

## LOCATION BASED CRIME NEWS ANALYSIS ANDROID APPLICATION

**Jagdhane Shubham<sup>\*1</sup>, Khatane Vishal<sup>\*2</sup>, Darade Nilesh<sup>\*3</sup>, Kadam Akshay<sup>\*4</sup>**

<sup>\*1,2,3,4</sup>Dept of Computer Science and Engineering, Sanjivani Rural Education Society,  
College of Engineering, Kopergaon, Maharashtra, India

### ABSTRACT

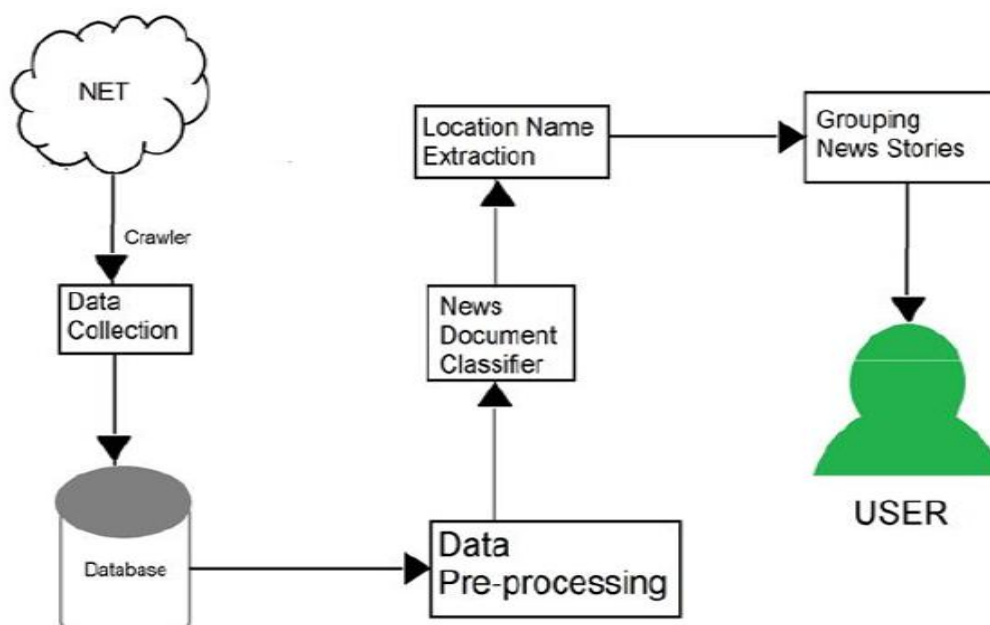
In intelligent and law enforcement organizations crime analysis is very important activity all over the world. Collect domestic and foreign crime related data to prevent future attacks. It utilizes a limited number of law enforcement resources. The major challenge by most of the law enforcement and intelligence organizations in Crime analysis is accuracy and efficiency. The organizations should accurately analyze the growing volumes of crime related data. Crime data mining has a auspicious future for increasing the usefulness and coherence of criminal and intelligence analysis.

Keywords: Crime News, Organizations, SVM, Data

### I. INTRODUCTION

Multiple sources should be taken into consideration when analyzing crime. Newspaper gives massive amount of data related to any specific topic and hence they are good source of crime analysis but in reality they are not structured enough to make a definite decision. To analyze crime news from online newspaper using different data mining techniques is another method of crime analysis. These data mining techniques help to take out useful information about crime from unstructured crime news. The analysis will provide the definite information about a news whether it is a crime related news or not, crime location, similarity and grouping crime news as stories based on a crime news.

### II. METHODOLOGY



**Figure1:** System Architecture.

The system has a front-end web-based interaction which provides interaction with the user. There is a back end where all the processing is done. All the algorithms to collect data, and crawler collect plain text saving into database and data pre-processing algorithm create in structure format. Classifier divides data into two class's crime news and other news.

Entity Recognition can recognition or generate token documents converted into sentence lassify giving location based news and sorting order to perform location base

### Methodology/possible Algorithms

**1. Data Collection:-** As we are working with online newspaper, we have used online English-language news paper of Bangladesh for both training and testing purpose. We have collected news from The Daily Star” and New Age” newspaper’s online version.

**I. Crawler:-** We have implemented a crawler using simple tools and libraries like selenium and newspaper to collect plain text from online newspaper. The crawler only takes the title , the body contents and the date of a news. The crawler run within a specific domain to collect data. At a time it will only collect data from a single domain address.

**II. Saving Data:-** We have used MongoDB as a database back end for saving text data collected by the crawler. As MongoDB is a NoSQL database that provides data model convenient to design and promises high performance and scalability. So it is the best-suited database program for our system. We have used driver for MongoDB to handle database operations.

**2. Data Pre-processing:-** Data pre-processing is performed to make a clear structure of text data. The common techniques like word tokenization, removing stop words and special characters, case conversion, lemmatization performed on each document before saving in database for processing purpose. All of these operation is performed using NLTK library.

**3. News Document Classifier:-** To classify news collected from online newspaper we built a two-class classifier. The classifier classifies news article as crime-related news and non-crime related news. We have used support vector machine (SVM) classifier. We have tried both linear and radial kernel to compare the performance. SVM can give better performance when hyper parameter is tuned properly. We have used grid search with 5-fold cross-validation to find the best value for gamma and C. We have used Scikit-Learn library on this purpose. 20 percent of total data set is used for validation purpose.

**4. Location Name Extraction :-** Crime location extraction is performed on documents using named entity recognition (NER) and sentence classification. We have extracted location using NER methods and collected sentences that contain location names. Then a sentence classifier is used to classify sentence with crime locations.

**5. Grouping News Stories:-** Searching similarity based on a news is performed at first then grouping similar news as a story based on a crime news is performed.

### SVM

For predicting and classifying data various machine learning algorithms are used according to the dataset. SVM is a linear model used for classification and regression problems. The idea of SVM is simple: The algorithm creates a line or a hyper plane which separates the data into classes. In this blog post I plan on offering a high-level overview of SVMs. I will talk about the theory behind SVMs, its application for non-linearly separable datasets and a quick example of implementation of SVMs in Python as well. In the upcoming articles I will explore the math’s behind the algorithm and diagram under the hood SVM

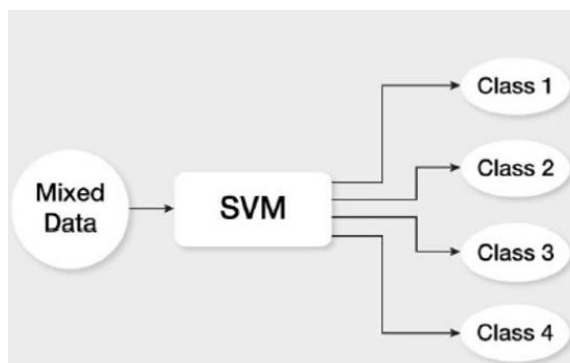


Figure 2: SVM

### Validation

SVM doesn’t give us the probability, it directly gives us the resultant classes. The methods of validation like specificity, sensitivity, ROC, AUC and cross validation are used for validation.

### III. RESULTS AND DISCUSSION

#### A. Location based news:

This application will provide local news according to the user's location. Every area will be ranked from 1 to 5. Where rank above 3 will be considered as RED zone. And the ratings will be decided according to the crime incident happened in that area. So the new person in that locality will come to know about that area. This app will make user aware about what happening near them.

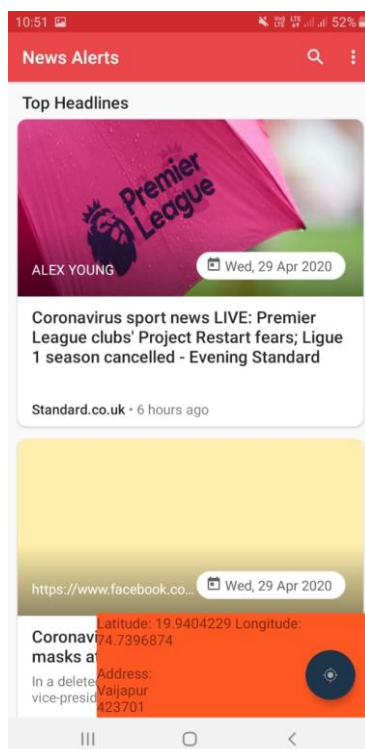


Figure 3: Location based crime news analysis

#### B. Category wise news:

This application will provide category wise news so that user can watch only the relative or the news in which he/she is excited about. The categories contains Crime and Non-crime news like sport, entertainment, politics, business, technology, educational etc.



#### **IV. CONCLUSION**

The system provides different features to analyze crime through the online newspaper. We can port these subsystems to any crime analysis system to perform a basic analysis. Crime news detection, crime location extraction and story detection may help to understand crime pattern in a single location. Ranking crime location shows the crime prone areas. Grouping stories show similar crime stories and linked news for the crime story to be concerned. It helps to understand the density of a single type of crime.

#### **V. REFERENCES**

- [1] Mehedee Hassan and Mohammad Zahidur Rahman, "Crime News Analysis: Location and Story Detection," in 2017 20th International Conference of Computer and Information Technology (ICCIT), 22-24 December, 2017
- [2] S. Sathyadevan, D. M. S, and S. G. S., "Crime analysis and prediction using data mining," in 2014 First International Conference on Networks Soft Computing (ICNSC2014), pp. 406412, Aug 2014.
- [3] B. Chandra, M. Gupta, and M. P. Gupta, "Adaptive query interface for mining crime data," in Proceedings of the 5th International Conference on Databases in Networked Information Systems, DNIS'07, (Berlin, Heidelberg), pp. 285296, SpringerVerlag, 2007[9]
- [4] U. Thongsatapornwatana, "A survey of data mining techniques for analyzing crime patterns," in 2016 Second Asian Conference on Defence Technology (ACDT), pp. 123128, Jan 2016.
- [5] Hsinchun Chen, Wingyan Chung, Yi Qin, Michael Chau, Jennifer Jie Xu, Gang Wang, Rong Zheng, Homa Atabakhsh, "Crime Data Mining: An Overview and Case Studies," AI Lab, University of Arizona, proceedings National Conference on Digital Government Research, 2003.
- [6] Hsinchun Chen, Wingyan Chung, Yi Qin, Michael Chau, Jennifer Jie Xu, Gang Wang, Rong Zheng, Homa Atabakhsh, "Crime Data Mining: A General Framework and Some Examples," IEEE Computer Society April 2004.
- [7] Y. Cai and T. Xu, "Design, analysis, and implementation of a large-scale real-time location-based information sharing system," In MobiSys, 2008.
- [8] C. U. Leirv.ag, "User-centred design of a location-sensitive mobile application," Master's thesis, University of Bergen, Dept. of Information Science and Media Studies, June 2009.
- [9] K. V. fiye, "Educating media students in new journalistic practices - an empirical study of location journalism," in Interweaving technologies, Copenhagen, Denmark, oct 2010.
- [10] P. Bellavista, A. Corradi, R. Montanari, and C. Stefanelli, "A mobile computing middleware for location- and context-aware internet data services," ACM Trans. Internet Techn., vol. 6, no. 4, pp. 356380, 2006. 8