Building a Learning Machine Classifier with Inadequate Data for Crime Prediction

Trung T. Nguyen, Amartya Hatua, and Andrew H. Sung

School of Computing, The University of Southern Mississippi, Hattiesburg, MS 39406, U.S.A.

Abstract*—In this paper, we describe a crime predicting method which forecasts the types of crimes that will occur based on location and time. In the proposed method, the crime forecasting is done for the jurisdiction of Portland Police Bureau (PPB). The method comprises the following steps: data acquisition and pre-processing, linking data with demographic data from various public sources, and prediction using machine learning algorithms. In the first step, data pre-processing is done mainly by cleaning the dataset, formatting, inferring and categorizing. The dataset is then supplemented with additional publicly available census data, which mainly provides the demographic information of the area, educational background, economical and ethnic background of the people involved; thereby some of the very important features are imported to the dataset provided by PPB in statistically meaningful ways, which contribute to achieving better performance. Under sampling techniques are used to deal with the imbalanced dataset problem. Finally, the entire data is used to forecast the crime type in a particular location over a period of time using different machine learning algorithms including Support Vector Machine (SVM), Random Forest, Gradient Boosting Machines, and Neural Networks for performance comparison.*

Keywords: crime prediction, missing features, random forest, gradient boosting, SVM, neural networks

1 Introduction

Crime is a common problem in nearly all societies. Several important factors like quality of life and the economic growth of a society are affected by crime. There are many reasons that cause different types of crimes. In the past, criminal behavior is believed to be the result of a possessed mind and/or body and the only way to exorcise the evil was usually by some torturous means [1]. A person's criminal behavior can be analyzed from different perspectives like his/her socio-economic background, education, psychology, etc. Researchers have done exhaustive research on these factors. Data mining and analytics have contributed to the development of many applications in medical, financial, business, science, technology and various other fields. Likewise, to obtain a better understanding of crime, machine learning can be used for crime data analytics.

Analysis and forecasting the nature of crime has been done based mainly on the criminal's economic status, race, social background, psychology, and the demographics of a particular location. In the article by Gottfredson [2], the author discussed how to make a prediction whether a person will be criminal. On the other hand, he summarized and reviewed many of the previous works in order to identify general problems, limitations, potential methods and general nature of prediction problems of crime. The scope of that paper was limited to individual prediction, it did not address global prediction problems like predicting the number of offenses or offenders to be expected at a given time and place. In [3] Hongzhi et al. used improved fuzzy BP neural network to crime prediction. There is no mention of place, time and type of the crime. In [4,5] Mohler used crime hotspot to forecast a particular type of crime (gun crime) in Chicago; they did not address the other issues like other types of crime and occurrence time of those crimes. In [6] Tahani et al. focused on all the three major aspects of crime forecasting: place of crime, time of crime and type of crime. They performed the experiment using some machine learning algorithms for the state of Colorado and California of the United States. In their research, they used only the dataset with its information based on the National Incident Based Reporting System (NIBRS) [7], where information related to crime type, crime time and crime place was present but they did not consider any information about demographic, economic and ethnic details of criminals.

In order to forecast crimes successfully, we need to forecast the three main parameters of a particular crime: its type, location and time. Also, the methodology of crime prediction should consider the pattern of previously happened crimes and the other external factors like demographic, economic and ethnic details of criminals. In the present article, we have taken care of all the above mentioned factors. Our main objective is to forecast a crime along with its type, location and time.

In the following we describe data pre-processing, prediction methods, results and conclusion.

2 Proposed Methodology

Our proposed methodology can be broadly divided into four phases: Data acquisition, Data preprocessing, Application of Classification algorithm, Finding result and drawing the conclusion. Diagrammatic representation of proposed methodology is given in Fig 1.

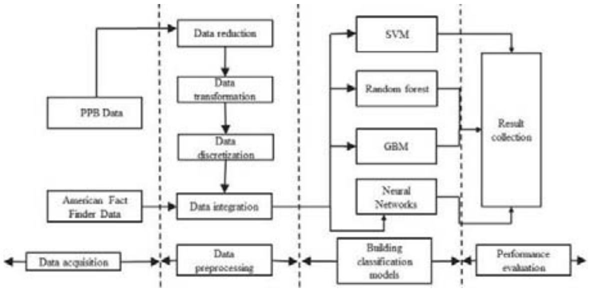

Fig. 1. Proposed methodology

The data is acquired from Portland Police Bureau (PPB) and the public government source American FactFinder. In data preprocessing phase we have performed feature engineering tasks and integrating the PPB dataset and the demographic dataset. Thereafter, we have applied several machine learning algorithms to build the crime prediction models. Finally, we have performed the testing on trained models and evaluated the performance. Detailed descriptions about each of the phases are provided below.

3 DATA PREPROCESSING

One of the key contribution of this article is data preprocessing. In previous researches, either only the crime occurrence data obtained from police were considered or data related to the criminals were considered, while in this research both aspects are considered at the same time. In addition, several other data preprocessing techniques are described next.

3.1 Description about Dataset

In our experiments, we have used data from two different sources: first, the dataset provided by the PPB is for the period of March 1, 2012, through September 30, 2016 [8], and the dataset from the American FactFinder website [9]. The data in PPB dataset is listed in calls-for-service (CFS) records giving the following information: Category of crime, Call group, Final case type, Case description, Occurrence date, X and Y coordinate of the location of the crime and Census tract. The data in American FactFinder is the census data of Portland area. From this data, we obtained information about economic, education, employment and racial background of people in this area.

3.2 Data Reduction

When we examine the data from PPB, there were some missing values in census tract information. Those data points are ignored as we have enough data to perform our experiments. In the dataset, four different types of parameters that describe crime type are Category, Call groups, Final case type and Case descriptions. Out of these four types, we are going to forecast only the final case type. Other parameters are not important in this case. So, we have not included those parameters in our experiments. Thereafter, we have performed dimensionality reduction and obtained a dataset with a reduced number of dimension. Our reduced dataset has a total of five parameters (Final case type, Case description, Occurrence date, X and Y coordinate of the location of the crime and Census tract).

3.3 Data Transformation

In the PPB dataset, one of the fields is census tract. The state and county information is removed from this field. Therefore, we have to convert that field into eleven-digit census tract, because in the later part of our experiments it is necessary to integrate PPB data with data from the American FactFinder dataset using census tract as joined key. As the PPB data was collected from Multnomah county of Oregon state so the first five digits of census tract are 41051. The last six digits of a census tract is formed by leading zero padded with the original census tract in the PPB dataset. Now we get the standard census tract with eleven digits number in which the first five digits are 41051 and last six digits are the census tract.

The total area of each of the census tract is different than others. So, we divided each of the census tract into small clusters with a fixed area of 0.71 square mile. The number of clusters in a particular census tract depends on the total area of the census tract. A new parameter named Area Code is derived, to identify the location of a crime using census tract of that place and the cluster number of that place. In our original dataset, X and Y coordinate of the crime location is provided, from this information we have created clusters of an area. If a crime is committed in cluster number "MM" of census tract "41051000XXX", then Area Code of that crime will be "41051000XXXMM". This is a unique id for location and in the rest of this article this parameter is used to identify the location of a crime.

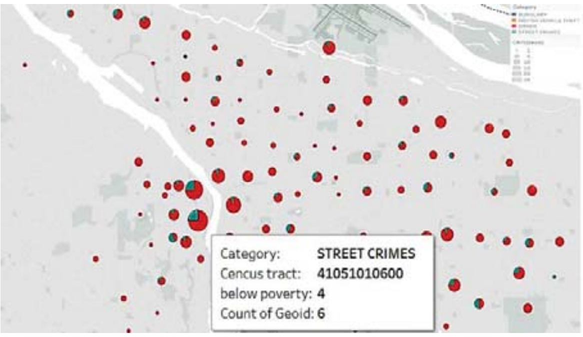

Fig. 2. Different crimes occur in different census tracts of Portland

Fig. 2. visualizes crime hotspots in different census tracts on Portland map. Each crime hotspot is represented by pie-chart. The size of the pie charts proportional to the numbers of crimes happened in a particular census tract.

3.4 Data Discretization

In the PPB dataset, every crime has a corresponding occurred time and date. However, our objective is to predict the crime within a span of seven days. Therefore, instead of a particular date, it is converted into corresponding week days and week number in the year. Then we can divide all the crimes into their occurrence day out of the seven days of the week, and occurrence week out of the 53 weeks of the year. That makes it easy to handle the data and it also helps to achieve our target more easily.

3.5 Data Integration

Data integration is one of the most important steps of this work on crime prediction. Economic, Demographic, Educational and Ethnic information about the people of Multnomah County of the state of Oregon are collected from the census data provided by American FactFinder and then integrated with the data provided by PPB. A total of 21 features are added, and description of the features are provided in Table 1 below.

TABLE I. FEATURES ADDED BY CENSUS DATA

Demographic ID	Description
HC01_EST_VC01	Total Population
HC02_EST_VC01	Below poverty level; Estimate; Population for whom poverty status is determined
HC03_EST_VC01	Percent below poverty level;
HC01_EST_VC14	Total population of One race - White
HC01_EST_VC15	Total population of One race - Black or African American
HC01_EST_VC16	Total population of One race - American Indian and Alaska Native
HC01_EST_VC17	Total population of One race - Asian
HC01_EST_VC18	Total population of One race - Native Hawaiian and Other Pacific Islander
HC01_EST_VC19	Total population of One race - Some other race
HC01_EST_VC20	Total population of Two or more races
HC01_EST_VC28	EDUCATIONAL ATTAINMENT - Total population of Less than high school graduate
HC01_EST_VC29	EDUCATIONAL ATTAINMENT - Total population of High school graduate
HC01_EST_VC30	EDUCATIONAL ATTAINMENT - Total population of Some college, associate's degree
HC01_EST_VC31	EDUCATIONAL ATTAINMENT - Total population of Bachelor's degree or higher
HC01_EST_VC36	EMPLOYMENT STATUS - Employed
HC01_EST_VC39	EMPLOYMENT STATUS - Unemployed
HC01_EST_VC51	Total population below 50 percent of poverty level
HC01_EST_VC52	Total population below 125 percent of poverty level
HC01_EST_VC53	Total population below 150 percent of poverty level
HC01_EST_VC54	Total population below 185 percent of poverty level
HC01_EST_VC55	Total population below 200 percent of poverty level

We observe that there exist certain relations between each of these features and the rate of crimes for a particular census tract. To figure out the effectiveness of demographic data integrated with the PPB data, we have compared the performance of our models on two kinds of datasets. Our first dataset contains only the preprocessed PPB data, in which there is no demographic information. In the second type of dataset, we assigned values randomly for those parameters in such way that, the overall percentage is the same. In the experiment, we have 92,715 data points. Based on the data distribution from the demographic data, we generated missing data using random variables with pre-determined distribution. For example, the HC01_EST_VC01 attribute represents the percentage of below poverty level for each census tract. In the present context, 9.1% of people are below poverty level at census tract 100. So, in the newly generated data, 9.1% data points of census tract 100 will be assigned as 0 (which denotes below poverty level) and the rest as 1 (which denotes not below poverty level). Similarly, for all other attributes, missing data will be assigned based on the percentage mentioned in census data.

The poverty level of all census tract is divided into four levels: from 1 to 4 in which the lower the poverty level, the poorer the people in such census tract. Fig. 3 below shows the percentage of census tracts in total that corresponding to different poverty levels. Similarly, the crime density of census tracts is also divided into six levels which denote the increasing crime density from level 1 to level 6. In order to find the relationship between poverty level and crime density, we compare the distribution of crime density that corresponding to poverty level in census tract. Fig. 4 below represents the crime density levels of all census tracts that have poverty level 2 while Fig. 5 below represents the crime density levels of all census tract that have poverty level 3. By observation, it can be figured out that in such census tract that have higher poverty level (the higher income level), the total of census tract and the density of crimes are less than the census tracts that have lower poverty level. So, there is a relation between poverty level and crime density in a particular census tract. Similarly, we have found some other relations between the parameters mentioned in TABLE I and numbers of crime in different areas of Portland.

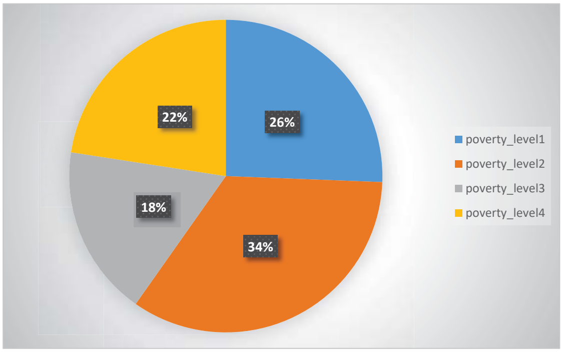

Fig. 3. Percentage of different poverty levels

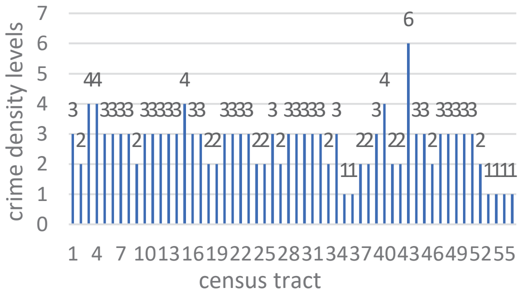

Fig. 4. Occurrences of crimes and in areas with poverty level 2

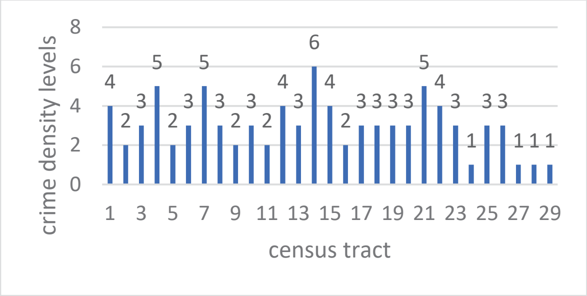

Fig. 5. Occurrences of crimes and in areas with poverty level 3

4 Learning Machine Models

For this crime prediction problem, we employed several machine learning algorithms to build models to accomplish the classification task and then compared the results of them. The learning machines used include Support Vector Machines or SVM [10, 11, 12], Random Forest [13], Gradient Boosting Machines [14], and multilayer neural networks (using MATLAB toolbox, employing scaled conjugate gradient training and resilient backpropagation training algorithms) [15, 16, 17, 18]. These learning machines are well-known and details about them can be found in a variety of references.

5 Results

All the models described in the previous section were trained and tested in our crime prediction tasks. The following sections present the result. The first subsection will discuss our solution of the imbalanced dataset we face. The second and third subsections describe the results for our two kinds of dataset we have after preprocessing.

5.1 Dealing with imbalanced dataset

Below is the distribution of our compilation dataset after preprocessing. In TABLE II Class 1, Class 2, Class 3, Class 4 represent the following classes: STREET CRIMES, OTHER, MOTOR VEHICLE THEFT, BURGLARY respectively.

TABLE II. NUMBER OF DATA POINTS FOR DIFFERENT CLASSES

Class 1	Class 2	Class 3	Class 4	Total
20,216	70,390	1,221	924	92,751

In Table II, we can see that there is a big difference between the numbers of each crime types because of the nature of crime occurrence probability. According to Chawla [19, 20], there are 4 ways of dealing with imbalanced data:

1) Adjusting class prior probabilities to reflect realistic proportions
2) Adjusting misclassification costs to represent realistic penalties
3) Oversampling the minority class
4) Under-sampling the majority class

In this project, we have applied under-sampling of the majority class technique. In our dataset, class 3 and class 4 are the minority classes and class 1 and class 2 are the majority classes because the number of samples of class 3 and class 4 are much smaller than the number of samples of class 1 and 2. Below are the steps to construct the new dataset T from the original dataset:

- We applied k-means clustering with number of cluster equals to k=2000 on two majority classes (class 1 and class 2)
- From each cluster of class 1 and class 2, select m random samples and put to new dataset T. In our experiment, we chose m that range from 1 to 5 to select representative samples from clusters of majority classes
- Put all samples of class 3 and class 4 into new dataset T

After new dataset T is constructed, 10-fold validation is used to divide this dataset into training set and test set and fed into selected learning machine to build classification models. These models will then be validated on the original dataset to benchmark the performance.

5.2 Prediction with only PPB dataset

After preprocessing original PPB data, we have 6 features dataset with the size of more than 92,000 records. Then we applied under-sampling technique (mentioned in 5.1 above) on our first dataset to create a training dataset. The new dataset T that was constructed from above section had the size of 18,145 samples (6,000 samples of class 1; 10,000 samples of class 2; 1,221 samples of class 3; 924 samples of class 4). This training dataset has the almost equal number of samples of 4 different CFS classes that were defined by the Portland Police Bureau. The CFS classes are corresponding to Burglary, Street Crimes, Theft of Auto and Others. The following subsection with discuss the results of applying different machine learning algorithms on our first dataset.

5.2.1 Support Vector Machines (SVM)

After the model were trained using SVM with Gaussian kernel on the above under-sampling training dataset, we tested the model run on all the available samples of our first dataset and got the overall 72.6% correct prediction of crime types. The confusing matrix of testing SVM model and the classification accuracy for each class are described in TABLE III.

TABLE III. RESULTS USING SVM ON FIRST DATASET

	Class 1	Class 2	Class 3	Class 4
Class 1	10242	19576	45	37
Class 2	6542	42154	57	45
Class 3	2038	5423	1100	13
Class 4	1394	3237	19	829
Precision	50.66%	59.89%	90.09%	89.72%

5.2.2 Random Forest

Next, we have applied Random Forest to complete the crime prediction task. We have trained random forest models using deep tree with minimum leaf size equals to 5 on our first training set, evaluated resulting models on all samples of our first dataset and got the results. At first, we have started from using 50 trees and then adding 50 trees more next loop. In our experiment, the best accuracy results occur when we set number of tree to 100, and increased the number of trees doesn't improve the results. Furthermore, TABLE IV showed the confusing matrix when evaluated random forest on our first dataset (number of tree = 100).

TABLE IV. RESULT OF RANDOM FOREST ON FIRST DATASET (NUMBER OF TREE IS 100)

	Class 1	Class 2	Class 3	Class 4
Class 1	11427	19375	583	455
Class 2	8704	50827	559	440
Class 3	78	153	78	3
Class 4	7	35	1	26
Accuracy	56.52%	72.2%	6.38%	2.81%

5.2.3 Gradient Boosting Machines

Finally, we have applied Gradient Tree Boosting to complete the prediction task. We have directly trained and evaluated this model using AdaBoost with training set and got the results. Generally, it will be better to employ more decision trees for higher prediction accuracy (lower mean square error). However, in our experiment, the best result occurs when we set number of tree estimators to 100, and it does not turn better as the number of trees increases. Furthermore, TABLE V showed the confusing matrix when evaluated Gradient Boosting Machines on our first dataset (number of tree = 100).

TABLE V. RESULT OF GBM ON FIRST DATASET (NUMBER OF TREE IS 100)

	Class 1	Class 2	Class 3	Class 4
Class 1	12019	20648	64	68
Class 2	6491	45685	71	47
Class 3	1125	2511	1077	13
Class 4	581	1546	9	796
Accuracy	59.45%	64.9%	88.0%	86.1%

5.2.4 Neural Networks

Lastly, we trained neural networks on our first dataset using two backpropagation training techniques: Scaled Conjugate Backpropagation (SCG) and Resilient Backpropagation (RP). We constructed our neural network using 2 hidden layers. Different combination of different number of nodes of two layers were tested and classification performance were recorded. With SCG, we obtained the best result of 65.7% classification accuracy with 30 nodes of hidden layer 1 and 50 nodes of hidden layer 2. With RP, we could obtain the best result of 66.46% classification accuracy with 45 nodes of hidden layer 1 and 50 nodes of hidden layer 2. In TABLE VI. and TABLE VII, the confusing matrix of classification using the two best neural network models that obtained from SCG and RP training methods.

TABLE VI. RESULT OF NEURAL NETWORK ON FIRST DATASET WITH SCG TRAINING

	Class 1	Class 2	Class 3	Class 4
Class 1	9459	18664	565	414
Class 2	10523	51454	630	493
Class 3	181	170	19	8
Class 4	53	102	7	9
Accuracy	46.79%	73.10%	48.32%	0.97%

TABLE VII. RESULT OF NEURAL NETWORK ON FIRST DATASET WITH RP TRAINING

	Class 1	Class 2	Class 3	Class 4
Class 1	8750	17309	529	406
Class 2	11337	52860	679	507
Class 3	96	177	12	5
Class 4	33	44	1	6
Accuracy	43.28%	75.10%	0.98%	0.65%

5.3 Prediction results with second dataset

After working on our first dataset, we decided to develop our second dataset which derived from original PPB data with demographic data based on census tract that we obtained from FactFinder. After preprocessing original PPB data, the dataset have 6 features. Based on demographic dataset from FactFinder, new 9 features have been introduced to form the second dataset. The 9 features are: below poverty level status, age, sex, race, education status, employment status, working time status, poverty level status, past 12 month working status. Based on the data distribution from the demographic data, missing data have been generated using random variable with pre-determined distribution. This resulting dataset contains 15 features with the same size with our first dataset (over 92,000 records). Under-sampling technique has also been applied for this dataset to create a training dataset size of 12,145 samples which contains 4,000 samples of class 1; 6,000 samples of class 2; 1,221 samples of class 3; and 924 samples of class 4. The following subsection with discuss the results of applying different machine learning algorithms on our second dataset.

5.3.1 SVM

After the model were trained using SVM with Gaussian kernel on above training dataset, we tested the model on the whole sample of the second dataset and got the overall 79.39% correct prediction of crime types. The confusing matrix of testing SVM model and the classification accuracy for each class are described in TABLE VIII. We got good results to classify class 2, class 3 and class 4 but very bad accuracy with class 1.

TABLE VIII. RESULT OF SVM ON SECOND DATASET

	Class 1	Class 2	Class 3	Class 4
Class 1	1398	2	0	0
Class 2	18816	70324	367	281
Class 3	2	34	854	1
Class 4	0	30	0	642
Accuracy	7.0%	99.9%	70.0%	69.5%

5.3.2 Random Forest

Next, we trained a model using Random Forest deep tree with minimum leaf size equals to 5 on second training dataset to complete the crime prediction task. We trained different Random Forest models which ranged from 50 trees and 300 trees, and then recorded the classification performance. In our experiment, the best accuracy results occur when we set number of tree to 250 (65.79% accuracy). In TABLE IX, the result confusing matrix and classification accuracy on each class when training Random Forest with 250 trees.

TABLE IX. RESULT OF RANDOM FOREST ON SECOND DATASET

	Class 1	Class 2	Class 3	Class 4
Class 1	11878	473	0	0
Class 2	8014	47647	122	133
Class 3	278	17978	1047	341
Class 4	46	4292	52	450
Accuracy	58.75%	67.69%	85.75%	48.7%

5.3.3 Gradient Boosting Machines

Next, we trained a model using Gradient Boosting using AdaBoost training technique on second training dataset to complete the crime prediction task. We trained different Gradient Boosting Tree models which ranged from 50 trees and 300 trees, and then recorded the classification performance. In our experiment, the best accuracy results occur when we set number of tree to 300 (61.67% accuracy). In TABLE X, the result confusing matrix and classification accuracy on each class when training GBM with 300 trees.

TABLE X. RESULT OF GRADIENT BOOSTING TREE ON SECOND DATASET

	Class 1	Class 2	Class 3	Class 4
Class 1	12869	459	0	0
Class 2	7016	42298	18	14
Class 3	236	16491	1167	45
Class 4	95	11142	36	865
Accuracy	63.65%	60.09%	95.57%	93.61%

5.3.4 Neural Networks

Lastly, we trained neural networks on our second dataset using two popular backpropagation training techniques which are Scaled Conjugate Backpropagation (SCG) and Resilient Backpropagation (RP). We constructed our neural network using 2 hidden layers. Different combination of different number of nodes of two layers were tested and classification performance were recorded.

TABLE XI. RESULT OF NEURAL NETWORK ON SECOND DATASET WITH SCG TRAINING

	Class 1	Class 2	Class 3	Class 4
Class 1	12659	952	4	1
Class 2	7170	52402	484	424
Class 3	295	12304	590	297
Class 4	92	4732	143	202
Accuracy	62.62%	74.44%	48.32%	21.86%

With SCG, we obtained the best result of 74.02% classification accuracy with 60 nodes of hidden layer 1 and 40 nodes of hidden layer 2. With RP, we could obtain the best result of 74.24% classification accuracy with 40 nodes of hidden layer 1 and 25 nodes of hidden layer 2. In TABLE XI. and TABLE XII, the confusing matrix of classification using the two best neural network models that obtained from SCG and RP training methods.

TABLE XII. RESULT OF NEURAL NETWORK ON SECOND DATASET WITH RP TRAINING

	Class 1	Class 2	Class 3	Class 4
Class 1	13373	1134	0	0
Class 2	6493	54785	548	494
Class 3	278	11513	581	308
Class 4	72	2958	92	122
Accuracy	62.2%	77.8%	47.6%	13.2%

5.4 Comparision of results

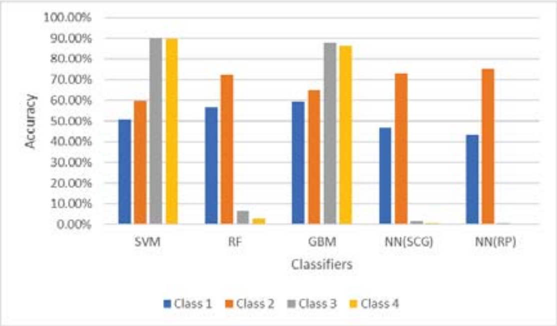

Fig. 6: Prediction accuracy of different classifiers on first dataset

Fig. 6. above displays the comparison of prediction accuracy of different classifiers on first dataset. Among those classifiers, SVM with Gaussian kernel AdaBoost gave 72.6% and 74.3% correct prediction of crime types while Random Forest (RF) and Neural Network perform badly. GBM handle the imbalanced data issue best among those classifier models. Fig. 7. is the comparison of prediction accuracy of different classifiers on second dataset. Among those classifier models, RF and GBM (AdaBoost) gave lowest overall performance with 65.79% for RF and 61.67% for GBM. However, SVM and Neural networks gave best overall prediction accuracy (79.39% for SVM, 74.02% for SCG and 74.24% for RP). Nevertheless, SVM and NN have problem with imbalanced classification between classes while GBM still handle well with this problem.

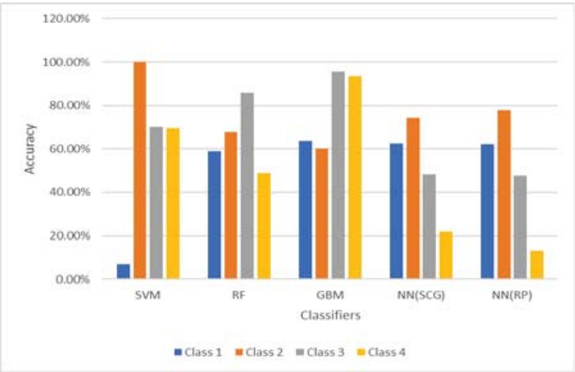

Fig. 7. Prediction accuracy of different classifiers on second dataset

6 CONCLUSION

For the crime prediction problem offered as a competition by the U.S. National Institute of Justice, we have started with preprocessing the dataset from Portland Police Bureau. Then, we have attempted to select some helpful features to represent the attributes of the samples in a proper manner. The total forecast area is divided into crime hotspots and integrated with demographic data from FactFinder. Thereafter, machine learning techniques are used to train our prediction models that calculated the probabilities of different categories of crimes. Because of the large dataset and problem of imbalanced data, we employ under-sampling technique on our dataset to reduce the training set size to less than 20,000 samples. According to the results, integrating demographic data from other public sources such as Fact Finder, U.S. census bureau has resulted in improving the performance of our models significantly.

With our first dataset in which demographic was not used, as shown in the results of previous section, SVM does not seem to be a suitable model for this task because of the bad classification accuracy when compared to other used methods. Ensemble methods such as Random Forest or Gradient Boosting turned out to be the two best models when compared the performance of them with SVM. These two methods can handle with big training sets and the training time is faster than the training time of SVM model.

With our second dataset that we use demographic data as reference to generate missing features, SVM and Neural Network show the best accuracy models when compared to other two ensemble methods of Random Forest and Gradient Boosting Machines. Overall, the classification accuracy of different machines on our second dataset are better than our first dataset. However, there are imbalanced classification accuracy between four desired classes where resulting models have large misclassification of one of four classes while the classification of other three classes are very good.

In the era of big data, analytics are increasingly being used for modeling, prediction, knowledge extraction, and decision making, etc. How to make the best use of datasets that are missing important features poses an often very challenging problem in data mining tasks. This research demonstrates a successful approach to build learning machine models with insufficient features. The authors are continuing the work to explore methods to handle imbalanced data, and to develop a more general model to predict crime type, time, and place using the best performing algorithms.

7 REFERENCES

[1] C. N. Trueman. (Mar 2015). "Why Do People Commit Crime." http://www.historylearningsite.co.uk/sociology/crime-and-deviance/ why-do-people-commit-crime.

[2] Don M. Gottfredson, "Assessment and Prediction Methods in Crime and Delinquency." *Contemporary Masters in Criminology,* pp. 337-372, Springer US, 1995

[3] Yu, Hongzhi, Fengxin Liu and Kaiqi Zou, "Improved Fuzzy BP Neural Network and its Application in Crime Prediction." *Journal of Liaoning Technical University (Natural Science)* 2 (2012): 025.

[4] Mohler, George, "Marked Point Process Hotspot Maps for Homicide and Gun Crime Prediction in Chicago." *International Journal of Forecasting*, vol. 30, Issue 3, pp 491-497, July-September 2014.

[5] Office of Justice Programs, "Why Crimes Occur in Hot Spots." http://www.nij.gov/topics/law-enforcement/strategies/hot-spot-policing/pages/why-hot-spots-occur.aspx.

[6] Tahani Almanie, Rsha Mirza, Elizabeth Lor, "Crime Prediction Based on Crime Types and Using Spatial and Temporal Criminal Hotspots", arXiv preprint arXiv:1508.02050 (2015).

[7] DENVER Open Data Catalog, http://data. denvergov.org/dataset /city- and-county-of-denver-crime. [Accessed: 20- May- 2015].

[8] National Institute of Justice, https://www.nij.gov/ funding/ Pages/ fy16-crime-forecasting-challenge. aspx#data

[9] American FactFinder, https://factfinder.census.gov/ faces/ nav/ jsf/ pages/index.xhtml

[10] Boser, Bernhard E., Isabelle M. Guyon, and Vladimir N. Vapnik, "A Training Algorithm for Optimal Margin Classifiers." *Proceedings of the fifth annual workshop on Computational learning theory*, pp. 144-152. ACM, 1992.

[11] Bottou, Léon, Corinna Cortes, John S. Denker, Harris Drucker, Isabelle Guyon, Lawrence D. Jackel, Yann LeCun et al., "Comparison of Classifier Methods: A Case Study in Handwritten Digit Recognition." *IAPR International Conference*, vol. 2, pp. 77-82, 1994.

[12] Tong, Simon, and Edward Chang, "Support Vector Machine Active Learning for Image Retrieval." *Proceedings of the ninth ACM international conference on Multimedia*, pp. 107-118, ACM, 2001.

[13] Breiman, Leo, "Random Forests." *Machine learning* 45.1 (2001): 5-32.

[14] Freund, Yoav and Robert E. Schapire, "A Desicion-Theoretic Generalization of On-line Learning and an Application to Boosting." *European conference on computational learning theory,* pp. 23-37, Springer Berlin Heidelberg.

[15] Møller, Martin Fodslette, "A Scaled Conjugate Gradient Algorithm for Fast Supervised Learning." *Neural networks* 6.4 (1993): 525-533.

[16] Orozco, José and Carlos A. Reyes García, "Detecting Pathologies from Infant Cry Applying Scaled Conjugate Gradient Neural Networks." *European Symposium on Artificial Neural Networks, Bruges (Belgium)*, pp. 349-354, 2003.

[17] Riedmiller, Martin, and Heinrich Braun, "A Direct Adaptive Method for Faster Backpropagation Learning: The RPROP Algorithm." *IEEE International Conference on Neural Networks,* pp. 586-591, 1993.

[18] MathWoks, Inc., Matlab.NeuralNetworksToolbox,V.9.1.0 https://www.mathworks.com/.

[19] Chawla, Nitesh V., "Data Mining for Imbalanced Datasets: An Overview." *Data mining and knowledge discovery handbook*, Springer US, vol. 4, pp. 853-867, 2005.

[20] Rahman, M. Mostafizur, and D. N. Davis, "Addressing the Class Imbalance Problem in Medical Datasets." *International Journal of Machine Learning and Computing* 3.2 (2013): 224.