**A Multilayer Perceptron for Event Detection and Violence Recognition from Textual News**

**Student name**

**Student Id: xxxxxxxxx**

**Student name**

**Student Id: xxxxxxxxx**

**Student name**

**Student Id: xxxxxxxxx**

**Student name**

**Student Id: xxxxxxxxx**

A project/thesis in the Department of Computer Science and Engineering presented

in partial fulfillment of the requirements for the Degree of

Bachelor/Master of Science in Computer Science and Engineering

****

United International University

Dhaka, Bangladesh

October 2018

©Imtiaz Mamun, 2018

**Declaration**

I/We, [Student Name (s)]\*, declare that this thesis titled, Thesis Title and the work presented in it are my own. I confirm that:

* This work was done wholly or mainly while in candidature for a [MSc/ BSc]\* degree at United International University.
* Where any part of this thesis has previously been submitted for a degree or any other qualification at United International University or any other institution, this has been clearly stated.
* Where I have consulted the published work of others, this is always clearly attributed.
* Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
* I have acknowledged all main sources of help.
* Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

[Name, Student ID and Department of the student]

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

[Name, Student ID and Department of the student]

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

[Name, Student ID and Department of the student]

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

[Name, Student ID and Department of the student]

**Certificate**

I do hereby declare that the research works embodied in this thesis/project entitled “**Thesis or Project Title**” is the outcome of an original work carried out by [Student Name (s)]\* under my supervision.

I further certify that the dissertation meets the requirements and the standard for the degree of [MSc/ BSc] in Computer Science and Engineering.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

[Name and designation of Supervisor]

Abstract

An unprecedented way is accomplished by using concept words derived from statistical context analysis between sentences which is better than traditional methods that uses 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 backpropagation.

Acknowledgement

Table of Contents

[LIST OF TABLES vii](#_Toc528803073)

[LIST OF FIGURES viii](#_Toc528803074)

1. [Introduction 1](#_Toc528803075)
2. [Background and Literature Review 2](#_Toc528803076)
3. [Materials and Methods 5](#_Toc528803077)

[3.1 Data Collection 5](#_Toc528803078)

[Algorithm 3.1 5](#_Toc528803079)

[3.2 Preprocessing 6](#_Toc528803080)

[Algorithm 3.2 6](#_Toc528803081)

[3.3 Feature procurement 7](#_Toc528803082)

[3.3.1 Murder (murder and attempt to murder) 8](#_Toc528803083)

[3.3.2 Kidnap (abduction and rape) 9](#_Toc528803084)

[3.3.2 Hassle (rioting and affray) 10](#_Toc528803085)

[3.3.2 Protest (unlawful assembly) 10](#_Toc528803086)

[3.3.2 Accident (accident and Injuring or defiling) 11](#_Toc528803087)

[3.3.2 Terror (persons concerned in criminal act) 11](#_Toc528803088)

[Algorithm 3.3 12](#_Toc528803089)

1. [Bottom up Hierarchical (BUH) Classifier 14](#_Toc528803090)
2. [Results, Diagrams and Flowchart 19](#_Toc528803091)

[5.1 Results 19](#_Toc528803092)

[5.2 Diagrams 22](#_Toc528803093)

[5.3 Flowchart 27](#_Toc528803094)

1. [Conclusion 28](#_Toc528803095)
2. [References 29](#_Toc528803096)
3. [Appendix A 31](#_Toc528803097)

## LIST OF TABLES

[Table 1: A data point after the collection of the news from star\_date to stop\_date 6](#_Toc528803098)

[Table 2: After procurement of the feature we isolated the dataset into two tables 13](#_Toc528803099)

[Table 3: Keywords and the following network 14](#_Toc528803100)

[Table 4**:** Weighted by total appearance of each keyword for their corresponding network 15](#_Toc528803101)

[Table 5: Binary presence of keywords for news id “11” 16](#_Toc528803102)

[Table 6: Maximum presence of keywords for news id “11” 16](#_Toc528803103)

[Table 7**:** Weighted by neural network classifier with binary presence 17](#_Toc528803104)

[Table 8: Weighted by neural network classifier with maximum presence 17](#_Toc528803105)

[Table 9**:** Weighted by KNN classifier with binary presence 17](#_Toc528803106)

[Table 10: Weighted by KNN classifier with maximum presence 17](#_Toc528803107)

[Table 11: The discrimination of accuracy for all 2018 news 21](#_Toc528803108)

[Table 12: The discrimination of accuracy for human labeled dataset 21](#_Toc528803109)

## LIST OF FIGURES

[Figure 1: Graphical representation of multi-layer perceptron 27](#_Toc528803110)

Chapter1

# Introduction

Violence are unfortunately showing an alarming increase. We need to recognize a 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 user friendly and dynamic.

Chapter 2

# Background and Literature Review

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: “ami bhat khai”, and “ami khai bhat” 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 “Karim valo chele” 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.

An 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 class. 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 were 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].

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

Chapter 3

# Materials and Methods

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 Data Collection

In Bangladesh there is a daily newspaper named Prothom Alo 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

**Collection of the news**

**Input:** start\_date, stop\_date

Declaration of news\_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 | উখিয়ার অনিবন্ধিত শিবিরের পাশে নতুন বসতি | স্ত্রী নুর জাহান ও আট সন্তানকে নিয়ে দুই মাস ধরে রাখাইনের বুড়াশিকদার পাড়ায় খালের ওপর মাচা বেঁধে দিন যাপন করছিলেন মোহাম্মদ হোসেন। কিন্তু গত তিন-চার দিনে ওই এলাকায় মিয়ানমারের সশস্ত্র বাহিনী ও নাডালা বাহিনীর নতুন অভিযান শুরুর পর প্রাণ বাঁচাতে তাঁরা বাংলাদেশে এসেছেন।………… |

Table 1: A data point after the collection of the news from star\_date to stop\_date

## 3.2 Preprocessing

After collecting the news we had to preprocess our data. In preprocessing we removed all English symbolic letters, and some Bengali characters (ex: ঁ, ঃ, ৡ etc.), and all the numbers. We also removed some intimations, 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 3.2

**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©⅐⅑⅒⅓⅔⅕⅖⅗⅘⅙⅚⅛⅜⅝⅞⅟↉¤¿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\_chars or 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

## 3.3 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: ঁ, ঃ, ৡ etc.), and all the numbers. We also removed some intimations, 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 keyword 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 measures 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", "দক্ষিণ সুরমা", "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.

### 3.3.1 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.*

### 3.3.2 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.*

### 3.3.2 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.*

### 3.3.2 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.*

### 3.3.2 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.*

### 3.3.2 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 3.3

**Balancing the weight of a news for each keyword network**

**Input:**

1. data\_set
2. 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 | rangpur | 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

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 3: Keywords and the following network

So each keyword network holds twenty keywords (Table 3: Keywords and the following network). This 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 an individual news.

By the previous method we get a single row for an identical news (Table 4**:** 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 4**:** 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 an 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 an individual news, which is the binary appearance and the maximum appearance.

In the following table (Table 5: 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 | 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 5: Binary presence of keywords for news id “11”

In the following table (Table 6: 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 6: 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 an 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 7**:** Weighted by neural network classifier with binary presence” and “Table 8: Weighted by neural network classifier with maximum presence” we have accordingly binary and maximum presence by neural network classifier and in “Table 9**:** Weighted by KNN classifier with binary presence” and “Table 10: 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 7**:** 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 8: 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 9**:** 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 10: 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 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 is 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 a news. For example different production should have different weights such as if a 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 is 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 classes appear 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 11: 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 12: The discrimination of accuracy for human labeled dataset”, which is demonstrating the same result as “Table 11: The discrimination of accuracy for all 2018 news” but performed in a human labeled dataset of various timeline. It has 593 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 11: 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 12: 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.

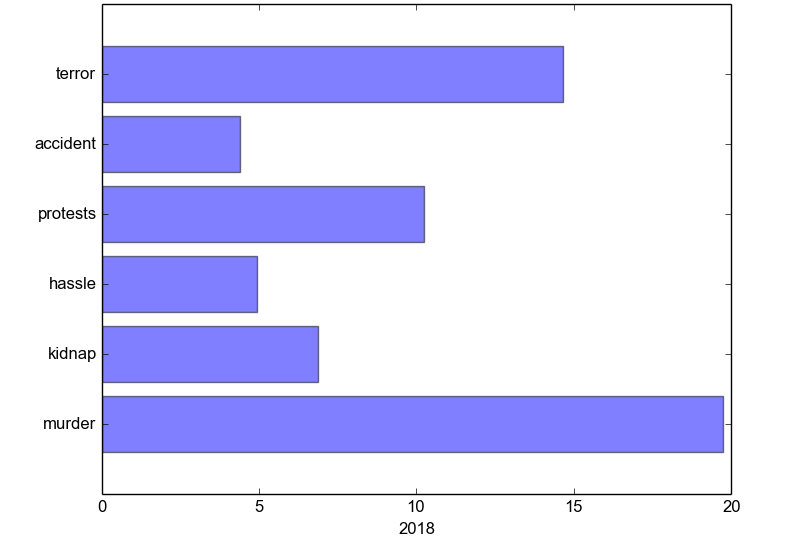
## 5.2 Diagrams

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).

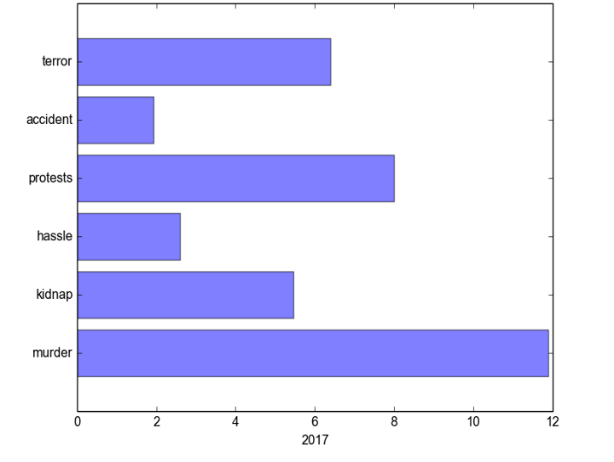
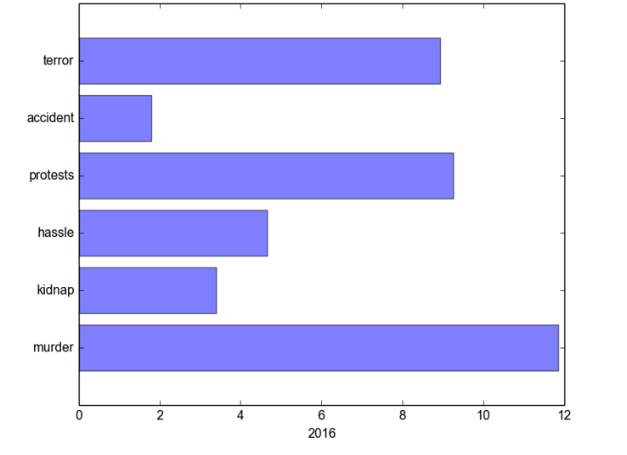
1. **Simple bar chart:**

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 in dataset.

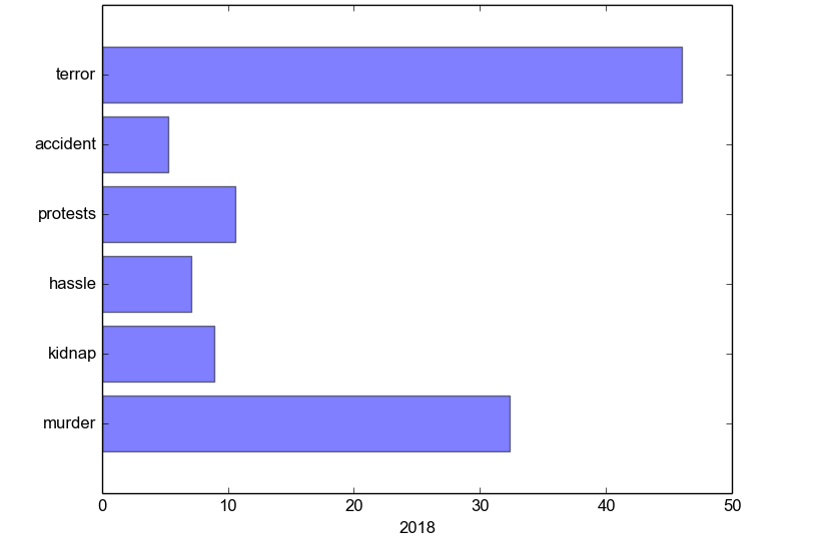


Graph 1: 2018 (January-May) violence performance in all area of Bangladesh

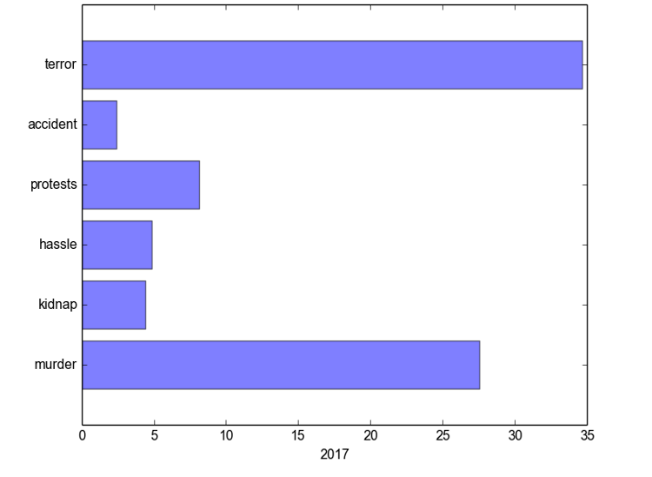
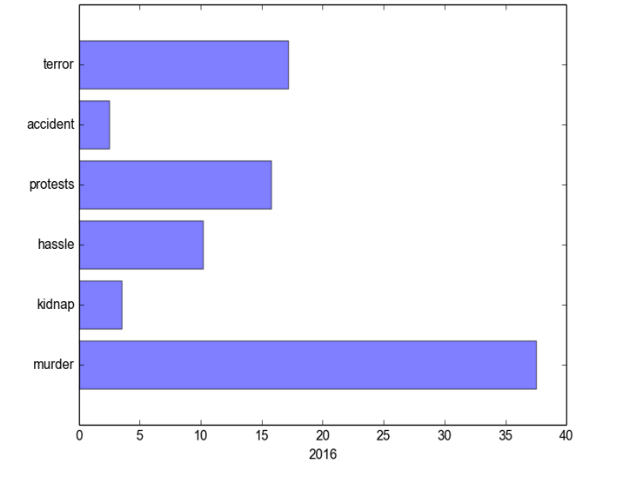


Graph 2: 2016 to 2017 violence performance in all area of Bangladesh

Graph 1: 2018 (January-May) violence performance 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.

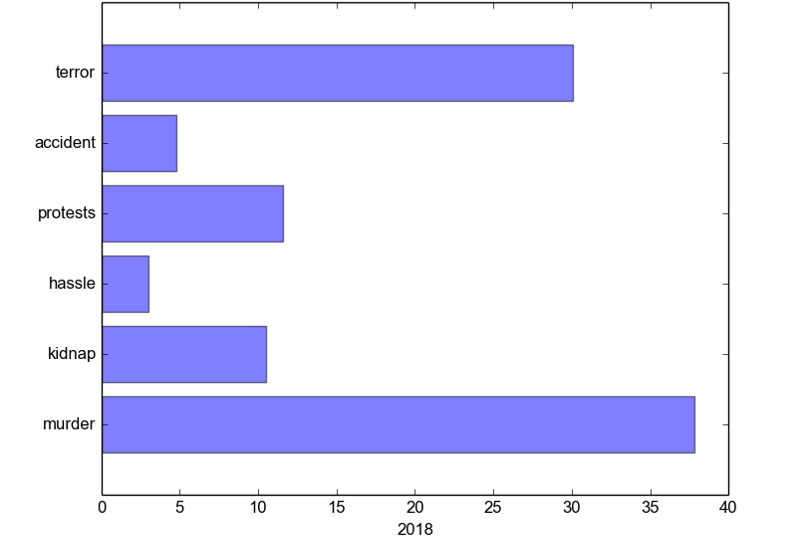


Graph 3: 2018 (January-May) violence performance in Dhaka

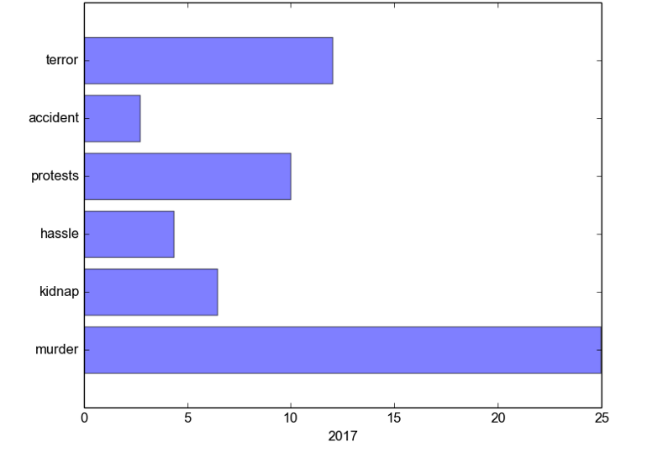
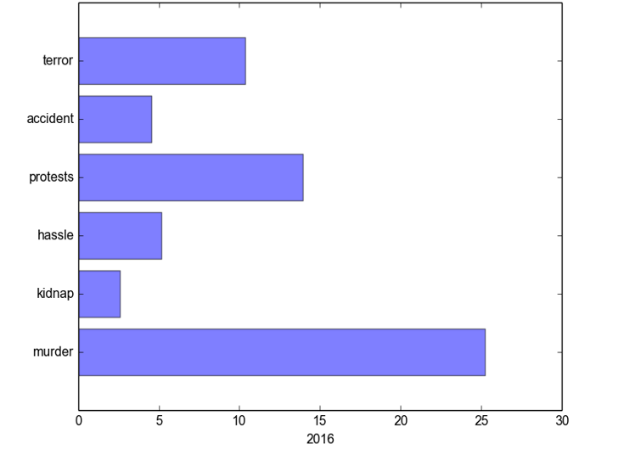


Graph 4: 2016 to 2017 violence performance in Dhaka

Graph 3: 2018 (January-May) violence performance 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 performance in Rajshahi

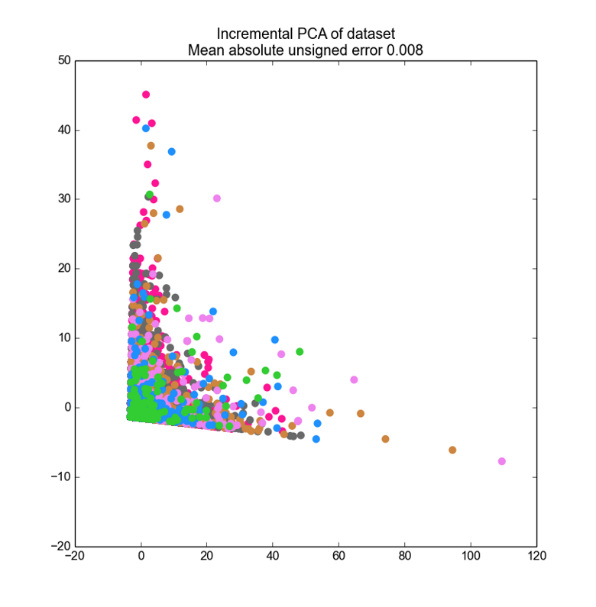
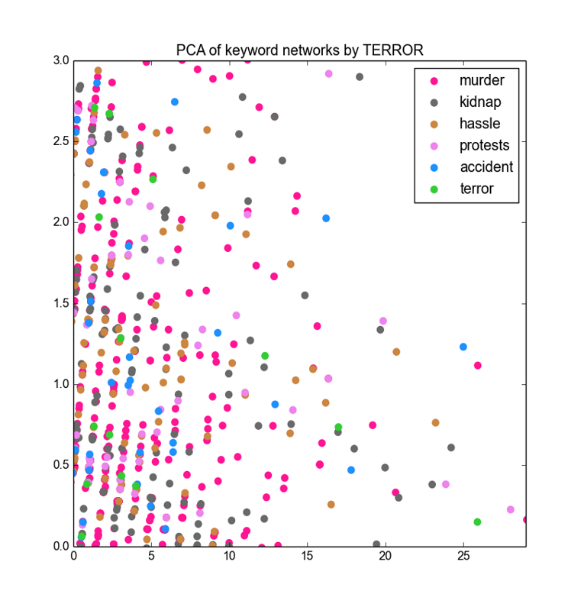


Graph 6: 2016 to 2017 violence performance in Rajshahi

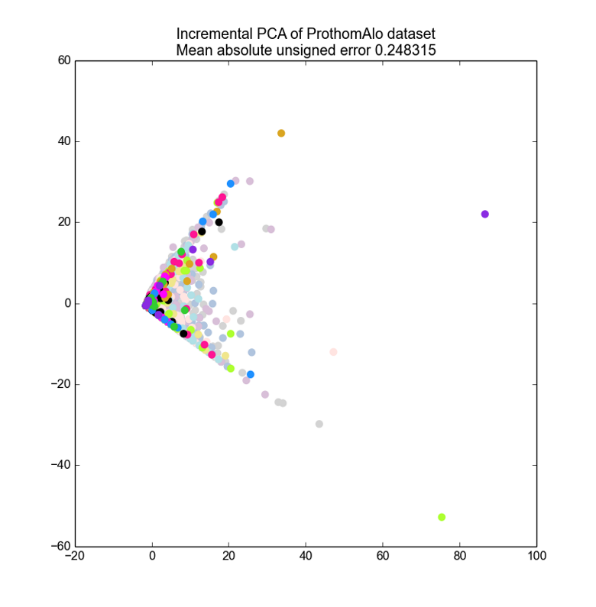
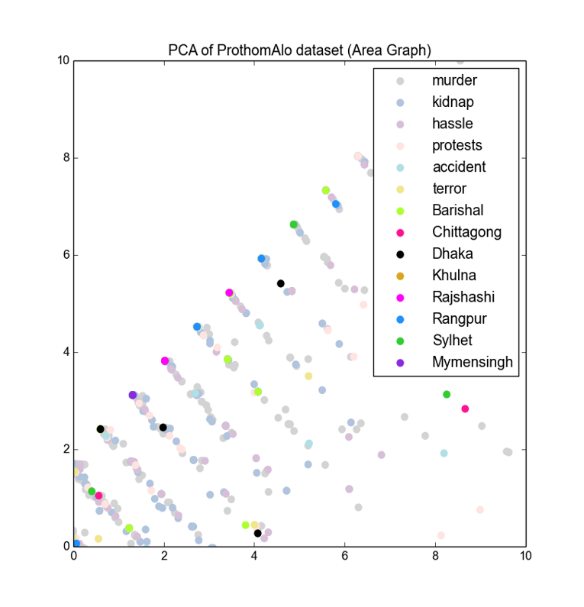
Graph 5: 2018 (January-May) violence performance in Rajshahi and Graph 6: 2016 to 2017 violence performance in Rajshahi represents the violence performance in percentage for Rajshahi area.

1. **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 is no other better options except PCA for multivariate analysis. PCA is for better perspective and less Complexity where having too many dimensions in features which 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

## 5.3 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.

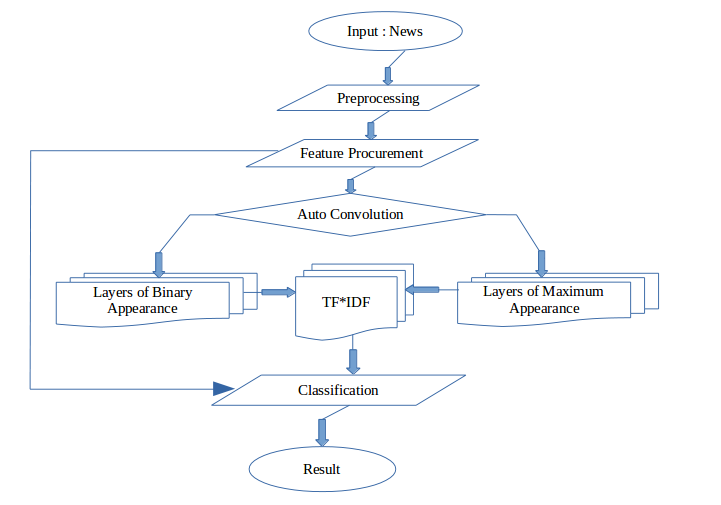


Figure 1: Graphical representation of multi-layer perceptron

Chapter 6

# Conclusion

This chapter could also be called “Conclusions and Recommendations” or “Conclusions and Implications.” In general, there should be no new information presented here. It should be a syn of information that you’ve already discussed.

# References

|  |  |
| --- | --- |
| [1] | F. H. M. N. H. Foyzul Hassan, "Bangla phonetic feature table construction for automatic speech recognition," in *IEEE*, 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, BARCLAY: IEEE SIGNAL PROCESSING 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. |
| [11] | 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, Uppsala, 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 Analysis 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. |

# Appendix A

Type or paste your appendices here. Appendices are a place to organize and include all of the “extra” material that is important to your research work but that is too detailed for the main text. Examples can include: specific analytical methods, computer code, spreadsheets of data, details of statistical analyses, etc. But, these materials do not speak for themselves. There should be a reference to these materials from the main chapters (complete details included in Appendix A) and there should be some text at the beginning of each appendix to briefly explain what the information is and means that is included in that appendix.