PREDICTION OF CRIME IN CHICAGO USING BAYESIAN AND NEURAL NETWORKS

ECE57000 PROGRAMMING LANGUAGES FOR ARTIFICIAL INTELLIGENCE

TEAM MEMBERS
Raghavan Naresh Sarangapani
Sreeram Venkitachalam

IUPUI

Date: 01-Dec-2016

CONTENTS

1 Abstract:

Crime occurrences are continuously being documented in datasets [1] and it is now possible to analyze, classify and predict the occurrences of the crime. The aim of this paper is to identify methods to cluster the crime data, identify patterns in it and run prediction algorithms to identify crime hotspots. Identifying crime hotspots at any time of the year can be used to prevent crimes.
The dataset with crime statistics from the City of Chicago [1], is used to identify patterns in crimes based on location and time of occurrence of Chicago. Then a prediction algorithm is used to predict the crime occurrence for the future.

Crime hotspots are areas on a map that have high crime intensity. They are developed for researchers and analysts to examine geographic areas in relation to crime [2]. These pocketed areas in the city carry with them around 50% of the happening crimes [16]. The representation of crime hotspots over the month of the year can give better understanding of the crime occurrences across Chicago.

The dataset [1] contains attributes such as Case ID, time and date of the crime occurrence, report update date, codes for beats, communities, wards and districts, location co-ordinates of the crime occurrence and the types of crime based on Illinois Uniform Crime Reporting Code [2].

This paper approaches the analysis of dataset [1] based on dividing the dataset into training and testing sets. The training set is used for classification based on both the Naïve Bayes classifier and the Neural Network. The performance of both the algorithms is analyzed and the best algorithm is used for further classification of the testing set. The classification of the dataset is done by dividing the time into various months of the year and the crime patterns are analyzed for each month of the year.

The prediction algorithm finds the probability of occurrence of a particular crime in the month of the year for a given location. This prediction can be verified within the test set to measure accuracy based on classification of actual crime occurrences. The feedback from this process is used to update the thresholds of crime prediction.

Finally, the output of the prediction of crime hotspots is depicted graphically in the map of City of Chicago. This prediction can be depicted for the past years for showing the algorithm's prediction efficiency.

2 Literature review:

There are several papers reviewed in relation to the crime rate prediction and hotspots creation. In paper [4] the authors Rasoul Kiani et al, propose methods to cluster crimes based on occurrence and to assign weights to the values for improving the accuracy of the dataset. Apart from this the authors use the Genetic Algorithm to improve the performance of Outlier Detection [4]. The paper [4] uses k-means algorithm, but the k-means algorithm has a possibility of settling into a local minimum. Secondly, the prediction method utilizes decision tree method which is prone to local-optimum decision and also result in over classification of data [5].

Authors Sathyadevan, Shiju, and Surya Gangadharan, in paper [6] propose usage of Naïve Bayes [7] classifier to identify the type of crime from various text reports. This method also used Named Entity Recognition to classify names, locations and organizations etc. For pattern recognition, the authors used the Apriori algorithm [8], and for prediction, the authors used decision tree process. Finally, the output is represented in a heat map to indicate the areas with higher crime rates. This paper similar to [4] utilizes decision tree which is prone to local optimum decisions.

In [9] authors by Tahani Almanie et al, discuss ways to find the temporal and spatial hotspots criminal hotspots in a city by reducing the classification models, to reduce algorithm complexity.

In paper [10] authors Nath, Shyam Varan, use a weighted k-means clustering algorithm to cluster the crimes. The distinction from other papers is that the authors of [10] use expert opinion to classify the crimes and court rulings to analyze the efficiency of the prediction.

In [11], the authors, Lawrence McClendon and Natarajan Meghanathan, compare the accuracy and effectiveness of Linear regression, Additive regression and the decision stump algorithms.

Yufei Zhang and Huifeng Ji, in paper [12], used a combination of Geographic Information Science (GIS) and Markov chain model to analyze crimes. But, the key of the Markov chain model is the reliability of transition probability matrix and for ensuring the accuracy of prediction, large and accurate statistics datum should be given [12]. As a result, lot of data in the data set is required for the accurate calculation of the prediction.

In paper [13], authors Hua Liu and Donald E. Brown, mainly discuss about predicting the crime location of the crimes in the coming week based on the previous week data using a multivariate prediction model for hotspots. The paper employs a class of cohesiveness measures that do not require any partitioning. An equation of the transfer function basically determines the error from the uniform pattern (hotspot). For getting the space time model prediction the transition density model is used which in the 1st step separates spatial and temporal transitions. This Research paper has a very similar research question as our research question. This can be expanded and the prediction of the crime can be extended to a year along with the addition of a safety index for each area we will be able to produce a better algorithm.

Authors Mohammad A. Tayebi et al, in paper [14] research the scope of detecting the number of crimes in "coldspots". The author creates anchor locations based on familiarity of frequent offenders and used the Gaussian distribution to analyze the human movement around particular points. But it fails to address the issue of offenders who are not from the same state and only regular offenders are taken into consideration.

3 Methodology:

3.1 Data Collection:

The dataset obtained from the City of Chicago [1], lists the crime occurrences in the city from 2001 to present. This dataset is updated frequently [1] and hence can also be used to analyse the efficiency of the prediction.
The attributes of this dataset can be classified based on their functionality:

3.1.1 Basic details of incident:

The basic details like ID of the incident and report, Date in which they occurred are given in the dataset. Two separate attributes "date" and "updated on" are used in the dataset, the former is used to mention the date and time of the incident, whereas the latter is used to mention the date and time when the report was updated.

3.1.2 Type of Crime:

The type of the crime is identified using the Illinois Uniform Crime Reporting Code [3], by using the attribute “iucr”. Further the type of the crime can also be elaborated and it is given by the attributes “primary_type” and “description”. Another way to classify the crime is by using the FBI's National Incident-Based Reporting System (NIBRS) [19], and this is done in the dataset [1] using the attribute “fbi_code”.

3.1.3 Location of Crime:

The location of the crime is limited to block level accuracy for privacy reasons [1], and hence a rough location of the incident is given in the dataset. Location is mentioned by using two ways:

3.1.3.1 Using Police districts:

The attributes “beat” and “district”, give out the location of the crime incident based on a map of the city. The city of Chicago is structured in a way that three-four beats form a sector, three sectors form a police district [16]. This classification is given at the beat level in [15] and at the district level in [16].

3.1.3.2 Using City Map:

The attributes “ward”, “community_area”, “x_coordinate”, “y_coordinate” and “block”. The attributes “ward” and “community_area” are classified using [17] and [18] respectively. The attribute “block” gives the address of the incident, but the block number is not stored for privacy reasons. Finally, the attributes “x_coordinate” and “y_coordinate” give out the co-ordinates based on the state plane projection [1].

3.1.3.3 Using GPS co-ordinates:

The attributes “latitude” and “longitude” are used to give a latitude and longitude co-ordinates and the attribute “location” is a combination of both in a format suitable to be applied on a map [1].

3.1.4 Action Taken:

The final classification is used to describe the action taken for the incident. The attribute “arrest” mentions whether an arrest was made and the attribute “domestic_violence” describes whether a domestic violence crime was committed [1].

3.2 Preprocessing Data:

To achieve better results from the clustering as well as prediction algorithm, we can assign weights to types of crimes based on their severity. Crimes like Murder, Aggravated assault could have better weights than public interference, intimidation etc. Initially these weights will be given with basic values, later this can be updated based on the user preferences and feedback from the efficiency analysis.

As the entire dataset consists of incidences from 2001 to 2016, we can divide based on the year of incident occurrence to two types:

1. The training set is obtained by the incidents occurred from the year 2001 to 2012. This is because having a larger training set gives more inputs for the prediction algorithm. Out of the total 6172977 individual crime occurrences, the training set contains 5133020 incidents. This training set is stored into a new csv file named “Trainingdata.csv”.

2. The test set is obtained by dividing the remaining incidents in the dataset, i.e. the individual years from 2013 to 2016. The algorithm is used to predict the crime occurrence in 2013. The number of individual events in this set is 1039957 as at the end of September 2016 [1].

Currently we have classified the data according to 3 primary types "HOMICIDE", "ARSON", "CRIMINAL DAMAGE". These are saved into separate files "HOMICIDE.csv", "ARSON.csv", "CRIMINAL DAMAGE.csv"

3.3 Classification of Data:

To classify the dataset, we plan to compare the performance of two classifiers namely:

1. Naïve Bayes Classifier.
2. Time series prediction with neural network

3.3.1 Naïve Bayes Classifier:

The Naïve Bayes classifier, which is based on the Bayes theorem, calculates the posterior probability [7] of each class to identify the outcome of the classification [20]. It involves forming a frequency table and a likelihood table to classify the dataset based on the given classes [20].

3.3.2 Time series prediction with neural network:

The Time series prediction using the neural network is used for the prediction of the next incoming data by sampling an input set. The activation values of the input values are weighted and accumulated at each node in the first hidden layer. The total is then transformed by an activation function, for example the sigmoid function $f(x) = 1/(1+e^{-x})$, into the node's activation value. It in turn becomes an input into the nodes in the next layer, until eventually the output values are found [23]. However, the Gradient descent method used in the backpropagation has a disadvantage that it may be stuck in a local minimum and will never converge onto the global minima.

For our dataset [1], initially we have classified the training set initially based on the attributes "year", "month of occurrence", "district". This is classified for homicide data at present, which can be extended to other Primary types like Arson, Criminal damage.

Further runs through the dataset [1], can classify the areas of crime occurrences based on the blocks in which they occur. These blocks together form the beats and hence can be used to identify crime hotspots. These hotspots can be further classified by dividing into different months of the year to identify patterns.

3.4 Prediction of Crime:

Once the crimes are classified the task of the prediction phase is to predict the likelihood of occurrence of a crime based on past occurrences. By taking a specific date of the year, and the type of the crime we can predict the probability of event occurrence by using the past occurrences. The specific crime dataset over all the months of the years 2001-2012 is used to train the prediction algorithm. The crime distribution over a month over all the years from 2001-2012 is used to predict the occurrence rate of the crime of the same month in 2013. Currently the homicide rate as a probability distribution is calculated for each district and each month separately.

For predicting the event occurrence by area, we can take in the block or beat level values and classify them based on the time of the year and type of the crime to identify crime hotspots. These crime hotspots change during different times of the year, and hence we calculate the hotspots for each month for a specific type of crime.

3.5 Training and testing:

In the training phase, the training dataset is used for crime classification based on various attributes. The training dataset is used to compare the performance of classification based on Naïve Bayes and Neural Network algorithms. The algorithms are analyzed based on efficiency, space and time complexity, and based on this analysis the better algorithm is used for the classification test set.

In the testing phase, the prediction is run for the year of 2013 by calculating the expected probability of crime occurrence for each month of the year. This is then compared with the classification of the test set. Any feedbacks are noted and the next year prediction is performed with better accuracy.

3.6 Feedback from the prediction results:

The advantage of this dataset is that it is updated frequently [1], and hence this gives a real-time feedback that can be used from the system to improve the classification and prediction of the crime.

Feedback to the algorithm is classified as:

1. Accuracy/Failure of calculated probabilities for crime occurrence in a particular area during a particular month. If the calculated probability does not match the actual event occurrence, then this can be noted as an aberration and updated while calculating the probability for the same month in the next year.
2. Identifying crime patterns from the classification can be useful in setting up thresholds to limit the number of crimes being analyzed. For example, if during a particular month the occurrence of murder is more than burglary in a certain area, then we can assign more importance to predicting the murder.

3.7 Output representation:

The output obtained by the classification and prediction algorithm is used to graphically represent crime hotspots during various months. For mapping this we use the combination of Location co-ordinates provided in the dataset [1] and the mapping of beats [15], districts [16], wards [17] and community areas [18].

Based on this the output can be represented in the following ways:

1. For the training set (year 2001-2012), we use the classification of the dataset to represent various hotspots identified during every month of the year.
2. For the testing set (year 2013-2016), we use both classification and prediction to represent two representations of hotspots to show the accuracy of the prediction for the months of a given year.
3. For the future years, we show the probable hotspots based on the month of year.

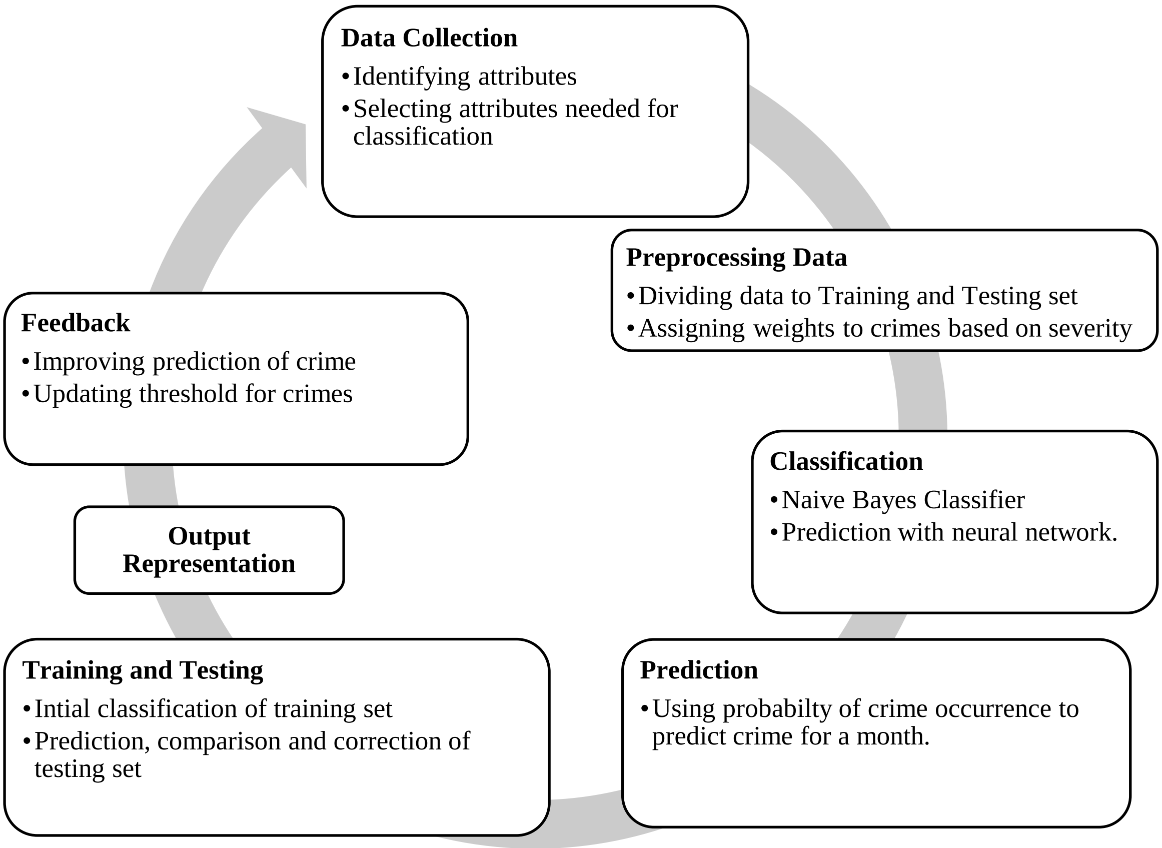

Figure 1 Flow-cycle for crime prediction

4 Implementation

For the implementation, the entire dataset [1] was downloaded in the Comma Separated Values(CSV) format. Next for the preprocessing phase the entire dataset was divided into Training and Testing sets. The criteria for selection of the training set was arbitrary and it was chosen as 12 years from 2001 to 2012, and these values were stored in a separate dataset. The testing dataset consist of the years 2013-2016 separated on a year basis into different datasets. The prediction algorithm will be first run on the training dataset to train, and after performance analysis the necessary additions are performed. Once this is done, the prediction algorithm can be used to predict the crimes of the year 2013 on a monthly basis.

Once the training dataset is produced, we separate them to further datasets based on the type of the crime, and initially this is done for three crimes, arson, homicide and criminal damage. Initially these three crimes will be used for the pre-processing phase and later other crimes will also be added for analysis. The source code for preprocessing the data is in source code

Based on the individual crime datasets for the years 2001-2012, the data for each month is also obtained. The prediction of crimes at present has been limited to homicides which can be extended to others crimes. The preprocessed data for homicides from the period 2001-2012 is saved inside HOMICIDE.csv which is used for processing.

Once the data has been classified we can verify it with the year 2013 data from the dataset [1]. For this we have developed another program that analyzes the efficiency of prediction of the algorithm over 2013. This program, at present calculates the probability of months of the year and districts. It will further also calculate the efficiency of the prediction by Least square error method. Based on the analysis by the testing program we can identify the areas where the classification went wrong and adjust the same

4.1 Neural Network Implementation:

For working on the neural network, the training data was divided into 12 sets of data, split according to the years. The probability of the crimes are for each district and each months are calculated for each year. This data is then fed into the neural network as the input. This data is then fed into the hidden layer along a path with some weights. The hidden layer is a sigmoid function. The output of the hidden layer is then given to the output node where the error is calculated with respect to the Testing data. This process is called the feed forward. The error is calculated using gradient descent method which converges on the global minimum. This error is used to calculate a delta value which is then back propagated to modify the weights to get minimum error. The future work will be mainly on the input features which is to be given as input to the neural network. The function for the implementation of the neural network is given in def NeuralNetwork.

4.2 Naïve Bayes Implementation:

In the Naïve Bayes based implementation, initially the dataset is explored to get the count of number of crimes over the years 2001-2012 distributed across 12 months and 25 districts. So now we get a combination of 300 individual counts to all the month and district combinations. For each year prediction, the entire 300 possibilities are ordered using Normal Distribution. These normalized values are mapped to 10 classes based on their z score as shown in Figure 2. Next individual class probabilities are obtained for each class, based on their number of occurrences over the 300 combinations. Now the District and Month are classified as two separate sources for crime occurrence and hence their conditional probabilities with respect to those of the class is found. Finally, as per Naïve Bayes method, for each class in each month and in each district, the final probability is obtained by multiplying the conditional probabilities of District and Months with the Class probability of the specific class. Based on these probabilities the class with the highest probability is chosen to represent the number of crimes for the current district and month combination. Apart from this the crime incidents are sampled for each fixed value of month and district and variable values of years (2001-2012 in first iteration). Using this sampling the Mean and variance are calculated, and hence the class value obtained previously is used to substitute in the Normal distribution formula to get the actual number of incidents for the give month and year.

Once this is done for one year, a weight matrix is prepared comparing the predictions with the actual values, this prediction matrix is added to the number of crimes in the subsequent years to and is also updated with each iteration. The source code for this is available in section 7.2 Naïve Bayes Implementation.

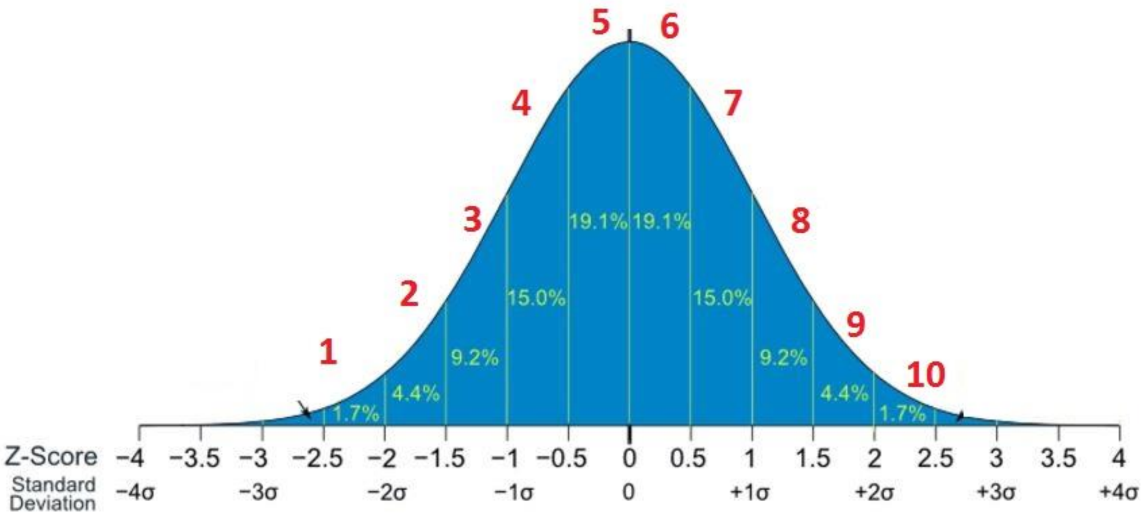

Figure 2 Normal Distribution Class division [24]

We can also compare the performance of both the algorithms, Naïve Bayes and the neural network. Based on this analysis the prediction algorithm probability and weights can be adjusted to predict the next year 2014 correctly.

Apart from this for each crime we can obtain a classification based on the location of the beats and the month of the year. This classification can be used to identify crime hotspots over the months of the year in the training dataset. This can then be extended for the year 2013 and the performance can be analyzed again by the testing program.
Finally, this can be used to represent graphically on a map to obtain crime hotspot trends for the past years and the prediction for the next year. This must be done with the help of location data available in the dataset.

5 Result

5.1 Neural Network Implementation:

1. During the first case the input to the neural network was the probability of the Criminal damage in each month and the probability of Criminal Damage in each district. The neural network had 1 hidden layer with 2 nodes. The nodes had sigmoid activation function. The following observations were made:

    a. The accuracy for the data was observed to be 88.59% when the neural network was run 100000 cycles. The RMS error for the probability was calculated to be 3.90. The graph representing percent error value for probability of criminal damage happening at each district at a specific month is shown below:

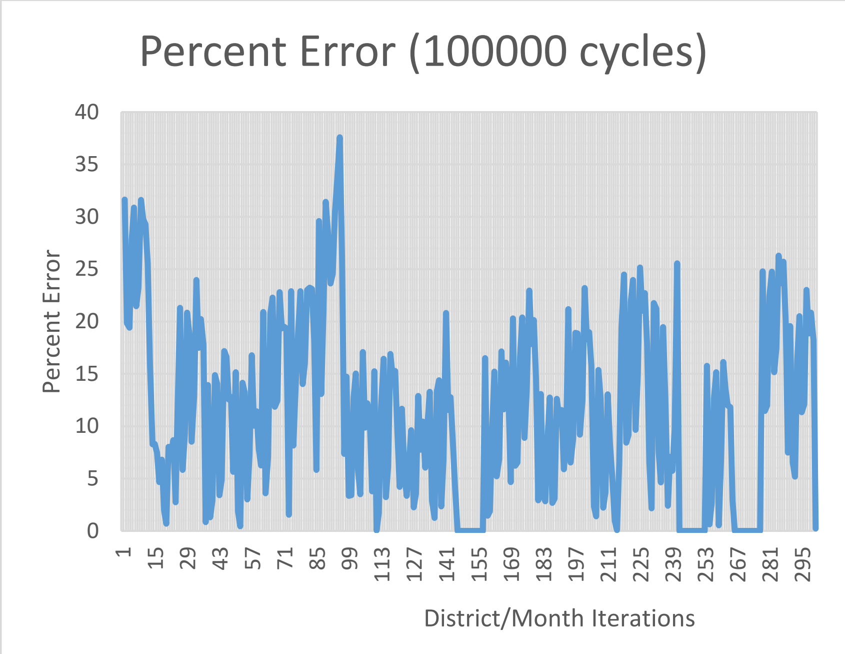

Figure 3: Neural Network: Percent Error (100000 cycles)

b. The accuracy decreased to 86.52 when the number of cycles was reduced to 1000 cycles It was also observed that the RMS value rose to 5.45. The graph representing percent error value for probability of criminal damage happening at each district at a specific month is shown below:

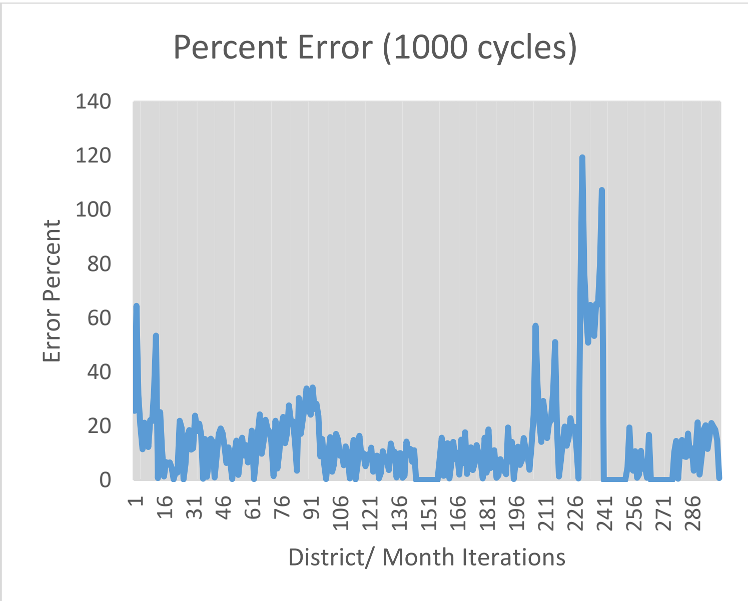

Figure 4: Neural Network - Percent Error (1000 cycles)

2. In the second case, the input to the neural network was changed to month, districts and year with the single hidden layer consisting of 5 nodes. It was expected that the network should predict the probability of a crime happening in the year 2013. For this particular neural network, the results were converging on to the same value each time. The weights and the learning rate also had to be adjusted so that the same result is not produced for all inputs. This

did not produce the expected result. The conclusion was that the number of hidden layers and the neurons were insufficient for the network to learn the pattern.

5.2 Naïve Bayes Implementation:

Using the Naïve Bayes method, the test was performed for the crime of Homicide and its occurrence was predicted over the year 2013-2015.

1. The Naïve Bayes implementation was able to predict the number of homicides to occur each year accurately after 3 sets of prediction. As seen in Figure 5, the prediction error was decreasing as we progressed to the next year prediction. This shows that Naïve Bayes classifier with a weighted matrix was able to increase the accuracy of prediction for the total number of crimes.

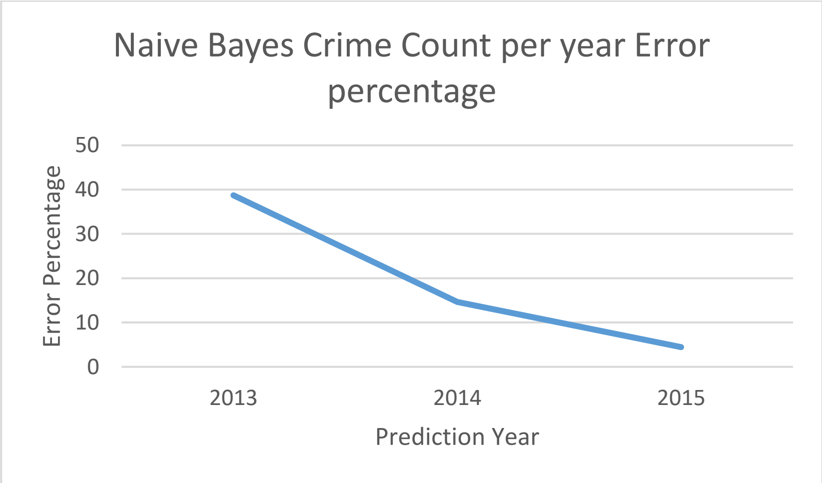

Figure 5: Naive Bayes - Number of Homicides in year - error percentage

2. In the same prediction for the year of 2015, the predicted number of incidents for month district combinations were compared with the actual values. On individual comparison some of the districts had larger error percentage, ie. in some cases the method predicted that 10 crimes will happen but only 5 occurred. These cases need further optimization in the future to detect the outliers and increase the accuracy.

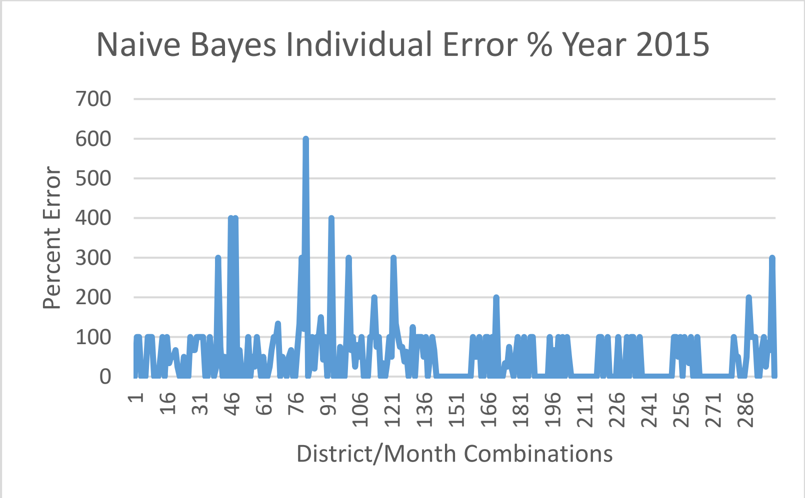

Figure 6: Naive Bayes Individual Error % year 2015

6 Conclusion

In summary, we focused on the prediction of a particular crime in a district per month using Naïve Bayes and Neural Network.
In Naïve Bayes method, it was understood that method was successful in reducing the error of predicting the number of the crimes in a year but further optimization is needed to improve the accuracy for the individual district month combinations.

It was understood that the neural network, with a single hidden layer will be insufficient for prediction of the crime if more than 2 inputs are given to it and a deep learning is required to process the amount of data given. There were also occasions in the neural network were the gradient descent was stuck to a local minimum. The key to improving the prediction is increasing the number of hidden layers in the network. To avoid a local minimum every time it is required to start the weights at random values at the start. This way the minimum can be avoided and the optimum output can be obtained.

7 References

1. data.cityofchicago.org (October 1, 2016,) “Crimes - 2001 to present” [Data file] https://catalog.data.gov/dataset/crimes-2001-to-present-398a4
2. Crime hotspots. (2016, May 13). In Wikipedia, The Free Encyclopedia. Retrieved 12:07, October 2, 2016, from https://en.wikipedia.org/w/index.php?title=Crime_hotspots&oldid=720032184
3. data.cityofchicago.org (October 4, 2016) “Chicago Police Department” - Illinois Uniform Crime Reporting (IUCR) Codes Retrieved from: https://data.cityofchicago.org/Public-Safety/Chicago-Police-Department-Illinois-Uniform-Crime-R/c7ck-438e
4. Kiani, Rasoul, Siamak Mahdavi, and Amin Keshavarzi. "Analysis and Prediction of Crimes by Clustering and Classification." Analysis 4.8 (2015).
5. Decision tree learning. (2016, September 10). In Wikipedia, The Free Encyclopedia. Retrieved 05:47, September 13, 2016, from https://en.wikipedia.org/w/index.php?title=Decision_tree_learning&oldid=738708257
6. Sathyadevan, Shiju, and Surya Gangadharan. "Crime analysis and prediction using data mining." Networks & Soft Computing (ICNSC), 2014 First International Conference on. IEEE, 2014.
7. Naive Bayes classifier. (2016, August 29). In Wikipedia, The Free Encyclopedia. Retrieved 05:55, September 13, 2016, from https://en.wikipedia.org/w/index.php?title=Naive_Bayes_classifier&oldid=736783747
8. Apriori algorithm. (2016, August 12). In Wikipedia, The Free Encyclopedia. Retrieved 05:56, September 13, 2016, from https://en.wikipedia.org/w/index.php?title=Apriori_algorithm&oldid=734148868
9. Almanie, Tahani, Rsha Mirza, and Elizabeth Lor. "Crime Prediction Based On Crime Types And Using Spatial And Temporal Criminal Hotspots." arXiv preprint arXiv:1508.02050 (2015).
10. Nath, Shyam Varan. "Crime pattern detection using data mining." Web Intelligence and Intelligent Agent Technology Workshops, 2006. WI-IAT 2006 Workshops. 2006 IEEE/WIC/ACM International Conference on. IEEE, 2006.
11. McClendon, Lawrence, and Natarajan Meghanathan. "Using Machine Learning Algorithms to Analyze Crime Data." Machine Learning and Applications: An International Journal (MLAIJ) vol 2.
12. Zhang, Yufei, and Huifeng Ji. "Using GIS to analyze and forecast the Chinese Crime rate." The 2nd International Conference on Information Science and Engineering. IEEE, 2010.
13. Liu, Hua, and Donald E. Brown. "Criminal incident prediction using a point-pattern-based density model." International journal of forecasting 19.4 (2003): 603-622.
14. Tayebi, Mohammad A., Uwe Gla, and Patricia L. Brantingham. "Learning where to inspect: Location learning for crime prediction." Intelligence and Security Informatics (ISI), 2015 IEEE International Conference on. IEEE, 2015.
15. data.cityofchicago.org (October 4, 2016) “Boundaries - Police Beats (current)” Retrieved from: https://data.cityofchicago.org/Public-Safety/Boundaries-Police-Beats-current-/aerh-rz74
16. data.cityofchicago.org (October 4, 2016) “Boundaries - Police Districts (current)” Retrieved from: https://data.cityofchicago.org/Public-Safety/Boundaries-Police-Districts-current-/fthy-xz3r
17. data.cityofchicago.org (October 4, 2016) “Boundaries - Wards (2015-)” Retrieved from: https://data.cityofchicago.org/Facilities-Geographic-Boundaries/Boundaries-Wards-2015-/sp34-6z76
18. data.cityofchicago.org (October 4, 2016) “Boundaries - Community Areas (current)” Retrieved from: https://data.cityofchicago.org/Facilities-Geographic-Boundaries/Boundaries-Community-Areas-current-/cauq-8yn6
19. gis.chicagopolice.org (October 4, 2016) “FBI's National Incident-Based Reporting System (NIBRS)” Retrieved from: http://gis.chicagopolice.org/clearmap_crime_sums/crime_types.html

20. Sunil Ray (2015 September 13) “6 Easy Steps to Learn Naive Bayes Algorithm (with code in Python)”, Retrieved from: https://www.analyticsvidhya.com/blog/2015/09/naive-bayes-explained/
21. Rob Schapire (2016, October 1) “Machine Learning Algorithms for Classification” Retrived from: http://www.cs.princeton.edu/~schapire/talks/picasso-minicourse.pdf
22. Iqbal, Rizwan, et al. "An experimental study of classification algorithms for crime prediction." Indian Journal of Science and Technology 6.3 (2013): 4219-4225.
23. An, Nguyen Hoang, and Duong Tuan Anh. "Comparison of Strategies for Multi-step-Ahead Prediction of Time Series Using Neural Network." 2015 International Conference on Advanced Computing and Applications (ACOMP). IEEE, 2015.
24. MathsIsFun, "Normal distribution," 2014. [Online]. Available: https://www.mathsisfun.com/data/standard-normal-distribution.html. Accessed: Dec. 2, 2016.