Developing a Predictive Model for On-Campus Crime Using Machine Learning Algorithms and Reporting via Mobile App

Ghatak, Aratrika

On-Campus crime is a leading issue in US Colleges. US Department of Education captures data related to various types of crimes in US colleges. In my project, I wanted to leverage that dataset, collect other related data and then run various Machine Learning algorithms to find the best predictive model. Finally, I will empower students with a mobile application to easily access the trend and prediction. First, I gathered on-campus crime data at each individual crime type level, demographic, social and economic census data from various websites. I assigned each college a 'Crime Severity Index'(CSI), which is a weighted score considering severity of each type of crime. Each college was assigned a safety grade ranging from A+ to F based on their CSI. The next phase was to run various Machine Learning algorithms with my training dataset, compare the key metrics and find up with a best fit. I ran Decision Tree, Bayes Net and Logistic Regression on my training dataset (approximately 1.5M data points). The last phase was to develop a mobile app for users to access the historic trend and predicted results. Based on the key measures of various algorithms, Decision Tree was found to be the best fit and could predict on-campus safety grade correctly for 87%. It also had better accuracy, precision and recall values. This experiment was done using 10-fold cross-validation technique. Moreover, it was important to remove the outliers so that the model does not over-fit. In conclusion, my project has demonstrated that Machine Learning can effectivity be used to create a predictive model for on-campus 'Safety Grade'. My project also educates the prospective students by providing a mobile application to check detailed crime statistics, trends and future prediction.

Awards Won: