Indian Journal of Science and Technology

An Experimental Study of Classification Algorithms for Crime Prediction

Rizwan Iqbal[1*], Masrah Azrifah Azmi Murad[2], Aida Mustapha[3], Payam Hassany Shariat Panahy[4], and Nasim Khanahmadliravi[5]

Faculty of Computer Science and Information Technology, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia.
mail.rizwaniqbal@yahoo.com[1], masrah@fsktm.upm.edu.my[2], aida@fsktm.upm.edu.my[3], payam_shp49@yahoo.com[4], nasim_khanahmad@yahoo.com[5]

Abstract

Classification is a well-known supervised learning technique in data mining. It is used to extract meaningful information from large datasets and can be effectively used for predicting unknown classes. In this research, classification is applied to a crime dataset to predict 'Crime Category' for different states of the United States of America. The crime dataset used in this research is real in nature, it was collected from socio-economic data from 1990 US Census, law enforcement data from the 1990 US LEMAS survey, and crime data from the 1995 FBI UCR. This paper compares the two different classification algorithms namely, Naïve Bayesian and Decision Tree for predicting 'Crime Category' for different states in USA. The results from the experiment showed that, Decision Tree algorithm out performed Naïve Bayesian algorithm and achieved 83.9519% Accuracy in predicting 'Crime Category' for different states of USA.

Keywords: Crime Prediction, Crime Category, Algorithm.

1. Introduction

The primary goal of data mining is to discover interesting and hidden knowledge in the data and summarize it in a meaningful form [6, 11, 14]. One of the most commonly used and important technique in data mining is Classification. Classification is a supervised class prediction technique. It allows predicting class labels which should be nominal [5]. Classification has been previously used for many domains including weather forecasting, health care, medical, financial, homeland security and business intelligence [9, 11]. This research will focus on applying different classification algorithms on the real crime data and compare the accuracy of their results in predicting the crime categories.

The birth and growth of crime in a community is based on many characteristics related to the community and society. These characteristics are — different races in a society , different income groups, different age groups, family structure (single, divorced, married, number of kids), level of education, the locality where people live (cheap or expensive housing, size of houses, number of rooms), number of police officers allocated to a locality, number of employed and unemployed people and etc.

In this research, a real crime dataset is used for data mining [13]. The attributes of this dataset are the characteristics related to a community or a society, some of which already discussed above. The two different classification algorithms are used to perform classification on the dataset, namely–Decision Tree and Naïve Bayesian. By experiment, results of both the algorithms will be compared and studied, and the most efficient algorithm in predicting the goal class (crime category) will be identified. There are many tools available for data

* *Corresponding author:*
Rizwan Iqbal (mail.rizwaniqbal@yahoo.com)

mining, for this research WEKA is chosen. It is an open source tool written in JAVA [16].

The organization of this paper is as follows. Section 2, covers preliminary discussion and experiment preparation, classification methods, UCI communities and crime dataset, crime dataset pre-processing, and measures for performance evaluation. Section 3, discusses the experimental results of the classification algorithms for predicting the 'Crime Category' attribute, in different states of USA. Finally, section 4 covers conclusion and future works.

2. Preliminary Discussion and Experiment Preparation

Classification is a class prediction technique, which is supervised in nature. This technique possesses the ability to predict the label for classes, provided that sufficient numbers of training examples are available [10]. There is a variety of classification algorithms available, including Support vector machines, k Nearest Neighbours, weighted voting and Artificial Neural Networks. All these techniques can be applied to a dataset for discovering set of models to predict the unknown class label. In classification, the dataset is divided into two sets, namely the training set (dependent set) and a test set (independent set). The data mining algorithm initially runs on the training set, than later the predicting model is applied on the test set [5, 12].

The dataset used in this experiment contains 128 attributes. From a large list of attributes, only twelve attributes are chosen. The chosen attributes are namely US state, population of community, median household income, median family income, per capita income, number of people under the poverty level, percentage of people 25 and over with less than a 9th grade education, percentage of people 25 and over that are not high school graduates, percentage of people 25 and over with a bachelor's degree or higher education, percentage of people 16 and over in the labour force and unemployed, percentage of people 16 and over in the labour force and unemployed, total number of violent crimes per 100K population.

There are different methods available for attribute or feature selection. For this experiment, manual method was chosen for attribute selection [16] based on human understanding and intellect. It is practical, especially when dealing with a large number of attributes. It was also taken in account that only those attributes are chosen, which do not contain any missing values. Classification was applied using Decision Tree and Naïve Bayesian classifier. In the first step, the model is built on the training set with known class label and in the second step; the proposed model is applied by assigning class labels on the test set. After performing the experiment using the above mentioned classification algorithms, accuracy of both the algorithms is evaluated, for predicting the 'Crime Category' attribute.

2.1 Crime Dataset Collection

The dataset used for this experiment is real and authentic. The dataset is acquired from UCI machine learning repository website [13]. The title of the dataset is 'Crime and Communities'. It is prepared using real data from socio-economic data from 1990 US Census, law enforcement data from the 1990 US LEMAS survey, and crime data from the 1995 FBI UCR [13]. This dataset contains a total number of 128 attributes and 1994 instances. All data provided in this dataset is numeric and normalized. The data in each instance belong to different states of the US. The states are represented in the form of number, every number representing its respective US state [15]. The complete details of all 128 attributes can be acquired from the UCI machine learning repository website [13]. For the sake of saving space the list of attributes used in this experiment are mentioned in Table 1.

2.2 Crime Dataset Pre-processing

The dataset used for the experiment consists of a total of 1994 instances which contain some missing values. In order to perform data processing, it is essential to improve the data quality [5]. There are a few techniques in practice, which are employed for the purpose of data pre-processing. The techniques are data cleaning, integration, reduction and transformation [5, 11]. Before applying a classification algorithm usually some pre-processing is performed on the dataset.

In the first step, data reduction is performed by selecting the most informative attributes in a dataset, while attempting to lose no critical information for classification. Only twelve attributes are selected from

Table 1. Crime Dataset Attributes

Attributes	Data Type	Description
State	Numeric	US state (by number)
population	Numeric - decimal	Population for community
medIncome	Numeric - decimal	Median household income
medFamInc	Numeric - decimal	Median family income (differs from household income for non-family households)
perCapInc	Numeric - decimal	Per capita income
NumUnderPov	Numeric - decimal	Number of people under the poverty level
PctLess9thGrade	Numeric - decimal	Percentage of people 25 and over with less than a 9th grade education
PctNotHSGrad	Numeric - decimal	Percentage of people 25 and over that are not high school graduates
PctBSorMore	Numeric - decimal	Percentage of people 25 and over with a bachelor's degree or higher education
PctUnemployed	Numeric - decimal	Percentage of people 16 and over, in the labor force, and unemployed
PctEmploy	Numeric - decimal	Percentage of people 16 and over who are employed
ViolentCrimesPerPop	Numeric - decimal	Total number of violent crimes per 100K population.
Crime Category	Nominal	Crime categorization in to three categories, namely Low, Medium, High. GOAL attribute (to be predicted)

a large collection of 128 attributes. There are different methods available for attribute or feature selection [16]. For this experiment, manual method was chosen for attribute selection [16] based on human understanding and intellect. It is practical, especially when dealing with a large number of attributes. It was also taken in account that only those attributes are chosen which do not contain any missing values.

In the second step, a new attribute was added in the dataset called 'Crime Category'. This added attribute is based on the values of 'Violent Crimes Per Pop' attribute, which depicts the total number of violent crimes per 100K population. The reason to add this new attribute is that, in order to perform prediction, the class (goal) attribute should be nominal in nature. In the case of the original dataset, all the original attributes are numeric [13], so a new attribute has to be added, to enable prediction. As mentioned earlier, this attribute is based on the data values in 'Violent Crimes per Pop', this dependency also retains the integrity of the dataset. The new attribute is just acts, as a means for providing different nominal labels for the values in 'Violent Crimes Per Pop', for prediction purposes.

The new added nominal attribute have three values, which are 'Low', 'Medium', and 'High'. If the value in 'Violent Crimes Per Pop' is less than 25 percent than the value of 'Crime Category' is 'Low', If the value in 'Violent Crimes Per Pop' is equal to or greater than 25 percent and less than 40 percent, than the value of 'Crime Category' is 'Medium', If the value in 'Violent Crimes Per Pop' is equal to or greater than 40 percent than the value of 'Crime Category' is 'High'. All the values were added in the newly created attribute carefully for the 1994 instances, and cross checked multiple times by all authors, to eradicate any chances of errors.

2.3 Building Classifiers and Measurements for Performance Evaluation

Bayesian classifiers adopt a supervised learning approach. They have the ability to predict the probability that a given tuple belongs to a particular class [1]. The strength of Naïve Bayesian classifier, as a powerful probabilistic has been proven for solving classification tasks effectively [3]. For any given instance, $X = (X_1, X_2, \ldots X_n)$, where, X_1 is the value of attribute X_1, $P\langle C|X\rangle$ is calculated by Bayesian classifier for all possible class values C and predicts $C^* = \text{argma}_{xc}\ p\langle x|c\rangle$ as the class value for instance X. Hence, estimating a $P\langle X|C\rangle$ which is proportional to $P\langle X|C\rangle P(C)$ is the key step of a Bayesian classifier.

Decision Tree is also a famous and commonly used predictive model, following the supervised learning approach [5], [17]. As the name suggests, Decision Tree forms a tree like structure, where each node in the tree denotes a test on an attribute value. The leaves represent classes or class distribution that predict model for classification. The branches represent conjunctions of features, which lead to classes. The tree structure carries high potential to easily produce classification

rules for the applied dataset [1]. The algorithm treats all the dataset as a large single set and then proceeds to recursively split the set. The algorithm applies the top down approach to construct the tree until some stopping criterion is met. Gain in entropy is used to guide the algorithm for the creation of nodes [4, 7].

Both the classifiers used for this experiment have some pros and cons. Naïve Bayesian requires short training time, fast evaluation and is more suitable for real world problems. If talking about solving complex classification problems, then Naïve Bayesian is not a recommended choice. In order to handle complex classification problems, Decision Tree is a better choice. It can produce reasonable and interpretable classification trees, which can be used for making critical decisions. However, it does not work well on all datasets. The results from both the algorithms will be evaluated on three performance measurements, which are defined below:

a. Precision and Recall are two significance performance measures for evaluating classification algorithms [2]. In this experiment, Precision refers to proportion of data which is classified correctly using classification algorithm. Here, Recall refers to percentage of information which is relevant to the class and is correctly classified.
b. Accuracy is the percentage of instances which is classified correctly by classifiers [5].
c. F-Measure is another performance measure which combines Recall and Precision into a single measure [8]. This measure is commonly used in classification.

3. Experiment Results, Analysis and Performance Evaluation

In the experiment, a comparison between Naïve Bayesian and Decision Tree algorithms was performed, over the crime dataset [13]. During experiment, the pre-processed dataset was converted to .ARFF file, which is the standard file type for WEKA input [16]. 10 fold cross-validation was applied to the input dataset in the experiment, separately for both Naïve Bayesian and Decision Tree algorithms. The Accuracy for 10 fold cross-validation for Naïve Bayesian and Decision Tree is 70.8124% and 83.9519%, respectively. Hence, Decision Tree out performed Naïve Bayesian and manifested higher performance. Moreover, confusion matrixes for Naïve Bayesian and Decision Tree are shown in Table 2 and Table 3. Figure 1 shows the comparison between the two algorithms.

Table 2. Confusion Matrix Using Naïve Bayesian

	Low	High	Medium	
Low	**1198**	70	47	1315
High	183	172	31	386
Medium	179	72	**42**	293

Table 3. Confusion Matrix Using Decision Tree

	Low	High	Medium	
Low	**1264**	70	19	1315
High	86	290	10	386
Medium	125	48	**120**	293

Table 2 illustrates the classification of low, high and medium classes using Naïve Bayesian algorithm. The result from the confusion matrix is discussed for each class below:

There are 1315 items classified in to class Low.

- 1198 of these items are correctly classified into class Low.
- 70 of these items are wrongly classified into class High.
- 47 of these items are wrongly classified into class Medium.

There are 386 items classified in to class High.

- 172 of these items are correctly classified into class High.
- 183 of these items are wrongly classified into class Low.
- 31 of these items are wrongly classified into class Medium.

There are 293 items classified in to class Medium.

- 42 of these items are correctly classified into class Medium.
- 179 of these items are wrongly classified into class Low.
- 72 of these items are wrongly classified into class High.

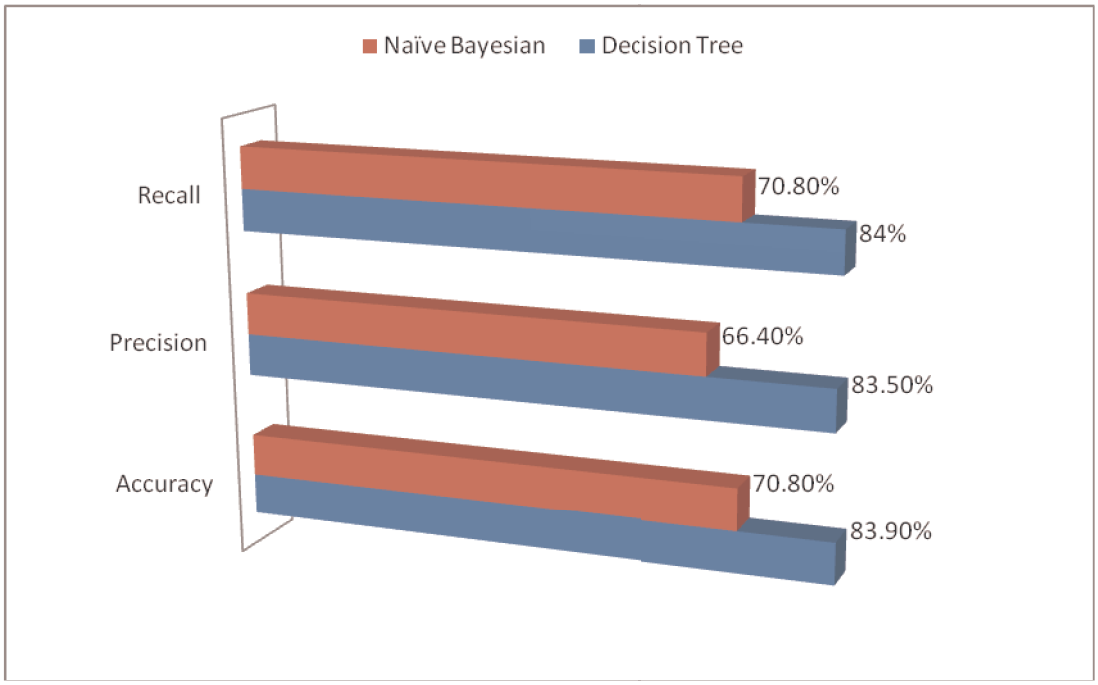

Figure 1. Comparison between different measures of the two algorithms.

Table 4. Accuracy, Incorrectly classified instances, Precision, Recall and F-measure for Both Algorithms

Method	Accuracy (Correctly classified instances)	Incorrectly classified instances	Precision	Recall	F-Measure
Decision Tree	83.9519%	16.0481%	0.835	0.84	0.826
Naïve Bayesian	70.8124%	29.1876%	0.664	0.708	0.675

Similarly, Table 3 illustrates the classification of low, high and medium classes using Decision Tree algorithm. The result from the confusion matrix is discussed for each class below.

There are 1315 items classified in to class Low.

- 1264 of these items are correctly classified into class Low.
- 32 of these items are wrongly classified into class High.
- 19 of these items are wrongly classified into class Medium.

There are 386 items classified in to class High.

- 290 of these items are correctly classified into class High.
- 86 of these items are wrongly classified into class Low.
- 10 of these items are wrongly classified into class Medium.

There are 293 items classified in to class Medium.

- 120 of these items are correctly classified into class Medium.
- 125 of these items are wrongly classified into class Low.
- 48 of these items are wrongly classified into class High.

The results from the confusion matrixes and its explanation above, clearly shows that Decision Tree performed better than Naïve Bayesian. Decision Tree performed better in predicting all the classes, namely Low, Medium and High. Table 4 illustrates the Accuracy (correctly classified instances), incorrectly classified instances, Precision, Recall and F-measure for both the algorithms used in the experiment. Figure 1 shows a comparison between different measures of the two algorithms. The values of Accuracy, precision and

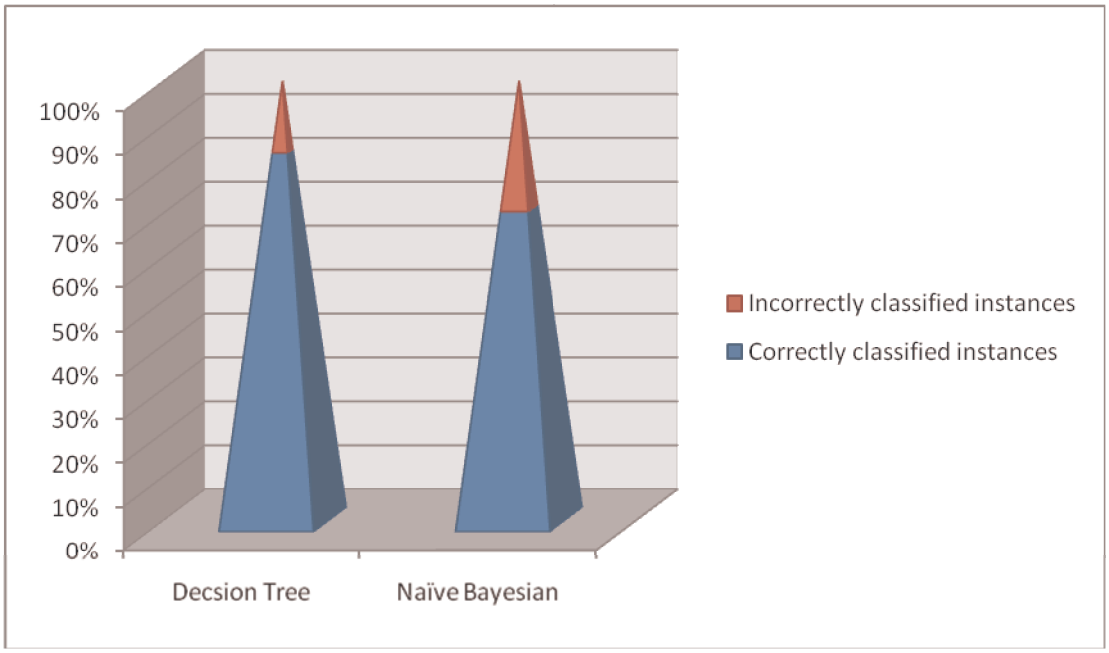

Figure 2. Percentage comparison of correctly and incorrectly classified instances.

Recall for Decision Tree is almost the same. Whereas, in Naïve Bayesian the Accuracy and Recall is almost the same but the value of precision is a little less. The percentage comparison for correctly classified instances for the two algorithms is demonstrated in Figure 2.

4. Conclusions and Future Work

This paper presents a comparison between two classification algorithms namely, Decision Tree and Naïve Bayesian for predicting the 'Crime Category' attribute, having labels, namely 'Low', 'Medium', and 'High'. For Decision Tree, the Accuracy, Precision and Recall are 83.9519%, 83.5% and 84%. On the other hand, Accuracy, Precision and Recall values for Naïve Bayesian are 70.8124%, 66.4% and 70.8%, respectively. Experimental results for both the algorithms manifest that, Decision Tree performed better than the Naïve Bayesian for the crime dataset, using WEKA. This experiment was performed using 10-fold cross- validation. It is evident that law enforcing agencies can take great advantage, using machine learning algorithms like Decision Tree to effectively fight crime and war against terrorism. For future research, there is a plan to further apply other classification algorithms on the crime data set and evaluate their prediction performance. Another direction for future work is to use other techniques for feature selection, and study their effect on the prediction performance of different algorithms.

5. References

1. Batchu V, Aravindhar D J et al., (2011). A classification based dependent approach for suppressing data, IJCA Proceedings on Wireless Information Networks & Business Information System (WINBIS 2012), Foundation of Computer Science (FCS).
2. Cios J, Pedrycz W et al., (1998). Data Mining in Knowledge Discovery, Academic Publishers.
3. Geenen P. L, van der Gaag L C et al., (2011). Constructing naive Bayesian classifiers for veterinary medicine: A case study in the clinical diagnosis of classical swine fever, Research in Veterinary Science, vol 91(1), 64–70.
4. Hamou A, Simmons A et al., (2011). Cluster analysis of MR imaging in Alzheimer's disease using decision tree refinement. International Journal of Artificial Intelligence, vol 6(S11), 90–99.
5. Han J, and Kamber M (2006). Data mining: concepts and techniques, Morgan Kaufmann Publishers, San Francisco, CA.

6. Kochar B, and Chhillar R (2012). An Effective Data Warehousing System for RFID using Novel Data Cleaning, Data Transformation and Loading Techniques. Arab Journal of Information Technology, vol 9(3), 208–216
7. Kováč S (2012). Suitability analysis of data mining tools and methods, Bachelor's Thesis.
8. Kumar V, and Rathee N (2011). Knowledge discovery from database using an integration of clustering and classification, International Journal of Advanced Computer Science and Applications, vol 2(3), 29–32.
9. Li G, and Wang Y (2012). A privacy-preserving classification method based on singular value decomposition, Arab Journal of Information Technology, vol 9(6), 529–534.
10. Ngai E W T, Xiu L et al., (2009). Application of data mining techniques in customer relationship management: A literature review and classification, Expert Systems with Applications, vol 36(2), 2592–2602.
11. Santhi P, and Bhaskaran V. M (2010). Performance of clustering algorithms in healthcare database, International Journal for Advances in Computer Science, vol 2(1), 26–31.
12. Selvaraj S, and Natarajan J (2011). Microarray data analysis and mining tools, Bioinformation, vol 6(3), 95–99.
13. UCI Machine Learning Repository (2012). Available from: http://archive.ics.uci.edu/ml/datasets.html
14. Wahbeh A H, Al-Radaideh Q A, et al., (2011). A comparison study between data mining tools over some classification methods, International Journal of Advanced Computer Science and Applications, Special Issue, 18–26.
15. WikiPedia (2012). WIKIPEDIA, Available from: http://en.wikipedia.org/wiki/List_of_U.S._states_by_area.
16. Witten I, Frank E et al. (2011), Data Mining: Practical Machine Learning Tools and Techniques. Morgan Kaufmann.
17. Xu Y, Dong Z. Y et al., (2011) A decision tree-based on-line preventive control strategy for power system transient instability prevention, International Journal of Systems Science.