# Modelling Machine Learning For Analyzing Crime News

submitted in partial fulfillment of the requirement
for the award of the Degree of

**Bachelor of Engineering**
in
**Computer Engineering**

by

**Surabhi Ghankutkar**
**Neelabh Sarkar**
**Pooja Gajbhiye**
**Sanyukta Yadav**

under the guidance of

**Dr. Dhananjay R. Kalbande**
**Ms. Nida Bakereywala**


**Department of Computer Engineering**
Bharatiya Vidya Bhavan's
Sardar Patel Institute of Technology
(Autonomous Institute Affiliated to University of Mumbai)
Munshi Nagar, Andheri-West, Mumbai-400058
University of Mumbai

November 2019

# Certificate

This is to certify that the Project entitled "Modelling Machine Learning For Analyzing Crime news" has been completed to our satisfaction by Ms. Surabhi Ghankutkar, Mr. Neelabh Sarkar, Ms. Pooja Gajbhiye and Ms.Sanyukta Yadav under the guidance of Dr. Dhananjay R. Kalbande and Ms. Nida Bakereywala for the award of Degree of Bachelor of Engineering in Computer Engineering from University of Mumbai.

**Certified by**

**Dr. Dhananjay R. Kalbande**
**Ms. Nida Bakereywala**
**Project Guide**

**Dr. Dhananjay R. Kalbande**
**Head of Department**

**Principal**


**Department of Computer Engineering**
Bharatiya Vidya Bhavan's
Sardar Patel Institute of Technology
(Autonomous Institute Affiliated to University of Mumbai)
Munshi Nagar, Andheri(W), Mumbai-400058
University of Mumbai
November 2019

# Project Approval Certificate

This is to certify that the Project entitled "Modelling Machine Learning For Analyzing Crime news" by Ms. Surabhi Ghankutkar, Mr. Neelabh Sarkar, Ms. Pooja Gajbhiye and Ms.Sanyukta Yadav is found to be satisfactory and is approved for the award of Degree of Bachelor of Engineering in Computer Engineering from University of Mumbai.

**External Examiner** **Internal Examiner**

**(signature)** **(signature)**

**Name:** **Name:**

**Date:** **Date:**

**Seal of the Institute**

## Statement by the Candidates

We wish to state that the work embodied in this thesis titled "Modelling Machine Learning For Analyzing Crime news" forms our own contribution to the work carried out under the guidance of Dr. Dhananjay R. Kalbande and Ms. Nida Bakereywala at the Sardar Patel Institute of Technology. We declare that this written submission represents our ideas in our own words and where others' ideas or words have been included, we have adequately cited and referenced the original sources. We also declare that we have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in our submission.

**Name and Signature:**
**1.Surabhi Ghankkutkar**
**2.Neelabh Sarkkar**
**3.Pooja Gajbhiye**
**3.Sanyukta Yadav**

# Acknowledgments

It is really a pleasure to acknowledge the help and support that has gone to in making this project. I express my sincere gratitude to my guide Dr. Dhananjay Kalbande and co-guide Prof. Nida Bakereywala for their invaluable guidance. I also thank them encouraging us to work in Machine Learning and Natural Language Processing on the topic like Crime analysis. Without his encouragement this work would not be a reality. With the freedom they provided, we really enjoyed working under them.

I thank my examiners, Prof. Surekha Dholay for the word of advice. I thank HOD and staff of Computer Engineering Department for giving me all the facilities required to carry out this research work.

I would like to thank all my family members and well wishers for their constant encouragement for all these years, without which I could not have completed this work.

# Contents

# List of Figures

# List of Tables

# List of Abbreviations

| | |
|---|---|
| SVM | Support Vector Machine |
| ARIMA | Auto Regression Integrated Moving Average |
| GATE | General Architecture for Text Engineering |
| ANN | Artificial Neural Network |
| NLP | Natural Language Processing |

**Abstract**

Increasing crime rate and recurring crimes in India constitutes towards analyzing crime. Crime analysis is a way of observing the patterns of crime happened recently. Crime analysis is very important for the law and detectives to find the criminals. It can also be useful for predicting future crimes. One can be aware of the criminal activities and can take safety measures to protect his or her family and friends. Traditional methods of analyzing crime news includes either reading the complete news and then analyzing which is very time consuming and vulnerable to human errors or work on the old data sets of crime which will not be assistive enough to provide any analysis of the real time crime news. The proposed system is a model which analyzes real time crime data in the form of news article and gives a report of the news related to crime. The contents of the websites of newspapers will be the input to the system. The website will be crawled using a crawling program written in python language and the data will be stored in the database. Using classifier the data will be classified into crime related data and non-crime related data. Final analysis will be displayed to the user in a tabular format.

# Chapter 1

# Introduction

Crime is an act which is punishable and which causes harm to other innocent people and cities. Crimes are of different type – robbery, murder, rape, assault, battery, false imprisonment, kidnapping, homicide. Analysis is the the method of deeply inspecting the constituent elements or structures of the object or subjects under consideration. Analysis is performed with the intention of gaining a thorough understanding of a field. though methods of analysis may differ, the goal is always to derive a fruitful result or conclusion which can be used further. Crime analysis is a systematic way of identifying and analyzing patterns and trends in crime.

The technologies encountered currently are models developed for crime analysis. Some of the key technologies that have been explored previously include Web crawling, SVM based classifier GATE(General Architecture for Text Engineering), ANNIE, LingPipe OpenNLP JavaScript and AJAX, Apriori Association Mining, k-Means Clustering,Naive Bayes method, Classification Correlation and Regression Named Entity, Recognition Natural Language Toolkit. A Box-Jenkins Model with Exponential Smoothing and 3. ARIMA (Auto Regression Integrated Moving Average) Model. Fuzzy Theory, Artificial Neural Network (ANN), Multivariate Time Series.

The crime rate in India is rising rapidly. The rise of forms of technologies have provided more avenues for more complex crimes. Dealing with this increase in crimes is very difficult without having a proper basis for classifying crimes cases that have been committed previously .The crimes that are recorded, need to be visited again to get all the details of a crime. Using data old databases may lead to inaccuracies as the data is outdated and is not reliable in the current scenario.

With current progress in recording and storage of data, crime analysis is possible on a greater scale than before. Crime analysis is useful for agencies, detectives to solve their cases and identify the suspects. It helps to solve the case in a more effective manner. People must have some idea about crimes happening in our country. People should be made familiar with the crime cases happening. So they will be alert as a precaution and future crimes can be avoided. Crime analysis must focus on the various factors relevant to a particular crime. for example: motive, location, type, time of occurrence, frequency, etc, this is necessary in order to understand the nature of the crime, and come up with specific countermeasures to avoid similar crimes in the future. The step that follows after crime analysis is to accurately classify the crime so as to get a numerical estimate about the crimes occurring. If the crimes are successfully classified, then efficient ways to respond can be designed.

There is a need of a system that can analyse real time news. Following paper describes the implemented system that classifies news from online newspaper sites into crime and non-crime new using Machine learning techniques and Natural Language Processing. The input to the system is the news crawled from online news sites. The data that is retrieved after crawling the news is given as an input to three classifier models i.e. Support Vector Machine, Random Forest and Multinomial Naive Bayes that classify the news into crime and non-crime. The three classifiers are trained using a dataset which contains around 6700 crime and non-crime news approximately. The new records to train the models are taken from the News Category dataset which is a publicly available dataset. A detailed description of the following steps has been given in the paper.

This paper has various sections. The first section is the introduction which briefly covers the field of crime analysis. The technologies that have been previously implemented in this domain as well as the problems that are seen. The next section covers the related work that has been done in the the fields of crime analysis and prediction using various concepts and models of machine learning. The fourth section covers our proposed model for crime analysis. The implementation of web crawlers, and other classification techniques have been presented.

## 1.1 Motivation

- The crime rate in India is growing at a rapid pace.
- The types of crimes committed in India is seen to be highly complex and have many relevant details.
- Most crimes require in-depth analysis which is very difficult to perform on short notice.
- Law enforcement agencies often have to deal with multiple crimes simultaneously and need information to be provided in a quick and efficient manner.
- If a system for analyzing crime through online news papers is deployed effectively, it could reduce further incidences of crimes and help the citizens.

## 1.2 Objectives

- To give accurate analysis of real time data from online newspaper articles.
- Accessible and efficient to any user interested in crime analysis.

## 1.3 Problem Statement

To design a machine learning model for analysis of newspaper for classifying the news into crime and non-crime.

## 1.4 Contributions

- The implemented system allows users to perform analysis on the real-time crime data in India and then stores it in a database.
- Since the data is retrieved from recently published news paper articles, it has more relevance to the current environment.
- This system also structures the raw data. Thus it will perform better than conventional solutions which use databases that are often outdated.

## 1.5 Layout of the Report

A brief chapter by chapter overview is presented here.
Chapter 2: A literature review of various crime classification and analysis model is presented.
Chapter 3: System Architecture , Activity diagram and Use case diagram.
Chapter 4: In this chapter, the whole implementation process for the project has been presented.
Chapter 5: This chapter includes the results and discussion.
Chapter 6: This chapter includes the conclusion.
Chapter 7: The future scope of the project has been presented in this chapter.

# Chapter 2

# Literature Survey

An intelligent crime analysis web based system which takes the real time data from the electronic newspapers and generates a graphical report for the user. It makes use of web crawler to extract information from the world wide web and stores the data in the database. The web-crawler is solely responsible for parsing the required required websites. The crawler used in this paper is crawler4j. The documents are then preprocessed so that further classification can take place Features are assigned and stop words are removed. This is done by using Weka libraries One issue that was encountered was the inability of Crawler4j to parse websites Support Vector Machine is the classifier that is used for document classification. The data is then classified into two categories, crime data and non crime data. Important entities are extracted from the database followed by removal of duplicate data.It extracts important parameters such as crime date, location of the crime, police personnel and victims involved. An analyzer module performs hot spot detection, crime comparison and crime pattern visualization. The hot spot detection is given as a map with colored regions indicating crime regions. Crime comparison is given as a pie chart where the percentage of the occurrence of the crime is shown. Crime pattern visualization is given as a time series plot showcasing frequency in crime [1].

A model for predicting crime rate in various states to help local police stations in crime suppression. The paper uses four classification algorithms. Using these four algorithms crime prediction is performed, making it possible for relevant parties such as specialists and crime analysts to study crime patterns. This increase in the accuracy for accurate crime forecasting is always desirable. Apriori algorithm is used on the output generated from K means clustering. K means algorithm generates two clusters based on high or low crime rates. After using Weka tools for visualizing the data, it is seen that the conviction rate is inversely proportionate to the number of arrests. A regression model is created on the previous 14 years data crime related data. The crime rate can be predicted by using this model. The regression model showed dependencies, positive negative correlation are observed. It gives the range high or low for people convicted, people acquitted by using Apriori Algorithm. It estimates the number of crimes that will happen for particular input as a prediction. It then calculates the number of crime occurrences in the state and for every crime cases the number of people convicted by using regression model. For example, it is seen in 10 rape cases, the number of people convicted on an average is 3 [2].

A web-based system to analyse online papers using different data mining techniques has been implemented. Data mining techniques like Document Classification,

Named entity extraction clustering, measure of similarity have been examined in the context of news paper analysis. In comparison to the other systems that implement crime news classification and location extraction, this paper also proposed how to extract crime story from different newspaper articles based on a specific news. The system crawls web pages to collect plain text. Then the news is classified into crime or non-crime using Support Vector Machine classifier. The paper uses Named Entity Recognition to extract the location by first using sentence classification. The NER kernels that are implemented show similar values but the chunk parser gives a marginally better overall result for precision. The next step is further ranking the locations based on the crime occurrence. Then groups same type of stories together. The system was demonstrated by crawling two websites. The web crawler parses the sites and crime news and non-crime news are both collected. The accuracy of successful news classification lies in the range of 80-85% [3].

The Cluster analysis by using K-means clustering algorithm on criminal dataset taken from the The National Crime Records Bureau(NCRB). The dataset used does not have labels for different classes. To overcome this, they uses clustering to create groups of data. The Weka tool is used for the purpose of visualization of data. Weka takes the database as input. The weka tool is then able to form clusters from the data. An important factor in the formation of the groups is that the sum of squares of distances between each data point in a cluster and its centroid are minimized. These groups are further categorized based on through their attributes. The states are classified in high, low and medium crime zones. The crime map is created using CustomMap, an online tool that helps in visualizing clusters . Male and female crime information is also classified. They intend to increase the number of attributes under consideration for the future. [4].

A k-means clustering based system which analyzes news data sets and shows the areas having high crime rates. It takes the crime data set from the government of the respective country. Metadata of the data set contains additional attributes such as the location, area, type of crime and the year. The have used the Rapid Miner Tool in which various machine learning and data mining techniques can be used. Preprocessing is done to replace missing values and normalize the data for clustering. The data then is given to the clustering algorithm which gives results in clusters. Each cluster has lists the maximum number of crimes committed by category. These clusters indicate the cities with high crime rates. A map with colored regions is shown for the same. It was observed that highest concentration of crime was in the the central and south west region of the country. The crimes that were committed most recently were crimes such as drug abuse, and threatening behaviour [5].

A time series model for forecasting the annual crime rate in India using Auto-Regressive Integrated Moving Average (ARIMA) and Exponential Smoothing. The data is taken from the National Record Bureau of India. The dataset is given to the Box-Jenkins Model. The role of Box-Jenkins Model is to select appropriate and best mode. It consists of three phases and they are identification, testing and estimation and application. Further Exponential Smoothing is applied. It consists of three methods visualizing Single Exponential Smoothing, Holt Linear Method and Holt Winters Method. Later ARIMA model is used. In this paper it is conveyed that time series model can be used for crime forecasting [6].

A model to predict and classify the crimes that have occurred in the state of

Tamil Nadu, to aid law enforcement agencies. The identification of crime is based on the KNN classification methods. Methods of clustering such as K-means, Agglomerative and DBscan are compared to see which provides greater accuracy for a particular data set. It clusters local areas on the basis of the number of crimes that have occurred. For example, blue for low crime rates, yellow for moderate, red for high. The database used is National Crime Records Bureau (NCRB) of India. The clustering methods are further compared in terms of precision, recall and a new derived quantity known as F-measure. The range of F-measure helps in determining the efficiency of the system. When all three quantities were measured, it was seen that the system which uses the DBscan algorithm for clustering had the the best accuracy [7].

A system to qualitatively and quantitatively predict crimes. The data sets used utilize official records as a base. Methods such as Support Vector Machine (SVM) for generating hotspots, Fuzzy Logic, Artificial Neural Networks (ANN), and Multivariate Time Series have been applied. Each of the methods have been compared on the basis of efficiency and the time required to generate appropriate results. It discusses that having large data sets leads to better accuracy and builds a better model for both prediction and analysis. ANN is the most suitable method for huge data sets [8].

A system that provides an efficient prediction model that helps in reducing the crime rates on the basis of data retrieved by analyzing crime. The data about recent crimes is provide by local news papers A focused web-crawler implementation, Crawler4j is used to parse online news papers. Next, Document classification is applied to classify the articles as criminal or non-criminal. SVM classifier is used for this purpose as it provides the best results when applied on datasets that are uneven. Entity extraction which is a technique of text mining extracts relevant parameters from articles. For example events, names, locations. Duplicate detection is carried out so that same articles or similar articles are not considered again. This allows further increase in efficiency. Hotspot are generated that indicate crime rates in an area. Crimes are compared by types, and changes in crime frequency are shown by a time series Crime analysis allowed for accurate predictions of crimes [9].

A novel framework that uses large crime related datasets from various sources such datasets maintained by several law enforcement agencies. The features are extracted and added to different matrices. The matrices were then used to train a predictive model. The model is trained using parameters from one locality (borough). these parameters included spatial and temporal patterns across the city. The model was used to predict the the parameters for other localities. The paper also discusses further improvements to the system. They proposed mining each locality for a specific crime and Determining the times during which crime rates were highest. Another idea was to determine which crime was more likely to occur during a particular time in a locality [10].

A range of different machine learning models are been used for classifying and predicting the crime datasets which has been extracted from the official portal of Chicago police. The data is first preprocessed by removal of the fields having null values. The algorithms used for classifying and predicting are KNeighborsClassifier, GaussianNB, MultinomialNB, BernoulliNB, SVC, Decision Tree Classifier to check the accuracy of different models. It is seen after testing the model that the highest accuracy is provided by the KNN classifier Later the results are presented using

different data visualisation techniques by grouping using different features [11].

A review has been done on various data mining algorithms and tools to determine which one performs the most efficiently and helps reduce crime by the greatest margin. The method they have employed to calculate the crime rates takes into account the number of registered cases and the number of people arrested over a five year time frame. It is noticed that the Z-crime tool and the the ID3 algorithm allows the prediction and mitigation of criminal activities. The Hidden link algorithm is seen to help in prediction and possible prevention of crime before it occurs. The Naive Bayes algorithm has also shown an accuracy of over 90 percent while classifying crimes [12].

The paper explores machine learning algorithms such as deep learning to help prevent future crimes in Taiwan. The Broken windows theory is taken into consideration along with spatial analysis, the Broken windows theory suggests that when there is no response to minor crimes, a significant rise in crime rates for that particular area is seen. Using this theory as basis, predictions are made for drug related crimes which depend upon other crimes occurring in that area. in the implementation, the map has been split into individual sections, and features are assigned to all the previous crime occurrences. These occurrences are divided across multiple types of crimes. spatial and temporal features are additionally assigned. The models were prepared to analyze different classification algorithms including Naive Bayes classifier, Decision tree algorithm, and Deep learning algorithms. The parameters for evaluation were f-measure, precision and recall. Among the three, Deep Learning algorithms show significant improvement [13].

A model for analyzing crime using machine learning algorithms. The algorithms used were KNN along with Decision Tree were used to train the model. The Vancouver city dataset set which is publicly available, was used. In this paper, thy have used two different approaches for data processing and the results were compared. Categorical variables were converted to binary variables. The variables that had the value "1", were set as correct variables, the value "0" assigned to all other variables. This was to increase the accuracy and avoid false positives. The second approach assigned individual ID's to different category variables, such as locality and crime type. Cross-validation was performed to avoid the over-fitting problem. Both the approaches were used for each algorithms. When KNN algorithm was used it was observed that the first approach was marginally more accurate then the the second, but required significantly greater training time. For the decision tree model, they have employed the AdaBoost method. The second approach was seen to have better overall accuracy and training time when compared to the first approach [14].

# Chapter 3

# Design

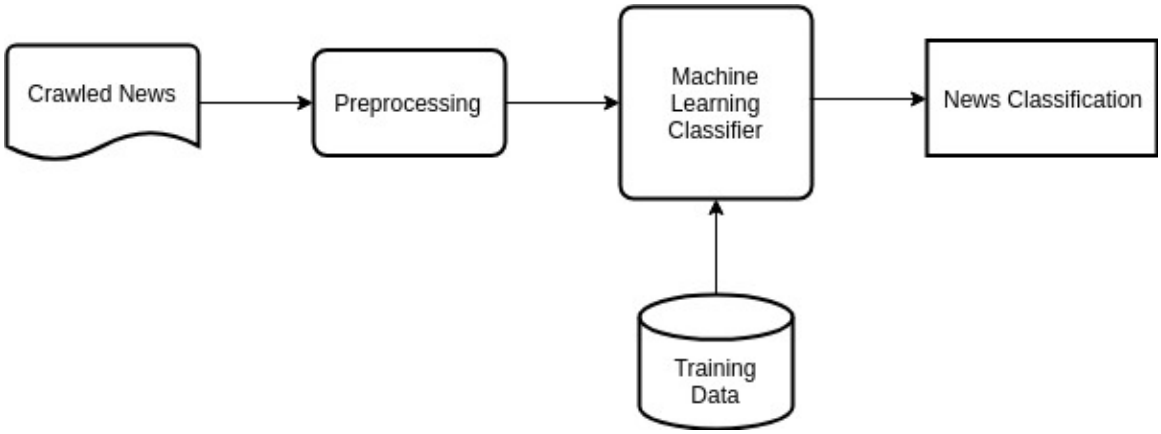


Figure 3.1: System Architecture

**Fig: Activity Diagram**

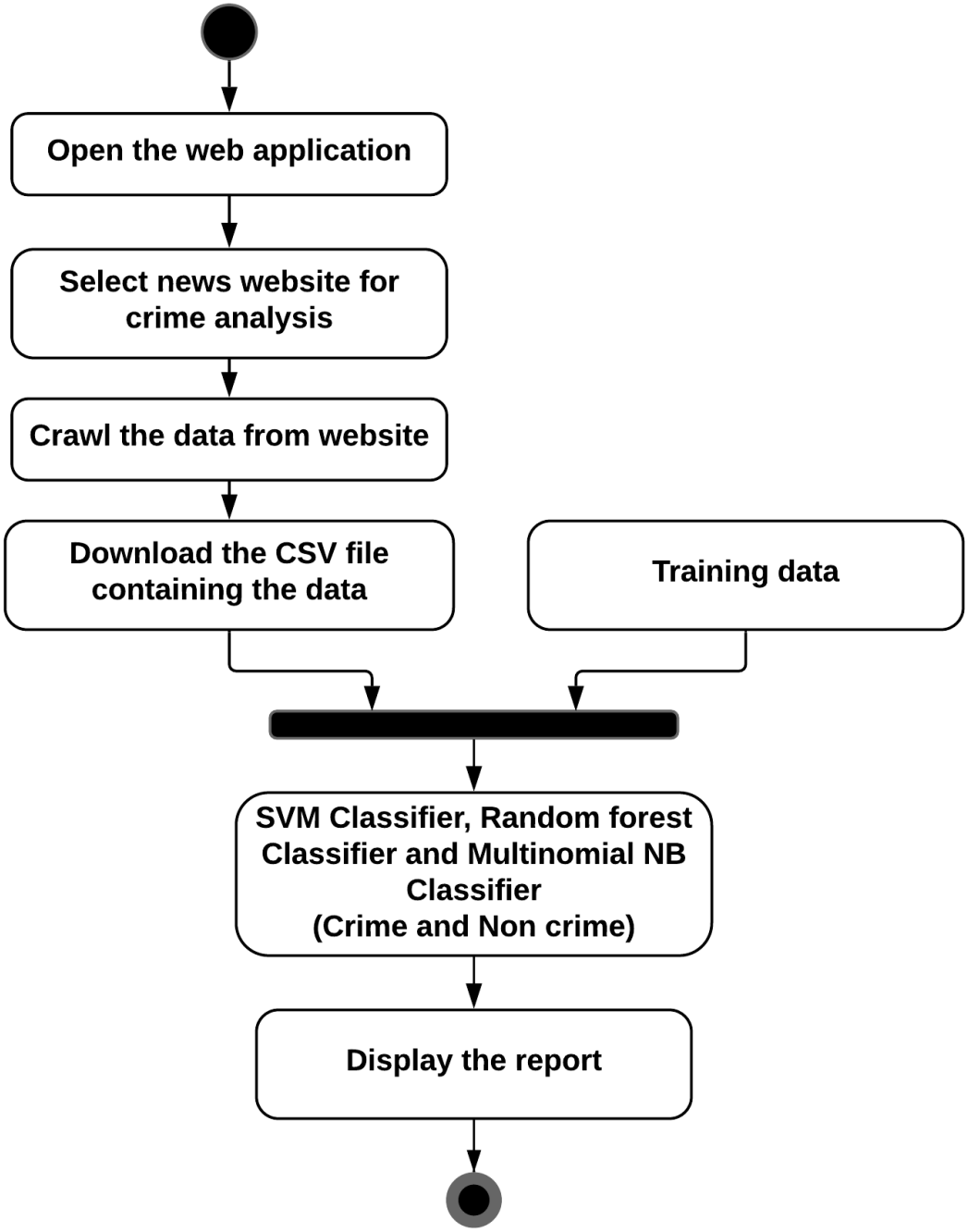


Figure 3.2: Activity Diagram

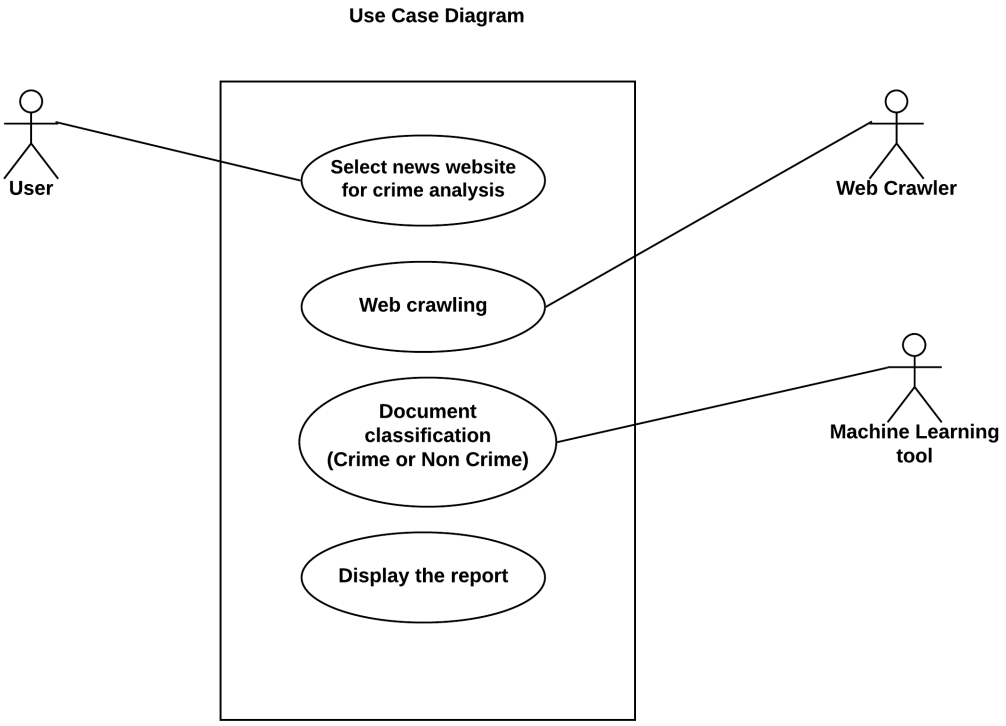


Figure 3.3: Use Case Diagram

# Chapter 4

# Implementation

### 4.0.1 Dataset and Data Pre-processing

The underlying machine learning model is trained using 6700 news records taken from the News Category dataset. The News Category dataset contains around 200k news from the year 2012 to 2018 taken from HuffPost. HuffPost is an American news and opinion website and blog. The News Category dataset consist of the following attributes: category (type of article), headline, authors, link (Link to the post), short description and date. The different categories of news include politics, wellness, entertainment, travel, healthy living, sports,a comedy, business, women, crime, arts and culture, environment, education, world news, science, style, religion, parents, tech, taste, money and many more. This dataset consisted of around 3400 records of crime.

To create the training dataset from the News Category dataset to train the machine learning model for our system we took these crime news records and around other 3300 records of other category of news which included comedy, business, education, entertainment, politics, religion, science, sports, tech, travel and world news. The training dataset consisted of the following attributes: headline, category(crime or non-crime), date.

### 4.0.2 Base Model

Classifying news into crime and non-crime is a type of text classification problem in machine learning. Here we have implemented classification of news using three classifiers: Random Forest, SVM( Support Vector Machine) and MultinomialNB. SVM is a fast and dependable classification algorithm that performs very well with a limited amount of data. Also Naive bayes serves as a good model for document classification. Random Forest combines the output from different weak classifiers to output the final category of news. We extract the headings from the training dataset and tokenize it into words using nltk library functions. NLTK( Natural Language Toolkit) is package of natural language processing libraries written in python to process human language. Further lemmatization is performed on the tokenized words and stop words are also removed. The final text obtained is vectorized using TF-IDF Vectorizer. TF-IDF stands for term frequency inverse-document frequency. This is done to find how important a word in document is in comparison to the corpus(collection of documents).

### 4.0.3 Web Crawling

The solitary purpose of Web Crawling is to gather information available on the Web. The proposed system uses Web Crawling for extracting information from the online newspaper websites. Web Crawling allows the proposed system to store information adequately. Through Web Crawling the system gathers real time data from the website and stores it into a Comma Separated Values (CSV) file. It immensely depends on the HTML code of the website.

For accomplishing Web Crawling, the system uses python programming language and some of the python libraries. It is easy to code and debug in python language that is why python programming language has been used. Libraries used for crawling are requests, lxml and BeautifulSoup4.

The *requests* library in Python is used for making HTTP requests. Hypertext Transfer Protocol (HTTP) is to used to establish connection between the client and the server. GET method gets data from the resource and POST method provides data to the server in order to create or update a resource. In order to use the library, it is first installed using pip and then imported in the python file. Using the *get* method, a connection is established between the python code and the website of the news source. The *lxml* parser is affluent in feature and uncomplicated when it comes to usage. It is used for processing Hypertext Text Markup Language in Python Programming language.

Since the connection is made, it is time for getting the data. BeautifulSoup is used for extracting data from the website. BeautifulSoup works with the parser of any type and provides colloquial manners for modifying, navigating and searching the parse tree. BeautifulSoup takes two parameters, first parameter is the source we get from the *get* method of *requests* library and the second parameter is the *lxml* parser. There can be different parsers such as *html.parser* and more. The proposed system uses BeautifulSoup4 for crawling website.

In the python code, URLs of the website is taken. URLs are earmarked as the source of the information. Classes of the HTML code of the website are targeted for information gathering. The system extracts headlines, summary and date of news (when the website posted the news). Each of the above entity could be present in one single class or different classes of the HTML code of the website. After targeting classes, tags of the HTML codes are targeted and text form of the tag is extracted from the website. One by one the news is extracted and stored into the csv file according to the entities. Hence, a csv file is created which consists of significant information extracted from the website using Web Crawling.

### 4.0.4 Real Time Classification

The crawled news from the online sites is used as an input to the base model for making predictions. Random Forest, SVM and Multinomial NB classifiers were used to classify the crawled news into crime and non-crime. Classification was done using the headlines of the crawled news. The headlines are first processed by tokenizing them into words, then performing lemmatization on it and removing the stop words. Then it is vectorized using TF-IDF Vectorizer. These vectorized headlines are clas-

sified into crime and non crime using Random Forest, SVM and Multinomial NB classifiers. The predicted output was manually reviewed. Here sometimes Support Vector Machine classifiers gave a better accuracy than Multinomial NB and Random Forest classifier. The model sometimes wrongly classifies the non-crime news into crime due to occurrence of certain type of words such as 'die, death' in the news.

# Chapter 5

# Results and Discussion

Following table shows the accuracy percentage for cross-validation dataset.

| Classifier | Accuracy Percentage |
|---|---|
| Support Vector Machine | 90.34 |
| Multinomial NB | 91.57 |
| Random Forest | 88.02 |

The predicted output obtained from the classification of crawled news was manually reviewed. Here sometimes Support Vector Machine classifiers gave a better accuracy than Multinomial NB and Random Forest classifier. The model sometimes wrongly classifies the non-crime news into crime due to occurrence of certain type of words such as 'die, death' in the news.

# Chapter 6

# Conclusions

Due to the emergence of multiple types of crime, it becomes necessary to analyse the trends and patterns of the crimes crimes that are committed. Newspapers are a major source of information. however they are mostly seen to have unstructured data. Structuring the information can lead to better analysis. Use of machine learning is trending in every field. Hence here we used machine learning for classifying the news into crime and non-crime. Classification of real time news was performed. The most important part was that it provided analysis for the news at the current moment which is of far more importance as compared to the past news.

# Chapter 7

# Future Scope

In this system, we have implement document classification to classify each news instance as crime or non-crime. The news from the website is collected in real-time and changes daily.

For future work, the news from each day day can be stored and monthly or yearly crime statistics can be generated to determine how the crime rate varies for a location. Furthermore, if datasets containing different crime types with individual features are utilized to train the machine learning model, classifying the news data on the basis of crime types is possible. This could be performed using Natural Language Processing techniques such as Entity extraction.

# Bibliography


[1] Jayaweera, C. Sajeewa, S. Liyanage, T. Wijewardane, I. Perera and A. Wijayasiri, "Crime analytics: Analysis of crimes through newspaper articles," 2015 Moratuwa Engineering Research Conference (MERCon), Moratuwa, 2015, pp. 277-282https://ieeexplore.ieee.org/document/8614828. doi: 10.1109/MERCon.2015.7112359

[2] S. Yadav, M. Timbadia, A. Yadav, R. Vishwakarma and N.Yadav, "Crime pattern detection, analysis prediction," 2017 International conference of Electronics, Communication and Aerospace Technology (ICECA), Coimbatore, 2017, pp. 225-230. doi: 10.1109/ICECA.2017.8203676.

[3] M. Hassan and M. Z. Rahman, "Crime news analysis: Location and story detection," 2017 20th International Conference of Computer and Information Technology (ICCIT), Dhaka, 2017, pp. 1-6. doi: 10.1109/ICCITECHN.2017.8281798

[4] L. S. Thota, M. Alalyan, A. A. Khalid, F. Fathima, S. B. Changalasetty and M. Shiblee, "Cluster based zoning of crime info," 2017 2nd International Conference on Anti-Cyber Crimes (ICACC), Abha, 2017, pp. 87-92. doi: 10.1109/Anti-Cybercrime.2017.7905269

[5] A. Joshi, A. S. Sabitha and T. Choudhury, "Crime Analysis Using K-Means Clustering," 2017 3rd International Conference on Computational Intelligence and Networks (CINE), Odisha, 2017, pp. 33-39. doi: 10.1109/CINE.2017.23

[6] M. Kumar et al., "Forecasting of Annual Crime Rate in India: A case Study," 2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI), Bangalore, 2018, pp. 2087-2092. doi: 10.1109/ICACCI.2018.8554422

[7] S. Sivaranjani, S. Sivakumari and M. Aasha, "Crime prediction and forecasting in Tamilnadu using clustering approaches," 2016 International Conference on Emerging Technological Trends (ICETT), Kollam, 2016, pp. 1-6. doi: 10.1109/ICETT.2016.7873764

[8] N. H. M. Shamsuddin, N. A. Ali and R. Alwee, "An overview on crime prediction methods," 2017 6th ICT International Student Project Conference (ICT-ISPC), Skudai, 2017, pp. 1-5. doi: 10.1109/ICT-ISPC.2017.8075335

[9] D. V. Rohini and P. Isakki, "Crime analysis and mapping through online newspapers: A survey," 2016 International Conference on Computing Technologies and Intelligent Data Engineering (ICCTIDE'16), Kovilpatti, 2016, pp. 1-4. doi: 0.1109/ICCTIDE.2016.7725331

[10] X. Zhao and J. Tang, "Exploring Transfer Learning for Crime Prediction," 2017 IEEE International Conference on Data Mining Workshops (ICDMW), New Orleans, LA, 2017, pp. 1158-1159. doi: 10.1109/ICDMW.2017.165

[11] Alkesh Bharati, Dr Sarvanaguru RA.K, Crime Prediction and Analysis Using Machine Learning, International Research Journal of Engineering and Technology (IRJET) Volume: 05 Issue: 09 — Sep 2018

[12] C. Chauhan and S. Sehgal, "A review: Crime analysis using data mining techniques and algorithms," 2017 International Conference on Computing, Communication and Automation (ICCCA), Greater Noida, 2017, pp. 21-25. doi: 10.1109/CCAA.2017.8229823

[13] Y. Lin, T. Chen and L. Yu, "Using Machine Learning to Assist Crime Prevention," 2017 6th IIAI International Congress on Advanced Applied Informatics (IIAI-AAI), Hamamatsu, 2017, pp. 1029-1030. doi: 10.1109/IIAI-AAI.2017.46

[14] S. Kim, P. Joshi, P. S. Kalsi and P. Taheri, "Crime Analysis Through Machine Learning," 2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), Vancouver, BC, 2018, pp. 415-420. doi: 10.1109/IEMCON.2018.8614828

Attach your plagiarism report here.