Performance Comparison of Data Mining Techniques to Analyse Crime against Women

Author
Aarti Bansal
M. Tech, Department of Computer Engineering, Punjabi University, Punjab, India
Email- bansal.aarti.03@gmail.com

ABSTRACT

Crime against women is an age old phenomenon. A total of 3,09,546 cases of crime against women were reported in the country in the year 2013 as compared to 2,44,270 in 2012, thus showing an increase of 26.7% during the year 2013. With the increase in reporting of crimes against women, there is urgent need to analyse the data and develop tools & techniques that can help the concerned authorities to suitable measures to mitigate increasing crime against women. A number of algorithms have already been designed in the data mining field. The objective of this paper is to study and analyze the performance of prominent data mining techniques viz. Decision trees, Apriori and K- Nearest Neighbor for analyzing crimes against women. The performance is measured in terms of time taken, correctly and incorrectly classified instances and accuracy. From the experimental results, it was found that Decision tree algorithm is better than other two algorithms.

Keywords- Decision Tree, Apriori, K-NN, Data Mining, Crime against Women.

1. INTRODUCTION

1.1 CRIMES AGAINST WOMEN

Crime against women has become a prominent topic of discussion in India especially after Nirbhaya’s incident on 16th of December, 2012 in Delhi. The issue comes forth time to time in the form of gang-rape, sexual harassment, acid attack, dowry death, domestic violence, human trafficking and forced prostitution, marital rapes, honour killings, stalking etc. It is deeply rooted in our Indian society despite increasing literacy rate. The major reasons of it are male dominated social and political structures, inefficient legal justice system and weak rules of law. Apart from this, social negligence of women's survival, her development and economic rights, and women's own ignorance and disregard of their own rights are also among the major reasons.

According to the National Crime Records Bureau of India, reported incidents of crime against women increased 6.4% during 2012, and a crime against a woman is committed every three minutes. In 2012, there
were a total of 244,270 reported incidents of crime against women, while in 2011, there were 228,650 reported incidents. Of the women living in India, 7.5% live in West Bengal where 12.7% of the total reported crime against women occurs. Andhra Pradesh is home to 7.3% of India's female population and accounts for 11.5% of the total reported crimes against women [29].

With the increase in reporting of crimes against women, there is need for accurate and timely information to react to women crime such as identifying the age group of those who are mostly involved in crime, relation of the accused with victim. Analysis can be made regarding which age groups of girls are the main target of criminals. Apart from this, there is need to recognize public areas especially the dark areas which have high probability of crime rate so that suitable steps can be taken to prevent the same. By analyzing previous similar crime cases, we can identify the criminal or his attributes such as age group, relation etc. in new crime cases. Thus there is urgent need to analyze the data and develop tools & techniques that can help the concerned authorities to suitable measures to mitigate increasing crime against women.

1.2 DATA MINING

In modern world a huge amount of data is available which can be used effectively to produce vital information. The information achieved can be used in many real life applications especially in the field of Medical science, Education, Business, Agriculture and so on. As huge amount of data is being collected and stored in the databases, traditional statistical techniques and database tools are no longer adequate for handling this huge amount of data [10].

1.2.1 Introduction

Data mining refers to extracting or “mining” knowledge from large amounts of data. Data mining which is also called as “Knowledge discovery from data or KDD” is the process of discovering interesting patterns and relations from voluminous amount of data. It is an essential process in today’s world because it uncovers hidden patterns for evaluation. These patterns can then be used for marketing analysis, making strategies, taking decisions, to increase revenues etc. Data mining provides a number of analytical tools and algorithms for analysing data. It provides various functionalities to data like multidimensional views of data, pre-processing of data, classifying data into classes according to their features, clustering the data etc.

1.2.2 Why use data mining?

Two main reasons to use data mining:

- Too much data and too little information.
- Need to extract useful information from the data.

Dealing huge volumes of data with no special tools make human analysts’ work very difficult. Data mining is used especially in science and business areas where there is need to analyze voluminous amount of data to discover patterns which they could not otherwise find. Besides these, data mining can be used in any field like banking, finance, retail, engineering, medical, web etc [18].
1.2.3 Data Mining Process

Data mining consists of five major elements as explained in figure 1.1

- Extraction and transformation of data onto the data warehouse system.
- Run data on multidimensional database system in a managed way
- Providing data access to business analysts and other professionals
- Data analyzing
- Presentation of data in useful and required formats such as tables and graphs.

![Figure 1.1 Data Mining Process](image)

Since data mining is highly application-driven, it is not possible to enumerate all applications where data mining plays a critical role. Some of the notable applications of data mining are science and engineering mining, business mining, spatial data mining, visual data mining, sensor data mining, pattern mining, medical data mining, web mining etc.

1.3 SCOPE OF DATA MINING

Given databases of huge size and quality, data mining technology provides new opportunities in the research field.

1.3.1 Automation in prediction of behaviour and trends

Data mining automates the process of finding information in large databases. Traditionally methods of data mining required extensive analysis by humans’ hands and with data mining it has become direct to answer the predictions. An example of the same is targeted marketing where it uses data of past promotional mailing system to identify the customers who will probably maximize the return on investment in future mailings. Other examples include insurance analysis for prediction and decision making, income tax department of government for fraud discovery.

1.3.2 Automated discovery of previously unknown patterns

Data mining tools sweep through databases and identify hidden information in them. An example of pattern discovery is to identify items that are often purchased together from sales data. Other examples include detecting fraudulent credit card transactions. Data mining techniques can also be implemented on new
systems as well as on existing platforms. Data mining tools can analyse massive databases in minutes. Larger databases, in turn, yield improved predictions [11].

1.4 GOALS OF DATA MINING

- **Prediction** -- How certain variables within the database will behave in the future
  - Predict what consumers will buy under certain discounts
  - Predict how much sales volume would generate in a given period
  - Predict an earthquake based on certain seismic wave patterns

- **Identification** -- Use previous data pattern to identify the existence of an item, an event, or an activity
  - Intruders trying to break a system may be identified by the programs executed, files accessed, and CPU time per session.

- **Optimization** – Optimize the use of limited resources such as time, space, money, or materials and maximize output variables such as sales or profits under certain constraints.

- **Classification** -- Partition data into classes and categories based on combinations of parameters. For example, customers may be categorized into
  - Discount-seeking shoppers
  - Shoppers in a rush
  - Loyal regular shoppers
  - Infrequent shoppers

1.5 APPLICATIONS

Data mining has been applied to countless areas, from medicine, biology, Engineering Management, Digital content, to the most flourished commercial area. The industries include: finance and insurance, retail, telecommunication, manufacture, and security investment. The most popular business applications are listed as follows:

- **Finance and Insurance**
  - Financial product cross-selling: For example, a customer of a laptop manufacturing company would like to buy other products of the same company.
  - Telemarketing: Data mining figures out the most responsive and valuable customers for certain products.
  - Product mix analysis: considers customer’s needs and marketing strategy to find out the best product mix.

- **Retail**
  - Analyzes customer’s transaction records to further introduce them other products.
  - Promotional product mix: discovers the items that often be purchased together by customers, so these products can be placed closely or be sold in pairs.
- Direct marketing: needs analytical software to figure out the most responsive and valuable customers for marketing.
- Inventory analysis: analyzes inventory and sales data to calculate the trend of future requirement. This information can be the accordance of stock replenishing and marketing decision making.

- **Manufacturing**
  - Production yield analysis: uses prediction methods to detect defective products quickly to improve the yield and to reduce the production failure cost.
  - Production process control: finds out the factors in the processes that cause failure, delay in production or identify and analyze the quality issues.

- **Security investment**
  - To build a simulation model to predict the fluctuation of stock price and foreign exchange.
  - Direct marketing

- **Fraud detection**
  When a transaction occurs, system will compare current transaction with existing profile to analyze where this transaction is dubious or not. If transaction has a high probability to be a fraud, so that transaction is not accepted and an alarm will be sent to the manager and the customer.

- **Stock movement prediction**
  Stock movement prediction model also utilizes data mining methods to do analysis of previous year’s data. The stock market’s performance is the result of analysis of all participator’s collaboration and Investor’s behavior over a period of time.

### 1.6 DATA MINING TECHNIQUES

In this report, three techniques (decision tree, Apriori and K-Nearest Neighbor) have been studied, analyzed and implemented. These algorithms are explained as follows:

#### 1.6.1 Decision Tree

Decision tree is a powerful classification technique. The decision trees, take the case described by its features as input, and outputs a decision. Each leaf node corresponds to a class label. A new example is classified by following a path from the root node to a leaf node, where at each node a test is performed on some feature of that example. The leaf node reached is considered the class label for that example. Figure 1.2 shows an example of decision tree representing root node, leaf nodes and internal nodes. It is a flowchart-like structure in which each internal node is a test on an attribute, each branch is an outcome of test and each leaf node represents class (decision taken after computing all the attributes). A path from root to leaf represents classification rules [16]. Figure 1.3 represents a decision tree for analyzing crimes against women.
Some of the major advantages of decision tree algorithm are:

- **Simple to understand and interpret.** People are able to understand decision tree models easily due to its flowchart like structure.

- **Requires little data preparation.** Other techniques often require data normalization, dummy variables etc.

- **Able to handle both numerical and categorical data.** Other techniques are usually specialized in analyzing only one type of variable.

- **Possible to validate a model using statistical tests.**

- **Robust.** Performs well even if its assumptions limited.

- **Performs well with large datasets.** Large amounts of data can be analyzed using standard computing resources in reasonable time [16].
• K-Nearest Neighbors
• Memory-Based Reasoning
• Example-Based Reasoning
• Instance-Based Learning
• Case-Based Reasoning
• Lazy Learning

Figure 1.4: Flowchart of Case-based learning (K-NN Approach) [24]

Figure 1.5 shows an example of KNN classification technique where a new case represented with a green circle is to be classified either as Blue Square or red triangle. If k = 3 (solid line circle) it is labeled as red triangle because there is 1 square and 2 triangles inside the inner circle. If k = 5 (dashed line circle) it is labelled as blue squares (3 squares vs. 2 triangles inside the outer circle).

Figure 1.5: Example of K-NN Classification Algorithm

1.6.3 The Apriori Algorithm

1.6.3.1 Association Rules

Association rules means finding frequent co-associations among a set of items. It is sometimes referred to as “Market Basket Analysis” i.e., set of items that are often purchased together.
Use of association rules in day to day life:

- Shopping centers use association rules to place the items next to each other so that users buy more items.
- Google auto-complete, where when we type in a word it searches frequently associated words that user type after that particular word [27].

1.6.3.2 Apriori Algorithm

Apriori algorithm is an association rule mining algorithm used in data mining. It is used to find the frequent item set among the given number of transactions [21]. Figure 1.6 shows an example of Apriori Algorithm using transaction data set.

- Advantages
  - Uses large dataset.
  - Easily parallelized
  - Easy to implement

- Disadvantages
  - Larger memory space for larger number of candidates.
  - Multiple database scans for generating candidate sets.
  - Execution time is wasted in producing candidates in every scan.

![Figure 1.6 Example of Apriori using transaction data set [23]](image-url)
The Figure 1.7 shows the comparison among the three data mining techniques based on different criteria.

<table>
<thead>
<tr>
<th>Fields</th>
<th>Decision Tree</th>
<th>Apriori</th>
<th>K-Nearest Neighbor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Major</td>
<td>Decision tree algorithms are used to build tree, if-then else rules to classify the data items.</td>
<td>It is used for mining frequent item sets for association rules.</td>
<td>It is a distance based algorithm.</td>
</tr>
<tr>
<td>Classifier</td>
<td>It recursively partitions data set using depth-first approach or breadth first approach. It will repeat till all data items are not assigned to some class.</td>
<td>It generates candidates of frequent item sets: scans database, calculate support of each candidate frequent item set and add those which satisfies the minimum support requirement.</td>
<td>For unknown sample, it searches the pattern space from training samples which is close to the unknown samples.</td>
</tr>
<tr>
<td>Different phases</td>
<td>Tree building Tree Pruning</td>
<td>Joining &amp; Pruning.</td>
<td>Finding distance. Assigning class to maximum class amongst k.</td>
</tr>
<tr>
<td>Advantages</td>
<td>Domain Knowledge not required. Works with huge dimensional data.</td>
<td>Uses large item set property. Easy to implement.</td>
<td>Implementation easy for parallel implementation work with local info.</td>
</tr>
<tr>
<td>Disadvantages</td>
<td>Categorical output. One output attribute.</td>
<td>Larger memory space for larger number of candidates, multiple database scans, Execution time is wasted in producing candidates in every scan.</td>
<td>Requires large storage area. May be slow in classifying tuples.</td>
</tr>
</tbody>
</table>

![Figure 1.7: Comparison of Data mining techniques](image)

### 2. LITERATURE REVIEW

Xiong Xiangchun et al. (2005) [1] combines some statistics and data mining techniques that can help in uncovering patterns hidden in the breast cancer database. PCA, PLS linear regression analysis with data mining methods such as select attribute, decision trees and association rules have been combined to find the unsuspected relationships. These can be used as reference for doctor’s decision-making and organization’s research and so on.

Alhajj Reda et al. (2005) [3] describes an interactive graphical tool called Visual Apriori for discussing two promising frequent item set generation algorithms, viz. Apriori and Eclat. It provides an active participation from the end user. Visual Apriori provides a new experience for a naive user by providing user-friendly and easy to understand features.

Yu Hong et al. (2010) [4] compares four data mining techniques - logistic regression (LR), decision tree (C4.5), support vector machine (SVM) and neural networks (NN) on the basis of efficiency by applying them to two data sets of credits. The results show that the LR and SVM techniques produce the best classification accuracy, and the SVM shows the higher robustness as compared to other algorithms. On the
other hand, the neural network (NN) technique performs relatively poor. C4.5 algorithm varies according to input data, and the classification accuracy is unstable.

Peiying Wang (2010) [5] discusses both objective and subjective reasons for increasing women crimes in China. The prime reasons are the overall negligence of women's survival and education, her development and economic rights, and women's own ignorance and disregard of their rights. The characteristics and causes of female crimes in China are analyzed first and then appropriate strategies have been proposed with the aim to reduce female crimes.

Kumar S et al. (2010) [13] have compared Apriori and Frequent Pattern (FP) Growth algorithms for web mining based on memory usage and time usage. Apriori algorithm is easier to implement but becomes costlier with large number of patterns whereas FP Growth algorithm uses complete and compact data structure and eliminates the repeated scan problem. But the main drawback is that it lacks a good candidate generation method with the explosive quantity of data. The research has combined both the algorithms to solve the disadvantages of Apriori and FP Growth algorithms.

Gupta Anish et al. (2012) [11] have explained the meaning of data mining and its process, scope and various techniques of it. The author has also discussed various security concerns of data mining and its security aspects and measures related with the databases for data mining. The security measures are very important for its applications. It has been suggested that a security measure should be implemented on behalf of the company policies.

Huang Shin-Chen et al. (2013) [6] conduct a comparative analysis on the accuracy of data mining classification techniques namely, support vector machine, decision tree, neural network and logistic regression for credit check in banking and reduce the credit risk. The support vector machine model has higher accuracy rates and therefore outperforms other classification methods in the context of credit risk in banking.

Shah Chintan et al. (2013) [7] have used three different data mining classification algorithms for prediction of breast cancer namely decision tree, Naïve bayes, and K-Nearest Neighbor with the help of WEKA (Waikato Environment for Knowledge Analysis), which is a open source software. Different parameters have been compared for prediction of cancer. But, for superior prediction, accuracy and lowest computing time have been focused. It has been concluded that Naïve Bayes is a superior algorithm compared to the two others because it takes the lowest computing time and at the same time provides highest accuracy.

Uppal Veepu et al. (2013) [10] describes the method to solve the problems faced in library because of the huge growth of library data and to improve the quality of managerial decisions. In this paper, various data mining techniques have been used that are helpful in predicting the allocation of books in library, need of the department, analysis of book circulation by time series and pattern identification of inventory loss. The main motive is that book occurrences in frequent sequences, layout of books should be arranged such that readers can easily find the books.
Bansal Divya et al. (2013) [14] has elaborated the use of association rule mining for extracting patterns within a dataset. The implementation of Apriori algorithm on a dataset containing crimes against women has been shown. For this, WEKA tool has been used for extracting results. A comparison analysis between Apriori and Predictive Apriori Algorithm has been done. The Apriori Algorithm is more efficient than Predictive Apriori Algorithm.

Singh A K et al. (2014) [8] discuss the comparison of two prominent data mining algorithms viz. Apriori and FP-growth for Web Usage data on frequent item set sequences. The comparison concentrates on web usage mining using both the algorithms and focuses on determining the patterns of web usage from the server log files. The results show that FP growth has a better performance than Apriori Algorithm. It is efficient for mining both long and short frequent patterns.

Ngaruiya Njeri et al. (2015) [2] uses two promising data mining tools (R Environment and WEKA) to derive patterns in Prostate Cancer. A tool has been built for identification of the Gleason score. It helps in deciding the treatment technique for Prostate Cancer. The WEKA and R Algorithm used gives almost the same results but the R Algorithm is an easier tool to learn and its representation of data is much efficient and easy to read. In developing a predictive model for prostate cancer, three popular algorithms including Logistic Regression, Decision trees (CART algorithm) and ANN were applied. The results show that ANN is better predictor compared to other two in all the instances. The patterns achieved will assist the GOK (Government of Kenya) for correctly placing the cancer diagnosis and treatment equipment which were launched by the National Government of Kenya in early 2015.

Ram Shrawan et al. (2015) [15] have done a comparative study and evaluation of decision tree and Artificial Neural Network with the help of Statlog Heart Diseases Database collected from UCI machine learning repository. These algorithms have been compared on the basis of classification accuracy and performance matrices.

3. PROBLEM FORMULATION

Crime against women is an alarming public issue not only in India but worldwide too. There has been a massive increase in reporting of crime rate against women. There is need for accurate and timely information to react to women crime such as identifying the age group of those who are mostly involved in crime, relation of the accused with victim, whether accused is stranger or known to the victim etc. can be of immense help. By analyzing previous similar crime cases, we can identify the criminal or his attributes such as age group, relation etc. in new crime cases.

Analysis can be made regarding which age group of girls are the main target of criminals. Apart from this, there is need to recognize public areas especially the dark areas which have high probability of crime rate so that suitable steps can be taken to prevent the same. Such information can be helpful for the Government, society and police to suggest measures to be taken towards creating a peaceful society. It will also help in
improving the appalling situation of women in society. Thus, a comparative study of Data classification algorithms has been proposed for analyzing crime against women.

3.1 Objective and Methodology

3.2.1 Objective: To study and compare some of the promising data mining algorithms for analyzing crimes against women.

3.2.2 Methodology: The tentative process followed during the course of Research Project: (as shown in Figure 3.1)

Step 1: Understanding the algorithms.
Step 2: Implementing a first draft of the algorithm step by step.
Step 3: Testing with the input files.
Step 4: Cleaning the code.
Step 5: Optimizing the code.
Step 6: Comparison of the performance with other algorithms.

![Figure 3.1 Methodology of Proposed Work]

4. IMPLEMENTATION OF THE TECHNIQUES

4.1 K-NEAREST NEIGHBOR ALGORITHM

In KNN algorithm we have found result on two bases:-

-- In first case we find minimum Euclidean distance with input values and value of relation with victim is found out and,

-- In second case we pick first 10% entries with minimum Euclidean distance. After that we pick entry of relation with the victim with maximum frequency.
Steps followed for implementing KNN Algorithm:-

1. Read dataset
2. Read input set
3. Repeat for iv (Input Variable) from 1 to inputset_size
4. Repeat for i from 1 to dataset_size
5. Find Euclidean distance using formula

\[
eludiff_{i} = \sqrt{(state_{iv} - state_i)^2 + (type_{iv} - type_i)^2 + (ageAccused_{iv} - ageAccused_i)^2 + (ageVictim_{iv} - ageVictim_i)^2}
\]

6. Sort eludiff array and find minimum euclidean distance.
7. Relation with victim of minimum Euclidean distance is found out.
8. If it matches with RelationI iv then increment correctlyClassifiedInstances
9. Pick 10% entries with minimum Euclidean distance.
10. Find maximum frequency of relation of victim.
11. If it matches with RelationI iv then increment correctlyClassifiedInstancesTen.

4.2 DECISION TREE ALGORITHM
Steps followed for implementing Decision Tree Algorithm:-

1. Read dataset
2. Read input set
3. Repeat for iv (Input Variable) from 1 to inputset_size
4. Find entries from dataset set in which state is same as stateI iv and store them in array a1.
5. Now find entries from a1 in which type of crime is same as typeI iv and store them in array a2.
6. Find entries from array a2 in which age of accused is same as ageAccusedI iv and store them in array a3.
7. Find entries from array a3 in which age of victim is same as ageVictimI iv
8. Now find maximum frequency of relation of victim.
9. If it matches with RelationI iv then increment correctlyClassifiedInstances.

4.3 APRIORI ALGORITHM
Steps followed for implementing Apriori Algorithm:-

1. Read dataset
2. Read input set
3. Repeat for iv from 1 to inputset_size
4. Repeat for I from 1 to dataset_size
5. If \( \text{state}_i = \text{state}_{iv} \) \&\& \( \text{type}_i = \text{type}_{iv} \) \&\& \( \text{ageAcussed}_i = \text{ageAcussed}_{iv} \) \&\& \( \text{ageVictim}_i = \text{ageVictim}_{iv} \) then store such entries in array \( a_1 \)

6. If array \( a_1.\text{size} = \text{null} \) then

7. If \( \text{state}_i = \text{state}_{iv} \) \&\& \( \text{type}_i = \text{type}_{iv} \) \&\& \( \text{ageAcussed}_i = \text{ageAcussed}_{iv} \) then store such entries in array \( a_2 \)

8. If array \( a_2.\text{size} = \text{null} \) then

9. If \( \text{state}_i = \text{state}_{iv} \) \&\& \( \text{type}_i = \text{type}_{iv} \) then store such entries in array \( a_3 \)

10. Find maximum frequency of relation with victim in \( a_3 \) or \( a_2 \) or \( a_1 \) accordingly.

11. If it matches with \( \text{Relation}_{iv} \) then increment correctlyClassifiedInstances.

5. RESULTS AND DISCUSSIONS

The implementation of the three promising data mining techniques namely, Decision Tree, Apriori and K-Nearest Neighbor (KNN) Algorithm has been shown using MATLAB (R2015a). The implementation has been carried out using Windows Operating system on a PC with Intel(R) Core CPU running at 2.00 GHz, with 4 GB of RAM. Figure 5.1 shows the working of Decision Tree Algorithm in Command Window of MATLAB. The elapsed time for running Decision Tree algorithm is 0.2868 seconds. With training data of 50 instances, the correctly classified instances are 38 and incorrectly classified instances are 12. Thus the accuracy of decision Tree algorithm is 76%. Figure 5.2 represents the working of Apriori Algorithm in Command Window of MATLAB. The elapsed time for running Apriori Algorithm is 0.3705 seconds. With training data of 50 instances, the correctly classified instances are 38 and incorrectly classified instances are 12. Thus the accuracy of Apriori algorithm is 76%. Similarly Figure 5.3 represents the working of K-NN algorithm. The elapsed time is 44.7031 seconds. The correctly classified instances are 36 and incorrectly classified instances are 14. Its accuracy is 72%. However the correctly classified instances for matching 10% of cases with minimum Euclidian distance are 28 and incorrectly classified instances are 22.

Figure 5.1:- Working of Decision Tree Algorithm
5.1 PERFORMANCE OF ALGORITHMS BASED ON DIFFERENT PARAMETERS

Figure 5.2: Working of Apriori Algorithm

Figure 5.3: Working of K-NN Algorithm

Figure 5.4: Performance of Algorithms on the basis of Time taken (In seconds)
6. CONCLUSION

Several data mining algorithms have been compared by researchers using various real life applications. In this dissertation report, three prominent data mining techniques (Decision Trees, Apriori and K-NN) have been studied, analyzed and compared for analyzing crimes against women using MATLAB (R2015a).
Figure 6.1 shows the experimental results. From the results, it has been found that Decision tree is better than other two techniques. The elapsed time for decision tree is the minimum. Apriori Algorithm is also one of the good techniques. The accuracy of both decision tree and Apriori is same. However the performance of K-NN Algorithm is less in comparison to Apriori and Decision tree with the given training set.

<table>
<thead>
<tr>
<th>Classifier</th>
<th>Time Taken (in seconds)</th>
<th>Correctly Classified Instances</th>
<th>Incorrectly Classified Instances</th>
<th>Accuracy(in %age)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Decision Tree</td>
<td>0.2868</td>
<td>38</td>
<td>12</td>
<td>76</td>
</tr>
<tr>
<td>Apriori</td>
<td>0.3705</td>
<td>38</td>
<td>12</td>
<td>76</td>
</tr>
<tr>
<td>K-Nearest Neighbor</td>
<td>44.7031</td>
<td>36</td>
<td>14</td>
<td>72</td>
</tr>
</tbody>
</table>

Figure 6.1 Performance study of Algorithms

7. FUTURE SCOPE
With the increase in reporting of crimes against women, there is urgent need to develop such tools and techniques that will help the concerned authorities to get the attributes of the accused person. At the same time it will also help in taking suitable measures to mitigate increasing crime rate against women.

In Future, work can be done on the following points:-
- To study and compare other data mining classification algorithms.
- To extend the algorithm for large data set.
- To reduce its complexity.
- To improve the performance of the algorithms.
- To test the applicability of the algorithms in other real life applications.

REFERENCES


