**IBM Data Science Capstone:**

**Car Accident Severity Report**

Kwak hyunwoo

1. description of the problem & discussion of the background

Traffic accidents occur constantly every year, and traffic accidents are accompanied by property damage or human casualties, so it is necessary to prepare for them. The extent of the damage varies depending on the surrounding situation, such as the lack of visibility due to fog or the sudden appearance of young pedestrians on the road due to the large number of schools nearby. Therefore, the purpose of this project is to prevent traffic accidents by predicting the extent of traffic accident damage according to various circumstances and reminding drivers of it.

2. A description of the data

컴퓨터, 옅은, 노트북, 키보드이(가) 표시된 사진

자동 생성된 설명

The data used the Collision Data provided by the Coursera. This data contains information about traffic accidents since 2004. If you look at the characteristics of Dataset, the location of the collision (alley, block, intersection), the severity of the traffic accident, and the number of people involved in the traffic accident (walkers, cyclists, injuries, serious injuries, etc.) each have a collision.

3. Methodology section

스크린샷이(가) 표시된 사진

자동 생성된 설명

To identify the characteristics of the data, the correlation was identified by representing the Heatmap between collision severity, which is the y data, and the the remaining numerical data. In addition, time series data were categorized to suit model training. For example, for the time data, four parts were categorized into dawn, morning, afternoon, and evening, Monthly data were extracted ,assuming that severity could vary with season. The purpose of this project was to classify how serious the damage caused by the collision was, so three representative models, KNN, Decision Tree and SVC, were used to classify.

4. Results section

The results of the evaluation indexes for each model are as follows. 스크린샷이(가) 표시된 사진

자동 생성된 설명스크린샷이(가) 표시된 사진

자동 생성된 설명

스크린샷이(가) 표시된 사진

자동 생성된 설명

5. Discussion section

Based on the evaluation indexes of the models used in the project, it has about 80% accuracy, so it is quite a good model. However, because there were only 1 and 2 values of Y\_data that were intended to be used, there were many difficulties to be used directly in the industry, and even though it could not tuning hyper-parameters in the process of training the model, it could not increase the accuracy. Finally, in the process of training data, the number of data was over 100,000, so SVM model is unable to learn all of the data for time reasons, so they could not design a better model. There is something to be desired in this regard.

6. Conclusion section

So far, we have designed a simple machine learning model that categorizes the severity of a collision in different situations. Although it showed a slightly disappointing 80 percent accuracy, I think a much better classification model will come out if we make up for what was mentioned in the discussion section a little bit.