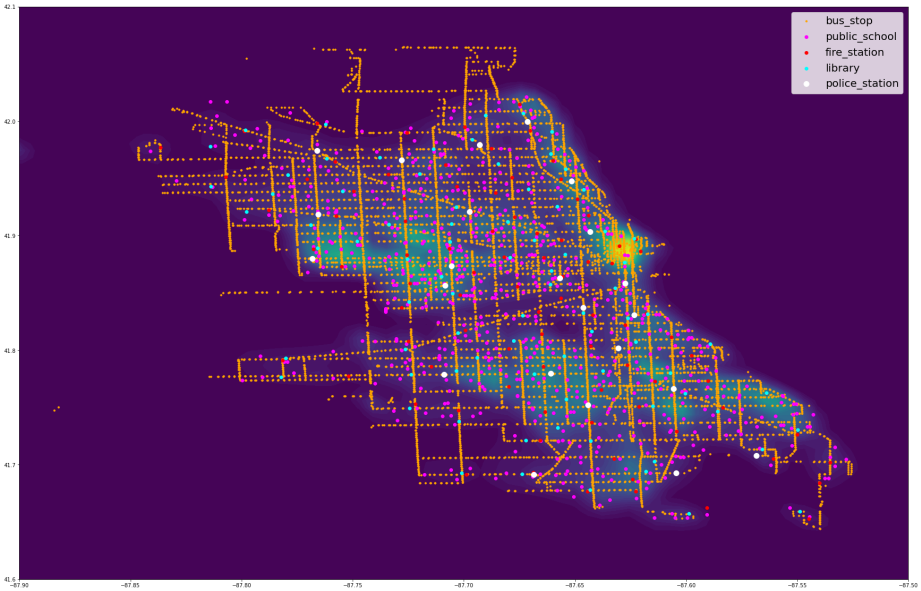

Fig 1. Crime density with scatter plot of facility and request.

- Data analysis is conducted with 2 methods.
- Method 1. To find the correlation between distance from the nearest facility and service request, and the density of crime.
  - Distance from the nearest facility and service request, and the density of crime was calculated by formula 3.

$$distance = \sqrt{(x_{cell} - x_{facility})^2 + (y_{cell} - y_{facility})^2} \quad (3)$$

  - Figure 2 informs the correlations with each data. The correlation with facility and service request was negative, which means if the distance from coordinate of cell to facility gets closer, then crime density gets higher.

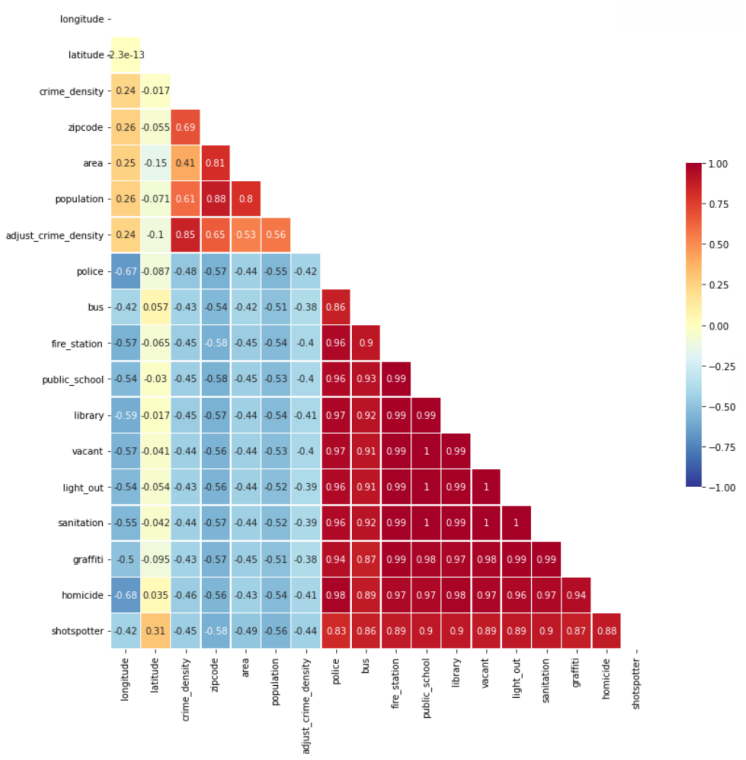

Fig 2. Correlation with Euclidean distance method

- Method 2. To find the correlation between density of facility & request, and density of crime.
  - KDE is applied to 12 types of datasets that could cause reduced or increased crime to calculate the density of each point and analyze correlation between density of facilities and density of crime.

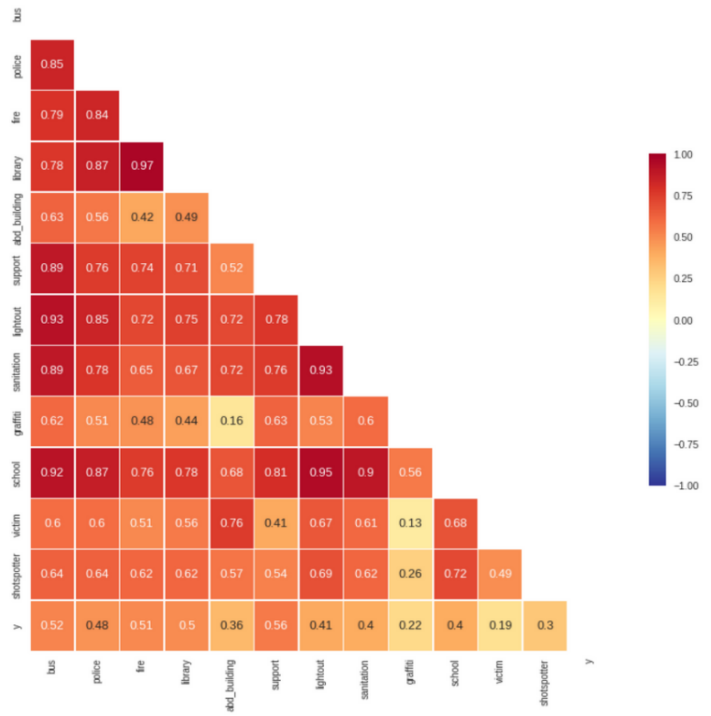

Fig3. Correlation between density of facilities and request and density crime.

- Among the 12 types of collected data features, AutoML was performed to identify the features that affect decision making for predicting the density of crime.
- The data used to predict the density of crime in AutoML is the density of 12 features.
- In the AutoML training, 17 types of regression models were calculated while the Feature importance and SHAP value(Figure 4) were calculated in the Extra Trees Regression Model, which was the best performing model based on MAE loss.

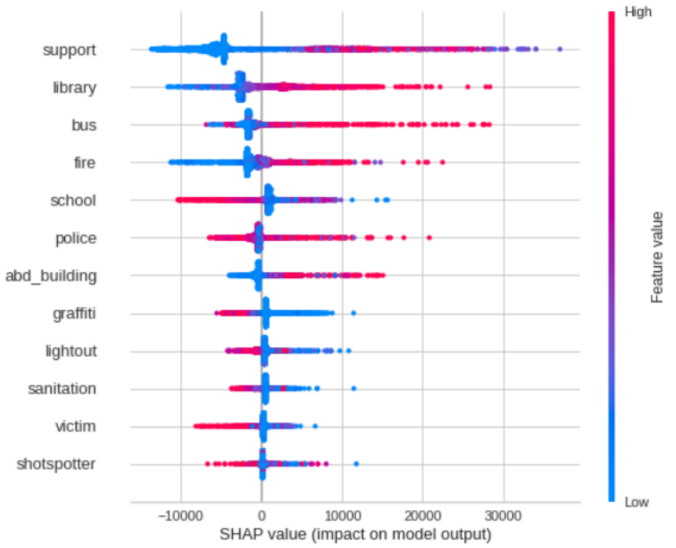

Fig 4. Shap value of best performing model

- **Back-end**
    - The back-end tools of Rabbit Application are Spring Boot, JPA, and MySQL. Spring Boot and JPA were used for creating API while MySQL was used for database systems.
    - The basic API related to user management is finished.
    - For example, 'GET | http://localhost:8080/api/user' API returns all the users in the DB.
- **Front-end**
    - Google API test of Prototype Application was successfully completed.
    - The main color of the application was decided to be orange because of the team name being 'ChasingFOX' and Rabbit is related to carrot.
- **Paper**
    - Investigation of related works
        - One of this paper's novelties was merging safety data sets with crime data sets. The concern was whether our novelty had a precedent or not. So about 50 related works were investigated. [1] - [4] Papers that met the condition related to were not found until now.
    - Paper draft
        - The draft of the abstract part was done except for explaining algorithms for risk estimating. The main structure of the Introduction part was done.

Things to do by next week

- Completing the Introduction part of the paper.
- Finish making the UI of application through Figma
- Server settings for database and application deployment
- Data analysis applying Eric's feedback
    - Need to consider the severity of crime in Chicago
    - Need to consider the density of population in Chicago
    - Need to compare crime density before COVID-19 and after COVID-19.

Problems or challenges:

- Algorithm for the navigator application is not determined, and therefore it is difficult to complete the abstract.
- Considering population density, data preprocessing for analysis was done. However, there were only a few differences from before. Expected reason is that the floating population and resident population are different. Data that can inform about the floating population is required, however it was not easy to find the data.

References

[1] E. Galbrun, K. Pelechrinis, and E. Terzi, "Urban navigation beyond shortest route: The case of safe
paths," *Information Syst.*, vol. 57, pp. 160–171, Apr. 2016.
[2] F. Mata *et al.*, "A mobile information system based on crowd-sensed and official crime data for Finding Safe Routes: A case study of mexico city," *Mobile Information Systems*, vol. 2016, pp. 1–11, Mar. 2016.
[3] Y. Pang, L. Zhang, H. Ding, Y. Fang, and S. Chen, "Spath: Finding the safest walking path in smart cities," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 7, pp. 7071–7079, May 2019.
[4] Y. Zhao, Y. Xie, and S. Ahvar, "On integration of any factor with distance for navigation : Walk safely and fast enough," *2019 IEEE 23rd International Enterprise Distributed Object Computing Workshop (EDOCW)*, 2019.