Event Detection and Violence Recognition from Textual News through Multilayer Perceptron and Supervised Learning

Tasmia Ishrat Alam

Student Id: 011 121 040

Imtiaz Mamun

Student Id: 011 133 061

Iftakhar Ali Khandokar

Student Id: 011 133 003

Zubair Ahmed Anas

Student Id: 011 133 013

A thesis in the Department of Computer Science and Engineering presented
Impartial fulfillment of the requirements for the Degree of
Bachelor of Science in Computer Science and Engineering



United International University
Dhaka, Bangladesh
October 2018

Abstract

An unprecedented way is accomplished by using concept words derived from statistical context analysis between sentences which is better than traditional methods that use only keyword representation. Through scaling to a very large dataset we proposed an algorithm which discovers, and describes events with effective keyword networks, based on their coexisting peripheral co-occurrences. In our experiment, we used real-world news, and supervised them into paraphrases by weighting for the all attempted events. We evaluated our scheme by a set of terms that maximally discriminated the percussion in news and which also keep the evidences. Here we are classifying the events with a multilayer perceptron by executing auto-convolution methodology in back propagation.

Acknowledgements

We would like to take this opportunity to show our gratitude to our supervisor Dr. SwakkharShatabda for guiding us throughout the process and making it smoother. We express our warm thanks to Dr. SwakkharShatabda for teaching us numerous things and for his aspiring guidance along with his enthusiastic assistance during the development of this project.

Table of Contents

LIST OF TABLES v
LIST OF FIGURES vi
1. Introduction 1
2. Background and Literature Review 3
2.1 Related Work 3
2.2 Benefits 5
2.3 Brief Review of the Learning Algorithms 5
3. Materials and Methods 12
3.1 Multilayer Preceptor 12
3.1.1 Data Collection 12
3.1.2 Preprocessing 13
3.1.3 Feature procurement 14
3.1.4 Murder (murder and attempt to murder) 15
3.1.5 Kidnap (abduction and rape) 16
3.1.6 Hassle (rioting and affray) 17
3.1.7 Protest (unlawful assembly) 17
3.1.8 Accident (accident and Injuring or defiling) 18
3.1.9 Terror (persons concerned in criminal act) 18
3.2 N-gram and Triggering Word 20
3.2.1 Data Collection 21
3.2.2 N-gram Selection 22
3.2.3 Dataset Representation 23
3.2.4 Feature Set Reduction 25
4. Bottom up Hierarchical (BUH) Classifier 26

5. Results, Diagrams and Flowchart 31

5.1 Experimental Analysis of Multilayer Perceptron Model 31

5.1.1 Results 31

5.1.2 Graphs 34

5.1.3 Flowchart 38

5.2 Experimental Analysis of N-gram & Triggering Word Model 39

5.2.1 Results 39

5.2.2 Graphs 42

5.2.3 Finalized Model on an Unlabeled Dataset 45

5.3 Auto-Convolution Vs N-gram Model 47

5.3.1 Percentage Split 47

5.3.2 Cross validation 48

6. Conclusion 50

7. References 51

LIST OF TABLES

Table 1:A data point after the collection of the news from star_date to stop_date 13
Table 2: After procurement of the feature we isolated the dataset into two tables 20
Table 3: Binary Representation of N-gram's and Triggering Word..... 24
Table 4: Keywords and the following network 26
Table 5: Weighted by total appearance of each keyword for their corresponding network 27
Table 6: Binary presence of keywords for news id "11" 28
Table 7: Maximum presence of keywords for news id "11" 28
Table 8: Weighted by neural network classifier with binary presence 29
Table 9:Weighted by neural network classifier with maximum presence..... 29
Table 10: Weighted by KNN classifier with binary presence..... 29
Table 11: Weighted by KNN classifier with maximum presence 29
Table 12: The discrimination of accuracy for all 2018 news 33
Table 13: The discrimination of accuracy for human labeled dataset..... 33
Table 14: Accuracy of Different Learning Algorithms in Train and Test Set Using Full Feature Set 40
Table 15: Accuracy of Different Learning Algorithms in Train and Test Set Using Reduced Feature Set 41
Table 16: Performance Comparison Table between Full & Reduced Feature Set on Train Set 41
Table 17: Performance Comparison Table between Full & Reduced Feature Set on Test Set 42

LIST OF FIGURES

Figure 1: KNN working procedure 6
Figure 2: Decision Tree Example 7
Figure 3: Logistic Regression Workflow 7
Figure 4: Binary Classification through Hyper-plane and Support Vector in SVM 8
Figure 5: Boosting Mechanism 9
Figure 6: Bagging Working Technique 10
Figure 7: Triggering Word Cloud 21
Figure 8: Data Collection process 21
Figure 9: N-gram Extraction Window 22
Figure 10: Extracted N-gram Cloud 23
Figure 11: News Article vector 25
Figure 12: N-gram Frequency Histogram 25
Figure 13: Graphical representation of multi-layer perceptron 39
Figure 14 : Accuracy Comparison between Auto-Convolution and N-gram Model Using Percentage Split 48
Figure 15: Accuracy Comparison between Auto-Convolution and N-gram Model Using Cross Validation 49

Chapter1

Introduction

Violence is unfortunately showing an alarming increase. We need to recognize violence as soon as possible so that reducing the impact of the violence spectrum and managing the threats of violence can be collaborated before it forms a large friction. It has probably always been a part of the human experience. Its impact can be seen, in various forms, in all parts of the world each year. In this research we detected the violence successfully by a few networks containing some keywords. Here we vastly followed some methods to get the best efficiency for violence classification. Information retrieval is not an easy task. So we followed the rules so that we do not get separated from our target stage. As example, for weighting our data points we followed tf-idf weighting approaches. We also used multiple classification algorithms. Whole process is a dynamic process as each classification algorithm provides a layer for the multilayer perceptron. We used auto-convolution methodology to stay and fit in semi-supervised region and make our system more users friendly and dynamic.

A newspaper is something that is regular in our daily life and day by day the number of textual news and articles are widely increasing. We wanted to know what kind of information news represents and what kind of event has occurred or is occurring according to the textual contents of any news and article.It's not a rational option to read all the news and objectify the events mentioned on the textual news. so we need an application to automatically extract information from any news and also detect what kind of event the news stand for. As an example what kind of state is our society going through or how is our infrastructure behaving is directly reflected through the vary news published daily on our local newspapers. The main objective of our work to make an automated tool that can identify violence related news and also identify the place of the violence occurring with time. The tool we want to develop will provide crime rates in each and every parts of the country with the type of crime happening daily. With the help of this tool we would have knowledge about what place or states are safe and are we heading towards a safer society by comparing the crime rates with the previous year or month or any duration of time. This purpose can be achieved through reading all the news manually and categorize the news articles according to their violence type but that will be

next to impossible for a human to read all the news one by one and classify the subjected news articles so for this baleful reason we decided to make an tool that can identify any violence happening in any news and also detect the place of its occurrence to make the process more rational as we think that this tool will be a beneficent contribution in recognizing and taking precaution against crime eradication.

Chapter 2

Background and Literature Review

2.1 Related Work

Although Bengali is the fourth largest language of the world having over 200 million native speakers but still now Bengali language does not accomplished any grammar checker for a Bengali sentence. Parsing the meaning of Bengali sentences is in a preliminary stage still now. Very few research work have been driven to parsing the Bengali sentences rather than that many research activities have been accomplished on the recognition of Bengali context-sensitivity. Phonological analysis of Bengali phrases is presented in several inquiries [1] [2] [3] [4]. As for building a keyword network we had to do the phonological analysis of Bengali phrases. In Bengali we do not have the concept of small or capital letters. Unlike English, every letter in Bengali word is capital only. For this reason we find difficulties in understanding whether a word is a proper noun or not. Bengali word also contains joint letters, and does not follow the regular grammar thus for subject-object-verb (S-O-V) structure of Bengali sentences does not fit precisely. For example: “amibhatkhai”, and “amikhaibhat” is meaningful in Bengali, and in English both means the same: “I eat rice” but it is not a meaningful sentence in English when it is “I rice eat”. In Bengali language often we do not use verb [3]. For example, in the Bangla sentence “Karimvalochele” has no verb but in English (Karim is a good boy) at least one verb must be present in a sentence.

Like some other languages, there is also a very intense intimacy with phonology in Bengali language. In various tone level of a Bengali phrase it can mean different things which can also help us to classify the identical emotion [1] [2] [3]. For example: “আচ্ছা” or “Okay” both has the same meaning but in different languages which basically means agreement. This word is used for multiple purposes like if it is meant like a confusion or question then it might be spoken in different tone than the primary tone level [4] [5]. There are also some other tone levels which is for expressing diverse meanings.

In paper they tried to detect flames and insulting statements from text for so they approached by detecting factive event which is when any kind of textual statements includes any says or told type of word they called it an factiveevent if any kind of factive event is detected the rest of the words after factive event declaring words are matched to any kind of insulting words and their synonyms if any kind of insulting word is matched with the statement word they objectified as it an insulting statement and it works more like an regular expression working technique.

Associative intrusion detection in database is complex, and a dynamic process. It has been verified that effective attributes selection improves the detection for a network intrusion detection using decision tree-based attribute [6]. In naive Bayesian tree, nodes contain, and split as regular decision-trees, but the leaves contain naive Bayesian classifier [7]. This algorithm maximizes the precision, and recall accomplishment for most of the circumstances [8].

In predictive analytics and machine learning, the concept drift means the statistical properties of the target variable, which the model is trying to predict, change over time in unforeseen ways [9]. The term concept refers to the quantity to be predicted. More generally, it can also refer to other phenomena of interest besides the target concept, such as an input, but, in the context of concept drift, the term commonly refers to the target variable [6] [9]. In law, time constraints are placed on certain actions and fillings in the interest of speedy justice, and additionally to prevent the avoidance of the ends of justice by waiting until a matter is argued.

In this approach Decision Tree is used to detect multiple novel classes. The basic decision tree algorithm ID3 builds decision tree. In this technique, we built a decision tree from training data points and calculate the percentage of number of data points in each leaf node with respect to data points in training dataset. Now apply cluster on each leaf node of tree based on similarity. In real time classification novel class is arrived if number of data point in leaf node of tree is increase than percentage calculated before. The idea of detecting multiple novel class is to construct graph, novel class [9] obtained is plotted on graph. After constructing the graph identify connected component, the number of connected component determines the number of novel class Procurement.

At the present moment, there are different lexicons for affect detection and opinion mining [10] [11]. The aim in the following evaluation is to test the different resources in the quote classification scenario and assess the quality and consistency of these lexicons. Each of the employed resources was mapped to four categories, which were given different scores – positive, negative, high positive and high negative. The assignment of these values was based on the intuition that certain words carried a higher affective charge and their presence should be scored accordingly. Its intuition was supported by experiments in which it used just the positive and negative categories and that scored lower [12]. A positive score leads to the classification of the quote as positive, whereas a final negative score leads to the system classifying the quote as negative.

For the auto-convolution we executed a multilayer perceptron depending on the smoothness of the keywords for each keyword networks which derivatives our solution into small arguments [13]. Auto-convolution was first analyzed in physics and later in function optimization as the problem of auto-convolution emerge [8] [14]. This methodology is also appeared in visual tasks to extract invariant patterns [15].

2.2 Benefits

Law enforcement teams for crime avoidance, evidences for the proper sense of justice and news publications for better filtered news.

2.3Brief Review of the Learning Algorithms

KNN:

Among all the learning algorithms KNN has the most simple working procedure where when a new instance appears KNN simply calculates the distance between the new instance and the entire training set instances and selects K number of closest instance from the training where K is a hyper parameter after selecting the K number of instances a voting technique is applied for the individual class labels and the majority voted label among the selected instances is assigned as the label for the new instance. For distance calculating various techniques is used mentioning some of them Manhattan distance , Euclidian distance , square root distance technique etc. the value of K is always selected an odd number so there can't be any draw situation and KNN doesn't actually fit an model whenever a new instance comes it just classify it by calculating the distance that is why KNN is called an lazy learner and the performance of KNN in primarily based on the

value of K and the authentication of the dataset. As we can see in the (Figure 1: KNN working procedure) below that the closest neighbors are 3 and among them two of them belong to the red class label so the new green instance will also belong to the red class according to KNN:

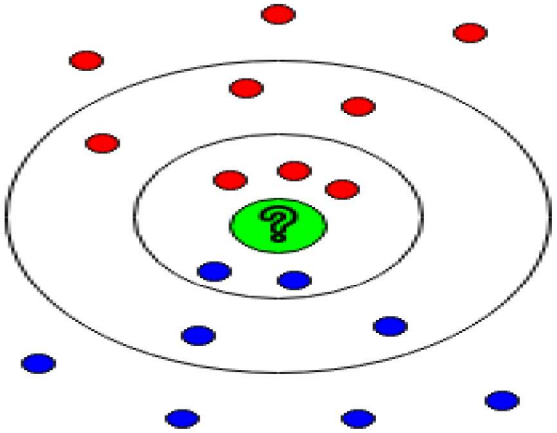


Figure 1: KNN working procedure

Decision Tree:

Decision tree working procedure goes something like this Among the available features in the training set information gain of each of the features present is calculated and after calculating we will construct a tree of this features where the root of the tree is the feature having the highest information gain and tree will expand according to the rest of the features information gains decreasing order. After constructing the full feature tree when an new instance appears we start with the root node of the tree and the new instance will follow the branches according to the value of the present node feature and it will go on until the new instance reaches the leaf node which means the class label node that is the predicted label for the new instance. The performance of the decision tree depends on the authenticity of the dataset so the right feature with the highest information gain is selected as the top most nodes an if the tree over fits the training set pruning have to be done to improve performance. According to the (Figure 2: Decision Tree Example) below the new instance will start with top node p and follow the branch according to the value of feature p of the new instance and in the next node is a leaf node the process is halted otherwise the entire process is repeated until leaf node is reached:

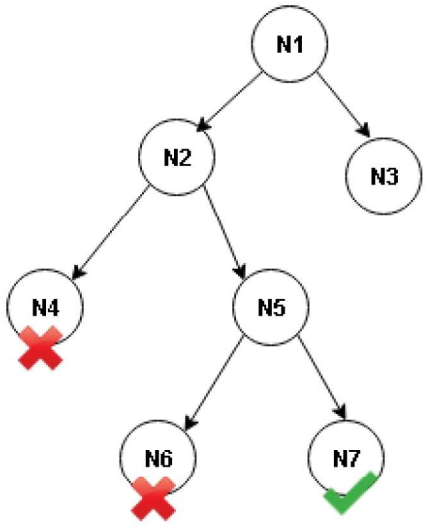


Figure 2: Decision Tree Example

Logistic Regression:

The Logistic Regression Algorithm is a regression algorithm that works almost like an regression algorithm procedure but doesn't fully behave like one the logistic regression algorithm doesn't predict any numeric value based on the trained test set model rather it gives an probability as an output. From the probability the class of any instance set is predicted or determined of any classification problem. The algorithm divides the data in two different line spaces that are very far from each other so the data have to be linearly separable.

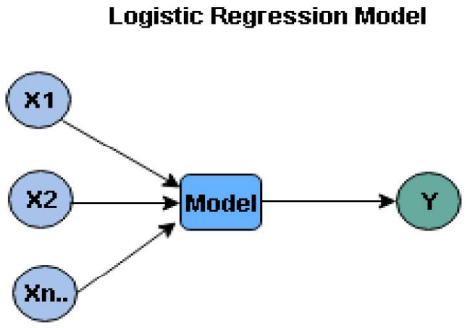


Figure 3: Logistic Regression Workflow

Above is the (Figure 3: Logistic Regression Workflow) working procedure of logistic regression where we can see that n number of features is given as inputs and this inputs goes through an single perceptron and y is the output of this

perceptron which is actually the probability of the instance set which will determine what side of the line space is the instance set on. At first the weights and the bias that we want to learn from the training set is initialized randomly or more commonly to zero after that we update the weights according to error in every iteration through gradient descent mechanism and for the output a sigmoid function is used which is the probability of the given input.

Support Vector Machine:

The Support Vector Machine is a supervised learning algorithm that can be used for both classification and regression problem but svm is more likely used in classification problem. The main agenda of svm is to separate the data as efficient as possible for which svm uses the hyper-planes. A hyper-plane can be described as a separating line among the data which can be derived from only two features but this hyper-plane can be deprived out of as many features available so the separating hyper-plane becomes multi-dimensional to separate the data more rationally and the classification is done by observing at what part of the hyper-plane is the new upcoming data points are going. Below is a (Figure 4: Binary Classification through Hyper-plane and Support Vector in SVM) of svm hyper-plane and support vector:

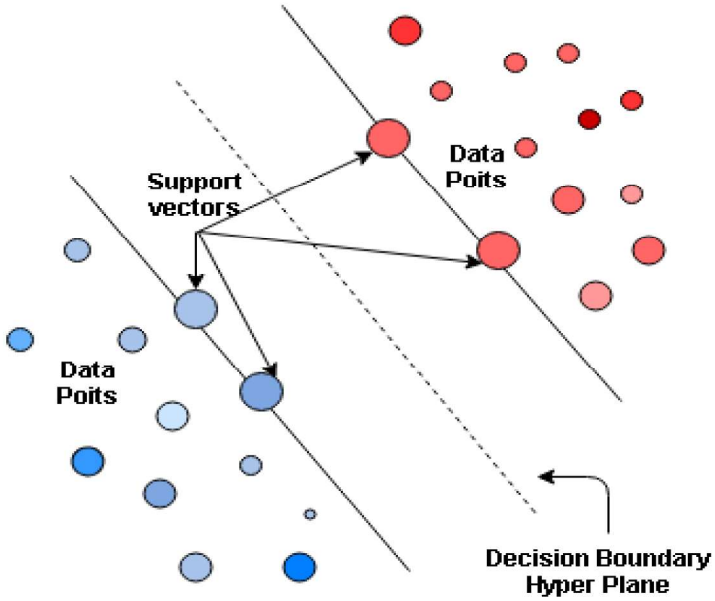


Figure 4: Binary Classification through Hyper-plane and Support Vector in SVM

From the (Figure 4: Binary Classification through Hyper-plane and Support Vector in SVM) above we can see the yellow dot line that is the hyper-plane and the dark straight lines on both side of the hyper-planes are the support vector lines. The support vectors are those data points that have the maximum distance from all of the data points among the available training dataset and the best rational hyper-plane is determined by which hyper-plane has the highest distance from the support vectors of all classes. The hyper-plane determination process requires an kernel which is used to map the features more like a function and this kernel can used to generate new features.

Boosting:

Boosting is an ensemble method to convert a weak classifier to an strong classifier. The main idea is that it learns from its previous mistakes and improves its performance so it uses multiple classifiers and each classifier gives more priority to the previous one's flaws.

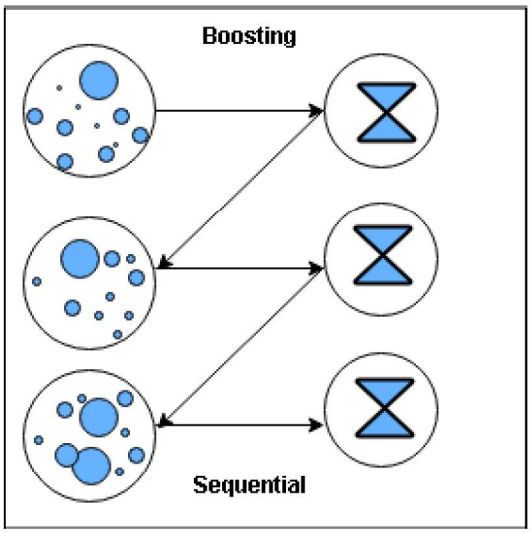


Figure 5: Boosting Mechanism

As we can see in the (Figure 5: Boosting Mechanism) boosting is a sequential process it begins with a classifier and based on the mistakes of the classifier it decides the next training set for the preceding classifier and gradually the performance of the classifier improves and the final outcome is decided based on the output of the classifiers and the classifier with the less error is given the highest priority respectively.

Adaptive-Boost:

Adaptive Boosting is a supervised ensemble learning algorithm that uses boosting mechanism for classification. First all the points in the training dataset are initially given a weight and all the weights of the points are same and the training set points are selected randomly and the highest weight points are given priority. If the error of any model is greater than 50 percent that it is dropped and the training set is again selected on the other hand if the model is accepted than it is assigned with a weight based on the accuracy of the model. After getting the final model set when a new instance appears all of the models vote for the new instance and the class or label with the maximum votes is assigned to the new instance and this is the working mechanism of Adaptive-Boost.

Bagging:

Bagging is an ensemble sampling technique to reduce variance in any machine learning model primarily the idea is to divide the full training dataset in k number of bags and all of this bags will have n number of instances from the main training dataset and we will select this instances randomly to fill the k number of bags any instance can have repetition in any bags multiple times. Below is a (Figure 6: Bagging Working Technique) of bagging working procedure:

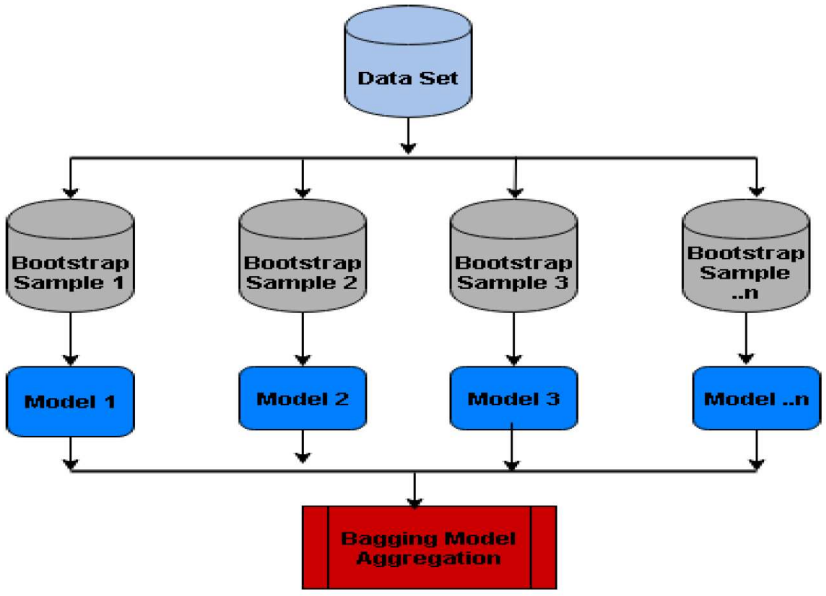


Figure 6: Bagging Working Technique

As we can see from the (Figure 6: Bagging Working Technique) that multiple Bootstrap set is derived out of the original training set randomly. After collecting the bags of instances we will learn a model from each and every one of the bags then we will have m number of models and during prediction task we will fetch the test set to each and every one of this model's and collect their output to average and determining the final result.

Random Forest:

Random Forest is an ensemble classifier that works gathering the outputs of multiple decision trees. Random Forest uses bagging technique to divide the dataset in k number of folds and learns a decision tree from each and every one of the folds. After learning all the k number of decision trees Random Forest stores the trees in an model set that is the final ensemble model after that if any new instance comes Random Forest uses all the trees in the model set to classify the new instance and each and every tree predict a label for the new instance to finalize any label Random Forest takes majority voting that mean the class or label with the majority votes conducted by the decision trees is selected as the label for the new instance. Random Forest works best if the bags or folds are weakly co-related that means they have very less similarity that is when random Forest performs best.

Chapter 3

Materials and Methods

3.1 Multilayer Preceptor

Newspaper has always been a part of our daily life, and the best public accessibility to know what is happening around us. It has always been a favor for multifarious people in our society. It is the source of data, and information that helps to increase awareness, and knowledge of the citizens in a country. Here we focused on classifying the news which is mainly the evidence for violence so that we can provide a good assistance to our law enforcement teams.

3.1.1 Data Collection

In Bangladesh there is a daily newspaper named ProthomAlo which is the most renowned, and reliable source of news. We collected a huge amount of news in Bengali font from this newspaper for testing our approach. All the news we collected are the real world events in various time divisions between 2016 and 2018, and this data is gathered through a portal. We have collected the news with their time of publishing and the title.

Algorithm 3.1.1

Collection of the news

Input:start_date, stop_date

Declaration ofnews_array as 2d array

For all news between start_date to stop_date

Do

    For each news of a day

    Do

        news_array = parse the date, title, and the news

    End for

End for

Draw news_array into a comma-separated values (CSV) file.

Date	Headlines	News
17 Oct	উখিয়ারঅনিবন্ধিতশিবিরেরপাশেনতুনবসতি	স্ত্রীনুরজাহানওআটসন্তানকেনিয়েদুইমাসধরেরাখাইনেরবুড়াশিকদারপাড়ায়খালেরওপরমাচাবেঁধেদিনযাপনকরছিলেনমোহাম্মদহোসেন।কিন্তুগতচার-চারদিনেওইএলাকায়মিয়ানমারেরসশস্ত্রবাহিনীওনাডালাবাহিনীরনতুনঅভিযানশুরুরপরপ্রাণবাঁচাতেতাঁরাবাংলাদেশেএসেছেন।.....

Table 1: A data point after the collection of the news from star_date to stop_date

3.1.2 Preprocessing

After collecting the news we had to preprocess our data. In preprocessing we removed all English symbolic letters, and some Bengali characters (ex: ঁ, ঃ, ঀetc.), and all the numbers. We also removed some intimation, and they are quotation, double quotation, exclamatory, question mark, colon, semicolon, comma, brackets, backslash, forward slash, percentage, equal and many more. We also removed Bengali full stop (।) from the news.

Algorithm 3.1.2

Preprocessing of the data set

Input:data_set [All news with the title and date]

Whitespace = u [\s\u0020\u00a0\u1680\u180e\u202f\u205f\u3000\u2000-\u200a]

bengali_digits = u [\u09E6\u09E7\u09E8\u09E9\u09EA\u09EB\u09EC\u09ED\u09EE\u09EF]

english_chars = u [a-zA-Z0-9]

Punctuation = u [(),$%^&*+={}\[\]:\"|\'\~`<>/,□¦!?½£¶□¼©□□⅒⅓⅔⅕⅖⅗⅘⅙⅚⅛⅜⅝⅞⅟□¤¿º;]

bangla_fullstop = u [\u0964]

punctuation_sequence = u ['\"“”‘’]+|[.?!,…]+|[:;]

For each news in data_set

Do

If news associate with whitespace or bengali_digits or english_charsor punctuation or bangla_fullstop or punctuation_sequence

cleaned_text = remove these characters from news

End if

Replace corresponding news from data_set by cleaned_text

End for

3.1.3 Feature procurement

After collecting the news we had to preprocess our data. In preprocessing we removed all English symbolic letters, and some Bengali characters (ex: ঁ, ঃ, ৯etc.), and all the numbers. We also removed some intimation, and they are quotation, double quotation, exclamatory, question mark, colon, semicolon, comma, brackets, backslash, forward slash, percentage, equal and many more. We also removed Bengali full stop (।) from the news.

So now our data set is ready we can accumulate feature from the news for the event classification. As we proposed the benefit is for the law enforcement teams, we settled our event on focusing only in violence so that we can appoint the most violent areas, time periods and crimes. At the origin of our approach we had to build a keyword network for balancing our features. Each feature is based on multiple keywords where every keyword illustrates their identification for the feature. A single keyword beneath a feature votes for their identification in a single news, and all the votes by each keyword turn into the weight for the feature. Likely we created six keyword networks by human expert assessment that represents the weight for individual feature which is relevant, and connected to a single type of violence. The keyword network representatives are murder, kidnap, hassle, protests, accident and terror. Here is the specimen of the keyword network named murder: "খুন", "নিহত", "ঘাত", "আঘাত", "হত্যা", "গুলি", "চাকু", "বন্দুক", "পিস্তল", "আগ্নেয়াস্ত্র", "ছুড়ি", "অস্ত্র", "সশস্ত্র", "রক্তপাত", "মার", "মেরে", "লাশ", "মৃত", "ঘাতক", "পিটিয়ে"।

We schemed our process as every news is an evidence for an event, and for the detection of the event location we had to create some keyword networks which measure

the location for a single event. We declared eight keyword networks which is the all divisions of Bangladesh. Each of this keyword network holds districts of the corresponding division as keyword. Here is the specimen of the keyword network named sylhet: "সিলেট", "মৌলভীবাজার", "হবিগঞ্জ", "সুনামগঞ্জ", "বালাগঞ্জ", "বিয়ানীবাজার", "বিশ্বনাথ", "কোম্পানীগঞ্জ", "ফেঞ্চুগঞ্জ", "গোলাপগঞ্জ", "গোয়াইনঘাট", "জৈন্তাপুর", "কানাইঘাট", "জকিগঞ্জ", "দক্ষিণসুরমা", "ওসমানী", "বড়লেখা", "কমলগঞ্জ", "কুলাউড়া", "রাজনগর", "শ্রীমঙ্গল", "জুড়ী", "নবীগঞ্জ", "বাহুবল", "আজমিরীগঞ্জ", "বানিয়াচং", "লাখাই", "চুনারুঘাট", "মাধবপুর", "বিশ্বম্ভরপুর", "ছাতক", "জগন্নাথপুর", "দোয়ারাবাজার", "তাহিরপুর", "ধর্মপাশা", "জামালগঞ্জ", "শাল্লা", "দিরাই"।

Each feature we procured has its own definition in the penal code of Bangladesh. A criminal code (or penal code) is a document which compiles all, or a significant amount of, a particular jurisdiction criminal law. Typically a criminal code will contain offences which are recognized in the jurisdiction, penalties which might be imposed for these offences and some general provisions (such as definitions and prohibitions on retroactive prosecution). Simply the penal code is a set of laws relating to crimes and the punishments for those crimes. Every expression which is explained in any part of this code is used in every part of this code in conformity with the explanation. Here is the all definitions of our features from penal code of Bangladesh for violence detection.

3.1.4 Murder (murder and attempt to murder)

Except in the cases hereinafter excepted, culpable homicide is murder, if the act by which the death is caused is done with the intention of causing death, or-

Secondly.-If it is done with the intention of causing such bodily injury as the offender knows to be likely to cause the death of the person to whom the harm is caused, or –

Thirdly.-If it is done with the intention of causing bodily injury to any person and the bodily injury intended to be inflicted is sufficient in the ordinary course of nature to cause death, or –

Fourthly.-If the person committing the act knows that it is so imminently dangerous that it must, in all probability, cause death, or such bodily injury as is likely to cause death, and commits such act without any excuse for incurring the risk of causing death or such injury as aforesaid.

Source: Under the section 300 of the penal code, 1860.

Whoever does any act with such intention or knowledge, and under such circumstances that, if he by that act caused death, he would be guilty of murder, shall be punished with imprisonment of either description for a term which may extend to ten years, and shall also be liable to fine; and, if hurt is caused to any person by such act, the offender shall be liable either to imprisonment for life, or to such punishment as is hereinbefore mentioned.

Source: Under the section 307 of the penal code, 1860.

3.1.5 Kidnap (abduction and rape)

Kidnapping is of two kinds: kidnapping from Bangladesh, and kidnapping from lawful guardianship.

Whoever conveys any person beyond the limits of Bangladesh without the consent of that person, or of some person legally authorized to consent on behalf of that person, is said to kidnap that person from Bangladesh.

Whoever takes or entices any minor under fourteen years of age if a male, or under sixteen years of age if a female, or any person of unsound mind, out of the keeping of the lawful guardian of such minor or person of unsound mind, without the consent of such guardian, is said to kidnap such minor or person from lawful guardianship.

Source: Under the section 359 of the penal code, 1860.

Whoever by force compels, or by any deceitful means induces, any person to go from any place, is said to abduct that person.

Source: Under the section 362 of the penal code, 1860.

A man is said to commit "rape" who except in the case hereinafter excepted, has sexual intercourse with a woman under circumstances falling under any of the five following descriptions:

Firstly.Against her will.

Secondly.Without her consent.

Thirdly. With her consent, when her consent has been obtained by putting her in fear of death, or of hurt.

Fourthly. With her consent, when the man knows that he is not her husband, and that her consent is given because she believes that he is another man to whom she is or believes herself to be lawfully married.

Fifthly.With or without her consent, when she is under fourteen years of age.

Source: Under the section 357 of the penal code, 1860.

3.1.6 Hassle (rioting and affray)

Whenever force or violence is used by an unlawful assembly, or by any member thereof, in prosecution of the common object of such assembly, every member of such assembly is guilty of the offence of rioting.

Source: Under the section 146 of the penal code, 1860.

When two or more persons, by fighting in a public place, disturb the public peace, they are said to "commit an affray".

Source: Under the section 159 of the penal code, 1860.

3.1.7 Protest (unlawful assembly)

An assembly of five or more persons is designated an "unlawful assembly," if the common object of the persons composing that assembly is

First.-To overawe by criminal force, or show of criminal force, Government or Legislature, or any public servant in the exercise of the lawful power of such public servant; or

Second.-To resist the execution of any law, or of any legal process; or

Third.-To commit any mischief or criminal trespass, or other offence; or

Fourth.-By means of criminal force, or show of criminal force, to any person to take or obtain possession of any property, or to deprive any person of the enjoyment of a right of

way, or of the use of water or other incorporeal right of which he is in possession or enjoyment, or to enforce any right or supposed right; or

Fifth.-By means of criminal force, or show of criminal force, to compel any person to do what he is not legally bound to do, or to omit to do what he is legally entitled to do.

Source: Under the section 141 of the penal code, 1860.

3.1.8 Accident (accident and Injuring or defiling)

Nothing is an offence which is done by accident or misfortune and without any criminal intention or knowledge in the doing of a lawful act in a lawful manner by lawful means and with proper care and caution.

Source: Under the section 80 of the penal code, 1860.

Whoever destroys, damages or defiles any place of worship, or any object held sacred by any class of persons with the intention of thereby insulting the religion of any class of persons or with the knowledge that any class of persons is likely to consider such destruction, damage or defilement as an insult to their religion, shall be punished with imprisonment of either description for a term which may extend to two years, or with fine, or with both.

Source: Under the section 259 of the penal code, 1860.

3.1.9 Terror (persons concerned in criminal act)

Where several persons are engaged or concerned in the commission of a criminal act, they may be guilty of different offences by means of that act.

Source: Under the section 38 of the penal code, 1860.

Algorithm 3.1.3

Balancing the weight of a news for each keyword network

Input:

a) data_set

b) keyword_network

For each news in data_set

Do

Initiate with day, month, year, and news from data_set

For each word of a news

Do

For each attribute of the keyword networks

Do

If news word associate with any keyword

Count the keyword attendance for the corresponding network

End if

Accomplish the weight of the networks for the corresponding news

End for

End for

Write values into a comma-separated values (CSV) file by the day, month, year and the accomplished weight of the networks for a news

End for

Table for time division

news_id	day	month	year	murder	kidnap	hassle	protests	accident	terror
0	8	5	2018	39	9	30	3	2	3
1	10	5	2018	7	1	22	3	5	1
2	24	5	2018	2	0	1	0	1	4

Table for area division

news_id	Barisal	Chittagong	Dhaka	Khulna	Rajshahi	Rongpur	sylhet	mymensingh
0	0	8	6	0	0	0	0	0
1	0	0	0	0	0	1	0	0
2	0	1	6	2	0	1	0	0

Table 2: After procurement of the feature we isolated the dataset into two tables

3.2 N-gram and Triggering Word

To detect the kind of violence occurred in any textual news we proposed two methods for extracting information from news and using it for the mentioned purpose. The first method was about using n-grams and other using auto –convolution based on some identified triggering words.

Before going any further first we needed to identify some words as to detect an event occurring from a textual article it's necessary for an article to have these words mentioned within it. So if this event related word is present in an textual article there is a good chance of the event occurring in that textual news that is why we have addressed this words as our triggering words cause when this words are present it triggers the chance of the event occurring. As we want to detect violence kind of events at first we classified the type of different violence event and after that we identified the words related to this type of events and also their synonyms. We would have to detect this kind of events based on the existence of this triggering words below is

a triggering word cloud (Figure 7: Triggering Word Cloud) to show what kind of events we are trying to detect:

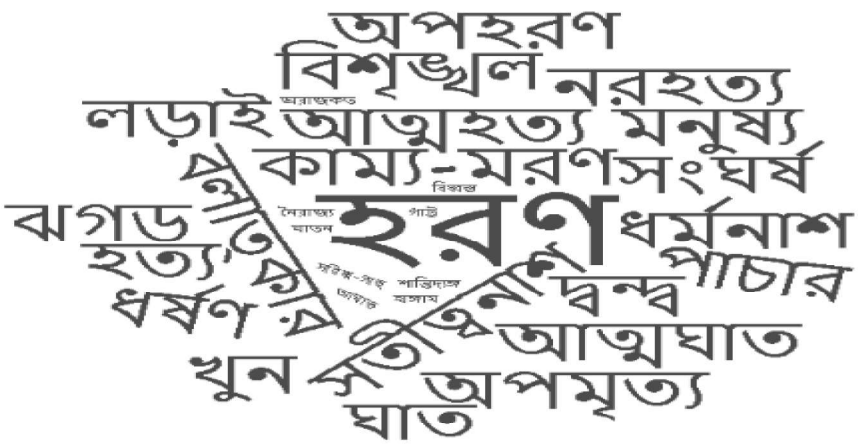


Figure 7: Triggering Word Cloud

3.2.1 Data Collection

At first for fulfilling the task we needed some organized and categorized news data but due to the lack of desired data we collected news from various online newspapers. We filled the dataset with news from all available online newspapers to make the dataset not to be biased with any specific one of them.

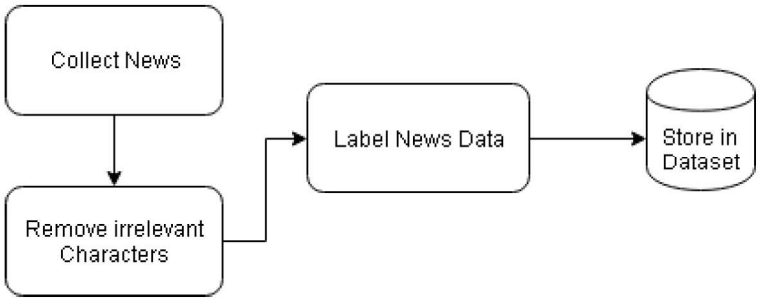


Figure 8: Data Collection process

The following (Figure 8: Data Collection process) is about the data collection process as we mentioned earlier we gathered news articles from various online

Bengali newspapers after collecting the news article's we removed some irrelevant character's and words that does not convey any important information. After getting the preprocessed article we had labeled each and every news by reading the whole content of the article which was later on stored in the training dataset along with its given label. Finally we managed to build a dataset containing the pair of raw news article and their corresponding label. We gathered about 600 news articles of six individual violence event each having 100 news respectively which gives our dataset 600 rows and 2 columns.

3.2.2 N-gram Selection

The most important task after collecting the news data was to extract the n-grams from news articles which included our triggering words. For doing so we maintained a window of words of size 7 which included preceding and following 3 words of the triggering words that makes the triggering word in the middle index of the window.



Figure 9: N-gram Extraction Window

The (Figure 9: N-gram Extraction Window) above is something what our word window looks like and we have extracted the n-grams from all the triggering word windows respectively. The extracted n-grams had a length range of 2 words to highest 5 words and the sequence of word's in the n-gram was similar to the actual news article it was taken form.

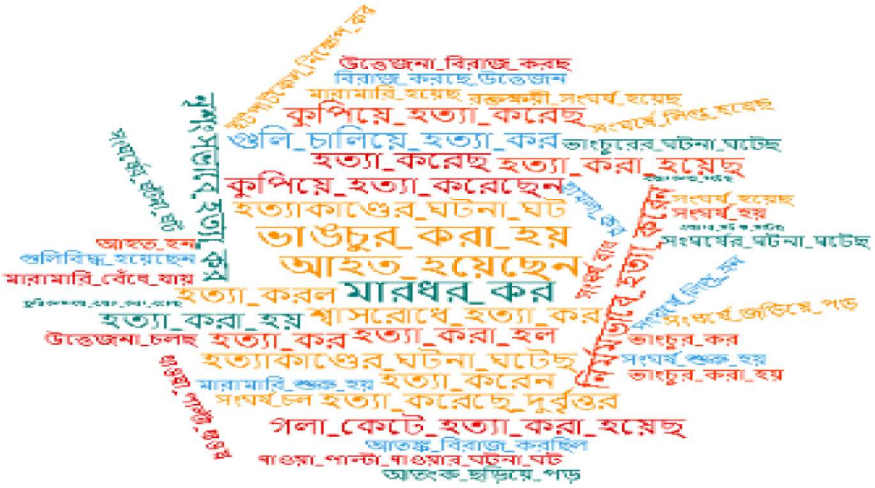


Figure 10: Extracted N-gram Cloud

The above is an n-gram cloud (Figure 10: Extracted N-gram Cloud) made from the extracted n-grams using the triggering word window. We managed to extract a bunch of n-grams from each and every triggering word individually which insures the occurrence of the events related with the triggering words. These n-grams might vary in lengths and they might have individual meanings but they share a co-relation and together these n-grams signify a particular violence event.

3.2.3 Dataset Representation

To fit a model on our training news article dataset we must represent the dataset in a way that conveys the existence of our extracted n-grams and triggering words. We proposed to use all the extracted n-grams and triggering words as feature column and place the news articles in rows as a binary vector signifying the existence of the extracted n-grams. The method works as below:

Algorithm 3.2.3

Dataset Representation

Input: News Article

```
For each news article:
    For each word in news article:
        If word in Triggering words:
            Find n-gram around word
            If true :=
                Set 1 to n-gram article index
            Else :=
                Set 0 to n-gram article index
        End if
    End for
End for
```

This method is designed to go through all the words that appear in the news article and find the occurrence of any triggering word further finding the feature n-grams through the detected triggering word. After going through all the article the dataset looks like something as below(Table 3: Binary Representation of N-gram's and Triggering Word):

News no	N-gram (1)	N-gram (2)	N-gram (3)	N-gram (n)	Label
0	0	1	0	1	1
1	1	0	1	0	2
2	0	0	0	0	0

Table 3: Binary Representation of N-gram's and Triggering Word

The dataset contains all the feature n-grams as column along with a hand labeled column that indicates the type of violence took place in the news articles marked as label or class column. We represented the news article as a large vector having length of the feature set and the components of the vector are binary to identify which n-gram or feature is subsistent in the news article normally marked by 1 and marked by 0 otherwise shown below in (Figure 11: News Article vector):

1	0	0	0	1	0	1	1	0	.....

Figure 11: News Article vector

3.2.4 Feature Set Reduction

As we have collected our feature n-grams by using an word window so it's quite natural that there might be some collected features that doesn't convey any important information that can help to achieve our desired goal so this kind of features are redundant and they increase the complexity of the process so it is better to get rid of this kind of not so useful features. Coming to a technique to identify this redundant features we used term frequency of this n-gram features where we compared the frequency of this individual features which was counted from various news articles after that we deducted the feature n-grams that had a very poor frequency eventually we got our reduced feature set. Below is an n-gram feature histogram that is sorted in decreasing order according to their term frequency in (Figure 12: N-gram Frequency Histogram):

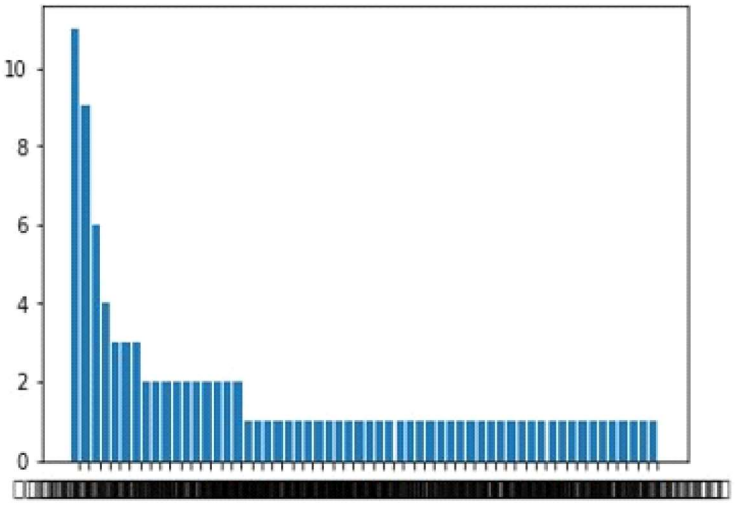


Figure 12: N-gram Frequency Histogram

Chapter 4

Bottom up Hierarchical (BUH) Classifier

In this experiment, our goal is to detect event with the recognition of exact violent news by their violence criteria's. After using many classifier algorithms we couldn't reach at our expectations. So we constructed a classification algorithm which prevents concept-drifting problem with better tf–idf evidence and utilizes a supervised learning technique. Our principal goal is to giving the proper importance for individual keyword network so that we can classify the event and recognize their criteria of violence. We have got very surprising results from this BUH classifier. In the beginning, we need to learn about the keyword networks. Here is our all keyword networks with all keywords.

Keyword Networks	Keywords									
Murder or "খুন"	খুন	নিহত	ঘাত	আঘাত	হত্যা	গুলি	চাকু	বন্দুক	পিস্তল	আগ্নেয়াস্ত
	ছুড়ি	অস্ত্র	সশস্ত্র	রক্তপাত	মার	মেরে	লাশ	মৃত	ঘাতক	পিটিয়ে
Kidnap or "অপহরণ"	অপহরণ	হরণ	তুলে	গুম	ক্ষিপ্র	জোর	পাচার	দখল	পীড়ন	নিপীড়ন
	শিকার	অত্যাচার	নির্যাতন	জুলুম	জবরদস্তি	বলাত্কার	নারীধর্ষণ	ধর্ষণ	ধর্ষিত	ধর্মনাশ
Hassle or "মারামারি"	মারামারি	হামলা	আহত	ধেয়ে	ধাওয়া	পাল্টা	আক্রমণ	হানা	ধোলাই	দ্বন্দ্ব
	চক্রান্ত	ষড়যন্ত্র	ঝগড়া	সংঘর্ষ	দাঙ্গা	লড়াই	কলহ	সহিংস	সংঘাত	শত্রুতা
Protest or "বিক্ষোভ"	বিক্ষোভ	সভা	সমাবেশ	হরতাল	ছত্রভঙ্গ	হল্লা	বিবাদ	বিরোধ	গোলমাল	প্রতিদ্বন্দ্বিতা
	সংগ্রাম্	আন্দোলন	মিছিল	অবরোধ	নাশকতা	ধর্মঘট	আলোড়ন	বিশৃঙ্খলা	ঝামেলা	প্রতিবাদ
Accident or "দূর্ঘটনা"	দূর্ঘটনা	ভিড়	আকস্মিক	দুর্দশা	সংকট	বাগিয়ে	কবল	এলোপাতাড়ি	ক্ষত	ক্ষয়
	বিপর্যয়	সর্বনাশ	আচমকা	হটাৎ	অপ্রত্যাশিত	দৃঢ়ভাবে	দুর্বিপাক	বিপাক	আপদ	আর্তনাদ
Terror or "আতঙ্ক"	আতঙ্ক	বিস্তার	হৈচৈ	ভয়	উত্তেজনা	রেশ	ক্রুদ্ধ	ক্রোধ	বিপদ	অশান্তি
	অস্থির	উপদ্রব	অসাধুতা	নিয়মভঙ্গ	উপদ্রব	প্রভাব	চক্র	চাঁচল্য	আশঙ্কা	নালিশ

Table 4: Keywords and the following network

So each keyword network holds twenty keywords (Table 4: Keywords and the following network). These keywords are trying to classify their corresponding networks which is helping to recognize the violence from the news data set by their criteria. We counted the presence of each keyword in two ways. One approach is binary appearance and other one is maximum appearance. So by the all consideration we should get three set appearance for individual news.

By the previous method we get a single row for identical news (Table 5**:** Weighted by total appearance of each keyword for their corresponding network). We have counted each keyword network importance with total appearance Here is the example of a news which holds the id as “11”.

news_id	murder	kidnap	hassle	protest	accident	terror
11	3	33	1	3	0	0

Table 5**:** Weighted by total appearance of each keyword for their corresponding network

In this system the weighting approach is not so strong for a classification algorithm. So we intend to give each keyword network a better rank by their corresponding keywords. In Details, for the requirement of the weight named as murder keyword network, we set the column murder as class variable and other networks as feature variables. Then we classified the characteristic for murder by a popular classification algorithm and set the training accuracy as the rank for the murder network. Same process occurs for each network for their proper ranking. Other networks also become the class variable when the ranking is needed. For individual news, class variable bends six times because we have six keyword networks.

In our system we ranked every attribute by two classification algorithm with two types of weighted keywords. We used Neural Network classifier which provides multilayer perceptron and we also used k-Nearest Neighbor classifier for both appearance approach of individual news, which is the binary appearance and the maximum appearance.

In the following table (Table 6: Binary presence of keywords for news id “11) we counted the keyword appearance as a Boolean variable. Such as if "খুন" exist in the news

then "খুন" keyword gets its importance 1 for murder network, If not then its importance is 0.

news_id	11	11	11	11	11	11		11	11	11	11	11	11	11	11	11	11	11	11	11
murder	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
kidnap	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	1	0	0
hassle	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
protest	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1
accident	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
terror	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Table 6: Binary presence of keywords for news id “11”

In the following table (Table 7: Maximum presence of keywords for news id “11) we counted the keyword appearance as much as it appeared in an individual news. Such as if "খুন" exist in the news then "খুন" keyword gets its importance the total amount it appeared in the news for murder network, If not then its importance is 0.

news_id	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11
murder	0	0	1	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
kidnap	0	0	0	0	0	0	0	0	0	0	5	0	19	0	0	0	0	9	0	0
hassle	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
protest	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	1
accident	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
terror	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Table 7: Maximum presence of keywords for news id “11”

Each news goes through this process and collects their rank for the corresponding keyword networks which exposes the importance of their violence criteria. After this process through two classifiers we achieve four types of importance for individual news. Two types of importance for Neural Network and K-Nearest Neighbor classifier attempts

to find the impact of each keyword network on the news. Now we can use any other classifier to acknowledge the violence of the news. In "Table 8**:** Weighted by neural network classifier with binary presence" and "Table 9:Weighted by neural network classifier with maximum presence" we have accordingly binary and maximum presence by neural network classifier and in "presence "and "Table 11: Weighted by KNN classifier with maximum presence" we have accordingly binary and maximum presence by K-NN classifier.

news_id	murder	kidnap	hassle	protest	accident	terror
11	1	1	0.05	1	1	0.1

Table 8**:** Weighted by neural network classifier with binary presence

news_id	murder	kidnap	hassle	protest	accident	terror
11	1	1	0.95	1	1	1

Table 9:Weighted by neural network classifier with maximum presence

news_id	murder	kidnap	hassle	protest	accident	terror
11	1	1	0.933	1	1	0.933

Table 10**:** Weighted by KNN classifier with binary presence

news_id	murder	kidnap	hassle	protest	accident	terror
11	1	1	0.933	1	1	0.933

Table 11: Weighted by KNN classifier with maximum presence

Algorithm: BUH Classifier

Input:data_set [Binary or Maximum appearance]

total_news = (data_set length)/20

For each news in the range of total_news

Do

news_tracking_id = get the id of the news from the data_set

For each 20 rows from the data_set by the news_tracking_id

Do

keyword_appearance_array= accumulate the appearance of the keyword for each news_tracking_id

End for

For each column in keywoard_appearance_array

Do

y_train = corresponding column as class variable

x_train = all columns except the class variable

clf = fit x_train and y_train into the classifier

Accuracy = get training score of clf by x_train and y_train of tracking_id

network_ranked_array = accomplish the accuracy as the rank of the class variable keyword network of the tracking_id

End for

End for

Write values of network_ranked_array into a comma-separated values (CSV) file.

Output: Table 7, 8, 9 and 10 type ranked network for their corresponding data_set.

Now we can fit this various ranked datasets into a classifier and observe the results in multilayer perceptron. In chapter 5 we exposed the results with comparing all approaches. Here we used decision tree for the detection of the event and recognized the violence, ranked by neural network and K-NN classifier. In this system R#1, R#2, R#3 and R#4 dataset stands for the solution of concept-drifting error which supports the decision tree classifier to detect the event.

Chapter 5

Results, Diagrams and Flowchart

5.1 Experimental Analysis of Multilayer Perceptron Model

5.1.1 Results

Here we meet the problem of detecting events from multiple and heterogeneous news. This heterogeneity makes the event detection task more challenging; hence we accomplished a very potential approach. We are able to automatically detect, and measure the violence weight when a new real world event has occurred, and also able to classify and cluster the news by any kind of events. In this module the textual similarity between the event keyword network and news words is measured potentially. This module can classify and cluster by any potential topics but the concernment must be switched to the corresponding event. We can also classify events into categories such as plane crashes, economic collapses and natural disasters. Some classifiers algorithm has been tested for better accuracy such as Naive Bayes Classifiers, k-Nearest Neighbors (k-NN) and Decision Tree. There are several steps of refinement which increases the exactness of the result of the classifiers. We can reduce the number of false positives by using heterogeneous classifier and cluster algorithm which is entitled as ensemble learning.

As we refine and improve this module we need to revise how we are calculating the importance of news. For example different production should have different weights such as if news contains more than one area we can increase the importance. The objective of event detection is to detect episodic related stories from a massive news collection. In information retrieval, tf–idf, short for term frequency–inverse document frequency, is a numerical statistic that is intended to reflect how important a word is to a document in a collection. It is mostly used as a weighting factor in searches of information retrieval, text mining and user modeling. The tf-idf value increases proportionally to the number of times a word appears in the document, and is offset by the frequency of the word in the corpus, which helps to adjust for the fact that some words appear more frequently in general. Here, tf-idf was applied to observe words with the purpose of conducting event matching as violence recognition.

Potential concept terms are the key terms which are united by the news. Concept terms can be used for dealing with the problems of lexicon altering, and accordingly the idea of using concepts has been applied for the quest of propagation. It combines the technique of global analysis and local feedback between quests. A major problem is ignored by most of the classification techniques, which is concept-evolution. That means the appearance of a novel class. In case of intrusion detection, a new kind of intrusion might go undetected by traditional classifier, but our approach should not only be able to detect the intrusion, but also deduce that it is a new kind of intrusion. This scheme would lead to an intense analysis of the intrusion by human experts in order to understand its cause, find a cure, and make the scheme more assured. The detection process can be done in unsupervised way, but supervision is necessary for classification. Without external supervision, two separate clusters could be regarded as two different classes although they are not. Conversely, if more than one novel classappears simultaneously, all of them could be regarded as a single novel class if the labels of those instances are never revealed. Furthermore, traditional novelty detection techniques simply identify data points as inconsistent that deviate from the normal class. But our scheme not only detects whether a single data point deviates from the existing classes, but also uncover whether a group of such outliers possess the potential of forming a new class by showing strong cohesion among themselves. Therefore, our scheme is a "multi-class" classification model and a novel class detection model.

In "Table 12: The discrimination of accuracy for all 2018 news" we can see the discrimination of all approaches with the following BUH classifier, which is performed only in the news of 2018. We have 2844 numbers of news following by their id and six keyword networks. We took 35% of our data as test set and the remaining 75% is training set. Here KNN classifier holds N=7 and NN classifiers holds iteration=1000.

We experimented this system in a different dataset to see the discrimination of all approaches with following BUH classifier. "Table 13: The discrimination of accuracy for human labeled dataset", which is demonstrating the same result as "Table 12: The discrimination of accuracy for all 2018 news" but performed in a human labeled dataset of various timeline. Ithas593 numbers of news following by their id and six keyword networks. We took 35% of our data as test set and the remaining 75% is training set. Here KNN classifier holds N=7 and NN classifiers holds iteration=1000.

Facts	Keyword network	Binary BUH of KNN	Maximum BUH of KNN	Binary BUH of neural network	Maximum BUH of neural network
Training set	0.542	0.653	0.619	0.635	0.638
Test set	0.485	0.661	0.238	0.621	0.613
misclassified news	418	275	619	308	496

Table 12: The discrimination of accuracy for all 2018 news

Facts	Keyword network	Binary BUH of KNN	Maximum BUH of KNN	Binary BUH of neural network	Maximum BUH of neural network
Training set	0.811	0.757	0.794	0.787	0.776
Test set	0.568	0.574	0.612	0.582	0.574
misclassified news	73	72	69	71	72

Table 13: The discrimination of accuracy for human labeled dataset

The auto-convolution with the autocorrelation provides a second-order description that discriminates between deterministic and stochastic signals, even those with equivalent spectrum. All classifiers and weighting approaches admit a multi-dimensional spectral representation that has unique and powerful properties, such as detecting deterministic event components in correlated stochastic noise with the knowledge of unification.

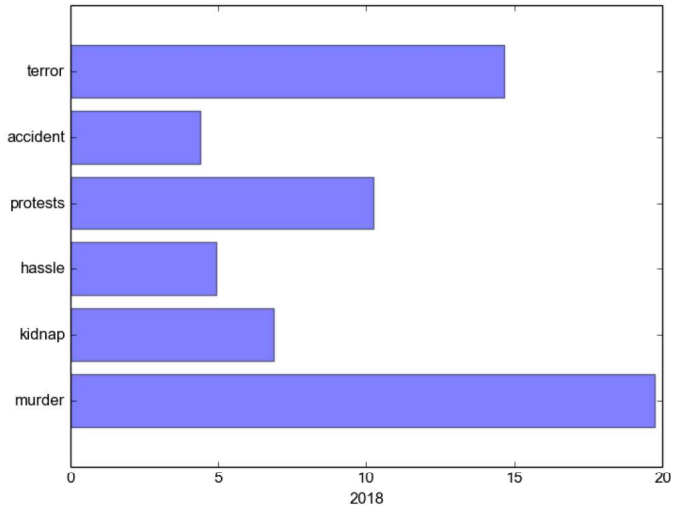
5.1.2 Graphs

Before we perform any analysis and come up with any assumptions about the distributions of the relationships between variables in our datasets, it is always a good idea to visualize our data in order to understand their properties and identify appropriate analytics techniques. Further, we should understand that basic statistic properties can often fail to capture real-world complexities such as outliers, relationships and complex distributions. We should choose the graph series which is actually useful to determine whether our dataset is biased or not and for better understanding of the intensity and periods of the dataset. Here, you will see the heuristic differences in conclusions that we can make based on (1) Simple bar chart, and (2) Principal component analysis (PCA).

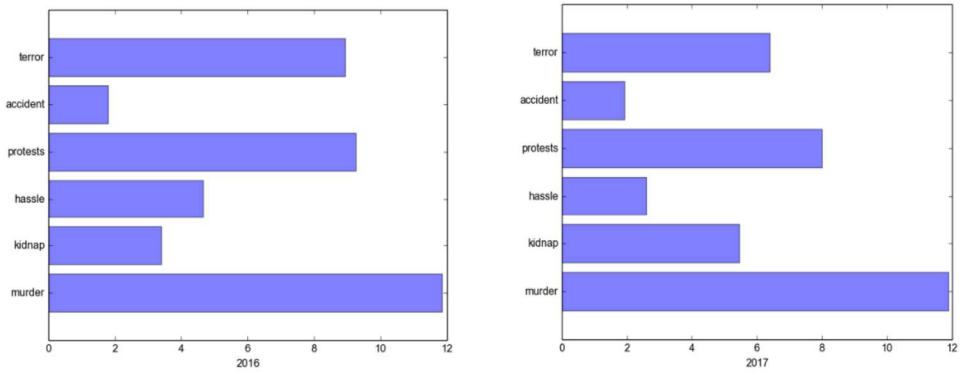
1) **Simple bar chart:**

Bar charts are a type of graph that are used to display and compare the number, frequency or other measurements such as mean for different discrete categories of data. Bar graphs are an extremely effective visual to use in presentations and reports. They are popular because they allow the reader to recognize patterns or trends far more easily than looking at a table of numerical data. Further, Bar graphs are an effective way to compare items between different groups. Bar charts are useful for displaying data that are classified into nominal or ordinal categories. Nominal data are categorized according to descriptive or qualitative information such as county of birth, or subject studied at university. Ordinal data are similar but the different categories can also be ranked, for example in a survey people may be asked to say whether they thought something was very poor, poor, fair, good or very good. However, this is not appropriate for ordinal data because the categories already have an obvious sequence. Bar charts are also useful for displaying data that include categories with negative values, because it is possible to position the bars below and above the x-axis.

In our bar chart you will see the mean percentage of each keyword networks. This bar chart makes it easy to compare the strength of each keyword networks indataset.

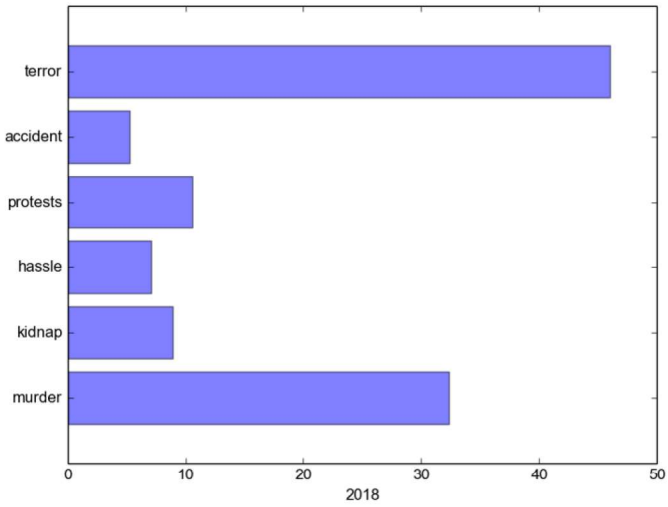


Graph 1: 2018 (January-May) violence performances in all area of Bangladesh

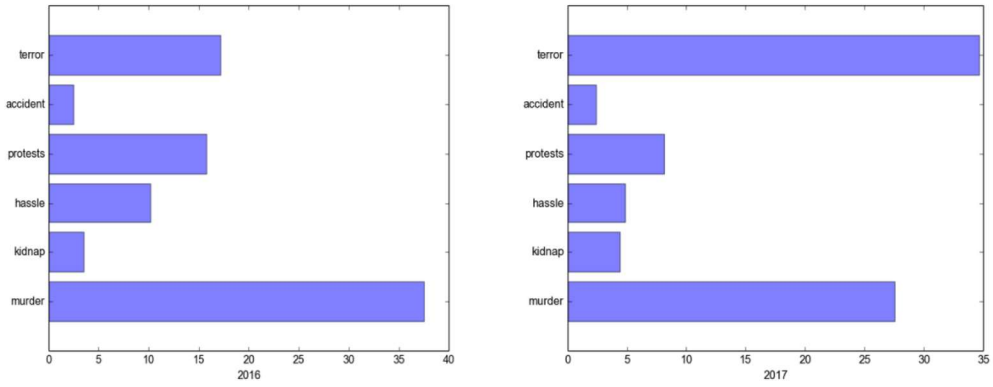


Graph 2: 2016 to 2017 violence performance in all area of Bangladesh

Graph 1: 2018 (January-May) violence performances in all area of Bangladesh and Graph 2: 2016 to 2017 violence performance in all area of Bangladesh represents the violence performance in percentage for all subdivision in Bangladesh.

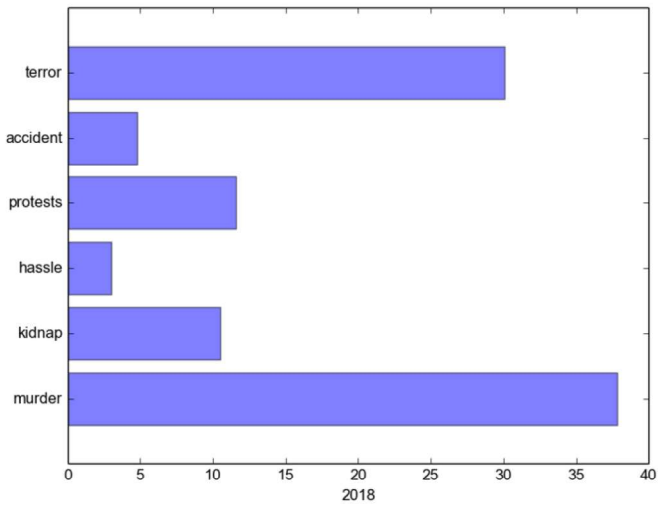


Graph 3: 2018 (January-May) violence performances in Dhaka

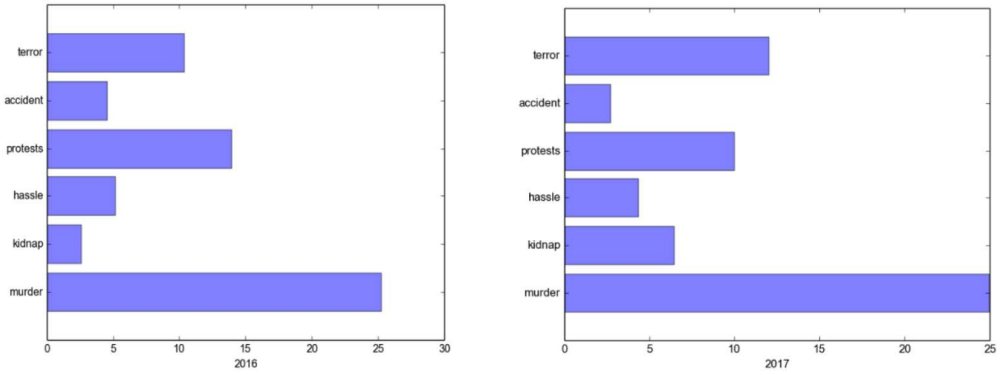


Graph 4: 2016 to 2017 violence performance in Dhaka

Graph 3: 2018 (January-May) violence performances in Dhaka and Graph 4: 2016 to 2017 violence performance in Dhaka represents the violence performance in percentage for Dhaka area.



Graph 5: 2018 (January-May) violence performances in Rajshahi



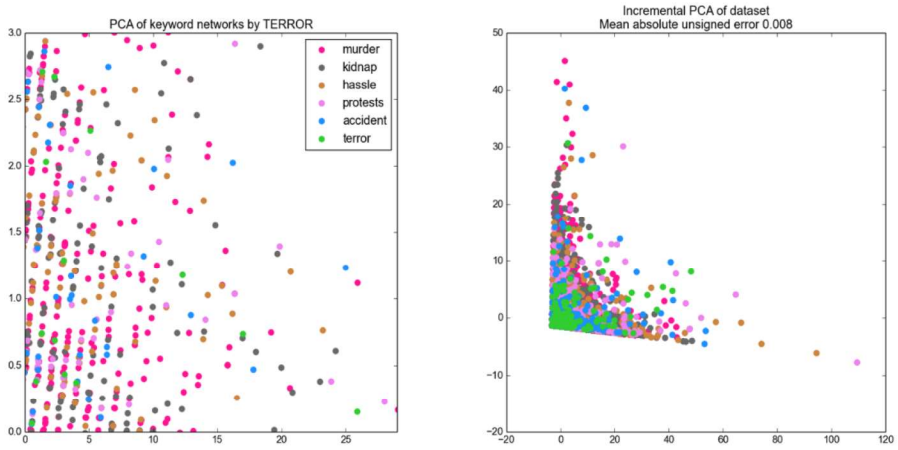
Graph 6: 2016 to 2017 violence performance in Rajshahi

Graph 5: 2018 (January-May) violence performances in Rajshahi and Graph 6: 2016 to 2017 violence performance in Rajshahi represents the violence performance in percentage for Rajshahi area.

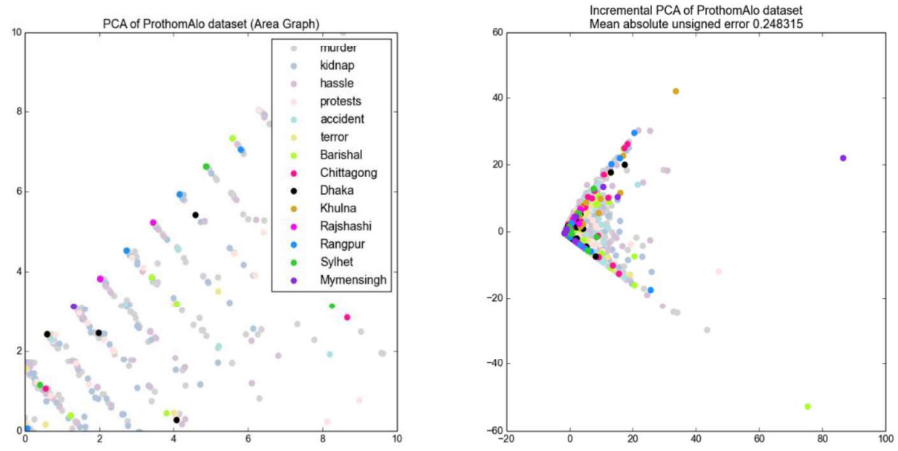
2) **Principal component analysis:**

PCA is most commonly used to condense the information contained in a large number of original variables into a smaller set of new composite dimensions, with a minimum loss of information.PCA should be used mainly for variables which are strongly correlated. If the relationship is weak between

variables, PCA does not work well to reduce data. Refer to the correlation matrix to determine. There are no other better options except PCA for multivariate analysis. PCA is for better perspective and less Complexity where having too many dimensions in featureswhich may hold different scales.



Graph 7: PCA of features where terror is the class



Graph 8: PCA of all features including area where terror is the class

5.1.3 Flowchart

A flowchart is worth a thousand words. Flowcharts are maps or graphical representations of a process. Steps in a process are shown with symbolic shapes, and the flow of the process is indicated with arrows connecting the symbols. Although there are

many symbols that can be used in flowcharts to represent different kinds of steps but the accurate flowcharts can be created using only a very few symbols.

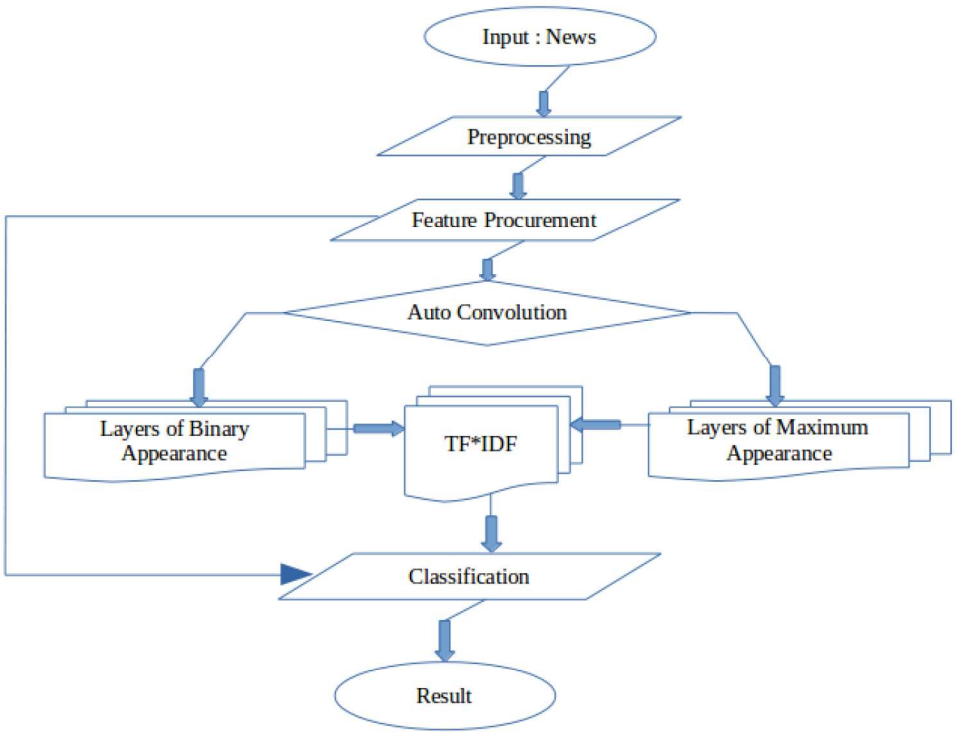


Figure 13: Graphical representation of multi-layer perceptron

5.2 Experimental Analysis of N-gram & Triggering Word Model

5.2.1 Results

Full Feature Set Outcome

After preparing the training dataset we have used several learning algorithms to fit a model on the training dataset and this is a multiclass classification problem where each class signifies its own violence category. To verify or evaluate the performance of the models we built a test dataset containing news of each violence category collected from those online Bengali newspapers which wasn't used in collecting the training dataset. We had labeled each and every news article in the test dataset indicating what kind of violence each news stand for just like the training dataset. After fitting the learning algorithms we used the models on the training dataset and the test

dataset for evaluation of the models and no other evaluation technique was used in the process. In the table 1 below performance comparison between different learning algorithms is shown through their accuracy on the training dataset and the test dataset as well and the accuracy shown on the table is achieved by tuning the best hyper parameter available. The algorithms are organized according to the increasing order accuracy on the test set because on the train set accuracy might increase due to over fitting in the (Table 14: Accuracy of Different Learning Algorithms in Train and Test Set Using Full Feature Set) shown below:

Algorithms	Train Set	Test Set
KNN	27.7	20.1
Random Forest	99.5	67.3
Decision Tree	99.5	67.7
SVM (Linear kernel)	98.5	68.6
Ada-Boost	72.7	69.3
Logistic-Regression	98.8	75.7

Table 14: Accuracy of Different Learning Algorithms in Train and Test Set Using Full Feature Set

2.0 Reduced Feature Set Outcome

After reducing the full feature set we conducted the same experiment on the train and test set using the same learning algorithms. Below is a (Table 15: Accuracy of Different Learning Algorithms in Train and Test Set Using Reduced Feature Set) of accuracy of different learning algorithms on train and test set:

Algorithms	Train Set	Test Set
KNN	64.4	48.9
Ada-Boost	81.9	76.4
SVM (Linear kernel)	98.3	78.2
Decision Tree	99.5	78.5
Random Forest	98.8	81.7
Logistic-Regression	95.3	83.4

Table 15: Accuracy of Different Learning Algorithms in Train and Test Set Using Reduced Feature Set

3.0 Full Feature Vs Reduced Feature Set on Train Set

Below is accuracy comparison (Table 16: Performance Comparison Table between Full & Reduced Feature Set on Train Set) between Full and Reduced Feature Set where we used both feature set on the same train set and most of the learning algorithms accuracy doesn't vary much but the main improvement occurred in KNN where the accuracy is much high in the Reduced Feature Set:

Algorithms	Full Feature Set	Reduced Feature Set
KNN	27.7	64.4
Ada-Boost	99.5	81.9
SVM (Linear kernel)	99.5	98.3
Decision Tree	98.5	99.5
Random Forest	72.7	98.8
Logistic-Regression	98.8	95.3

Table 16: Performance Comparison Table between Full & Reduced Feature Set on Train Set

4.0 Full Feature Vs Reduced Feature Set on Test Set

Below is accuracy comparison (Table 17: Performance Comparison Table between Full & Reduced Feature Set on Test Set) between Full and Reduced Feature Set where we used both feature set on the same test set and all of the learning algorithms Performed much better using the Reduced Feature Set rather than the Full Feature Set:

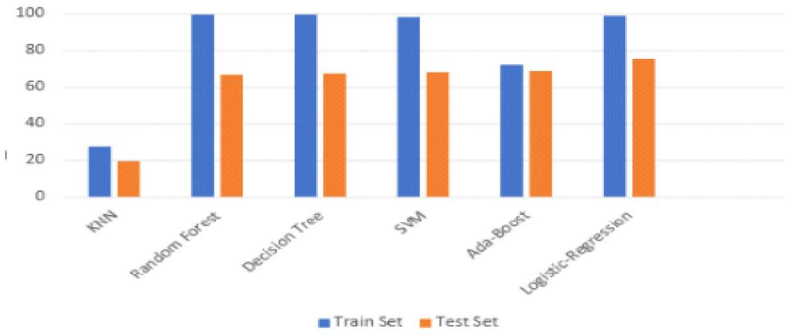
Algorithms	Full Feature Set	Reduced Feature Set
KNN	20.1	48.9
Ada-Boost	67.3	76.4
SVM (Linear kernel)	67.7	78.2
Decision Tree	68.6	78.5
Random Forest	69.3	81.7

Logistic-Regression	75.7	83.4

Table 17: Performance Comparison Table between Full & Reduced Feature Set on Test Set

5.2.2 Graphs

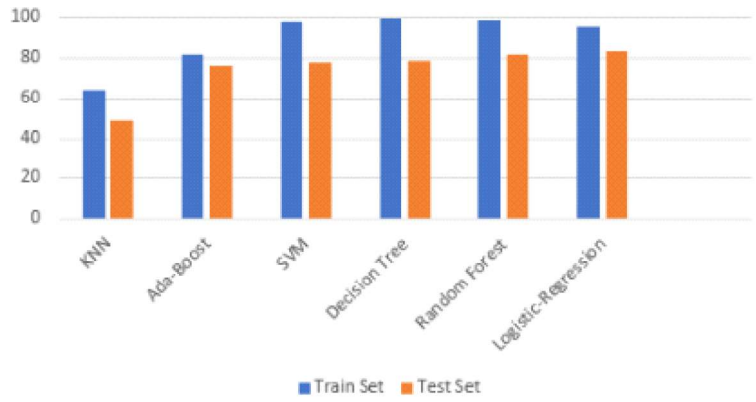
In (Table 14: Accuracy of Different Learning Algorithms in Train and Test Set Using Full Feature Set)results we can see that the accuracy achieved by different algorithms on the train set is quite a good figure but on the test set some algorithms performed very poorly and others had a very average score. Below is a learning algorithm accuracy comparison graph where we can see that KNN performed worse than rest of the learning algorithms in both train and test set on the other hand Random Forest and Decision tree makes a tie on the train set achieving highest accuracy among the algorithms and Logistic Regression out performed every learning algorithm used and if we look closely we can see that the accuracy of Logistic Regression on the train set is very closer to the highest achieved accuracy so we can say that Logistic Regression takes the first position among the learning algorithms in both train and test set shown in (Graph 9: Learning Algorithms Accuracy Comparison Graph in Train and Test Set Using Full Feature Set):



Graph 9: Learning Algorithms Accuracy Comparison Graph in Train and Test Set Using Full Feature Set

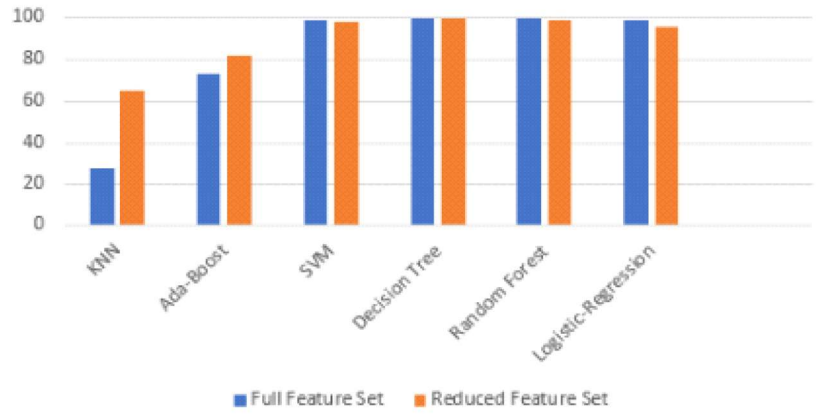
From (Table 15: Accuracy of Different Learning Algorithms in Train and Test Set Using Reduced Feature Set)we can see that almost all of the learning algorithms performed very well in the train set and as for the test set the performance is much higher than the full feature set results. Below we can see a algorithms performance comparison

graph where Decision Tree has the highest accuracy on the train set and on the test set Logistic Regression holds the highest accuracy to choose one we see that Logistic Regression performs better than any other algorithm both on the test and train set just like the full feature set outcome shown in (Graph 10: Learning Algorithms Accuracy Comparison Graph in Train and Test Set Using Reduced Feature Set):



Graph 10: Learning Algorithms Accuracy Comparison Graph in Train and Test Set Using Reduced Feature Set

As we can see from the (Table 16: Performance Comparison Table between Full & Reduced Feature Set on Train Set) result sections that the performance of the learning algorithms using full feature set and reduced feature set on the train set doesn't vary much in both feature selection techniques the almost all of the algorithms performed quite same so the reduction technique doesn't really improve anything or we can say that the models are getting over fitted with the train dataset so feature selection doesn't play a part. Below is a graph which shows the performance diversity in full and reduced feature set as mentioned above in both of the feature selection technique KNN achieved lowest accuracy and Logistic Regression remains at top of the heap among the learning algorithms as we can see in (Graph 11: Full Feature Set Vs. Reduced Feature Set Accuracy Graph inthe Train Set):



Graph 11: Full Feature Set Vs. Reduced Feature Set Accuracy Graph inthe Train Set

Now coming to the test set accuracy comparison in (Table 17: Performance Comparison Table between Full & Reduced Feature Set on Test Set) all of the learning algorithms performed far better using the reduced feature set other than the full feature set as it is quite natural that the full feature set may contain some features that are not essential and doesn't convey any important information so the model derived out of the full feature set is biased towards this kind of features in consequence making a poor performance. Just like the full feature set KNN had lowest and Logistic Regression had highest accuracy in both of the aspects only the performance increased in the reduced feature set exhibited in the below (Figure12: Full Feature Set Vs. Reduced Feature Set Accuracy Graph inthe Test Set):

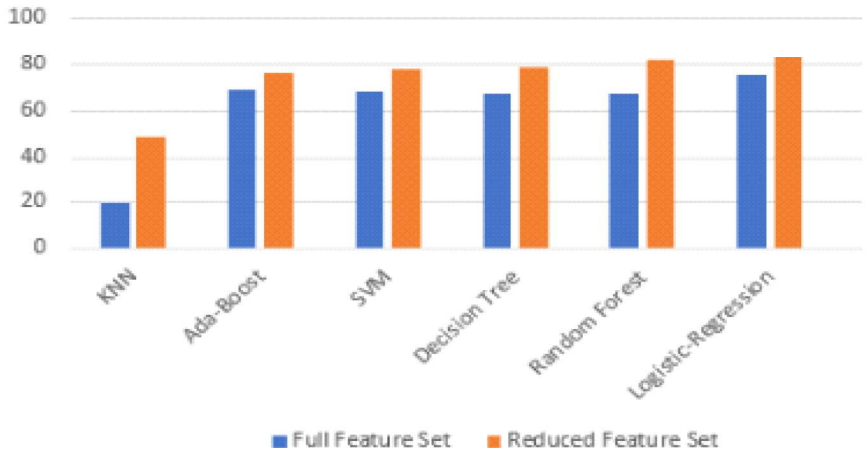
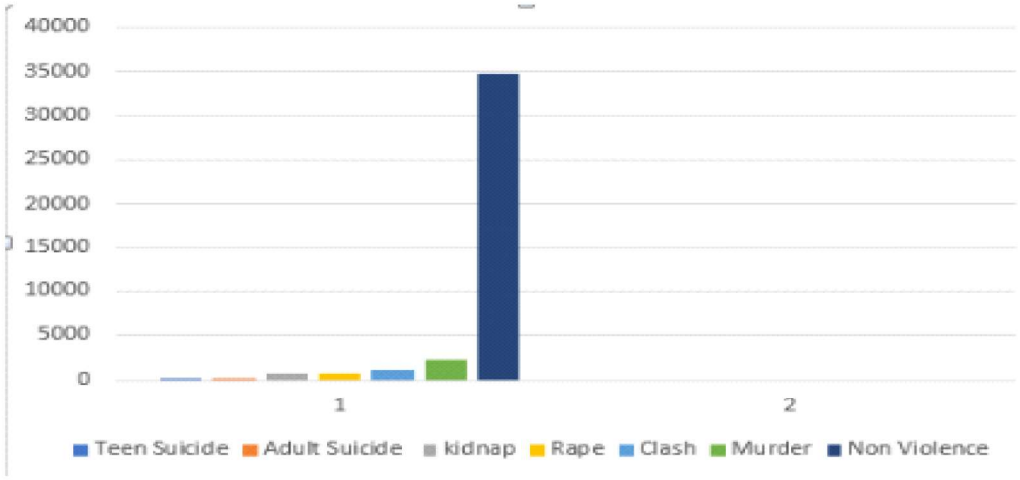


Figure12: Full Feature Set Vs. Reduced Feature Set Accuracy Graph inthe Test Set

5.2.3 Finalized Model on an Unlabeled Dataset

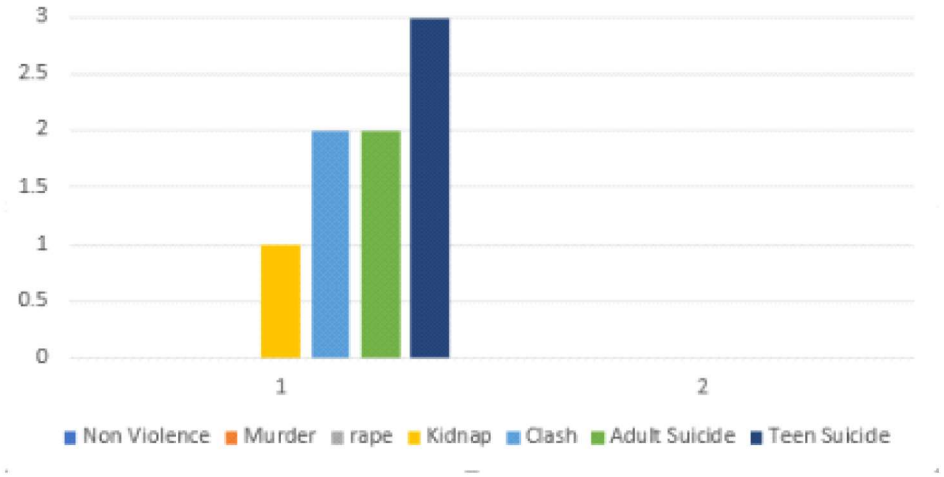
Comparing the results in both full and reduced feature set experiment we saw that Logistic Regression outperformed every learning algorithm so we finalized logistic regression as our final trained model which was derived out of the reduced feature set. Now we want to apply this model in a big unlabeled dataset to see how the model performs in an unseen and unlabeled dataset. The dataset consists of 40000 news articles that give our unlabeled dataset 40000 instances to test but before applying the model we preprocessed the dataset just like the training procedure. After cleaning we imparted the data to the Logistic Regression model and after observing the predicted result we saw that 34746 news articles wasn't about any violence occurrence we plotted an graph to show the amount of nonviolence news and the amount of different types of violence predicted from the unlabeled dataset. From the (Graph 13: Different Types of Violence Ratio Graph Predicted From the Unlabeled Dataset) below we can see that most of the news is about nonviolence and the most occurring violence is murder and the lowest type of violence is teen suicide predicted from the unlabeled dataset:



Graph 13: Different Types of Violence Ratio Graph Predicted From the Unlabeled Dataset

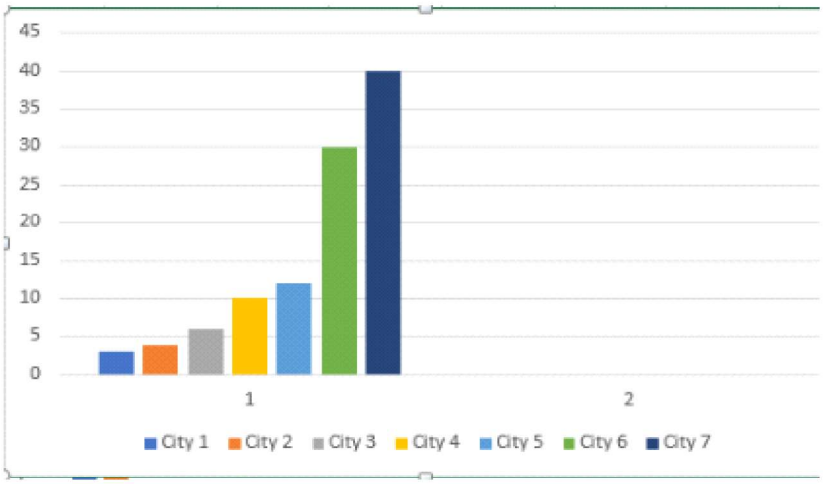
After witnessing the results we needed to conform that the outcome we got was valid so we conducted a test where we randomly picked some predicted news articles

and go through the whole article to verify that the predicted label is exact. We picked 50 random news articles of each class type and matched the articles with their respected predicted labels so after this validation test we observed that 341 out of the 350 articles were exact as their predicted label so the accuracy achieved is 97.4 percent. Below is an error (Graph 14: Error Rate Graph of Unlabeled Dataset in Each Class Type) that shows the error rates in each class type:



Graph 14: Error Rate Graph of Unlabeled Dataset in Each Class Type

Based on the prediction results we sorted different detected cities according to the rate of violence occurrence and this ratio is completely based on the model prediction so it might not be full authentic but we conducted some validation test as mentioned above so in reference to the test results that were found indicates this generated ratio is quite close to the original. Below is a (Graph 15: Crime Ratio Graph of Different Cities According to Prediction Results) of different cities sorted according to their crime occurring rates in increasing order:



Graph 15: Crime Ratio Graph of Different Cities According to Prediction Results

5.3 Auto-Convolution Vs N-gram Model

5.3.1 Percentage Split

After testing both of the proposed methods we wanted to compare this two model to see which model fulfills the task more efficiently in order to do that we tested both of the models with the same test set which had 600 labeled news articles. After conducting the test we found that the Auto-Convolution had 70.1 and the N-gram Model had 84.4 percent accuracy on the test set shown in the (Figure 14 : Accuracy Comparison between Auto-Convolution and N-gram Model Using Percentage Split)

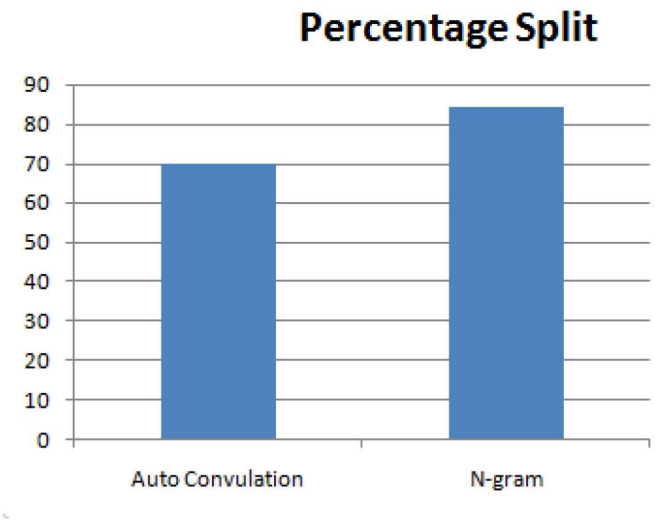


Figure 14 : Accuracy Comparison between Auto-Convolution and N-gram Model Using Percentage Split

5.3.2 Cross validation

After the percentage split test we again compared Auto-Convolution and N-gram Model in the same test set using cross validation accuracy measurement technique and again N-gram model performed better than the Auto-Convolution Model having accuracy 80.1 and 69.7 respectively shown in the (Figure 15: Accuracy Comparison between Auto-Convolution and N-gram Model Using Cross Validation):

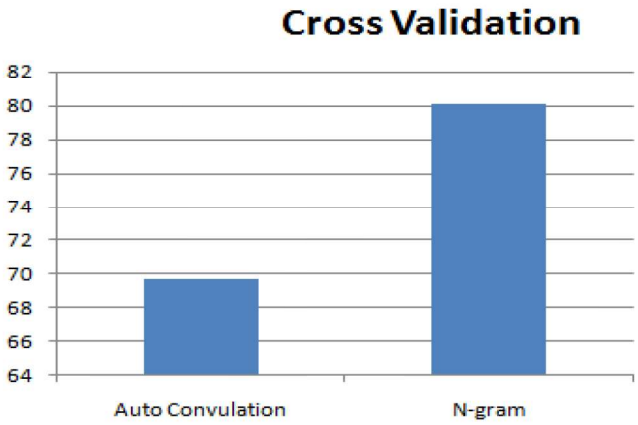


Figure 15: Accuracy Comparison between Auto-Convolution and N-gram Model Using Cross Validation

In both performance evaluation technique N-gram model outperformed Auto-Convolution model as we know that the Auto-Convolution model is trained on the basis of word frequency so it's natural that it will have a poor performance in comparison of the N-gram model which works with the context of any article but the performance of the Auto-Convolution model can be evaluated automatically through the N-gram model and with the increase of data the model's performance can also be enhanced.

Chapter 6

Conclusion

In our work we have proposed to method to detect any kind of violence event one is based on auto-convolution architecture and another is based on ann-gram model architecture. The auto-convolution method consists of both tf-idf and the n-gram model consisted only of tf and we developed two models based on these architectures collecting data from various online Bengali newspapers. Both of the models performed well in predicting an event occurring in any textual news or article and we managed to achieve almost 84 percent accuracy in detecting violence events. After the completion of developing an model we conducted an test which was about gathering one years news data which was unlabeled and from that data we identified different type of crime ratio of occurring throughout the entire year and after that we sorted different cities according to their safety level or crime occurring rates to validate the predictions we performed an test picking random news and matching those news with their labels. Finally we developed a web application that works based on the model we constructed that can classify any violence occurring from a textual Bengali news or article.

As for our future work for this topic we would like to add more category to detect almost all kind of events and observing all news through our final model to also detect if any new kind of events occurring and find relation among these events to predict any upcoming incident. We would also like to use our model in detecting graphic contents as we know that if any social media post includes any kind of graphic content the text along with the graphic content must have any kind of mention in the text. With the increase of daily textual data would like to gather more data to develop a model using neural network to increase the accuracy of our model to make the predictions more precise and authentic and also detecting more specific areas of crime occurring along with the time to detect the safety level of each area across the country.

References

[1] F. H. M. N. H. Foyzul Hassan, "Bangla phonetic feature table construction for automatic speech recognition," in *International Conference on Information and Communication Technology (ICCIT)*, Khulna, Bangladesh, 2014.

[2] J. M. P. a. W. B. Croft, "A language modeling approach to information retrieval," Melbourne, Australia, 1993.

[3] A.-M. A. M. A. S. K. M. Azharul Hasan, "Recognizing bangla grammar using predictive parser," *International Journal of Computer Science & Information Technology (IJCSIT) Vol 3, No 6,* 2011.

[4] N. L. a. D. D. Wu, "Using text mining and sentiment analysis for online forums hotspot detection," *Decision Support Systems,* 2009.

[5] M. M. A. Mohammed Moshiul Hoque, "Context-Sensitive phrase structure rule for structural".

[6] a. J. D. M. Z. R. Dewan Md. Farid, "Attribute weighting with adaptive nbtree for," *International Journal of Computer Science and Information Security,Vol. 8, No. 1,* 2010.

[7] K. D. W. Y. H. H. a. A. M. S. Dan Li, Detection, classification, and tracking of targets, *IEEE Signal Processign Magazine*, 2002.

[8] S. C. D. &. D. P. Mandic, "Autoconvolution and panorama: augmenting second-order signal," in *IEEE*, 2014.

[9] Mohammad M. Masud, Jing Gao, Latifur Khan, Jiawei Han, Bhavani Thuraisingham, "Classification and novel class detection in concept-drifting data streams under time constraints," *IEEE Transactions on Knowledge and Data Engineering, Vol. 23, NO. 6,* pp. 859-873, 2011.

[10] R. S. E. v. d. G. B. P. M. K. Alexandra Balahur, "Opinion mining on newspaper quotations," in *IEEE/WIC/ACM International Joint Conferences on Web Intelligence and Intelligent Agent Technologies*, 2009.

B. P. a. L. Lee, "Opinion mining and sentiment analysis," *Foundations and TrendsR in Information Retrieval Vol. 2, Nos. 1–2 (2008) 1–135,* 2008.

[12] V. J. H. &. J. AUSTIN, "A survey of outlier detection methodologies," *Artificial Intelligence Review 22: 85–126,* 2004.

[13] E. B. a. T. M. B. M. Boris Knyazev, "Recursive autoconvolution for unsupervised learning of convolutional neural networks," *Institut für Neuro- und Bioinformatik,,* 2017.

[14] P. R. &. C. W. Sajith Kecheril Sadanandan, "Automated training of deep convolutional neural networks for cell segmentation," in *Springer*, SciLifeLab, Up 0020psala, Sweden, 2017.

[15] S. Behnke, Hierarchical neural networks for image interpretation, California: Springer, 2003.

[16] A. T. P. P. Jay Gandhi, "Study on the different technique of concept drift and novel," *International Journal of Innovative and Emerging Research in Engineering, Volume 2, Issue 1,* 2015.

[17] *. H. M. L. M. K. L. W. J. C. H. Y. W. Lam, "Using contextual anaylsis for news," *International Journal of Intelligent Systems, Vol. 16, 525546,* 2001.

[18] S. S. Markos Markou∗, "Noveltydetection: a review—part 1: statistical approaches," *Signal Processing 83, 2481 – 2497,* 2003.

[19] S. S. Markos Markou∗, "Noveltydetection: a review—part 2: neural network based approaches," *Signal Processing 83,2499 – 2521,* 2003.

[20] S. V. Nath, "Crime pattern detection using data mining," in *Proceedings of the 2006 IEEE/WIC/ACM International Conference*, Florida, 2006.