Crime Prediction and Analysis System

Peiris M.D.P.T , Shaminda M.B.D , Senarathne H.K , Dr. Gayana Fernando

DOI: 10.29322/IJSRP.10.12.2020.p10862
http://dx.doi.org/10.29322/IJSRP.10.12.2020.p10862

Abstract *- Crimes are one of the major threats to society and also for civilization. The traditional crime solving techniques are unable to live up to the requirement of existing crime scenarios. Crime analysis is one of the most important activities of the majority of the intelligent and law enforcement organizations all over the world. The purpose of this study is to build a solution to support the criminal investigation process by providing a technological analysis in justifying criminal cases. This study gives guidelines to crime pattern detection, crime predictions and crime judgement suggestion processes. The feasibility of using machine learning techniques to build such platforms has been explored before. The vast geographical diversity and the complexity of crime patterns have made the analyzing and recording of crime data more difficult. Data mining is a powerful tool that can be used effectively for analyzing large databases and deriving important analytical results. The implemented system is a solution to be used in intelligent crime analysis, predictions systems which is designed to overcome the above-mentioned problems. The goal was to analyze the past crime data and past text case summaries and build applications to analyze which will support the investigation process in a wide range of angles to decision making.*

Keywords- Machine Learning, Lstm, Data Mining, PowerBI, Firebase, Real time Analytics, Visual Analytics, Visualization, Criminal Cases

I. INTRODUCTION

Introducing a 'CCMinner' which includes Crime predictions, Crime pattern analysis along with analysis of case documents in identifying patterns from past case summaries to support in making judgements. This can be approached in a more analytical point of view where a technological solution would help to find solutions for unsolved investigations, and support in decision making for the upcoming cases filed. Since Sri Lanka is a country that used outdated investigation methods and faces frequent problems, this system can be used to minimize certain circumstances. The final system is specially targeted to be used by a responsible person who carries out criminal investigations. This can be used as a package for criminal investigators which help their investigation process.

One of the most unfair situations is giving incorrect judgment for criminals and other legal matters. The average of grave crimes recorded from 2010 to 2018 in Sri Lanka, according to crime statistics reports released by Sri Lankan Police, is 45962. Percentage of crime increase in the past 3 years is 47.46%. Since, advanced technologies are used for crimes presently, investigation of crimes is not an easy task as in the 20th century. Sri Lankan crime statistics show that many of the crimes are pending on investigation. This gives us some bad impression of the investigation system of our country. The average of pending investigations from 2010 to 2018 is 31798. It is 70.31% of the average of recorded grave crimes with pending cases.

The remainder of this research paper will describe the requirements of the implemented crime prediction and analysis system overall description of them and their perspectives and functionalities. Furthermore, it will discuss performance requirements, design constraints, technological approach and Methodology of the component. Crime pattern detection and Crime judgment suggestions. This research paper also gives a detailed description of the implemented system and it will explain background information and overview of previous work based on literature survey, identification and significance of the problem, research questions, objectives, hardware and software requirements, detailed research methodology, sources for test data analysis, anticipated benefits, what the system will do the constraints under which it must operate in order to obtain optimal results and how the system will interact with the external factors.

II. RELATED WORK

High or increased crime-levels make communities decline, as crimes reduce house prices, neighborhood satisfaction, and the desire to move in a negative manner [1]. To reduce and prevent crimes it is important to identify the reasons behind crimes, predict crimes, and prescribe solutions. Due to large volumes of data and the number of algorithms needed to be applied to crime data, it is unrealistic to do manual analysis. Therefore, it is necessary to have a platform which is capable of applying any algorithm required to do a descriptive, predictive, and prescriptive analysis on a large volume of crime data. Through those three methodologies, law-enforcement authorities will be able to take suitable actions to prevent crimes. Moreover, by predicting the highly likely targets to be attacked, during a specific period of time and specific geographical location, police will be able to identify better ways to deploy the limited resources and also to find and fix the problems leading to crimes.

Regional crime analysis program and law enforcement data management systems are already developed for crime analysis [2] [3]. Most of these tools are developed to help the police to identify different crime patterns and even to predict criminal activities. They are complex software which

needs a lot of training before use. Designing a tool which is easy to use with minimal training would help law-enforcing bodies all around the world to reduce crimes.

Hence humans are incapable of comprehending with millions of data Hazwani et al. came up with a powerful computer-based system [4] which are able to predict crimes. This research identifies current implementations of crime prediction methods and the possibility to enhance it for future work. This is implemented using a support vector machine. fuzzy theory, artificial neural network and also multivariate time series. By using these models, the prototype got able to find out some helpful information but unfortunately there is no standard model that can solve the problems in different data sets. At the end of their finding Hazwani et al. report on results of an extensive comparison of crime prediction models. This got able to compare different crimes and show the capabilities of different crime prediction models but unfortunately Hazwani et al. didn't get able to develop this as an application level outcome.

In India Yadav et al. came up with a Crime Pattern Detection, Analysis & Prediction system with the capability of crime pattern detection, analysis and prediction [5]. The system is trained by feeding different crime records within 14 years in India. With the implementation of this regression model Yadav et al. got able to predict the crime rates in some selected areas. They used supervised semi-supervised and unsupervised learning techniques on crime records to knowledge discovery and increase predictive accuracy. Yadav et al. used some techniques such as Association mining, Clustering, Classification techniques and Correlation & regression. Authorities got able to depict results graphically and also the developed model will reduce crime by helping crime detection fields if implemented. Major disadvantage of this work is the developers didn't use any type of software metrics to check the Accuracy.

Sivaranjani et al. came up with a crime prediction and forecasting using clustering Approaches [6]. They used various clustering approaches of data mining to analyze the crime data of *Tamilnadu*. Three major clustering techniques are used in this K-Mean clustering, Agglomerative clustering and Density–Based spatial clustering. Sivaranjani et al. compared the clustering models and selected the best technique based on the result. K-Mean clustering is visualized using Google maps and the K-nearest neighbor algorithm is used for crime prediction. Each performance of clustering algorithms are evaluated using metrics such as Precision, Recall, F-mean and the results are compared. With the use of this system Sivaranjani et al. got able to visualize crime hotspots using google maps. But Sivaranjani et al. failed to implement a crime pattern recognition function.

Sivanagaleela et al. implemented a Crime Analysis and Predicting system by using Fuzzy C- Mean Algorithm for crime analysis and prediction [7]. This system mainly focused on which area will the crime occur rather than the criminal identification. *Naïve Bayes* classification was used to classify the data and Fuzzy C-Means algorithm is used to cluster the crime data. This system predicted outputs by region and crime types which will occur frequently. States that using Fuzzy c-means algorithm crime-prone areas can identify with less time. The results were able to map by using a chart. One of the major cons the system had is it showed some low accuracy outcomes.

Gandotra et al. came up with an innovative idea which made them implement a text mining method to support criminal case judgment Referential prison term generation for the judge's sentence [8]. The major idea of this paper is used for police investigation document of criminal case send from district police office, parse this document into keyword table, then match this data with the trained database in the form of keyword table format of court's judgment. This paper tries to use a text mining approach to solve problems faced by judges. The system is programmed to use the police's investigation document of criminal cases sent from the district police office, parse this document into a keywords table, then match with a trained database. With the use of Cosine similarity algorithm of text mining technology to calculate the coefficient of similarity, based on the highest coefficient. The model reached nearly 87.5% of accuracy.

Keyvanpour et al. came up with a general crime matching framework which is able to detect and investigate the crime by means of data mining [9]. This research mainly focuses on using crime matching with behavioral burglary crime variables by identifying crime characteristics. Crime matching process is used for assigning the previously unsolved crimes to the arrested offenders. The system itself compares the clustering models and selects the best technique based on the result. SOM neural networks are used for clustering and the MLP neural network is used for classification. In this research, some of the most significant capabilities of data mining techniques were leveraged through a multi-purpose framework for intelligent crime investigation. But this system still got unable to recognize crime patterns

Investigating Crimes using Text Mining and Network Analysis is introduced by Elyezjy et al. [10]. This paper has been presented to the Crime Detection System, which has developed to discover a new relationship between offenders and communities using Arabic investigation documents and visualize the results to assist crime data analysis. Also a system is developed to extract useful information in the Arabic crime domain from unstructured investigation data in order to mine. The major approach of their research is basically based on building a predefined dictionary that contains some important keywords that can be used to classify the crime domain. Using a *Dracula* Graph Library for data visualization the Elyezjy et al. implemented a JAPE rule-based algorithm using the GATE tool to improve nominating the correct names from unstructured text.

A crime pattern detection is invented for investigations by Jain et al. [11]. The major objective of this research is to predict the probable suspects for the unsolved cases or test cases from the past criminal record, criminal record, and crime over past years' records. Subdividing the data into

different types of crime allowing the user to get information about those crimes easily. The system core is implemented using K-means clustering and HYB algorithm. This system has advantages such as providing a single all-embracing database in which all the data has been extracted, linked, cleaned and augmented and the system works like an interactive analytical instrument in which every part of the results is clickable to 'zoom in'. In this way, the user can simply embark on an analytical journey without the need to first design the process. As pros this system had some cons such as it is typically hard to extract data from police source systems because of old and diverse database systems with data models based on transactions instead of analysis. Also the existing systems usually generate static reports that do not allow interaction. They cannot be used to find the explanations behind the numbers they present.

III. METHODOLOGY

The crime prediction and analysis system is implemented by integrating three major components. They are Crime prediction, Crime case suggestion and Real time crime case pattern analysis. Each component plays a major role for the final accurate outcome. The implementation of this product is done by using agile software development methodology. For the prototyping model the developers used the evolutionary prototype model.

Crime Prediction

Crime prediction function is implemented to predict crimes division wise. *Jupyter* notebook and Visual studio code are the main developing Ids used for this function. The developers gathered data from police stations from the relevant division. These data are saved on *csv* format for further data processing. Before saving these data all the data is converted to float variable type. Then the developer normalized the data manually to stop more duplicate iterations. The normalized data set is divided into two data sets as training data set and test data set. These sets of data are converted an array of values into a data matrix. This matrix is reshaped into $X = t$ and $Y = t+1$. After this transformation the reshape input looked as *3D (num_sampales, num_timesteps, num_features).*

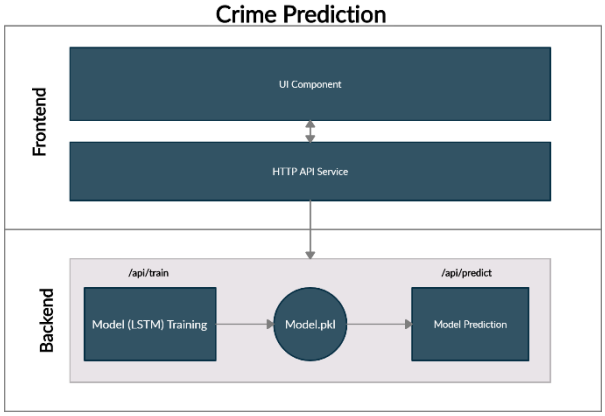

Figure 1. Architecture Diagram of crime prediction

After the data pre-processing these data are used to construct the LSTM model. The LSTM is defined with hundred neurons in the first hidden layer and with one neuron in the output layer to predict the pending and crime cases. The input shape is formed as a once time step with thirty features. This process is having a 20% of dropout rate. The MSE loss function and efficient Adam version of stochastic gradient descent is used to configure the model architecture. The implemented model was able to fit 20 training epochs with a batch size of 70. The model output is passed to the frontend by using HTTP API services.

Crime case suggestion

This component is implemented on focusing court judges as the target users. Previous case document details are extracted from the *supremecourt.lk* and the *courtofappeal.lk* websites.

In the initial stage of this function keywords, terms and cluster similar case summaries are extracted from the dataset using *Python Natural Language Toolkit Library.* The clustered term frequency table is stored for future references.

Once a new case summary is fed to the system the similarity algorithm is used to match the new case summary with the existing data storage in the form of key word table format to find the similar cases from the stored past case documents. The cosine similarity algorithm of text mining is used to calculate coefficient of similarity and based on the highest coefficients value the researcher finds out the closest past criminal cases for a particular given case.

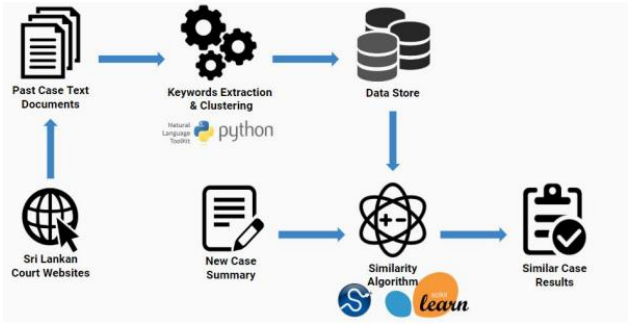

Figure 2. Architecture Diagram of crime case suggestion

Text analytics plays a huge role in this scenario. Text analysis process is based on linguistic, statistical and machine learning techniques. Information are retrieved from the unstructured data and the process of structuring the input text to derive patterns and trends to evaluate and interpreter the output data.

The cases are clustered using the keywords included in it. Keywords are extracted from the data set stored in clustered term frequency table. When a unique case is pushed to the system cosine similarity is used to calculate a statistical score for the comparison between the case summaries which are having a high similarity for the given unique case.

Real time crime case pattern analysis

This function does the major visualization parts in the user interfaces. This functions are implemented to gather real

time data from island wide police officers. These data are displayed real time in the power BI tool.

Backend is fully implemented on firebase and the front end implementation is done in visual studio code Id.

When the data is entered in the front end text field these data is posted to the backend using the HTTP API services. Then these data are saved in the *firebase* data collection. The data written in the *firebase* is pulled back using get methods to read and visualize data in the power BI tool.

In the front end user interface the power BI tool is called as an *I-frame.* The developers have implemented the firebase call through the power BI tool.

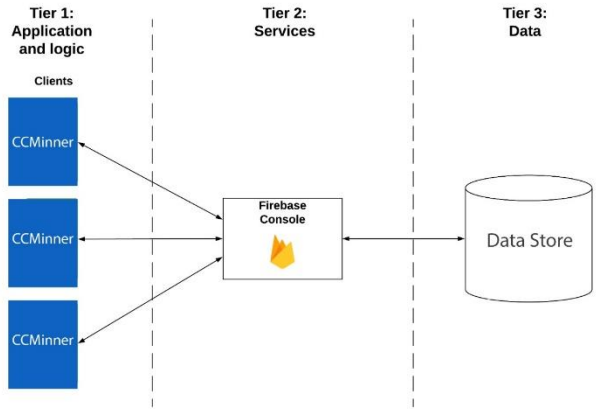

Figure 3. 3-tier-architecture diagram

3-tier-architecture diagram figure gives the brief idea of the 3-tier-architecture of the real time crime case pattern analysis functionality.

IV. EXPERIMENTS AND RESULTS

As mentioned before the implemented system contains three major components. Crime prediction, Crime case suggestion and Real time crime case pattern analysis. Below session widely explains the results of the system components.

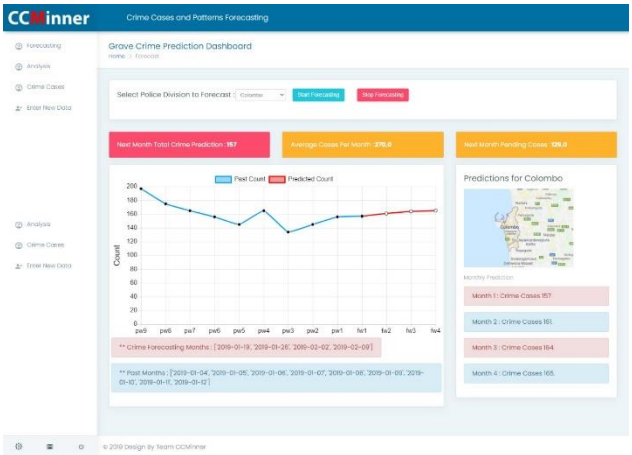

Figure 4. BI tool crime prediction Dashboard

BI tool crime prediction dashboard interface (*Figure 4)* shows the crime case prediction interface. In the drop down there are police divisions for some specific regions. There are data sets related to each division. Once the user clicks on the start forecasting button a graph will be generated. The blue line shows the past data count and the red line shows the prediction count. this also shows a summary of how many crimes may happen in the next month, Average cases per month, Pending cases etc.

The root mean square error rate is in a low amount such as 0.14545373.

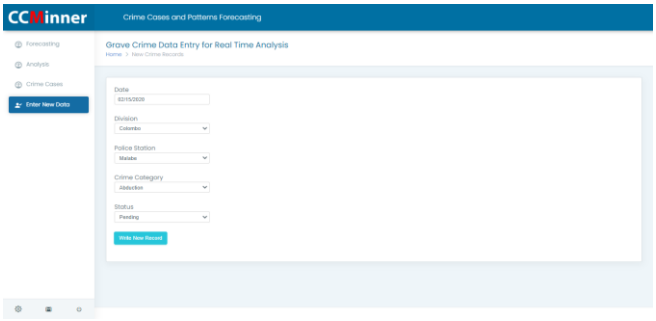

Figure 5. Crime case data input UI

In the crime case and pattern analysis function users can input date, division, police station, crime category and pending status (Figure 5). when the user input these data all of these data will be displayed in the below dashboard (*Figure 6)*. It will graphically show the location, and how it happens. all of these things are shown in the power BI. BI tool represent 100% of accuracy real time data.

Figure 6. Real time data display Dashboard

In the crime case suggestion functionality there is also a trained machine learning model, which is training by using many crime cases. Once a user input keywords in the interface (Figure 7) these words are extracted from the text box for the text mining procedure. then these mined texts are analyzed and all the related crime scenes which include these keywords in its documents will be displayed in the interface. All of these are displayed after being sorted according to the related score point. The score rate depends on the type of query length.

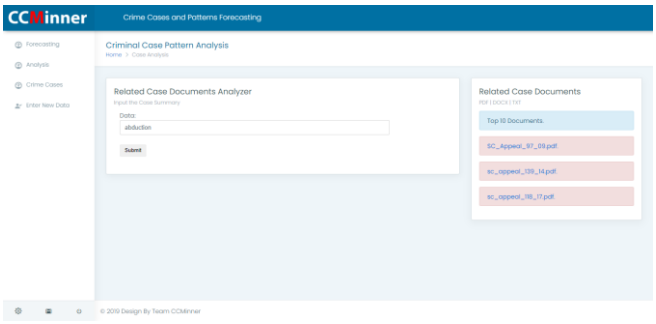

Figure 7. Crime case suggestion UI

V. CONCLUSION

The major objective of this research is to identify the patterns from past case summaries to give support and guidance in decision making for the users. Altogether the major three functional components have successfully brought the research objective into the reality.

The crime prediction and analysis system has passed through several test case criteria to make sure that the probability of accuracy is above 0.85 rate. All the generated outputs are compared and tested using real data to certify the accuracy of the relevant outputs. In the future developers are planning to add more data for its data stream to make the crime prediction and analysis system more accuracy.

There is some limitation with the crime prediction and analysis system. There are some limitations with the user session handling. The crime prediction and analysis system cannot handle an extreme amount of user loggings in a certain time so the system itself cancel some session to make more availability for some users.

ACKNOWLEDGEMENT

The research team is quite thankful to everyone who helped with ideas and providing relevant accurate data sets for this research and very special thanks goes to Dr Gayana Fernando for guiding in every aspects of the research from the very beginning until the end.

REFERENCES

[1] R. Krishnamurthy and J. S. Kumar, “Survey of data mining techniques on crime data analysis,” International Journal of Data Mining Techniques and Applications, vol. 1, no. 2, pp. 117–120, 2012.

[2] D. E. Brown, “The regional crime analysis program (recap): a framework for mining data to catch criminals,” IEEE Intl. Conf. on Systems, Man, and Cybernetics, vol. 3, pp. 2848–2853, 1998.

[3] H. Chen, D. Zeng, H. Atabakhsh, W. Wyzga, and J. Schroeder, “Coplink: managing law enforcement data and knowledge,” Communications of the ACM, vol. 46, no. 1, pp. 28–34, 2003.

[4] N.Hazwani, M.Shamsuddin, N.A.Ali, R.Alwee, “An Overview on Crime Prediction Methods”, Malaysia.

[5] S.Yadav, M.Timbadia, A.Yadav, R.Vishwakarma and N.Yadav, “Crime Pattern Detection, Analysis & Prediction” International Conference on Electronics, Communication and Aerospace Technology ICECA, India, 2017

[6] S.Sivaranjani, Dr.S.Sivakumari, Aasha.M, “Crime Prediction and Forecasting in Tamilnadu using Clustering Approaches”, International Conference on Emerging Technological Trends [ICETT], 2016
29

[7] B. Sivanagaleela, S. Rajesh, “Crime Analysis and Prediction Using Fuzzy C-Means Algorithm”, Proceedings of the Third International Conference on Trends in Electronics and Informatics (ICOEI) ,2019

[8] Riya, Namita Gandotra, “Text Mining on Criminal Documents”, Shoolini University, Solan, Himachal Pradesh 173229, India, International Journal of Advances in Electronics and Computer Science, ISSN: 2393-2835 Volume-3, Issue-9, Sep.-2016

[9] Keyvanpour, Mohammad Reza; Javideh, Mostafa; Ebrahimi, Mohammad Reza, "Detecting and investigating crime by means of data mining: a general crime matching framework", Procedia Computer Science 3 (2011), 872-880

[10] Nael T. Elyezjy, Alaa M. Elhaless “Investigating Crimes using Text Mining and Network Analysis”, International Journal of Computer Applications (0975 – 8887) Volume 126 – No.8, September 2015

[11] Nikita Jain, Anushree Pai, Yatharth Sharma, “Criminal data investigation and crime pattern detection”, Jaypee University of Information Technology, Himachal Pradesh, India.