CRIME PREDICTION SYSTEM

ARYAN AGARWAL: 17BCE0600

NAMAN DUTTA: 17BCE2379

DATABASE MANAGEMENT SYSTEMS (CSE-2004)

SUBMITTED TO- Prof. GOVINDA K

Vellore Institute of Technology, Vellore, Tamil Nadu

ABSTRACT

In this Project, we have implemented a Crime Prediction System(CPS). We first created hypothetical datasets samples of major city areas and different crimes taking place and then we used the algorithms to analyse it. For this project, we used HTML and CSS along with PHP, while wamp as a web server to this application. The main purpose of this project was to analyse the city and predict the chance of a crime happening using Apriori algorithm. In addition, we used Decision Tree as a searching algorithm and Naïve Bayesian classifier to predict the crime that can happen in a particular area at a certain time. The result of this can be used to raise people's awareness regarding the dangerous locations and to help agencies to predict future crime in a specific location within a particular time.

1. INTRODUCTION

Crimes are a typical social issue influencing the quality of life and the financial development of a society. It is viewed as a fundamental factor that decides if individuals move to a new city and what spots ought to be evaded when they travel. With the expansion of violations, law requirement organizations are proceeding to request progressed geographic data frameworks furthermore, new data mining ways to deal with enhance crime examination and better secure their networks.

In spite of the fact that crimes could happen all over the place, usually, hoodlums deal with crime openings they look in most recognizable zones for them. By giving an information mining approach to decide the most criminal hotspots and discover the sort, area and time of carried out crimes, we would like to raise individuals' mindfulness with respect to the perilous areas in certain eras. In this manner, the proposed arrangement can possibly enable individuals to avoid the areas at a certain time. What's more, having this sort of learning would assist individuals with improving their living spot decisions. Then again, police can utilize this solution to increase the level of crime prediction and counteractive action. Besides, this would be valuable for police assets portion. It can help in the resource allocation of police at in all probability crime places for some random time, to give a productive utilization of police assets. By having the majority of this data accessible, we plan to make the community more secure for the general population living there and furthermore for other people, who will go there.

2. PROBLEM STATEMENT

We aim to find the criminal hotspots using a hypothetical dataset of the city Vadodara. We will try to find the type of crime that can happen in a certain area and the chances of it to actually occur. The algorithms we will be using for the project are Apriori, Decision Tree and Naïve Bayes classifier.

3. LITERATURE REVIEW

There has been incalculable of work done regarding crime. Extensive datasets have been reviewed and data such as area and the kind of crime have been analysed to enable individuals to pursue law authorizations. Existing techniques have utilized these databases to recognize crime hotspots depending on areas. Despite the fact that crime areas have been recognized, there is no data accessible that incorporates the crime event date and time alongside systems that can precisely anticipate what crimes will happen later on.

We analysed some of the past work and the research papers regarding crime prediction. Below is a brief description of some of the past works.

An overview on crime prediction methods

2017 6th ICT International Student Project Conference (ICT-ISPC)

This paper has introduced various crime prediction methods. It discuss about “Support Vector Machine”, “Fuzzy Theory”, “Artificial Neural Network” and “Multivariate Time Series”. This research work focuses on reviewing a crime prediction analysis tool for many scenarios using particulars crime prediction methods that can help law enforcement to efficiently handle crime incidents. *Support Vector machine* (SVM) has performed well in prediction of time series crimes because they can model nonlinear relations in a stable and efficient way. SVM is applied to predict crime hotspots resulting in a global solution. *Fuzzy Theory modelling* requires numerical inputs and uses IF-THEN regulations to form vague prediction and quadratic combination of the presumptive variables. It requires less computational complexity and has well in learning abilities. The characteristic of the fuzzy modelling is used to improve the prediction efficiency. *Artificial Neural Network* (ANN) depends on the prediction by keenly investigating the pattern from an effectively existing voluminous historical set of

data. *Multivariate Time Series* is one of the statistical tools to study the behaviour of time-dependent data and predict the future values based on history of variations in the data. This method was applied in testing of statistical significance of multivariate time series analysis technique [17].

Crime prediction and forecasting in Tamilnadu using clustering approaches

2016 International Conference on Emerging Technological Trends (ICETT)

This research implements *KNN classification* that searches through the data to highlight similar instance when an input is given to it. The paper discusses about *K-Means Clustering* method to provide a large criminal data and simplify the records and ease in handling, searching and retrieving. *Agglomerative Hierarchical Clustering* assigns each objects to its cluster and then integrates these clusters to form a larger cluster. DBSCAN is based on density clustering method. The algorithm develops areas with appropriately high density into clusters and finds clusters of arbitrary shape in spatial databases with noise. It defines a cluster as a maximal set of density-connected points [16].

Crime tracer: Activity space-based crime location prediction

2014 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2014)

The research concludes through Crime Pattern Theory that offenders, rather than venture into unknown territory, frequently commit opportunistic crimes and serial violent crimes by exploiting openings they experience in spots they are most comfortable with as a feature of their activity space, which includes the most frequently visited places as determined by a person's daily routine activities, such as commuting patterns. *Random walk* over a graph shows how offenders encounter criminal opportunities, given that the behaviour of a random walk model is local. *Starting Probabilities* provides anchor locations of all the offenders' location, linking them and thus compute probability of each anchor location of two primary factors. Offender-based CF. The intuition behind the offender-based CF approach (OCF) is that offenders who had similar behaviour in the past will have similar behaviour in the future [15].

A novel serial crime prediction model based on Bayesian learning theory

2010 International Conference on Machine Learning and Cybernetics

Geographic Profile of a Specific Factor uses Discrete Distance Decay Function and Gaussian model to calculate its probability distribution in the region. *Dynamic Prediction Model based on Bayesian Learning Method* for a specific offender, the hypothesis above may not stand. Therefore, it needs to adjust the effect function dynamically based on the dynamical known crime data, especially data of crime sites. The paper brings forward an adaptive adjustment algorithm and put it in a Bayesian learning framework [14].

Classify interval range of crime forecasting for crime prevention decision making

2016 11th International Conference on Knowledge, Information and Creativity Support Systems (KICSS)

The Decision Support System (DSS) is intuitive, computer-based frameworks and subsystems intended to help decision-makers in utilizing communication technologies, information, archives, learning or potentially models to finish decision process tasks, explain in. They also stated that DSS can improve the effectiveness and quality of decision making by processing a lot of data and providing alternative solutions to various problems. Crime analysis is done first by identifying crime incidents, victim profile and potential risk, monitoring crime threshold, and examine crime trends. The research develops and application tool using Input Data module, Statistic Module and Crime prevention Decision Module and uses If-Then rules to forecast and associate qualitative attributes of crime parameters [13].

[Kurt Hornik et al., (2012)] introduced the R-extension package that delivers a computational environment containing fixed point and genetic algorithm solvers with interfaces to two external solvers (CLUTO and Gmeans) for spherical k-means clustering. The large-scale benchmark has been used to analyse the performance of the solvers. The authors concludes that the solvers provides better solution with the interface Gmeans and CLUTO that gives fast and enhanced results.

Brian Kulis and Michael I. Jordan (2011) analysed the links rising among the DP mixture models and hard clustering algorithms by revisiting the k-means clustering algorithm from a Bayesian nonparametric viewpoint. They proposed

hierarchical Dirichlet process provides high precision results and reduce the time computation complexity.

[Navjot Kaur et al., (2012)] presented an overview to k-means clustering. They used the ranking method for k-means clustering and compared its performance with the traditional k-means clustering. As a result of the comparison, the authors conclude that the ranking based k-means clustering had less execution time and provide better results than the traditional method.

Xingan Li and Martti Juhola (2014) had applied self-organizing map (SOM) for mapping crime data of 56 countries with 28 different crime situations and conclude that SQM would be a new crime mapping tool that would process a large amount of data.

Malathi and Dr S. Santhosh Baboo (2011) developed a crime analysis tool for Indian scenario by using various data mining techniques like k-means clustering and DBSCAN clustering. The authors conclude that the tool can be used by the Indian police and law agencies for crime detection and prevention as it provides faster analysis results and identifies common crime patterns.

Raphael Obi Okonkwo and Francis O. Enem (2011) analysed various data mining techniques that been adopted by various law agencies to identify and prevent terrorism. The authors had also studied the restrictions of data mining in fighting crime in Nigeria and concluded that data mining can only be used to assist the law agencies to analyse crime.

Uttam Mande et al., (2012) presented a new methodology for identifying the criminals who committed the crime and proposed a method that uses Generalized Gaussian Mixture Model to maps the criminals based on the eyewitness specified features and concludes that the proposed model had given a unique and accurate result.

Malathi et al., (2011) analysed the crime data of police department using the data mining techniques such as clustering and classification and also identified the crime trends and suggested that this method can be used to reduce and prevent crime for the upcoming years.

4. PROPOSED METHOD

For this project, we used data mining techniques since we assumed that finding a relationship between criminal elements and frequent patterns can highly aid us in predicting potentially dangerous hotspots at a certain time in the future. That's why this proposed method is to focus on three main elements of the crime dataset, which are the type of crime, the crime period (morning, evening) and the crime location. First, for predicting the probability of a crime happening we found the frequent data item pattern based on the crime variable. Then we applied some classification algorithms to predict the potential crime type in a specific location within a particular time period.

In this part, we'll explain how we created the datasets and how we analysed the data using data mining and machine learning algorithms.

When the CPS (Crime prediction system) is started we are introduced to a basic login page. This is used to prevent any other person except the police to add, change or delete any kind of data prevent.

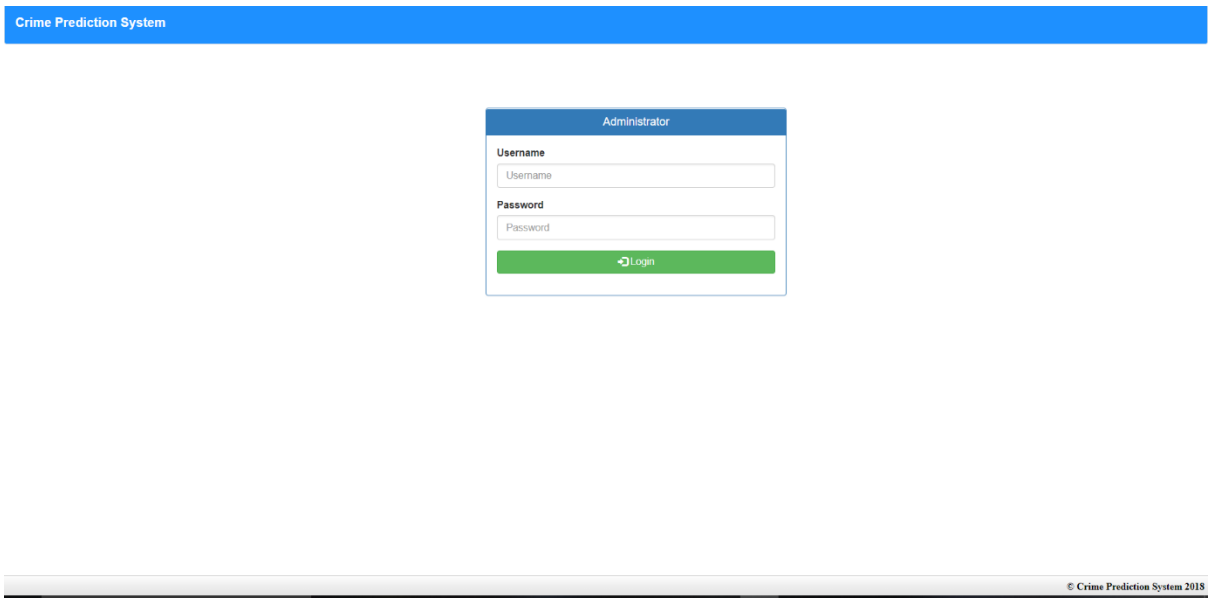

Fig. 1. Login page for CPS.

For this project we have used HTML, CSS as frontend and php as backend along with wamp server to maintain the database. After logging in, the dashboard contains the total number of FIRs registered along with the number of FIRs registered for each type of crime (fig 2).

Fig. 2. Dashboard.

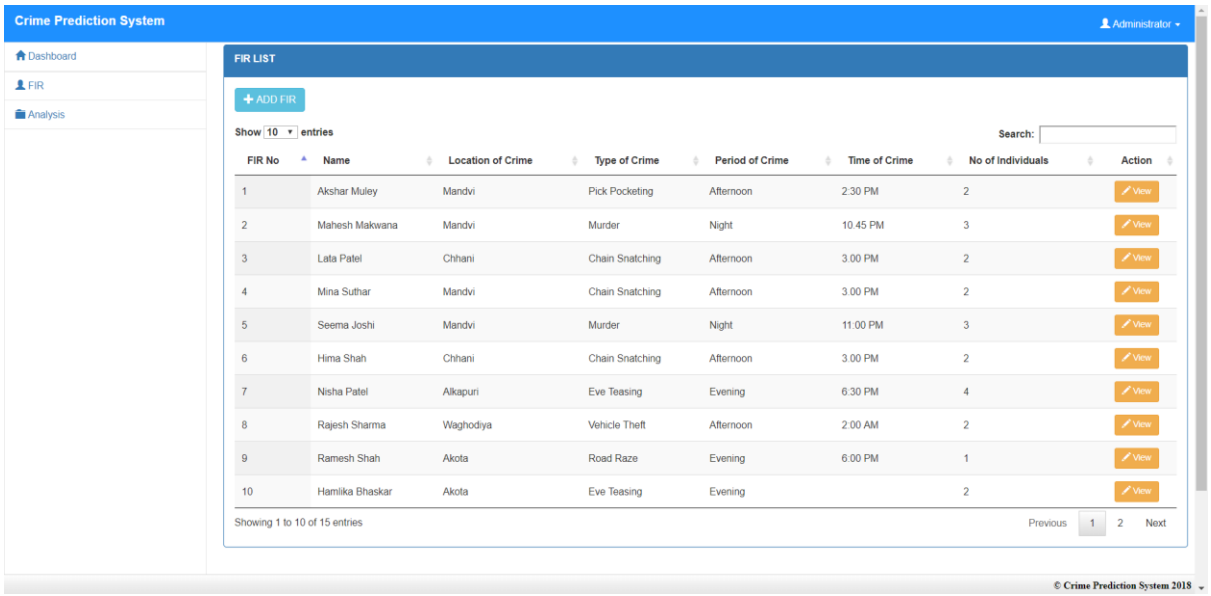

Fig. 3. Crime database.

4.1 Apriori Algorithm

Apriori is of the basic data mining algorithm for frequent patterns. It scans the database to find the item sets that satisfy a predefined minimum support. For this project we have taken support as equal to two. We have used it to find the frequent datasets which helps us in finding the probability or the chance of a particular crime happening in a certain area. We limited the probability or confidence to 25%, meaning when the probability of certain crime happening is more than 25% then only it will show in database. The Frequent Item sets (fig. 4) shows the most common pairs from the database, while the Association Rules shows the frequent patterns and correlation of crimes.

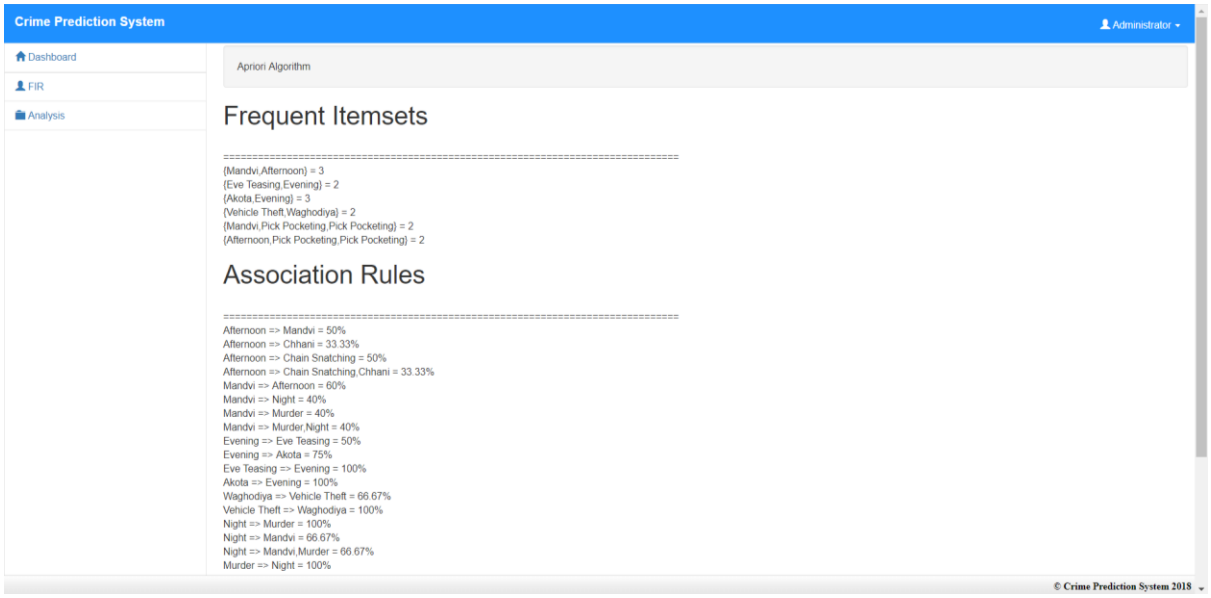

Fig. 4. Frequent Itemsets and Association rules.

4.2 Decision Tree

The decision tree is used to create a model to predict the class by using basic decision rules applied to the data. Since there is a possibility that the decision tree made might be complex so for this project we have used a predefined statement and have split the root node on the basis of the crime period. This algorithm is basically used as a searching algorithm to find names. For example, in fig 5 the pre-defined statement is: ‘Find all males who committed murder at night’ and in the result, it fetches the names of people from the database which have the attribute or keywords such as ‘murder’, ‘male’, ‘night’.

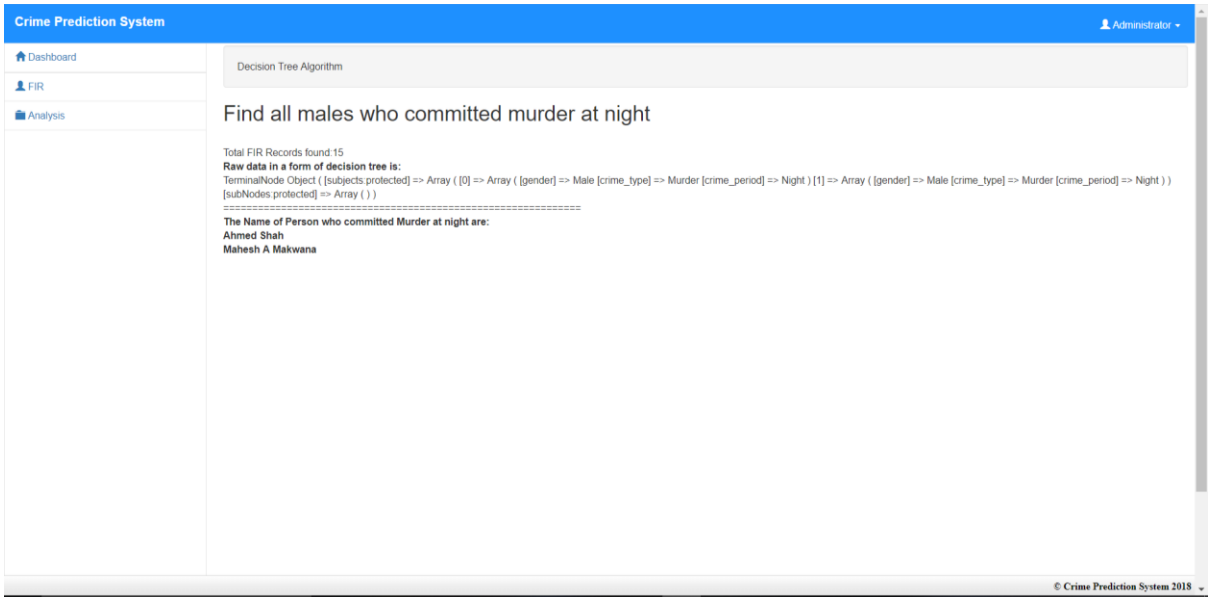

Fig. 5. Output for decision tree.

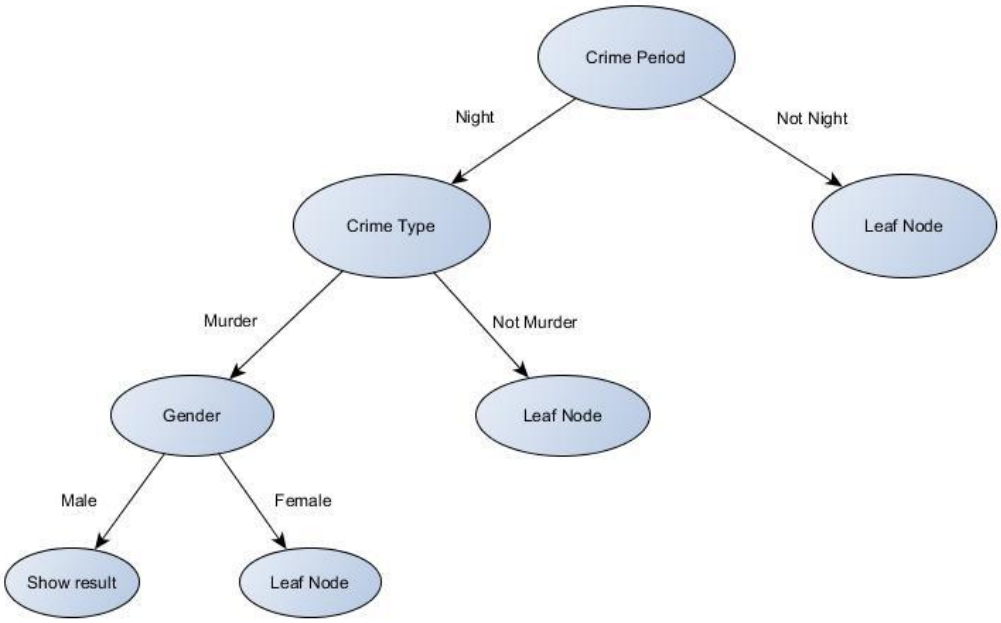

Fig. 6. Decision tree.

4.3 Naïve Bayes Classifier

Naïve Bayes is effective and widely used algorithm. It predicts class membership probabilities using Bayes theorem. The algorithm assumes all features to be conditionally independent which not always is the case so the prediction might not be hundred percent accurate. This is algorithm is also sort of used as a searching algorithm but instead of giving the no. of crimes or the names of the criminals it gives us the type of crime that can happen. As an example, when we search ‘alkapuri in evening’ we get the result ‘eve teasing’. The given fig. 7 and fig. 8 shows another example of Naïve Bayes Classifier.

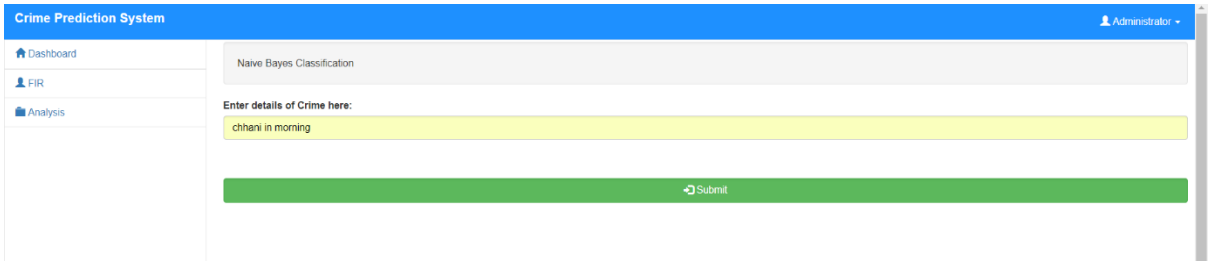

Fig. 7. Input for naïve Bayes classifier.

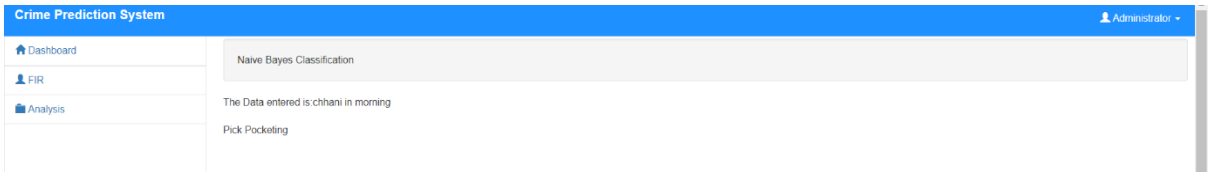

Fig. 8. Output for naïve Bayes classifier.

5. RESULTS

From this project, while comparing all the algorithms we learned that while the algorithms used are effective they are not fully reliable. First, we applied the apriori algorithm to find the frequent crime patterns meaning we analyse the database to get the possibilities of a crime happening. But, they are not 100% reliable since in one of the examples show that the chance of vehicle theft in waghodiya is hundred percent which is impossible to happen. So, we cannot completely rely on the result which gives 100% possibility. Next, we use Decision tree as a searching algorithm which helps us in finding names from a large-scale database, and Finally, we use the naïve Bayes classifier to find the type of crime that can happen at a particular area at a particular time. The naïve Bayes classifier only gives the type of crime and not the probability of that crime happening, so we can only use it to find the crime type. Another disadvantage of using naïve Bayes classifier in this project is that if a location has registered crime of the same number then the result will only show the most recent crime as output. By comparing the output of all the algorithms, we found that the most efficient and accurate algorithm used is decision tree followed by naïve Bayes and Apriori and that the efficiency of each algorithms depends on the type of application it is used in.

5.1 COMPARATIVE STUDY

We compared the main Apriori algorithm to another data mining algorithm FP growth. Although fp-growth is better than apriori in almost all the areas, we used apriori since it is very easy to understand and implement. While apriori scans the whole database to find the frequent itemsets, fp-growth creates a tree to do the same. The time complexity which is the time taken to execute is greater in apriori algorithm than Fp-growth.

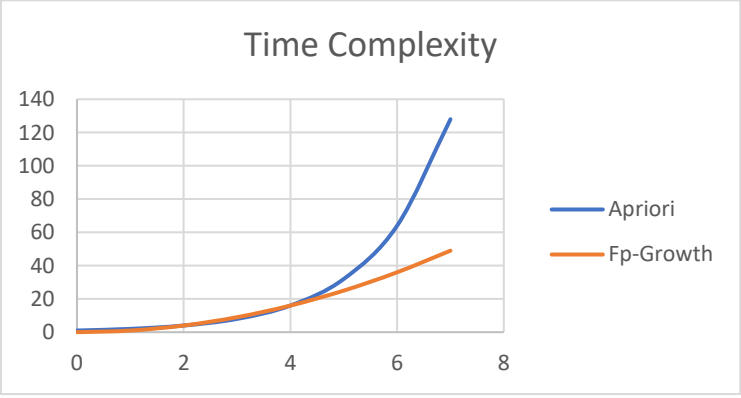

Fig. 9. Time Complexity comparison.

Time complexity of apriori algorithm is $O(2^d)$ where d is the number of unique elements.

Time complexity of fp growth is $O(n^2)$ where n is the number of unique items

We also studied and compared the performance between decision tree, naïve Bayes and kNN. Through an example from a reference paper [19], we learned that the F-measure is the harmonic mean of precision, meaning that calculating the value of F-measure (fig. 10) will tell us which is the best classifier. The average F-measure was largest among others meaning decision tree is best in terms of precision followed by Decision tree and kNN.

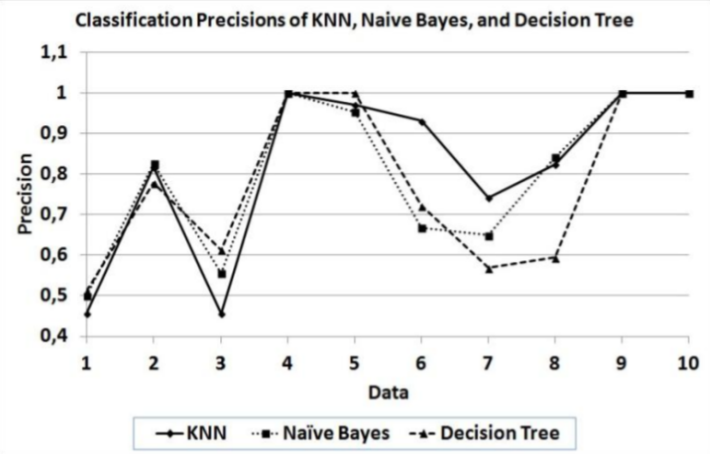

Fig. 10. Classification precision

6. CONCLUSION

In today's time where crime is increasing day by day, a major challenge faced by law enforcement is predicting crime to protect the citizens. As data science and technology progresses, tools of data mining and AI are now accessible to the law enforcement community. Computers can process a great many directions in seconds, sparing valuable time. Computers are likewise less prone to blunders than human investigators.

The project focuses on developing a crime prediction analysis tool for local society using data science and technology. This project enables law department to characterize and analyse the crime data to identify crime patterns and predict the possible future crimes. The system can say to be a small prototype but it can still be implemented to real-world database.

For the future extension of the project, we can apply more classification algorithms to increase the crime prediction accuracy and enhance the overall performance of the system.

7. REFERENCES

[1] Almanie, T., Mirza, R., & Lor, E. (2015). Crime prediction based on crime types and using spatial and temporal criminal hotspots. *arXiv preprint arXiv:1508.02050*.

[2] Bogomolov, A., Lepri, B., Staiano, J., Oliver, N., Pianesi, F., & Pentland, A. (2014, November). Once upon a crime: towards crime prediction from demographics and mobile data. In *Proceedings of the 16th international conference on multimodal interaction* (pp. 427-434). ACM.

[3] Iqbal, R., Murad, M. A. A., Mustapha, A., Panahy, P. H. S., & Khanahmadliravi, N. (2013). An experimental study of classification algorithms for crime prediction. *Indian Journal of Science and Technology*, *6*(3), 4219-4225.

[4] Nasridinov, A., Ihm, S. Y., & Park, Y. H. (2013). A decision tree-based classification model for crime prediction. In *Information Technology Convergence* (pp. 531-538). Springer, Dordrecht.

[5] Liao, R., Wang, X., Li, L., & Qin, Z. (2010, July). A novel serial crime prediction model based on Bayesian learning theory. In *Machine Learning and Cybernetics (ICMLC), 2010 International Conference on* (Vol. 4, pp. 1757-1762). IEEE.

[6] Yu, C. H., Ward, M. W., Morabito, M., & Ding, W. (2011, December). Crime forecasting using data mining techniques. In *Data Mining Workshops (ICDMW), 2011 IEEE 11th International Conference on* (pp. 779-786). IEEE.

[7] Laalmanac.com, 'City of Los Angeles Planning Areas Map', 2015. [Online]. Available: http://www.laalmanac.com/LA/lamap3.htm. [Accessed: 20- May- 2015].

[8] Varunon9, naïve-bayes-classifier, (2018), GithubGist, https://github.com/varunon9/naive-bayes-classifier.

[9] Dave Smith, PHP Apriori Algorithm Data Miner, (2015), https://www.phpclasses.org/browse/file/61953.html.

[10] VTwo-Group, Apriori-Algorithm (2014, November), https://github.com/VTwo-Group/Apriori-Algorithm.

[11] Julian Finkler, PHP Decision Tree Classifier: Compose decision trees and evaluate subjects, (2017, July), https://www.phpclasses.org/package/10385-PHP-Compose-decision-trees-and-evaluate-subjects.html.

[12] S. Ahmad, S. P. Simonovic, "An Intelligent Decision support system for Management of Floods", Water Resources Management, vol. 20, pp. 391-410, 2006

[13] 2016 11th International Conference on Knowledge, Information and Creativity Support Systems (KICSS)

[14] 2010 International Conference on Machine Learning and Cybernetics

[15] 2014 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2014)

[16] 2016 International Conference on Emerging Technological Trends (ICETT)

[17] 2017 6th ICT International Student Project Conference (ICT-ISPC)

[18] T. Efendigil, S. Onüt, C. Kahraman, "A decision support system for demand forecasting with artificial neural networks and neuro-fuzzy models: A comparative analysis", Expert Systems with Applications, vol. 36, pp. 6697-6707, 2009.

[19] Ashari, A., Paryudi, I., & Tjoa, A. M. (2013). Performance comparison between Naïve Bayes, decision tree and k-nearest neighbor in searching alternative design in an energy simulation tool. *International Journal of Advanced Computer Science and Applications (IJACSA)*, *4*(11).

[20] E. R. Groff, N. G. L. Vigne, N. Tilley, "Forecasting The Future of Predictive Crime Mapping" in Analysis for Crime Prevention, Lynne Rienner, vol. 13, pp. 29-57, 2002.

[21] R. B. Santos, "Effectiveness of Police in Reducing Crime and the Role of Crime Analysis" in Crime Analysis with Crime Mapping, Sage Publications, Inc., 2012.