

# Writing Assignment 4

Qi Mao

Haowen Luo

maoxx241@umn.edu

luoxx560@umn.edu

April 21, 2019

## 1 Introduction

Crime is always an important social security problem for every country. However, it can be studied and analyzed by using many different methods as Patricia L. Brantingham claimed.[1] Then we build a model with some algorithm to review the record of crime in a region and analysis it. What specific problem we should focus on is the next problem. “Crime cannot be predicted since it is neither systematic nor random.”, said by Shiju Sathyadevan.[7] However, Mr. Sathyadevan thinks crime prevention is a systematic method for identifying what kinds of crime happening. That is support for our crime prediction project. We think crime prediction would be very useful for social security department to prevent crime or protect the safety of officers. For this project, we would like to use Random Forest, Logistic Regression, Naive Bayes and decision tree. We will discuss them in detail in the description part. “USING MACHINE LEARNING ALGORITHMS TO ANALYZE CRIME DATA” is a good example project which analyzes the data of the FBI’s Uniform Crime Reporting (UCR) program.[5] We think it would be an example for inspiring us to finish our project by step and step.

## 2 Description

Logistic Regression was used to classify individuals in the target categories based on the logistic function and it forces estimated probabilities to lie within the range 0-1, which is more sensible than linear regression.[4] In Liu's paper, he builds three models for predicting violent re-offending. They provided the accuracy and AUC range of the three models. And they provided that the NNs was slightly better Than that Logistic Regression model, it did not demonstrate a significant improvement.

Random Forest can have profound effects on prediction quality as well as the to be introduced variable importance measures.[2]. In Chen's paper, he used the random forest to find the most important features. And they provided that the random forest can find several expected related features. We can use that to find something interesting.

Naive Bayes has the ability to predict the probability that a given tuple belongs to a particular class.[3] They also mentioned that if talking about solving complex classification problems, then naive bayes is not a recommended choice. In our project, we will clean up the data and make the question simple. So we can use Naive Bayes that can short training time and fast evaluation.

"we classify crime data using decision tree into two classes, such as danger and neutral.", said by Nasridinov. [6] They provided that decision tree can analyze are good enough to motivate us to use a decision tree in order to predict future crimes. So we can also try to simplify our question that we only consider the question only have two classes.

"An ROC curve is useful for comparing the relative performance among different classifiers."[8] And precision and recall are also widely used metrics employed. In our project, we will use them to compare which classifier has the best performance.

### 3 Conclusion

In our project, we discuss the precision, recall and ROC curve of four different algorithms we discussed above, then comparing the performance with each algorithm. We will find the relationship between the location of the crime and the type of crime.

### References

- [1] Patricia L Brantingham, Paul J Brantingham, et al. Mapping crime for analytic purposes: location quotients, counts, and rates. *Crime mapping and crime prevention*, 8:263–288, 1998.
- [2] Daniel L Chen, Markus Loecher, Nora Barry, Laura Buchanan, and Evelina Bakhturina. Events unrelated to crime predict criminal sentence length. Technical report, Technical report, 2016.
- [3] Rizwan Iqbal, Masrah Azrifah Azmi Murad, Aida Mustapha, Payam Hassany Shariat Panahy, and Nasim Khanahmadliravi. An experimental study of classification algorithms for crime prediction. *Indian Journal of Science and Technology*, 6(3):4219–4225, 2013.
- [4] Yuan Y Liu, Min Yang, Malcolm Ramsay, Xiao S Li, and Jeremy W Coid. A comparison of logistic regression, classification and regression tree, and neural networks models in predicting violent re-offending. *Journal of Quantitative Criminology*, 27(4):547–573, 2011.
- [5] Lawrence McClendon and Natarajan Meghanathan. Using machine learning algorithms to analyze crime data. *Machine Learning and Applications: An International Journal (MLAIJ)*, 2(1):1–12, 2015.

- [6] Aziz Nasridinov, Sun-Young Ihm, and Young-Ho Park. A decision tree-based classification model for crime prediction. In *Information Technology Convergence*, pages 531–538. Springer, 2013.
- [7] Shiju Sathyadevan, Surya Gangadharan, et al. Crime analysis and prediction using data mining. In *2014 First International Conference on Networks & Soft Computing (ICNSC2014)*, pages 406–412. IEEE, 2014.
- [8] Pang-Ning Tan, Michael Steinbach, Anuj Karpatne, and Vipin Kumar. *Introduction to Data Mining (2Nd Edition)*. Pearson, 2nd edition, 2018.