



CALIFORNIA STATE SCIENCE FAIR 2017 PROJECT SUMMARY

Name(s) Aratrika Ghatak	Project Number S1510
Project Title Developing a Predictive Model for On-Campus Crime Using Machine Learning Algorithms and Reporting via Mobile App	
<p style="text-align: center;">Abstract</p> <p>Objectives/Goals On-Campus crime is a leading issue in US Colleges, a serious concern for parents and a difficult challenge for the college authorities. 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 from various government agencies and then run several Machine Learning algorithms to find the best predictive model. Finally, I will empower prospective students and authorities with a mobile application to easily access the trend and prediction.</p> <p>Methods/Materials The project has been conducted in 3 phases. First, I gathered on-campus crime data at each individual crime type level and related demographic, social and economic census data from various websites. I delegated each college a Crime Severity Index(CSI), which is a weighted score considering severity of each type of crime. Depending on the CSI of each college, I assigned them a safety grade from A+ to F. The second phase was to run various Machine Learning algorithms with my training dataset, compare the key metrics and to come up with a best fit. I ran Decision Tree, Bayes Net and Logistic Regression on my training data set which is approximately 1.5M data points. The last phase was to develop a mobile app for users to access the historic crime statistics at detailed level, trend and predicted results. I used WEKA as Machine Learning Software, MIT AppInventor2 as app development platform and Google Fusion Database as cloud platform to store information.</p> <p>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 validation was done using 10-fold cross-validation technique. Moreover, it was important to remove the outliers so that the model does not over-fit.</p> <p>Conclusions/Discussion In conclusion, It is evident that law enforcing agencies, college security and college authority can take great advantage, using machine learning algorithms like Decision Tree to effectively fight crime on the campus and residence hall. My project also educates the prospective students, parents and other visitors in the college by providing a handy mobile application to check the detailed crime statistics, its influencing factors, historical trends and future prediction.</p>	
Summary Statement My project demonstrates that Machine Learning can effectively be used to create a predictive model for campus safety grade and provides an important tool to fight against on-campus crime.	
Help Received I have collected necessary data, consolidated and trained the model by myself. Thanks to my teacher, Mr. Higgins for supporting and sponsoring my project and thanks to The University of Waikato, New Zealand for making WEKA freely available for everybody to use.	