Predicting Crime in Minneapolis

Alnahari, Suhail   Goyal, Vedant

May 7, 2017

Abstract

A lot of taxpayers' money is being invested in keeping our neighborhoods safe. So, in an effort to help fight crime, there is increased interest in the use of AI to create predictive models for crime. These models can be used for better resource allocation and understanding of crime itself. Our project focuses on crime data from Minneapolis and compares how three different machine learning algorithms perform in predicting crime - Naive Bayes, Decision Trees and Artificial Neural Networks. The algorithms are then compared to each other to show the differences in predictions of the models. Our project shows that using these AI technologies is feasible for building a predictive model and can be a part of strategies for maintaining public safety.

1 Introduction

Our project is about creating a way to predict the level of threat in a particular area depending on data from the past. The prediction is based on attributes like location and time. It is a problem which has real world implications and also gives insights into how different attributes of crime are related and how to comprehend them for predicting crime.

We used Naive Bayes, Decision Trees and Neural Networks as the three machine learning algorithms for testing. To build our model, we used data from OpenDataMinneapolis[Min16] project which offers a full database of crime reported in the city of Minneapolis in different years. The data includes attributes such as x-coordinates, y-coordinates, time of report, type of offense etc. These attributes can then be used to predict crime in a particular area at a given time. We used the data from the year 2016 to build our models.

We wrote our own software to manipulate and pre-process the data from OpenDataMinneapolis. We then used our results from our software on Weka[Hal16]. Weka is an open source software offering many models for data classification and analysis.

After we created the models on Weka, we did some analysis and compared the results of different classifiers. Also, we applied some methods to make our models easier to interpret and make them more effective. In the end, we tried to visualize our prediction results and compare it to actual crime data to check the accuracy of our work.

2 Rationale for Predictive Policing

When police officers patrol places of crime hot-spots, they usually patrol places where crime has already happened. While previous crime locations can be a good indicator of future crime, a more comprehensive predictive approach can often lead to better resource allocation for the police. For that reason, investments in accurate ways that predict future crime locations, should be about more than just analyzing current crime locations [Cas11].

A method of future predictions is using Geo-spatial statistical analysis [Cas11]. The difference between Geo-spatial statistical analysis and density maps (hot spot maps), is that it characterizes the locations with past events and creates a model that incorporates environmental factors statistically associated with past incidents [Cas11]. Furthermore, the model that was created based on the characteristics of that locations, can identify other locations with the same characteristics, hence predicting future hot-spots. A certain study, that gives credence to this idea of policing, concluded that future crime is six times more predictable by address than by tracking high-risk individuals [LB89].

Several other techniques have been used and have had varied success. Often, the benefit of successful crime prediction tools is not only better allocation of city resources, but also insights in the factors surrounding crime that can actually be used in reducing overall crime. Either way, being correct more than 50% of the time in predictions is good enough to be investing in this field given that even small differences can lead to huge public savings. And, in this project, we came up with results that show an improvement over random chance which solidifies the arguments on why investments should be made in this field.

3 Previous Work

3.1 Discretization Techniques

One key process before we can apply Naive Bayes/Decision Tree algorithms to a data set, is discretizing continuous attributes. In the case of our crime data, we need to discretize the location and time of crime, in addition to the level of crime attributes. A review of basic discretization methods and their advantages/disadvantages is given by Petre, who compares techniques like equal width, equal frequency and entropy based methods[PET15]. The author applies Naive Bayes on a Credit Approval dataset (from UCI learning Repository) using RapidMiner data mining software which supports discretizing based on the previously mentioned methods. According to results from the paper, applying any discretization method improves the performance of the classifier with the best results achieved by using equal width based intervals (equal width performed slightly better). In our data-set, we have to apply two levels of discretization. First, we discretize the locations into separate partitions. Then, we aggregate the number of crimes occurring in the area over time. To get enough crime in a particular region, we can aggregate crimes that occur in the region during the same time of the day through the year. We can discretize the time into 4 parts such as morning, afternoon, evening, and late night. We will discretize the locations into a grid with equal sized cells. Similarly, time of day will be discretized into equal time intervals. The results from Petre show that equal width, even though simple, tends to work well. So, we have decided on this approach as compared to using more sophisticated discretization approaches.

3.2 Applying Naive Bayes to Crime Data

To choose classifiers, we look to some of the earlier works involving using different classifiers to get some insights. In a paper about crime prediction in LA, California and Denver, Colorado, the Naive Bayes Classifier had best accuracy of 51% in Denver and 54% in LA[TL15]. On the other hand, the decision tree Classifier was 42% and 43% respectively. Although this is a significant difference, both algorithms have proven to be useful in many domains. Given that our attributes would be different than the work done before and every data-set is different, we decided to use both Naive Bayes and Decision Tree for our comparison. Both are well established algorithms used in several AI domains.

Since there are many attributes in our crime data, we will try to filter information that is going to be valuable from other data for our calculations. For example, location, time of day and crime type will be included. On the other hand, the control number or the crimeID will be filtered out since they clearly do not add any valuable information. Also, we would remove some of the records that do not have complete information. After filtering and reducing our data to date, crime type and time of day, we will put our data into subgroups. For example, in work by Almanie[TL15], the categories were transformed to fall within subcategories. For instance, crime type was put into 6 categories instead of 14 and time of crime to 6 instead of 24.

3.3 Characteristics of algorithms for Predictive crime mapping

Previous work done in doing predictive crime mapping is useful to understand what factors were found to be useful for prediction and also give a sense of potential weaknesses of any approach, since many methods tend to suffer from similar flaws. Early work in the field and the different prevalent methods were summarized by Groff and Vigne [GV02] , which gave insights in what techniques were found to be more robust and which variables in prediction were more crucial. Groff found that simpler algorithms, like exponential smoothing or repeat victimization tracking, worked as well or even better than more advanced techniques, like Neural networks. Also, examining crimes over longer periods, a year for example, proved much more accurate than using shorter period like a week [Spe95]. This corresponded to the fact that there were long term crime hot-spots but significant swings over the short term.

Another important observation is that greater the volume of crime, the more accurate the prediction will be. In terms of the minimum area that should be considered as one cell of a grid, Gorr and Olligschlaeger suggest that at least 4000 ft square are the smallest area that should be used[Gor00]. When it comes to identifying factors for crimes, Adams-Fuller found in her examination of socio-economic factors that public housing and economic depression were key factors [Ada01].

3.4 Crime and Geography

Understanding the way that one geographic place affects surrounding places gives a better understanding on why certain incidents happen and the reason why they happen [Com97]. In other words, this explains how different characteristics and effects of a neighborhood and those of surrounding neighborhoods might affect the crime in the area. For example, a area close to another area which has many 311(non-emergency complaints) might experience more crime. So, crime visualization is not enough to fight crime. Criminal analysts have to asses neighborhood characteristics and socioeconomic properties of areas of high criminal activity. Crime analysts have to combine applied geography and research of the characteristics of a geographical area, along with criminal theories and reasons why criminal instances might occur in a given area [Dav08].

In order to analyze geographical areas, scientists must have a supply of incident data, police calls and reports, base data and physical characteristics of the crime's location [Dav08]. Together with crime theory, applied geography helps analysts formulate a close prediction that will help fight crime [Dav08].

Making correct predictions and finding areas of crime not only helps preventing it, but also helps governments predict what areas need more development in areas of education, economics, and developmental projects. If these predictions were used only for policing, then it's either that the problem disappears, diminishes, disperses, or creates other problems. So, only policing crime ridden areas more is not enough to prevent those crimes in the long term [Dav08].

3.5 Ethics of using AI for crime prediction and analysis

Although predictive policing is effective, like many Artificial intelligence applications, it faces some ethical issues. If police predicts that a certain area encounters crime more than other areas, officials will saturate the area with police patrols and enforce the law more strictly in those areas than other areas which can be problematic [Cas11]. There is rising concern over policing of minority populations. Crime prediction, as we previously explored, can often depend on socio-economic factors like income, unemployment etc. This will often expose minority and economically disadvantaged people to more policing which can be counter productive. There is even a greater danger that as human interpretation and influence is reduced in crime prediction, we might fall prey to hidden variables in complex algorithms that can have results that are discriminatory towards particular sections of society. Thus, it is important to not fall for the promise of predictive AI and completely yield control of crime resource allocation to algorithms whose working cannot be readily understood by the different stakeholders in law enforcement.

4 Initial Crime Data

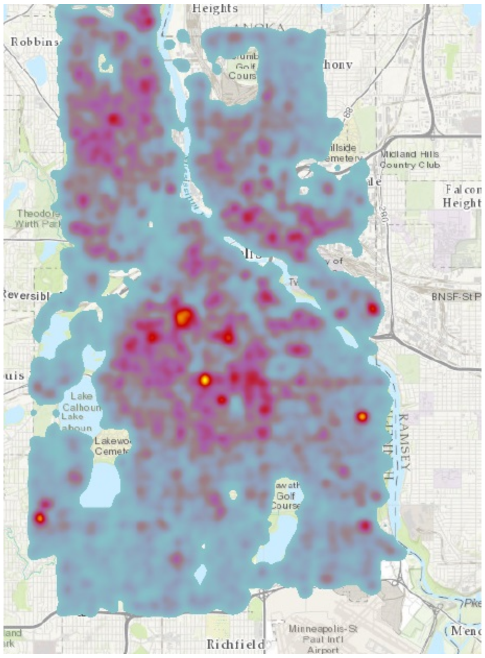

Figure 1: 311 Heat map

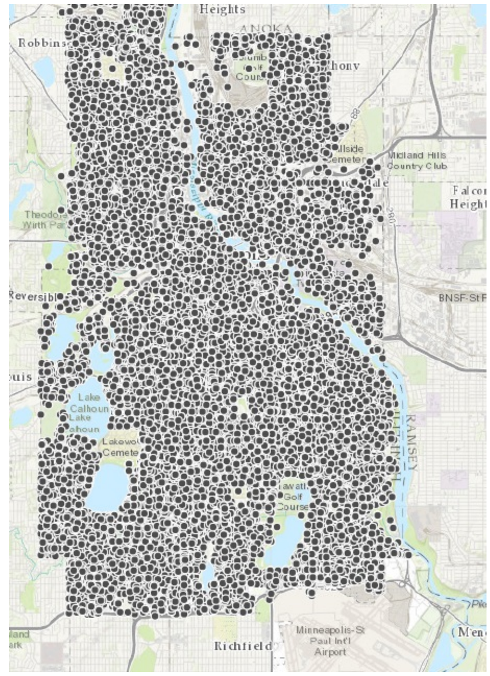

Figure 2: 311 instance map

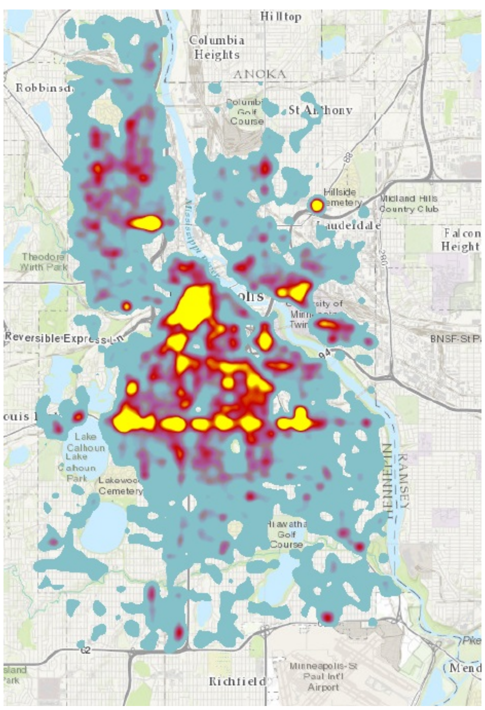

Figure 3: 311 Heat map

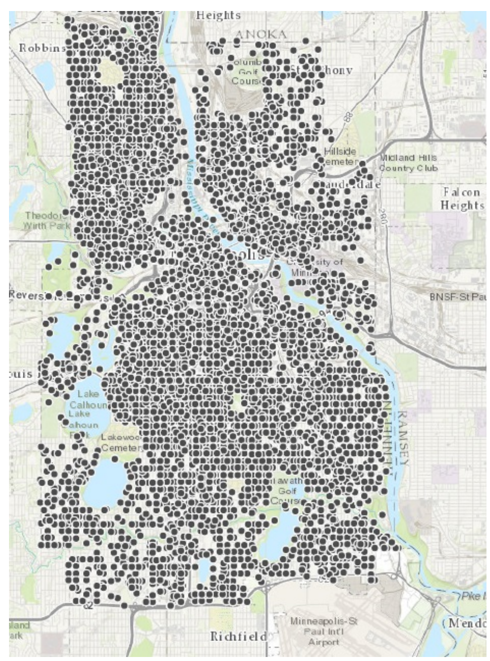

Figure 4: 311 instance map

4.1 Analyzing initial crime data

We needed to visualize the crime instances that happened in Minneapolis in 2016. So, we used ArcGis[ESR17] software, that takes spreadsheets and converts them to different kinds of maps. We visualized both 311 and 911 data to see if there are threats in the same areas where there are crimes. After getting the results, we noticed that, in 2016, crimes were concentrated at the center of the city and to a less extent towards the edges. However, 311 reports were concentrated towards the center and in the Northern side of the city. That means predictions based on 311 reports aren't as accurate to predict crime. So, logically, our classifiers need to be based on 911 calls. Although it might be the case that there is a more complex relationship between 311 and 911 instances which can only be leveraged by algorithms and not seen easily by visual inspection. Nonetheless, we focus in this paper on using only actual crime incidences for building the models.

5 Data Processing

We got our data from OpenDataMinneapolis, so, we had to filter columns that were unnecessary. We removed many of the non useful attributes like IDs and also removed records that did not have complete information. Only a few records had to be removed as most records were complete. After putting all the information from the file to our software, we scaled the longitude, latitude, time and date to fit a 4-dimensional grid. If we were to look at latitude and longitude of crime, the location of crimes in Minneapolis occur in a very small domain, so we made an educated guess of choosing a range of 0-12. The reason behind that is that 12 is large enough to provide some granularity for the location but not too large so as to make crime incidences very rare which can lead to class imbalance. While class imbalance can be handled by some algorithms, it can cause bad results for others and can make predictions more complicated.

One challenge when working with the data was that an output class did not exist on which predictions could be made. The data only indicated instances when crime occurred. To be able to run classification algorithms on the data, we had to transform the data so that we could have different levels of crime. To do this we chose to divide latitudes and longitudes in 12 by 12 partitions and then have 4 partitions for time and 12 partitions for months. In this way we were able to obtain 6912 unique instances. For each of these instances we aggregated all of the crimes that satisfied the attributes for that instance. We ended up with a portion of instances with no crime and the remaining portion with some level of crime. We kept the output class simply as No Crime or Some Crime to make the model simple. While we could have possibly made the output class a continuous attribute, it would have made interpreting the results harder. We would not have been able to get a percentage of accuracy but rather only the average error in predicting the number of crime which is a less directly meaningful statistic. Hence we chose to do a binary output class of No Crime or Some Crime. After putting the data into our grid, we made sure that our data is random by applying a randomizing function. This ensures that there is no bias between our training data-set and our testing data-set.

Then to use Weka, we had to convert our data to a specific format and ensure that all of our categories were appropriately aligned for the correct functionality of the software. Also, we divided our data into two halves so that we can use one half in creating and training a model on our classifiers. Then we based our final results on the model tests of the other half.

After using Weka on our data, we created another software that processes the result of the classifiers and draws a table of 12x12 longitude and latitude rows and columns, and, in each cell, indicates how many crime instances were predicted to occur. Then, using excel, we colored the crime instances based on frequency. The complete colored tables helped us in visualizing our results and verifying our predictions visually. We made the heat map tables for the actual crime, Naive-Bayes, Neural Network and Decision Tree results. Figures 5,6,7 and 8 show these results.

6 Results

6.1 Naive Bayes

Commonly used in Machine Learning, Naive Bayes is a collection of classification algorithms based on Bayes Theorem. It is not a single classifier but a family of classifiers that all share a common concept, which is that every feature being classified is independent of the value of any other feature. For instance, an animal is a Zebra if it has stripes, is about 6 feet tall, and has hooves. A Naive Bayes classifier interprets these characteristics to be independent in the probability that the animal is a Zebra, regardless of any correlations between features. Because of that, Naive Bayes is considered to be "Naive" since there may or may not be correlations between the characteristics[TSK05].

Although Naive Bayes looks simple, it outperforms some of the many other sophisticated algorithms in some scenarios. For that reason, we decided to use it as one of the algorithms in our project.

This algorithm works by combining the independent probabilities of different characteristics, which in our case are location, date and time. Below are the results of the predictions. Positives and negatives class represent if crime happened and didn't happen respectively.

		Prediction	
		Negatives	Positives
Actual	Negatives	1097	387
	Positives	203	1769

Table 1: Naive Bayes Prediction Result

Correct	82.93%
Incorrect	17.07%

Table 2: Naive Bayes Prediction Result

From the results we can see that the Naive Bayes does a good job at predicting whether crime happens for the given attributes. The predictions tends to skew more towards giving false positives as seen from the confusion matrix. This is expected since the data contains more cases of no crime vs cases of crime.

6.2 Decision Tree

Also used in Machine Learning, Decision Tree is a type of classifier that extracts a prediction by forming a tree of decision partitions. The way Decision trees work is by taking all the data and depending on a characteristic, form partitions that when tested, help make a prediction. For example, if we're trying to know if a fruit is an apple, we might ask is it round? If it is, then it narrows our search to only round fruits. We can then ask more questions that further narrow our search. Decision trees create models by asking differentiating questions. Decision trees have been proven to be effective in multiple applications. So, we decided to test it on our data. We used the J48 decision tree algorithm in Weka. Below are our results:[TSK05]

		Prediction	
		Negatives	Positives
Actual	Negatives	1159	325
	Positives	148	1824

Table 3: Decision Tree Prediction Results

Correct	86.31?%
Incorrect	13.69%

Table 4: Decision Tree Prediction Results

From the results we see that decision trees performs better than Naive Bayes. In fact it gives the best performance out of the 3 models that we test in our project.

6.3 Neural Networks

Neural Networks are more complex than the two previous models. Inspired by biological neurons, Neural Networks work by connecting nodes with different weights. Then, it adjusts the weights of the nodes until the rate of this decrease in error is minimized. This model comes up with a predictions based on the network of these weighted nodes. Neural networks have been found to have wide ranging applications and have been successful in making predictions in sophisticated scenarios. Hence, we used it to compare results with other models like Naive Bayes and Decision tree which are simpler and established earlier. Below are the results of our experiment.[TSK05]

		Prediction	
		Negatives	Positives
Actual	Negatives	1132	352
	Positives	200	1772

Table 5: Neural Network Prediction Results

Correct	84.03%
Incorrect	16%

Table 6: Neural Network Prediction Results

From the results we can see that neural networks perform only marginally better than Naive Bayes but not better than decision trees. One problem with using Neural Networks is that it is the most opaque algorithm in its working and hence law enforcement officials and the public at large might be hesitant to use an algorithm whose results are not easily understood especially for a sensitive area like crime.

6.4 Data comparison

5	20	15	26	15	22	16	12	19	14	4	0
20	19	9	19	21	20	19	7	20	24	14	0
16	14	8	25	22	21	18	16	28	20	12	0
17	8	24	21	21	25	25	29	25	21	13	0
12	17	23	26	26	23	16	23	27	12	3	0
4	16	20	21	19	22	23	25	19	22	11	0
4	23	27	22	24	23	23	24	31	19	2	0
1	22	28	21	21	22	18	20	21	12	1	0
5	24	27	27	16	20	22	15	9	1	0	0
7	24	19	20	4	8	8	15	16	0	0	0
9	19	25	24	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0

Figure 5: Actual crime instances

0	26	25	30	19	26	25	24	19	1	0	0
0	26	27	25	22	24	25	21	25	24	0	0
0	28	19	28	23	21	20	18	30	19	0	0
10	26	24	21	21	25	25	29	25	24	7	0
8	26	23	26	26	23	16	23	28	22	13	0
7	25	22	21	19	22	23	28	19	25	1	0
4	24	27	23	24	23	23	25	31	23	1	0
0	22	28	21	27	23	19	21	24	24	0	0
0	24	27	27	26	20	23	20	21	15	0	0
0	26	19	20	29	27	22	15	20	0	0	0
0	0	1	24	0	9	1	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0

Figure 6: Naive Bayes results

0	26	0	30	19	26	25	0	20	28	0	0
25	26	27	25	22	24	25	0	25	26	25	0
27	28	0	28	23	21	20	18	30	20	19	0
21	0	24	21	21	25	25	29	25	24	17	0
3	26	23	26	26	23	16	23	28	22	0	0
0	25	22	21	19	22	23	28	19	25	20	0
0	24	27	23	24	23	23	25	31	23	0	0
0	22	28	21	27	23	19	21	24	0	0	0
0	24	27	27	26	20	23	20	21	0	0	0
0	26	19	20	0	27	0	24	0	0	0	0
0	22	27	24	0	0	0	0	0	0	0	0
0	0	0	34	0	0	0	0	0	0	0	0

Figure 7: Decision Tree results

3	24	19	27	18	26	17	10	19	11	2	0
20	25	20	17	21	24	20	12	23	21	25	0
14	23	8	26	23	21	20	17	29	18	14	1
15	7	24	20	21	25	25	29	25	22	15	3
10	24	23	26	26	23	16	23	28	20	13	0
5	22	22	20	19	22	23	27	19	23	10	1
7	24	27	23	24	23	23	25	31	21	4	1
4	21	25	21	27	23	19	19	17	16	0	0
3	24	27	27	19	17	23	20	13	1	0	0
3	26	19	20	5	20	15	21	12	1	4	0
0	21	23	20	2	3	0	5	1	0	0	0
0	3	2	5	0	0	0	0	0	0	0	0

Figure 8: Neural Networks results

In Figures 5, 6,7 and 8, we notice that the edges of the tables are green in color, which proves our observation in the beginning that crimes in Minneapolis are concentrated in the center of the city. The columns represent the longitude and the rows are the latitude which have been scaled to 12 by 12 indices. The figures represent a heat map of the crime occurring in the actual data and the crime predicted by the models. To represent the results on the map, the dimensions have been reduced to the latitude and longitude by summing over the months and time because we wanted to visualize crime in the entire year rather than specific times or dates. Notice that the lower right of the map shows no crimes because of the natural shape of the Minneapolis area which does not fall into those longitude and latitude range.

From the maps we can see that the decision tree seems to get some areas completely wrong while doing other areas correctly. This is from the nature of decision tree which tends to make definite decisions based on certain attributes. Compare this to Naive Bayes which has a more spread out map, an artifact of its probability based model. In case of the neural networks we see a better distributed map. Due to the sophisticated pattern of weights which do not give the kind of certainty as decision trees, we see predictions of crimes even in areas where there never has been one. This flexibility of the model works both for and against the model. So, because of that, Neural Network, though has less correct predictions, is more resistant to making large errors in predictions, which may be more important where crime predictions are concerned .

7 Future Extensions

There is much room for improvement in the models and future extensions to the project. One shortcoming is that only four attributes were used to make the predictions. We can potentially add more attributes like type of offence, particular neighbourhood characteristics etc. Another particular property of this problem are the spatial and temporal auto-correlation. Crime patterns change over time and crime occurring over more recent past can be more relevant than crime occurring much before. Crime can possibly exhibit both of these properties and our model does not take advantage of these properties.

One of the immediate extensions of our models could be experimenting with the parameters such as number of divisions made of the attributes. Changing the granularity of the discretization can potentially improve the predictions. Another extension that can be made is to explore the relationship between crimes and 311 calls. Since 311 calls are much more numerous, they give a good data-set to augment the data for crime. As predicted by some social theories like broken windows, we could potentially exploit the relationship between general lack of order with crime.

Also, there could be plans to add potential felons into the model. This would of course involve the legal and ethical concerns that should involve a public debate over what would be the best criteria for using such a technology and where the boundaries should be drawn. The future of AI crime research must involve an ethically informed discussion over the issues that it influences.

Another aspect is considering the socio-economic factors of a particular area. For example, the percentage of unemployed people or the percentage of people between 26-40, because, according to the Federal Bureau of Prisons, is the most common age for criminals[Pri17].

8 Conclusion

Crime prediction is a problem which seems to yield well to the application of predictive AI and data analysis systems, because of the well documented and maintained data and also the usefulness of the application. At the same time, it is important to remember that crime is an extremely complicated phenomena which overlaps with several random, unpredictable and complex elements like human behaviour, coincidences, policy etc. We find that the best approach is to use simple techniques which can be interpreted easily. We compared Naive Bayes, Decision Tree and Artificial Neural Network algorithms. We found that decision trees performed best although other algorithms did only marginally worse. Looking at some of the other metrics like the heat maps, one could also make the case that Decision Tree did not perform as well because some of the predictions were very wrong which could be crucial when talking about resource allocation for crime. Furthermore, for future research it would be helpful to use data beyond just incidents of past crimes. This can include anything relevant, from environmental characteristics to socio-economic factors. At the same time we have to be cognizant of how our algorithm is using these factors and whether the conclusions can be used in an ethically appropriate way to fight crime. In conclusion, Artificial Intelligence and Machine learning have invented a completely new field for fighting crime and ensuring public safety. As a result, public officials need to integrate these fields into their forces.

9 Division of labor

Suhail: Introduction, Initial crime data, Analyzing initial crime data, Data processing,Results, Data Comparisons, Future Extensions, Conclusion, Crime and Geography, Applying Naive Bayes to Crime Data, Ethics of using AI for crime prediction and analysis, Applying Naive Bayes to Crime Data, Rationale for Predictive Policing

Vedant: Introduction, Analyzing initial crime data, Data processing,Results, Data Comparisons, Future

Extensions, Conclusion, Naive Bayes Discretization Techniques, Rationale for Predictive Policing, Characteristics of algorithms for Predictive crime mapping, Ethics of using AI for crime prediction and analysis

References

[LB89] P.R. Gartin L.W. and M.E. Buerger. “Hot Spots of Predatory Crime: Routine Activities and the Criminology of Place.” In: *Predictive Policing as a Part of 1989* (Jan. 1989), 27:27–56.

[Spe95] W. Spelman. “Criminal Careers of Public Places.” In: vol. 4. Monsey. 1995.

[Com97] National Research Council Rediscovering Geography Committee. “Rediscovering Geography: New Relevance for Science and Society”. In: *National Academies Press* (1997), pp. 29–57.

[Gor00] W.L. Gorr. “Approaches to Crime Predictive Modeling”. In: San Diego. Dec. 2000.

[Ada01] T. Adams-Fuller. *Historical Homicide Hot Spots: The Case of Three Cities.* Tech. rep. Howard University, 2001.

[GV02] Elizabeth R. Groff and Nancy G. La Vigne. “FORECASTING THE FUTURE OF PREDICTIVE CRIME MAPPING”. In: vol. 13. 2002, pp. 29–57.

[TSK05] Pang-Ning Tan, Michael Steinbach, and Vipin Kumar. *Introduction to Data Mining, (First Edition).* Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc., 2005. ISBN: 0321321367.

[Dav08] Carl R. Peed David W. Hagy. “Public Safety Geography A Quarterly Bulletin of Applied Geography for the Study of Crime & Public Safety”. In: *Geography and Public safety* 1.1 (Feb. 2008), pp. 1–7.

[Cas11] Tom Casady. “Police Legitimacy and Predictive Policing”. In: *Geography & Public Safety* 2.4 (Mar. 2011), pp. 1–14.

[PET15] Ruxandra PETRE. *Enhancing Forecasting Performance of Naïve-Bayes Classifiers with Discretization Techniques.* Tech. rep. 2. University of Economic Studies, Bucharest, Romania, Jan. 2015, pp. 1–7.

[TL15] Rsha Mirza Tahani Almanie and Elizabeth Lor. *Crime Prediction Based on Crime Types and Using Spatial and Temporal Criminal Hotspots.* Tech. rep. 4. Department of Computer Science, University of Colorado, Boulder, USA, July 2015, pp. 1–19.

[Hal16] Mark Hall. *The WEKA Data Mining Software.* `www.cs.waikato.ac.nz/ml/weka/`. Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, Ian H. Witten. 2016.

[Min16] City of Minneapolis. *Open Data Minneapolis.* `http://opendata.minneapolismn.gov/`. 2016.

[ESR17] Environmental Systems Research Institute (ESRI). *ArcGIS Geographical Mapping software.* `http://www.arcgis.com/features/index.html`. 2017.

[Pri17] Federal Bureau of Prisons. *Inmate Age.* `https://www.bop.gov/about/statistics/statistics_inmate_age.jsp`. 2017.