Application of Machine learning Algorithms in Crime Classification and Classification Rule Mining

Umair Saeed, Muhammad Sarim, Amna Usmani, Aniqa Mukhtar, Abdul Basit Shaikh and Sheikh Kashif Raffat
Department of Computer Science, Federal Urdu University of Arts, Sciences and Technology, Karachi, PAKISTAN

Available online at: www.isca.in, www.isca.me
Received 2nd December 2013, revised 6th May 2014, accepted 8th August 2014

Abstract

Nowadays crime is one of major threats faced by our government .Extensive research in criminology has been done on focusing the study of crime and criminal behavior scientifically. It is one of most important field where the applications of data mining techniques are producing fruitful results. Data mining has been using to model crime detection and classification problems. Manually addressing the large amount of the volume of crime that is being committed makes crime prevention strategies a time consuming and complex task. In this paper data mining techniques are examined to predict crime and criminality. We apply machine learning algorithms to a dataset of criminal activity to predict attributes and event outcomes. We will also do comparative analysis between different classification techniques..

Keywords: Crime, criminology, classification, classification rule mining, machine learning, expert systems.

Introduction

In current society the volume of data is being produced is growing rapidly. This data explosion causes a persistent rise in new challenges and possibilities. Information plays a critical role in key areas like law enforcement. Defiantly, the large volume of criminal data creates many problems in different domain for instance data storage, data warehousing and data analysis. Lots of technological efforts are in progress to achieve insights into this information and to discover the knowledge from it.

Versatile artificial intelligence and data mining tools are frequently and increasingly accessible by law enforcement community due to the revolutionary changes in science and technologies. Once reticent for major institutions of research and national intelligence agencies, data mining software tools are available for enhancement in analysis and decision making at the national level and as well as local levels also. Now a day some of the software packages are tremendously prompt, extremely powerful and highly user-friendly. Due to having these high quality features these packages are more productive for live environments such as operational Strategic sessions or task forces.

To explore the criminal behavior is a key issue in criminology. The factors that reinforce the violent criminal behavior are not taken seriously. Appropriate understanding and interpretation of this motivational procedure is critical. Sensations attach individuals to the social world and, so, are the reasons of many communal psychological phenomena, such as humanity, selfish behavior, and violence. To be able to recognize and categorize a behavior, one has to understand the behavior itself and the emotional states that concern to it. Violent illegal Behavior can be defined as any reactive action against others that may cause harm or distress to community. Violent Criminal Behavior has been linked to impulsive and disruptive behaviors, harassment, and in severe cases, school shootings. Various hypothetical and experiential models of crime have been developed by social experts and analysts. However there is still a gap in dynamic micro simulation models to estimate criminal behavior.

Law enforcement reserves and supplies are allocated on the basis of factors most notably the local, national, and international crime. These factors also put impact on investigative priorities over authorities and those in power. For example, crimes like terrorism and illegal trade involving the cultivation, production, distribution and sale of substances which are subject to drug prohibition laws demonstrate local and international proposition. Sometimes local crime activities become a trend in a community and they differ from surrounding communities. Likewise, regions with dense population and housing imply the probability of certain crimes. These local and regional crime patters can be discovered and allow law enforcement agencies and personnel to tackle large-scale crime trends.

Sometimes a crime is associated with complexities which revolve around theme and association with a particular location and other aspects which makes the analyst's job quite challenging. Furthermore, evidence can be loosely coupled while begin geospatially sparse, employing a more extensive investigation effort. Using the power of data mining techniques imparts the ability to examine, foresee, take into account, and take action against criminal activities and potential security risks.

Related Work

Crime analysts vary in their methods, techniques, and approaches of searching data sources to determine patterns. Rules and heuristic which determine and identifies important crime related information also differs from analyst to analyst. These concerns imply a difficult scenario for the development of an automated criminal act analysis system.

There exist several software systems of crime data analysis which have been studies by the research circle. Most effort focuses on nature, intensity, location, duration and frequency of the crime. The some current trends in crime analysis[1] are; i. Geospatial, map-based visualization, ii. Geographical clustering of crime activity, such as identifying hot spots, iii. Serial criminal behavioral pattern profiling and criminal career analysis, iv. Gang criminal network analysis, v. Data stream anomaly, novelty, or outlier detection, vi. Temporal analysis of crime patterns, such as crime sprees (temporal association of crime from an individual or group) and vii. Linking threats to risk of critical infrastructure based on vulnerability assessments.

Data mining is the underlying engine and key component in each of these crime data utilization systems. In the large volumes of crime data, "knowledge discovery" is a prominent tool/technique for identifying underlying novel patterns. "Knowledge discovery" can be performed on large transactional databases by several methods prominent one of which is association rule mining. Some other trendy data mining techniques in use for similar purposes are summarized as:

Semantic analysis and text mining for entity extraction from free-text narratives, police reports, and FBI bulletins[2-4]. Rule-based, expert systems established through knowledge engineering. The utility of this technique is limited due to the dynamic nature of crime. It is also difficult to quantify and adequately capture the knowledge of field experts with significant experience. Clustering and graph representations[5], both for identifying similar crimes and for visualization purposes. Cluster size, shape, and distribution can aid in inferring details about related crimes. Clustering is also utilized to group classes of criminals. Machine learning and classification for crime pattern recognition.

Case-based reasoning for identifying closed cases exhibiting characteristics similar to open cases. There are so many expert systems for crime analysis. There are so many expert systems for crime analysis. Common Integrated Police Application (CIPA) is developed to build the basic mechanisms and infrastructure for the Criminal Information System and Crime. It is based on CrPC, which is uniform throughout the country[6]. The Indian Government has developed Crime Criminal Information System [CCIS] to store and retrieve criminal records and crime[6]. The purpose of Crime and Criminal Tracking Network System (CCTNS) is to create a integrated and complete system for improving the effectiveness and performance of policing at all levels and especially at the Police station level through implementation of principles of e-Governance[6]. The expansion and spurt in the planned Crime, especially activities by the terrorists and Mafia activities and insurgent have attained significant level that need to be undertook with better proficient expertise and efficiency which requires co-ordination among different agencies in sharing criminal intelligence[6]. For this purpose, an Organized Crime Intelligence System [OCIS] was developed for gathering, storage and recovery of information on planned crime and criminals and provides synchronization amongst different law enforcements at State level and as well as Central level. Motor Vehicle Coordination System (MVCS) was implemented in all States. The main purpose of the system is to provide information to police, public and other agencies concerning the lost/ recovered motor vehicles[6]. The POLNET is an Indian Government project. It will be utilized for the communication of Crime and Criminal data, Voice Communication, Fax transmission, Finger print and Image-Photographs transmission all over the country[6]. Violent Crime Apprehension Program is a national program (ViCAP) was developed and funded by the FBI's National Center for the Analysis of Violent Crime. The NCAVC is structured into three components: Child Abduction Serial Murder Investigative Resources Center (CASMIRC), Behavioral Analysis Unit (BAU), and Violent Criminal Apprehension Program (ViCAP). Cases in the ViCAP database include sexual assault cases, missing persons, unidentified dead bodies and attempted homicides or homicides especially if they involved a kidnap[6].

COPLINK was deployed in 2001, it is an knowledge management and integrated information system created to share, confine and evaluate law enforcement-related information. National Institute of Justice and the NSF7 were financially supported in development of COPLINK and it was developed at the University of Arizona's Artificial Intelligence Lab. Phoenix police departments and Tucson collaborated during the development phase. COPLINK consists of two essential components. COPLINK Connect permits law enforcements to share information with each other. The interface of the component is very user-friendly. COPLINK Detect provides a variety of crimes patterns entered from databases. Four types of searches are available in COPLINK Connect[7]. We can search by location, by incident, by vehicle and by person. The police officers normally perform these kinds of searches and these are standard types[5]. COPLINK Detect component shares the identical instances information as Connect. However Detect component also utilizes a new set of artificial intelligence tools for its users[7]. Comprehensive criminal case records and the considerable terms related with each case are the foundation of the Detect module. The reports contain both structured data and unstructured data[7]. Another important module named COPLINK Collaboration is in the phase of being developed that will facilitate the sharing of crime information among team members. These features of data retrieving and detection make COPLINK a data mining tool also. That is, Machine learning and techniques are applied in exploring the crime data base[8].

Dataset

We utilized The Communities and Crime Data Set[10]. The data is obtained from the UCI Machine learning Repository. This dataset contains three types of data. Socio-economic data and law enforcement data is gathered from a Law Enforcement Management and Administrative statistics survey while the crime data is obtained from US FBI Uniform crime Report. The submission date of dataset to UCI Machine learning Repository is July 2009 containing 2215 total instances and 128 attributes for communities from each state

Preprocessing of Dataset: The data preprocessing is the critical step to improve the performance of machine learning algorithms. The removal of noise records is one of the most important and critical task in machine learning.

Instance Selection: Normally, instance selection techniques are differentiate between wrapper and filter[11,12]. Filter assessment just considers information removal but does not take into account actions. Whereas, wrapper techniques clearly highlight machine learning point of view and estimate results by using the precise machine learning algorithms to activate instance selection

Missing Feature Values: In lots of the real world data source the incomplete data is a key fact and unavoidable challenge. Process of disregarding instances with unidentified attribute values, most common attribute value, mean substitution, regression or classification methods[18,20], hot deck imputation and technique of entertaining missing feature values as special values are some of the methods for handling missing data[13] and expert can choose any one from them.

Discretization: Discretization is the process to convert the continues values into discreet value. As we know the large amount of possible features values may cause to ineffective and slow process of machine learning algorithms, so discretization is used to significantly reduce the number of continues feature values[14].

Normalization: For the purpose of scaling down the transformations of the features we use the Normalization process[14]. This is essential step for k-Nearest Neighborhood algorithms and neural network algorithms. There are two most common approaches for this purpose[14]:

Min-max process for normalization:

$$A' = \frac{A - min_v}{max_v - min_v}(newmax_v - newmin_v) + newmin_v$$

z-score process for normalization:

$$A' = \frac{A - Mean_v}{Standdev_v}$$

Feature Selection: The process to classify and eliminates as much inappropriate and redundant features as possible is called feature selection. This decreases the preventable dimensions of the data set and facilitates learning algorithm to work more efficiently and faster. The filter assessment functions may be categorized into four groups: consistency, dependence, information and distance[14].

Feature Construction: Constructing new-fangled features from the fundamental features set the problem of feature interaction can be resolved. This approach is defined as feature/ transformation/ construction. . The new-fangled constructed features may tend to the creation of more precise and concise classifiers[14].

Classification: Classification in Machine Learning is defined as the aptitude behavior conduct by a machine to enhance its document classification action based on prior outcome of document classification. Labels are attached to observations, measurements termed as training set, and these labels demonstrate the class of training data. Newfangled data classification is done on the basis of the training set. Classifiers can predict definite class labels which may be discrete or nominal. Classes are identified to construct the classifiers is known as Supervised learning. Classification widely using in different areas[17,19,21]. Classification is based on two steps; Model Construction: specified a set of data on behalf of samples of a goal perception, construct a model to "clarify" the concept and Model Usage: Model Construction is used for classifying destined or unidentified cases Estimate correctness of the model.

Classification algorithms that are used commonly are decision tree, neural network, Super Vector Machine, Naïve Bayes, Random Forest and KNN etc. Decision tree is considered to be one of the most popular data-mining techniques for knowledge discovery. In Decision tree algorithm using depth –first greedy and breadth-first approaches, the data set of instances are recursively partitioning. The recursive process finally ends when all the instances belong to a particular label are identified. Decision tree can entertain the data with high dimensions. Their demonstration of acquired knowledge in tree structure is intuitive and normally easy to understand by humans.
We can select the following splitting criterion:

Information gain utilized the entropy measure as the impurity measure. It is base on impurity criterion.

$$b_{i=t_{i,j}}Xy,\sigma \quad \sigma_{b_{i=t_{i,j}}X\vee\frac{}{|X|}}.EInformationGain(b,X) = E(w,X) - \sum_{t_{i,j}\in dom(b_i)}$$

Whereas: $\left(\sigma_{w \quad c_j}X \vee \frac{}{(X\vee)\sigma_{w \quad c_j}X} \vee \frac{}{X\vee}.log_2 - E(w,X) = \sum_{c_j\in dom(w)}\right)$ ini index computes the deviations between the probability distributions of goal feature's values. The Gini index has been used in a variety of works and it is defined as:

$$\left(X \vee @\sigma_{w} \quad {}_{c_j}X \vee \frac{}{G(w,X)} = 1 - \sum_{c_j \in dom(w)} \right)$$

As a result the assessment criterion for choosing the feature is described as:

$$\left(b_{i=t_{i,j}}Xw, \sigma \quad b_{i=t_{i,j}}X \vee \frac{}{X \vee}. G\sigma \quad GGain(b_i, X) \right.$$
$$\left. = G(w,X) - \sum_{t_{i,j} \in dom(b_i)} \right)$$

The gain ratio is described as follows:

$$GainRatio(b_i, X) = \frac{InformationGain(b_i, X)}{Entropy(b_i, X)}$$

When denominator is zero the ratio is not defined. The ratio also can tend to favor features for which denominator is extremely little.

The naive Bayesian classifier is another common classification method, and several researchers have done the theoretical and experiential results of this approach. It has been extensively applied in lot of data mining domains, and executes unexpectedly fine.

$$V_{map}(e) = argmax_{c} P(c) P(l1, l2, \cdots, ln \vee f)$$

$$P(l1, l2, \cdots, ln | f) = \prod_{i=1}^{n} P(li \vee f)$$

Validation: Validation is a prominent component to provide feasible and optimal confidence that any new-fangled methodology is successful and efficient to find out a variable solution, in this case a possible solution to malicious identification problem. Validation process is not only evaluating the results to what the estimated result should be, but it is also o comparative analysis with the results to other published approaches The values which consists of (FPR) false positive rate, accuracy, true positive rate (TPR), and precision[15]. FPR is a ratio of negative instances those were erroneously recognized[16]. Accuracy is a proportion of correctly identified number of positive instances, either false positive or true positive. TPR can also define as, "is the amount of significant data retrieved, computed by the ratio of the amount of relevant retrieved data to the entire quantity of relevant data in the data set." Basically we can say TPR is a ratio of concrete positive instances those were appropriately recognized. Precision is the ratio of retrieved data that are related, computed by the ratio of the mount of relevant retrieved relevance to the total number of retrieved applications, or a ratio of forecasted true positive instances that were recognized appropriately[15].

Following these values are derived from the truth table, the table-1. A truth table is actually a confusion matrix and it provides us the actual and estimated classification from the predicators[15].

Table-1
Truth Table for Validation Process

		Actual	
		Positive	**Negative**
Predicted	Positive	a	b
	Negative	c	d

$$TPR = \frac{a}{a+b}$$
$$FPR = \frac{b}{b+d}$$
$$Accuracy = \frac{a+d}{a+b+c+d}$$
$$Percision = \frac{a}{a+b}$$

where, "a" is the amount of wicked applications in the data set that were classified as wicked applications, "b" is the amount of benign applications in the data set that were classified as malicious applications, "c" is the amount of malevolent applications in the data set that were classified as benign applications, and "d" is the number of benign applications in the data set that were classified performance assessment that were used in effort for validation of the estimated results[15,16].

Methodology

The developed methodology has the following primary steps: All attributes with a large number of missing values were removed. We separate the all goals into different separate file and each goal have all the feature set. All the record with corresponding missing goal value were removed. Apply discretization method and normalize each dimension and goals. Applied the FCBF algorithm for each goal we select the feature sets for each Goals. Some features are common in all 10 goals. By applying fuzzy clustering algorithm on all goals we define the labels very Low (1), low (2) and High (3). Applied Naive Bayes, decision tree classifiers on selected features sets of each goals one - by - one. Validate the results with cross validation. Compute the performance and accuracy and compare each classifier.

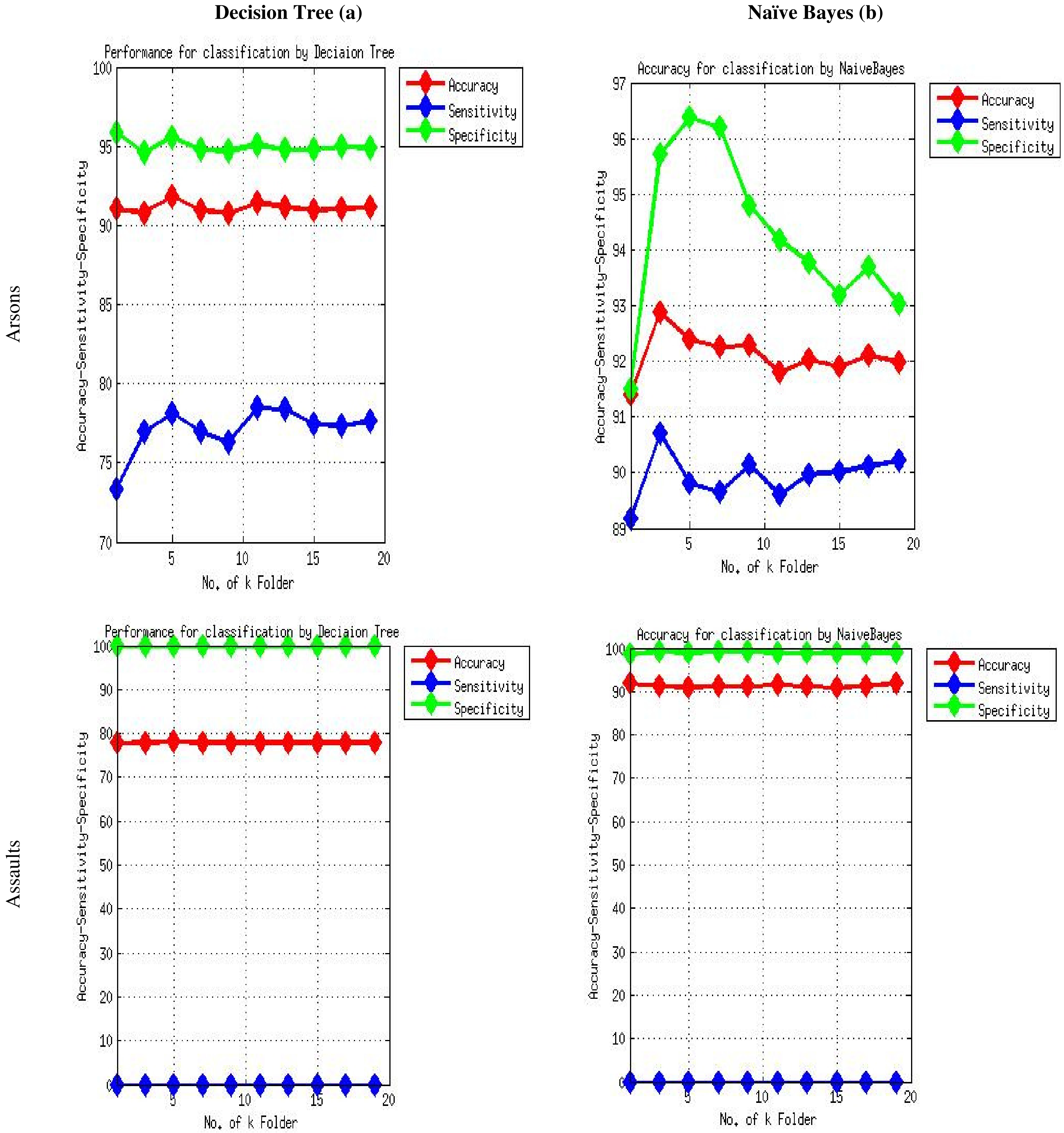
Conclusion

We compare the results of Decision tree and Naïve Bayes classifiers. According to figure-1 and figure-2 we observed that Naïve Bayes classification is more reliable and more accurate on Crime analysis. We can extract rules for classification using Naïve Bayes classifier.

References

1. Buczak A.L. and Gifford C.M., Fuzzy Association Rule Mining for Community Crime Pattern Discovery, *ACM SIGKDD Workshop on Intelligence and Security Informatics Held in conjunction with KDD-2010*, **(2010)**

2. De Bruin J., Cocx T., Kosters W., Laros J. and Kok J., Data Mining Approaches to Criminal Career Analysis, *Proc. of the Int. conf. on Data Mining*, *IEEE Computer Society Press*, 171-177 **(2006)**

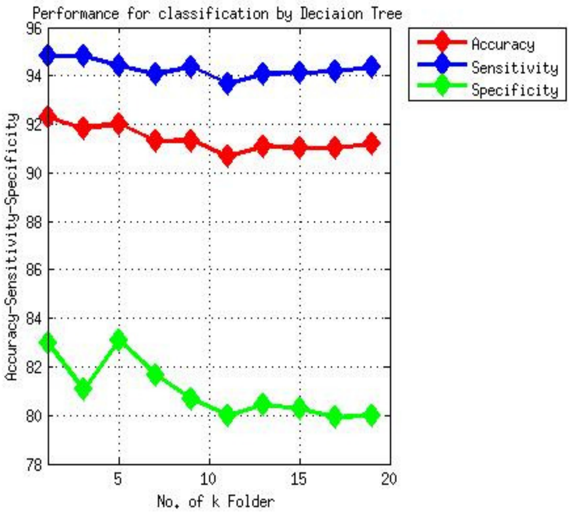

Decision Tree (a)

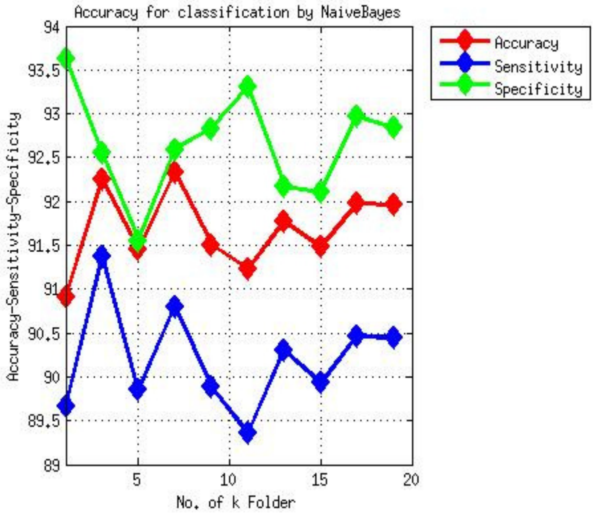
Naïve Bayes (b)

Arsons

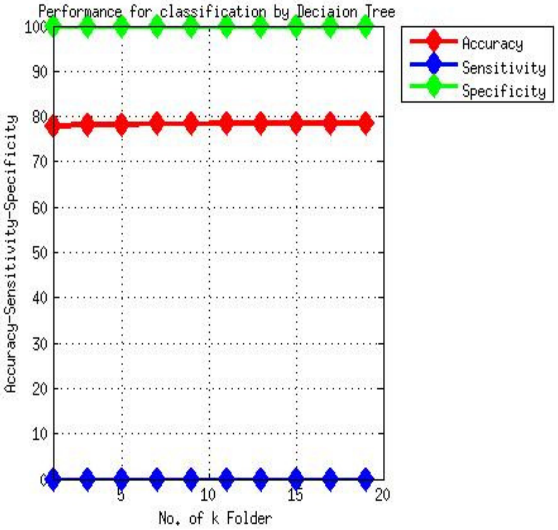
Assaults

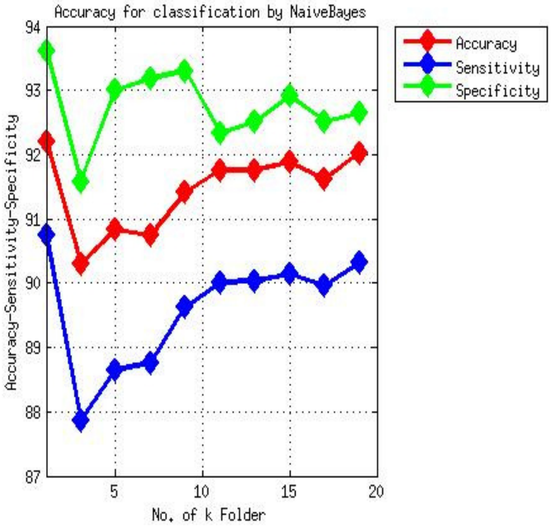
AutoThefts



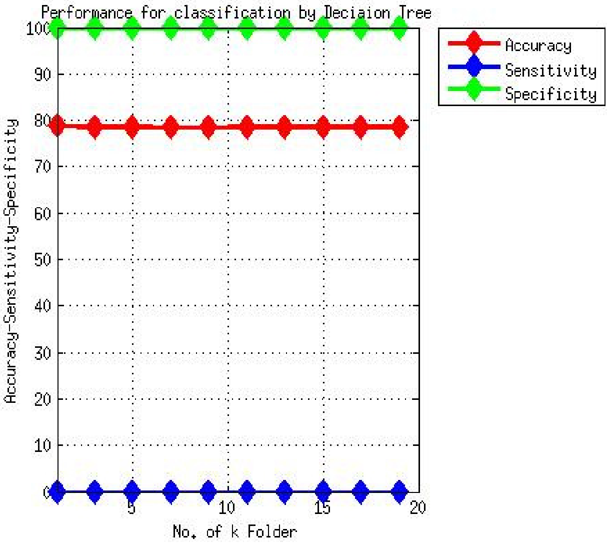


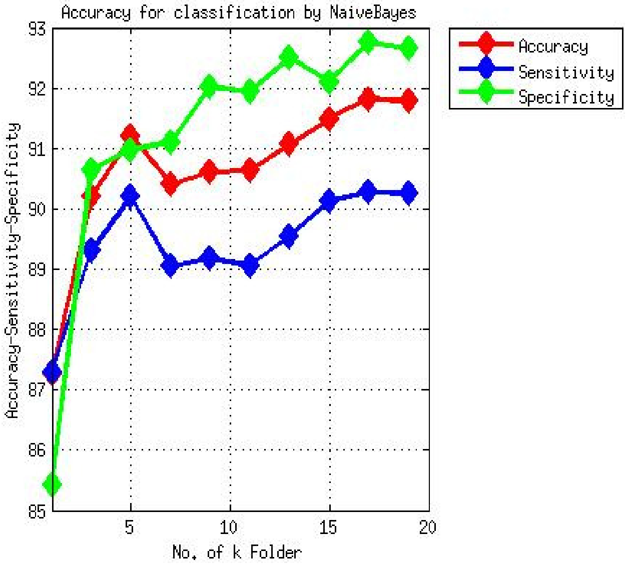
Burglaries



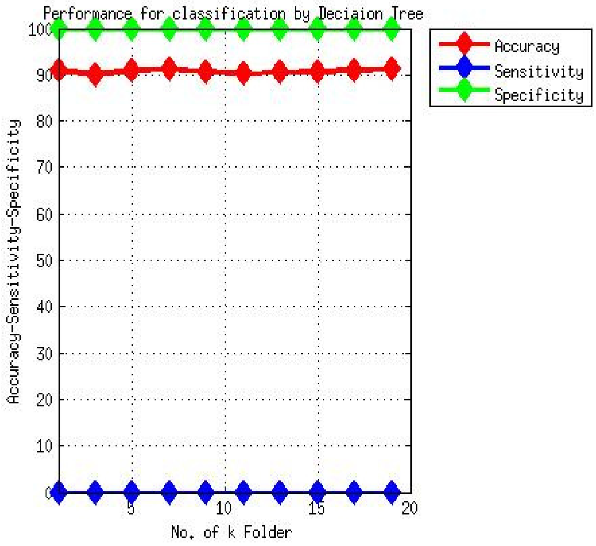


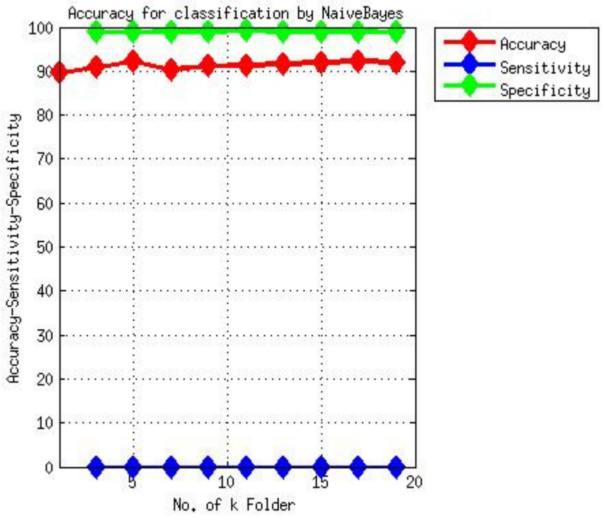
Larcenies



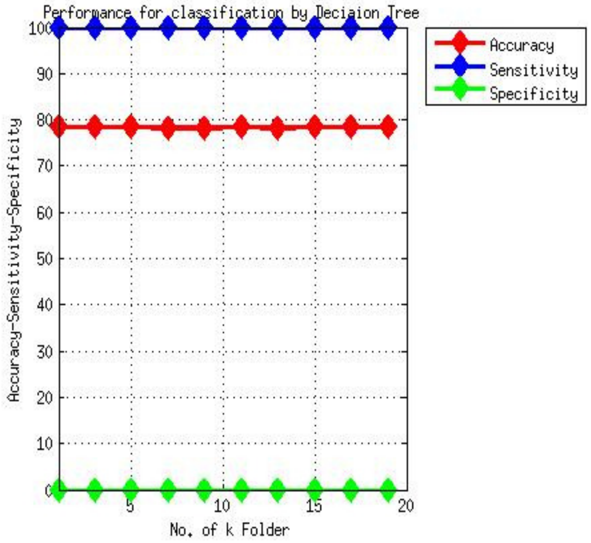


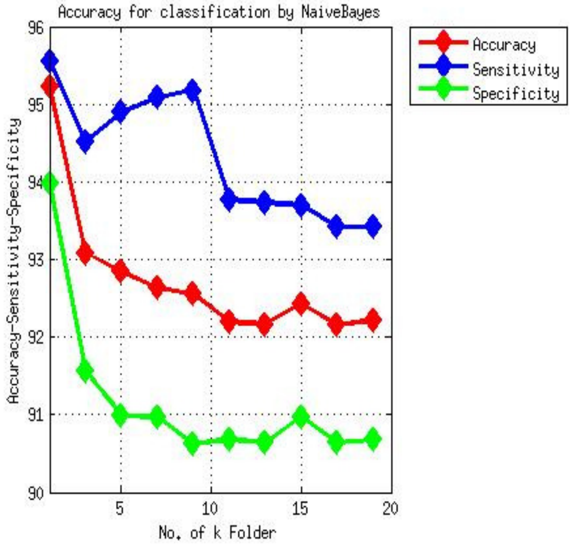
Murder



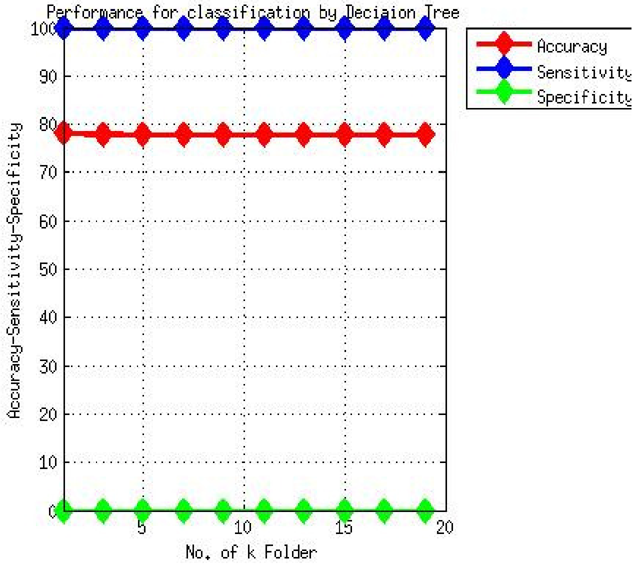


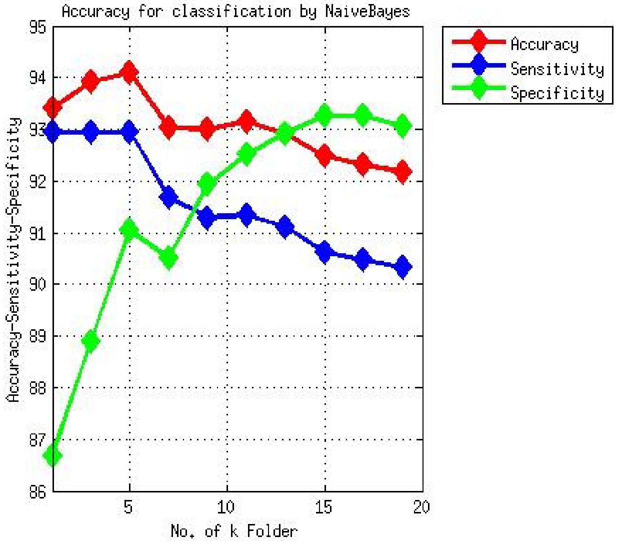
Non-violent crime





Rapes





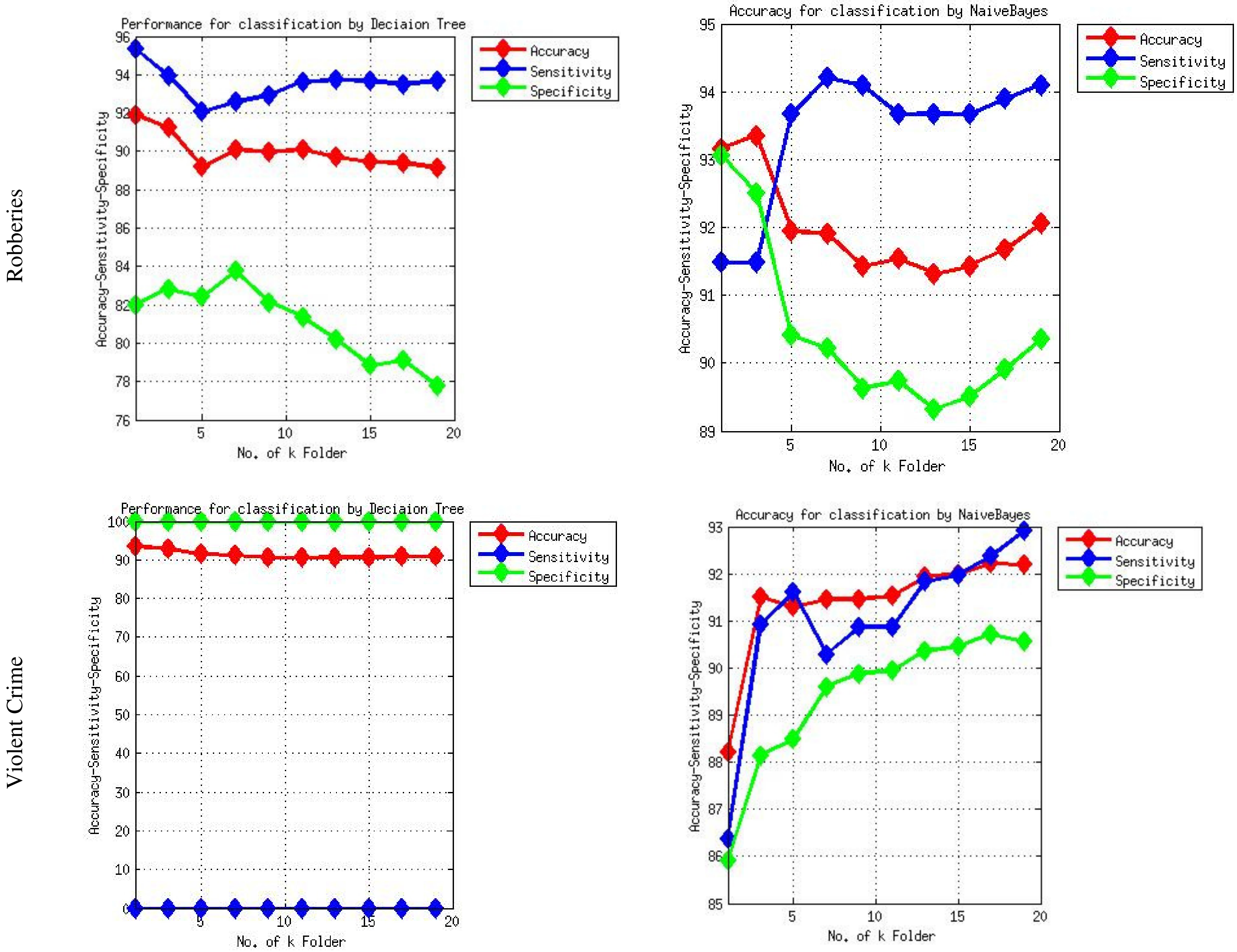


Figure-1
Accuracy – Sensitivity-Specificity Analysis

3. hau M., Xu J. and Chen H., Extracting Meaningful Entities from Police Narrative Reports, *National Conf. on Digital Gov. Res.*, 1-5 **(2002)**

4. Ku C., Iriberri A. and Leroy G., Crime Information Extraction from Police and Witness Narrative Reports, *IEEE Int. Conf. on Tech. for Homeland Security*, Boston, 193-198 **(2008)**

5. Phillips P. and Lee I., Mining Top-k and Bottom-k Correlative Crime Patterns through Graph Representations, *IEEE Int. Con. on Intell.and Sec. Informatics*, Dallas, 25-30 **(2009)**

6. Hanmant N. Renushe, Prasanna R. Rasal and Abhijit S. Desai, Data Mining Practices for Effective Investigation of Crime, *IJCTA,* **3(3), (2012)**

7. Hauck R., Atabakhsh H., Ongvasith P., Gupta H. and Chen H., Using COPLINK to Analyze Criminal-Justice Data, *IEEE Computer,* **35(3) (2002)**

8. Brown D., The Regional Crime Analysis Program (RECAP): A Framework for Mining Data to Catch Criminals, *Int. Con. on Sys., Man, and Cybernetics*, 2848-2853 **(1998)**

9. Asuncion A. and Newman D.J., UCI Machine Learning Repository, School of Information and Computer Science, University of California, Irvine, CA, 2007, http://archive.ics.uci.edu/ml/datasets/ Communities+and+ Crime (accessed, October 2013), **(2013)**

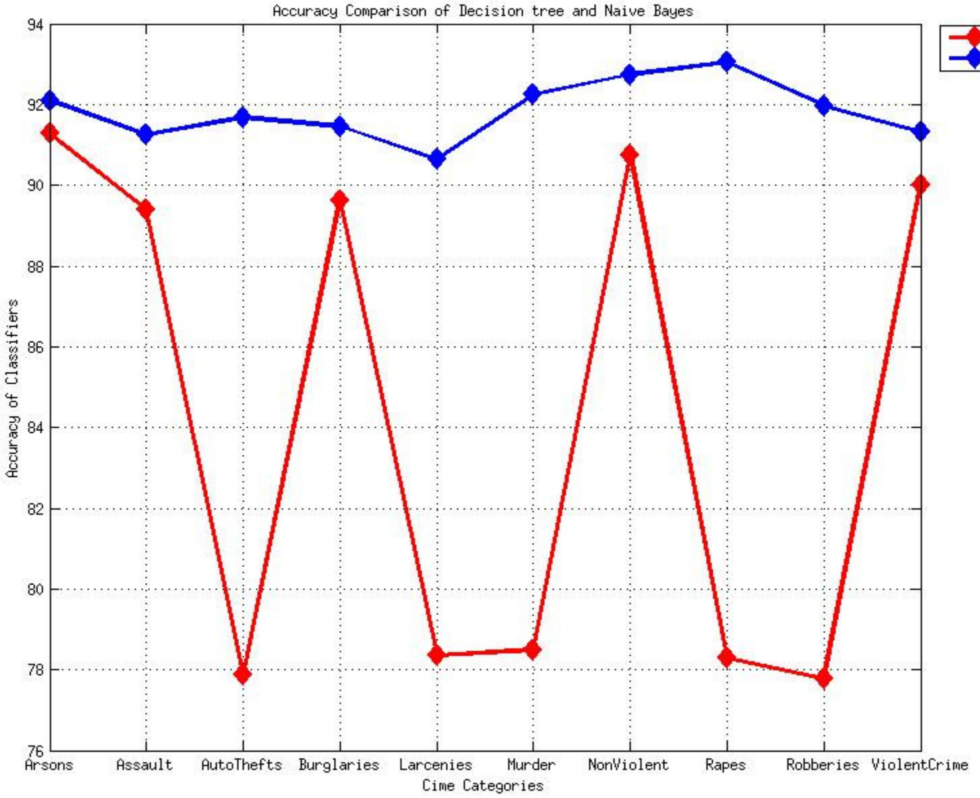


Figure-2
Accuracy Comparison analysis between Decision Tree and Naïve Bayes Classifier

10. Redmond M. A. and Baveja A., A Data-Driven Software Tool for Enabling Cooperative Information Sharing Among Police Departments, *Euro. J. of Operational Res.*, **141**, 660-678 **(2002)**

11. Grochowski M. and Jankowski N., Comparison of Instance Selection Algorithms II. Results and Comments, *ICAISC 2004a*, 580-585 **(2004)**

12. Jankowski N. and Grochowski M., Comparison of Instances Selection Algorithms I. Algorithms Survey, *ICAISC 2004b*, 598-603 **(2004)**

13. Lakshminarayan K.., Harp S. and Samad T., Imputation of Missing Data in Industrial Databases, *Applied Intelligence*, **11**, 259–275 **(1999)**

14. Kotsiantis S. B., Kanellopoulos D. and Pintelas P. E., Data Preprocessing for Supervised Leaning, *Int. J. of Comp. Sci.*, **1(2)**, 1306-4428 **(2006)**

15. Yamuna S. and Bhuvaneswari N. S., Datamining Techniques to Analyze and Predict Crimes, *The Int. J. of Engin. And Sci.*, **1(2)**, 243-247 **(2012)**

16. Polvi N., Looman T., Humphries C. and Pease K.., The Time Course of Repeat Burglary Victimization, *British J. of Criminology*, **31(4), (1991)**

17. Satish B. and Sunil P, Study and Evaluation of users behavior in e-commerce Using Data Mining, *Res. J. Recent Sci.*, **1(ISC-2011)**, 375-387 **(2012)**

18. Murangira B. and Jyoti B., DNA Technology: The Technology of Justice - Current and Future Need Thierry, *Res. J. Recent Sci.*, **1(ISC-2011)**, 405-409 **(2012)**

19. Alok A., Patra K.C. and Das S.K., Prediction of Discharge with Elman and Cascade Neural Networks, *Res. J. Recent Sci.*, **2(1)**, 279-284 **(2013)**

20. Movahedi M. M., A Statistical Method for Designing and analyzing tolerances of Unidentified Distributions, *Res. J. Recent Sci.*, **2(11)**, 55-64 **(2013)**

21. Kriti S. and Smita J., Artificial Neural Network Modeling of Shyamala Water Works, Bhopal MP, India: A Green Approach towards the Optimization of Water Treatment Process, *Res. J. Recent Sci.*, **2(1)**, 26-28 **(2013)**