NATURAL DISASTER TRACKING IN CONTEXT OF BANGLADESH USING DATA MINING

BY

Md. Nafiur Rahman Hridoy
ID: 192-25-785

This Report Presented in Partial Fulfillment of the Requirements for the Degree of Master of Science in Computer Science and Engineering

Supervised By

Mr. Shah Md Tanvir Siddiquee
Assistant Professor
Department of CSE
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY
DHAKA, BANGLADESH

APPROVAL

This Thesis titled **"Natural disaster tracking in context of Bangladesh using data mining"**, submitted by Md. Nafiur Rahman Hridoy to the Department of Computer Science and Engineering, Daffodil International University, has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of M.Sc in Computer Science and Engineering (MSc) and approved as to its style and contents. The presentation has been held on 9th July, 2020.

<u>BOARD OF EXAMINERS</u>
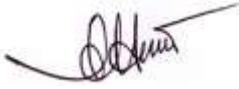
Dr. Syed Akhter Hossain   **Chairman**
Professor and Head
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Gazi Zahirul Islam   **Internal Examiner**
Assistant Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University
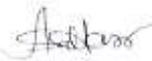
Abdus Sattar   **Internal Examiner**
Assistant Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Dr. Md. Saddam Hossain   **External Examiner**
Assistant Professor
Department of Computer Science and Engineering
United International University

©Daffodil International University   i

DECLARATION

I hereby declare that, this project has been done by me under the supervision **of Mr. Shah Md Tanvir Siddiquee, Assistant Professor, Department of CSE** Daffodil International University. I also declare that neither this thesis nor any part of this Thesis has been submitted elsewhere for award of any degree or diploma.

Supervised by:
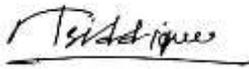
Mr. Shah Md Tanvir Siddiquee
Assistant Professor
Department of CSE
Daffodil International University

Submitted by:

Md. Nafiur Rahman Hridoy
ID: 192-25-785
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

First I express my heartiest thanks and gratefulness to almighty God for His divine blessing makes us possible to complete the final year project/internship successfully.

I am really grateful and wish my profound my indebtedness to **Mr. Shah Md Tanvir Siddiquee, Assistant Professor**, Department of Computer Science and Engineering Daffodil International University, Dhaka. Deep Knowledge & keen interest of my supervisor in the field of *"Data Mining"* to carry out this Thesis. His endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior draft and correcting them at all stage have made it possible to complete this project.

I would like to express my heartiest gratitude to **Dr. Sayed Akhter Hossain**, **Professor & Head**, Department of CSE, for his kind help to finish this project and also to other faculty members of CSE department, Daffodil International University.

Finally, I must acknowledge with due respect the continues support and patients of my parents.

ABSTRACT

The Title of this thesis is "Natural disaster tracking in context of Bangladesh using data mining". The life of people of Bangladesh has often been shaped by tragedy. We know that Bangladesh is a land of natural disaster. Natural disaster is a common problem of Bangladesh. Cyclones, floods, river erosion, storm, thundering are common phenomena for us. A lots of peoples, animal are injure and die for attacking those disasters. Huge number of properties such as home, road, bridge, education institutes are damage and destroy reason of natural disasters. Natural disaster give us a big number of loses value. Those are directly hit to economic point of our country. So that in this research work I collected last year data using web scraping algorithm then implement this data set using different type of Algorithm and prediction result are given by this system. This result are helpful for preparing purpose of upcoming natural disasters .By using this system our government take a necessary step before upcoming year natural disaster.

TABLE OF CONTENTS

CONTENTS **PAGE**

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 Introduction

Bangladesh is a natural disasters prone country. It has a land area of 147,570 square kilometers (56,980 sq. mi) with population nearing 164.7 million. In our country most of the land area are affected by natural disasters. Every year some of natural disasters such as floods, cyclones, river erosion, storm, thundering destroy the life and property every year. Those disasters the most familiar problem in Bangladesh.

1.2 Motivation

Bangladesh is land of natural disasters. Most of the time we don't know how many people, property and how much loses value that we losing in every year by attacking natural disasters. Reason of those issues I build a system for knowing about those information for betterment of country peoples.

1.3 Rationale of the study

In our country most of the time affected by natural disasters. Many people are victim of those disasters. They are loses their property and destroys their crops land. This is the big issue of our country which one directly affected to economic growth of our country. Rationale of the study of the research to find prediction of damaging rate of disasters.

1.4 Research Questions

Research questions of this topic is why I research about it? And how implement it in data mining? I want to build a system which is it track a natural disaster via using a data mining. I use different type of prediction algorithm and classifier algorithm for get a best output for this research. There were no research about natural disaster tracking in context of Bangladesh. This system really give a best result and also accurate result.

By using this system our government take a necessary step before upcoming year natural disaster. That is the reason are research and implementation about this topic.

1.5 Objectives

- ➢ Getting an information from newspaper for create a dataset by using web scraping.
- ➢ Organize dataset to target structural format from raw dataset.
- ➢ Use a prediction algorithm for data analysis.
- ➢ Finally getting a most affected disasters list, Affected area by specific disasters , Disaster affecting information and most affected disaster in specific place will performed by the proposed statistical model.

1.6 Expected Output

By using this system our government and many Non-government organization (NGO) are easily to know information about natural disasters and take a perfect step for peoples. They able to know what is loses value of any kind of disasters. They can find a most affected place by dataset using those algorithm.

1.7 Report Layout

Chapter 1: In this chapter I described about the motivation, objectives and expected outcome of this research.

Chapter 2: This chapter I described the background study related works which is similar works with others works research. I also provide a short summary about research and mention a challenges that I faced for researching about this concept.

Chapter 3: This chapter contains the Research Methodology. It also contains research subject and instrumentation, Data collection whole procedure, Statistical analysis of Algorithm and Implementation tools and requirement.

Chapter 4: In this chapter Discuss about experimental Result and Descriptive analysis of this research.

Chapter 5: In this chapter I described about conclusion and future work about this research.

CHAPTER 2

BACKGROUND

2.1 Introduction

This chapter describes about background of **Natural disaster tracking in context of Bangladesh using data mining** system. In this chapter I mention about study for this research and discusses about some similar related works. I analysis their work to finding new concept and scopes.

The life of people of Bangladesh has often been shaped by tragedy. We know that Bangladesh is a land of natural disaster. Natural disaster is a common problem of Bangladesh. Cyclones, floods, river erosion, storm, thundering are common phenomena for us. A lots of peoples, animal are injure and die for attacking those disasters. Huge number of properties such as home, road, bridge, education institutes are damage and destroy reason of natural disasters. Natural disaster give us a big number of loses value. Those are directly hit to economic point of our country.

So the concept of this research is tracking disaster to know about most affected disaster, most affected place of Bangladesh and to top affected place by this system from big dataset. This dataset are contain natural disaster data of last five year. Our government and Non-government organization are use this system they are perfectly know about disaster destroy rate and take a proper step for victim peoples.

2.2 Related Works

In this part I provided an information about similar works with this research.

1. A review on application of data mining techniques to combat natural disaster: This research are based on natural disaster. They use different type of techniques for their required output. This paper published by Saptarsi Goswami, Sanjay Chakraborty, Sanhita Ghosh, Amlan Chakrabarti, Basabi Charkarborty [1].

2. Damage tracker a cloud and mobile system for collecting damage information after natural disasters: The main purpose of this research create system mainly provide a tornadoes

information. They collect a real-time information of tornadoes. Then based on those data application give a output. This paper published by Chris Hodaap, Matt Robbins, Leff Gray and Andrew Graettinger [2].

Mentioned researches influence me for think about my concept. And those worked are really helpful for build my system of "Natural disaster tracking in context of Bangladesh using data mining".

2.3 Research summary

Bangladesh is a natural disaster prone country. In our country every year different kinds of natural disaster are attack such as flood, cyclone, river erosion, storm, thundering. Reason of attacking huge number of people are lose their home and destroy their crops land and they get a big number loses value. So that, I research about it for find a destroy rate of disaster, also find which disaster are mostly affected in our country in specific place and top affected area are also find using data mining.

For research purpose I need a big number of real time data about natural disaster. So that I collected data from online newspaper which are The Daily Prothom Alo and The Daily bdnews24.com Those newspaper are one of the most trusted in our country. I collect different type of data about natural disaster and create a csv file. After that I generalize data from raw data set to targeted data structure.

In Implementation part I use a different type of algorithm for getting an expected output that I already mention in above. I use a python language in Jupiter notebook. Jupiter notebook is a most popular open-source web application that allows me to create and share documents that contain live coding, visualization of statistical model and others required tasks. I analyze the dataset of the natural disaster by using different type of algorithm and machine learning techniques for getting a best output for tracking a disasters type form training dataset.

For getting an accurate output I implement dataset in different algorithm which are Data decision tree, Data Naive Bayes, Data random forest and DataKNN(K-nearest neighbors). From those

algorithm Data decision tree give best accurate output and worst output get from Data Naïve Bayes algorithm.

I use a different type of library and technique for get a visualization of statistical model of expected output those are required for this research purpose.

2.4 Scope of the problem

In this research there are so many scope for finding a different type of data analysis output by using data mining algorithm. It has a disaster tracking way by using analysis technique. By using this method can find a destroy rate of any disasters. It has a scope to find a most affected area with loses value. By using this technique can find a most damaging disaster in specific place. Find a disaster type using a machine learning from sample test data set.

But in this research also have some scope of the problem. It has don't have any application to find those information. For further development of this research using this data set can opportunity to develop an Android or web-based application to finding mentioned disaster information easily. This is the main scope of the problem of this research.

2.5 Challenges

When I starting this research i face some of difficulties and challenges. The big challenge is data collection. For collecting real data is the main challenges. There are so many source of data but trusted and valid are not available. I follow the trusted online newspaper which is The Daily Prothom Alo and The Daily bdnews24 for collecting data. When I get data it is got as a unstructured format. After get a data I have to create data set as targeted structure format it is a big challenging issue in this research.

Second challenge is data generalization processing as required as mentioned. This task are another difficult part of finding the actual required as different algorithm. After generalization I need to find and select the best algorithm for data analysis. I tried different type of algorithm for finding the accurate result. I use Data decision tree, Data naive Bayes, data random forest and DataKNN algorithm. So it's really hard job to find which one is the best Algorithm for this research.

And third and last challenge is create a result output as visualize in display. For create this type of system we need to find a best library in python and applied it in this research purpose for displaying the output as visualize.

So that, above problem that I faced for researching this topic. Mentioned challenges are the main and difficult tasks for me.

CHAPTER 3

RESEARCH METHODOLOGY

3.1 Introduction

In this chapter gives an outline of research method that were followed in the study. It provides information about research subject and also instrumentation that is the criteria for inclusion in the study. Which is the data collection procedure and what kind of data are required for this research this was mentioned in this chapter. In this chapter also provide the mathematical method and statistical analysis of collected data about natural disasters. Another and last part of this chapter is implementation requirement of this research. In this part provide which tools and algorithm was used for this study and describe about it. Moreover, this chapter discussed about every type of research methodology those was used for this study.

3.2 Research Subject and Instrumentation

In this part of chapter I discussed about research subject. The topic of this research is Natural disaster tracking in context of Bangladesh using data mining system. Bangladesh is a land of Natural disasters. We know that every year many types of disaster are attacking in our country. The reason of attacking disaster big number of loses are faced in every years. For this reason I tried to research about it and find the damaging rate and find the prediction of disaster type using damaging. I collected last five year real time data of natural disaster and analysis this data by using prediction algorithm and classification algorithm for this study. And finally find the actual result of research. I think it's really essential research for our government and other NGO for making a preplan about natural disaster for upcoming year. Above topic are the research subject and study material for this thesis.

Instrumentation of this research is the main important part of any thesis. A Research instrument is a tool used to collect measure and analyze data related to research work. The research instrument is usually determined by researcher and is tired to the study methodology. We know that there are different type of instrument are available for research purpose. So, that I use a web scraping technique and method for collection data about natural disaster. This is the best way and authentic

way to collect data from online newspaper. In the next point of this this chapter which is data collection procedure I described about web scraping method.

3.3 Data Collection Procedure

In this part I discuss about data collection procedure .I have studied different kinds of newspaper for scraping the natural disaster related data from them. I follow two online newspaper The Daily Prothom Alo and The Daily bdnews24.com as a trusted two. The newspaper section of The Daily Prothom Alo is organized in such way that it makes me believable on the other hand bdnews24.com is the first and top most online newspaper in our country those reason I can proceed to the further work.

3.3.1 Describe the Newspaper Scraping Algorithm

➢ Initially take the url of the newspaper section as str.

➢ Open targeted page with BeautifulSoup for parsing html Code with soup.

➢ After opening the page stored the selected section.

➢ Collect the headline and links in headline = [].

➢ Check the Natural disaster list = [] in headline.

➢ Open the opennews() if disaster list find in the headline.

➢ Open the paragraph tag of the news and stored in x.

➢ Collect data from x.

3.3.2 Data collection process

For collecting data we have executed the further step from the selected type paragraph stored in x. The further steps are:

➢ Convert the whole paragraph in lower case.

➢ After that convert the lower cased paragraph latter number to integer.

➢ Then focus on place of disaster affected area.

➢ After finding it search the damaging rate of disaster and other related information.

- ➢ Create a natural disaster list by date.
- ➢ Make a raw csv file.
- ➢ Then generalized data for target structural format.
- ➢ Make another csv file for training dataset for machine.

3.4 Statistical Analysis

After getting the data from scraping the newspaper then generalized data. Data generalization is a process by using this method dataset are get targeted structure as required statistical calculation.

So, that I have used the csv file to find the most affected disaster list according to the natural disaster. For getting an accurate output I used four different type of prediction algorithm and classifier algorithm those are Data decision tree Algorithm, Data random forest Algorithm this two are prediction algorithm on another hand Data naive bayes, Data KNN Algorithm this two are classifier algorithm. Now describes the algorithm started in below-

3.4.1 Data Decision Tree Algorithm

- ➢ Read the csv file.
- ➢ Store all natural disaster data in k_data = [] as a training dataset.
- ➢ Store test data are store in t_data = [].
- ➢ Select training data from x and total value of each attribute of x has been calculated.
- ➢ All target data are store on y.
- ➢ Test data are contain in x_pred .
- ➢ Testing data target are contain in y_pred.
- ➢ Define the classifier clf = tree.decisionTreeClassifier().
- ➢ Compare a training data with target data clf = clf.fit(x,y).
- ➢ Predicted the targeted output as an array.
- ➢ Give an Accuracy result of prediction.
- ➢ Get another output using a confusion matrix.

[N.B] Data random forest Algorithm follow the same mechanism as Data decision tree. In this algorithm generate a more tree for give an accurate results.

3.4.2 Data Naive Bayes Algorithm

- ➢ Read the csv file.
- ➢ Store all natural disaster data in k_data = [] as a training dataset.
- ➢ Store test data are store in t_data = [].
- ➢ Select training data from x and total value of each attribute of x has been calculated.
- ➢ All target data are store on y.
- ➢ Test data are contain in x_pred .
- ➢ Testing data target are contain in y_pred.
- ➢ Define the classifier clf = GaussianNB().
- ➢ Compare a training data with target data clf = clf.fit(x,y).
- ➢ Predicted the targeted output as an array.
- ➢ Give an Accuracy result of prediction.
- ➢ Get another output using a confusion matrix.

[N.B] Data KNN algorithm are provide a most nearest similar data as a targeted data set and this algorithm are not give satisfying result as I expect for this thesis.

3.5 Implementation Requirement

In this part I described about implementation requirement and tools. For implementing this research I used various type of tools such as Anaconda, Jupiter notebook, Microsoft excel in different purpose.

3.5.1 Operating Environment

Operating environment is important part for run all type of implementing tools. I used a windows OS for this purpose. Windows I good platform for implementing this research. I used windows 10 64-bit but 32-bit or 86x are also capable to run any kinds of tools for implementing. In linux OS those tools will also run.

3.5.2 Language for Implementation

Language selection is the challenging and most important part for implementation. In this research I use a python language. Python is most popular and most reliable programming language. Mostly

This type research work and implementation part are easily implementing by using python language. Data mining related algorithm are perfectly work if use this language. It give a best accurate output what I required for this research work. That's the reason I use this language.

3.5.3 Development Tools

I used anaconda and Jupiter notebook as a development tools. Anaconda is an open-source distribution of the python language which one we used as a program language. For All type of data mining related work are develop easily using anaconda. It provide big number of library that are required for developing purpose of code. It also provide a Jupiter notebook. Jupiter notebook is a web-based open-source application that allow an interactive computing notebook environment. Edit and run human-readable docs or csv file while describing the data analysis work.

3.5.4 Dataset

I had to create a dataset for implementation of natural disaster data. For containing all different type of data in individual columns. Create a csv file and use it in data analysis purpose. Because all type of Data analysis algorithm required a csv format file as a dataset. That's the reason I use csv file.

3.5.5 Browser Dependency

Browser dependency is another issue for any development work. In this research work don't have any browser dependency issue. We know that there are various kinds of browser are available such a chrome, Firefox, Opera, torch, Microsoft edge etc. All browser are capable the Jupiter notebook.

CHAPTER 4

EXPERIMENTAL RESULTS AND DISCUSSION

4.1 Introduction

In this chapter I provided all experimental results which had been done for this research purpose. I also provided various type of statistical model that are required for this thesis. Newspaper scraping result and raw dataset and generalizing dataset are gave on this chapter. Python code has implemented based on the different type of algorithm and methods stated on this chapter. The results and their analysis with the graph has been also stated.

4.2 Experimental Results

Experimental results of this thesis are provided on this part. Newspaper scraping result, implementation result of algorithm all are provided in this part.

4.2.1 Result of Newspaper Scraping

The scraping result I have got has been implemented in a csv file. Which contains the disaster name, Area affected, Crop land, Union affected, Died, Education affected, Home destroy, Homeless, Injured, Missing, Place, Relocated, Victim, Road destroy, Loss value, Agriculture loss value, Bridge Affected. There are 5 top most occurred disaster information are store in csv file which are occurred in last five year. There are total 483 rows of data was collected and store in csv file. I have tied to show some attributes from the csv file in the following figure.

Disasters_	Area affec	Crop Land	Union/Lar	Died	Eduaction	Home des	homeless	Injured	Missing	Place	Relocated	Victims	Road dest	Loss value	Agricultur	Bridge affected
Thunderir	0	0	0	0	0	0	0	1	0	Coxs Baza	0	1	0	0	0	0
River ersi	1	0	1	0	0	350	300	0	0	Jamalpur	0	500	0	0	0	0
Cyclone	1	20000	2	0	0	328	1350	0	0	Noakhali	0	1450	0	0	0	0
Cyclone	1	850	1	0	0	300	660	0	0	Chittagon	0	3300	0	1E+09	1000000	0
Cyclone	1	22836	16	0	0	4800	18200	0	0	Patuakhal	0	21000	0	0	2.63E+08	0
Cyclone	1	0	75	0	0	8788	132000	0	0	bagerhat	0	140000	0	0	0	0
Thunderir	0	0	0	2	0	0	0	0	0	Noakhali	0	2	0	0	0	0
Thunderir	0	0	0	2	0	0	0	0	0	Satkhira	0	2	0	0	0	0
Thunderir	0	0	0	1	0	0	0	1	0	Dinajpur	0	1	0	0	0	0
Thunderir	0	0	0	1	0	0	0	0	0	Dinajpur	0	1	0	0	0	0
Thunderir	0	0	0	1	0	0	0	0	0	Madaripu	0	1	0	0	0	0
Thunderir	0	0	0	4	0	0	0	3	0	Chadpur	0	7	0	0	0	0
River ersi	0	0	0	1	0	0	0	0	0	Patuakhal	0	1	0	0	0	0
Flood	1	0	2	0	0	0	0	0	0	Chapai na	0	300	0	0	0	0
Flood	1	0	3	0	0	0	0	0	0	Pabna	0	30000	0	0	0	0
Flood	1	2047	4	0	0	0	0	0	0	Pabna	0	1200	0	0	1E+08	0
River ersi	0	0	0	1	0	0	0	0	0	Patuakhal	0	1	0	0	0	0
River ersi	1	0	3	0	0	150	400	0	0	Patuakhal	400	450	0	0	0	0
Flood	1	0	0	0	0	0	0	0	0	Rajbari	0	16500	0	0	0	0
Flood	1	348	4	0	0	0	0	0	0	Faridpur	0	0	0	0	0	0
Flood	1	0	0	0	0	0	0	0	0	Chapai na	0	65000	0	0	0	0
Flood	1	0	0	0	0	150	0	0	0	Rajshahi	0	800	0	0	0	0
Flood	1	0	2	0	0	0	0	0	0	Kushtia	0	60000	0	0	0	0
Flood	1	0	3	0	0	0	0	0	0	Natore	0	12000	0	0	0	0
Flood	1	0	1	0	0	0	0	0	0	Natore	0	4800	0	0	0	0
Flood	1	30	1	0	2	0	0	0	0	Faridpur	0	2000	4	0	0	0
River ersi	1	0	1	0	0	12	58	0	0	Shariatpu	0	350	0	0	0	0
River ersi	1	100	2	0	0	100	400	0	0	Rajbari	0	1500	0	0	0	0
River ersi	1	0	1	0	4	350	0	0	0	Madaripu	0	1550	0	0	0	0

Figure 4.1: Raw dataset of newspaper scrapping before generalization in csv file.

We see that all natural disaster are store in csv file. There were no duplicate data in there. All data are real time data that are directly collect from Newspaper those are trusted source of data. I will work for the further implementation.

Disastersf	AreaAffec	CropLand	UnionLan	Died	Eduaction	HomeDes	homeless	Injured	Missing	Relocated	Victims	RoadDesti	LossValue	Agricultur	BridgeAffected
0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0
0	0	0	0	2	0	0	0	0	0	0	2	0	0	0	0
1	1	100	0	0	0	0	0	0	0	0	0	0	0	0	0
1	1	0	1	0	0	32	150	0	0	0	185	0	0	0	0
2	1	0	0	2	0	725	4050	0	0	0	5500	0	0	0	0
3	1	0	0	0	0	0	0	0	0	0	700	0	0	0	0
2	1	0	0	0	0	990	5500	0	0	0	6500	0	0	0	0
3	1	0	0	25	0	0	0	0	0	0	1	0	0	0	0
4	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0
4	1	0	0	1	0	40	200	0	0	0	250	0	0	0	0
1	1	0	1	0	0	50	300	0	0	300	400	0	0	0	0
1	1	0	2	0	0	50	250	0	0	0	350	0	0	0	0
1	1	0	1	0	0	150	750	0	0	0	950	0	0	0	0
1	1	70	1	0	0	150	650	0	0	0	750	0	0	0	0
1	1	0	2	0	0	120	550	0	0	0	650	0	0	0	0
2	1	6200	0	5	0	26500	450000	0	0	0	495000	0	0	0	0
2	1	823	0	3	0	3200	80000	0	10	0	960050	0	0	0	0
2	1	75	0	1	0	300	8000	0	1	0	12000	0	0	0	0
2	1	406	0	1	0	850	25000	0	0	0	35000	0	0	0	0
2	1	1929	0	0	0	6000	185000	0	0	0	190000	0	0	0	0
2	1	0	1	3	0	0	0	0	24	0	27	0	0	0	0
3	1	0	2	0	0	0	0	0	0	0	60000	0	0	0	0
3	1	0	3	0	0	0	0	0	0	0	12000	0	0	0	0
3	1	0	1	0	0	0	0	0	0	0	4800	0	0	0	0
3	1	30	1	0	2	0	0	0	0	0	2000	4	0	0	0
3	1	0	1	0	0	150	2500	0	0	0	3500	0	0	0	0
4	0	0	0	0	0	300	1500	0	0	0	2000	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	300	0	0	0	0
4	1	0	0	1	0	40	200	0	0	0	250	0	0	0	0

Figure 4.2: dataset of newspaper scrapping after generalization in csv file.

Generalization is a process by using it data are convert from string to numeric value for understanding machine. Because we know that, if we try to train any machine we must provide a numeric value to train the machine. Machine can't understand string data. In this dataset $0 =$ Thundering, $1 =$ Flood, $2 =$ River erosion, $3 =$ cyclone, $4 =$ storm. That's why I generalized dataset before implementing algorithm.

4.2.2 Result of Data Decision Tree Algorithm

Data decision tree is a prediction algorithm that use for data analysis work in data mining field. Data I choose this Data Decision tree algorithm because it's give a best result of this research work as I required. Its gives a best accurate result compare with others Algorithm that had I implement for accurate result. I give the result and accuracy of implementing this algorithm in showing via figure.

Figure 4.3: Result and accuracy of disaster tracking using prediction algorithm.

Show that:

➤ Define the Data decision tree algorithm.
➤ Calculate the damage value and compare with training data with targeted dataset (x,y).

- ➢ After comparing give a prediction result as an array wise.
- ➢ And accuracy given by Metrix.

See that, accuracy rate is 0.95 out of 1.00 those are the best result compared with others algorithm.

4.2.3 Result of Data Naive Bayes Algorithm

Data Naïve Bayes Algorithm is classification algorithm. This algorithm use for data analysis propose in Data mining field. Naïve bayes algorithm give a worst result in this research implementation work. It give a very low accuracy rate compared with others algorithm that had I implement in data analysis propose. I give the result and accuracy of implementing this algorithm in showing via figure.

Figure 4.4: Result and accuracy of disaster tracking using Classification algorithm.

Show that:

- ➢ Define the Data naive bayes algorithm.
- ➢ Calculate the damage value and compare with training data with targeted dataset (x,y).
- ➢ After comparing give a prediction result as an array wise.
- ➢ And accuracy given by Metrix.

See that, Accuracy is 0.25 out of 1.00.This algorithm give a very low accuracy and worst result in research work those are required. Data naive bayes algorithm follow one strategy which is when it's analysis any data set its search and target majority data. That's the reason it can't predict data accurately.

So that, Data decision tree algorithm gives a best prediction and accurate result compared with Data naive bayes algorithm.

4.3 Descriptive Analysis

In this part I provided a Statistical Model which I got via data analysis process in data mining field. I showing the bar chart of top affected area by specific disaster, pie chart of disaster affecting information of specific disaster and most affected disaster in specific place.

This time I select a river erosion as a disaster and Jamalpur as a place. For the further implementation I will use another diasrer and another place.

4.3.1 Result of Statistical model
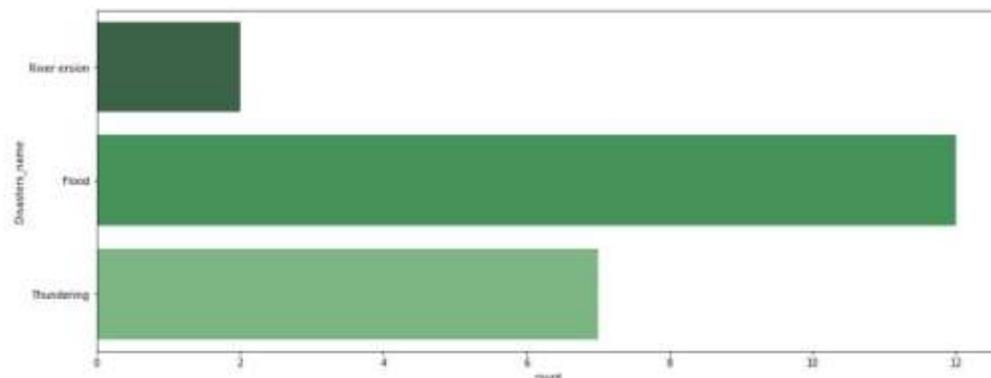
Figure 4.5: most affected disaster in specific place.

Show that, in jamalpur are mostly affected by Flood, second most is thundering and third one is River erosion.
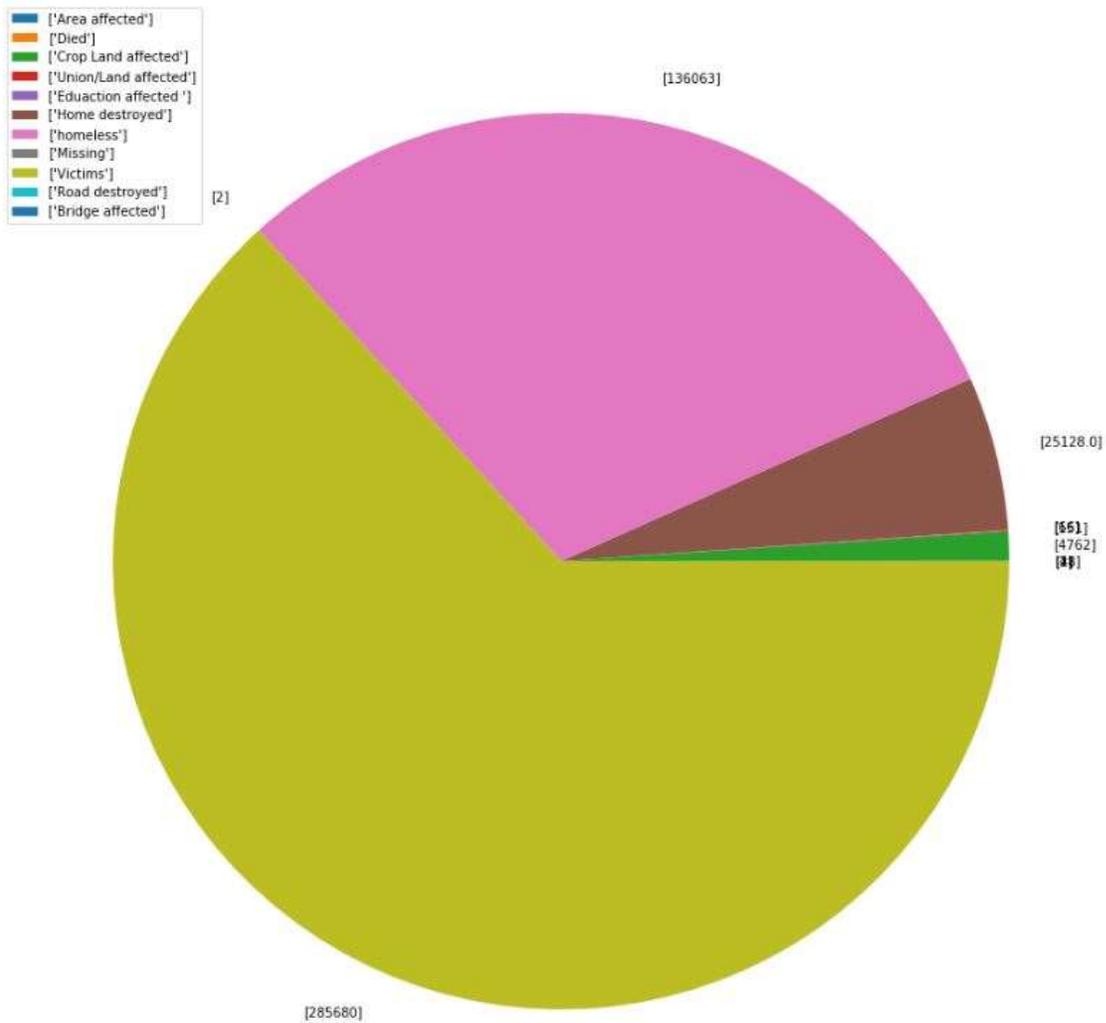

Figure 4.6: Pie chart of disaster affecting information of specific disaster.

Show that, this is the River erosion information of Jamalpur. See that huge number of people are victim of River erosion and second huge number of people are homeless. So many home are destroy by river erosion in last 5 year.
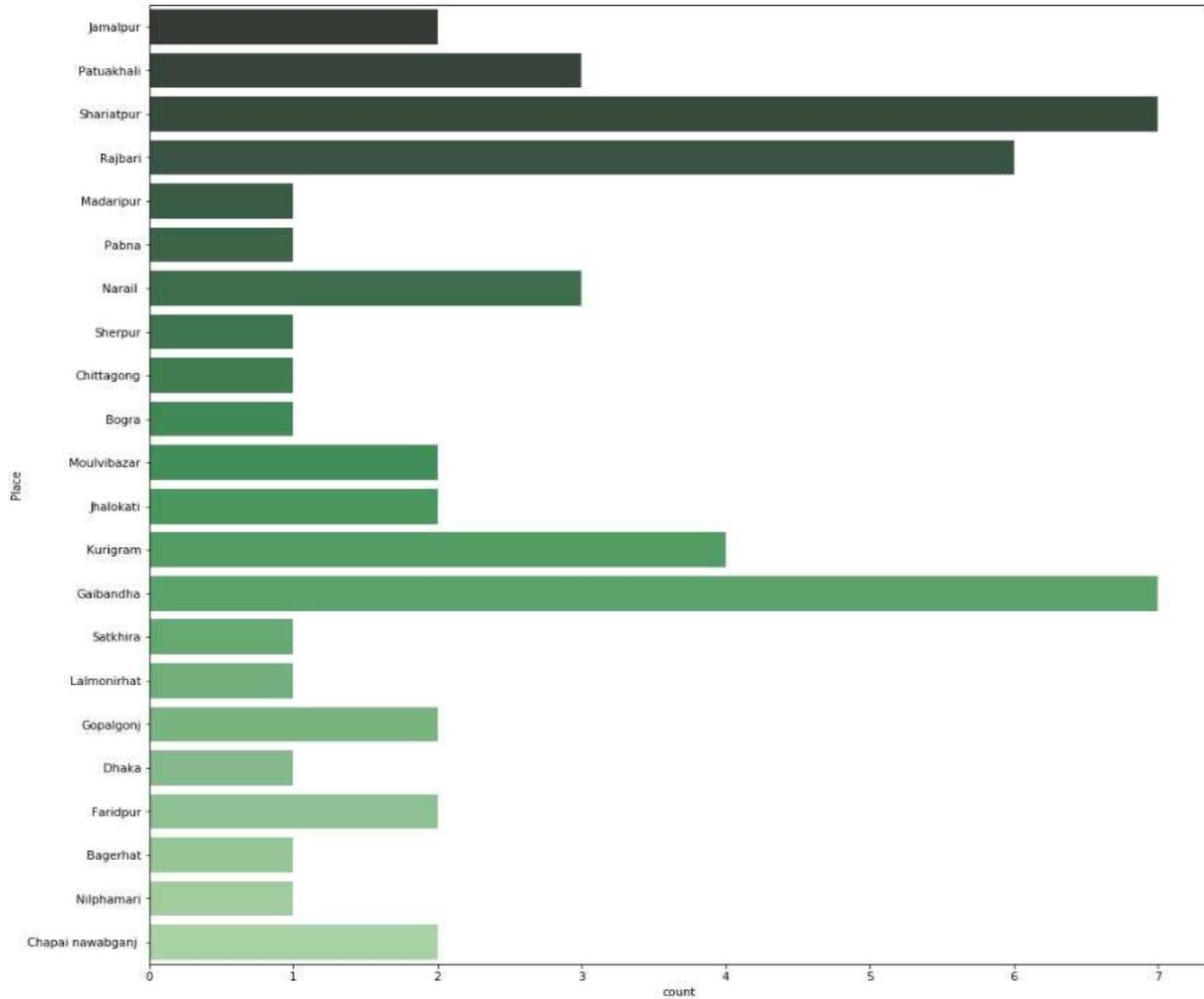
Figure 4.7: Bar chart of top affected area by specific disaster.

See that, this is the affecting result of River erosion of Bangladesh. We see that Shariatpur and Gaibandha are top and second top are Rajbari those area are most affected area by River erosion in last 5 year.

4.4 Summary

We know that, Bangladesh is a natural disaster prone country. So that, in this chapter I was show the implementation work and result and also show the accuracy of result. I provided web scraping result, Prediction Algorithm result and also provided Statistical model for better understanding about research.

CHAPTER 5

CONCLUSION AND FUTURE WORKS

5.1 Summary of the Study

In this part I summarize of this research work. The Natural disaster related data are collected by using web scraping. I select the top most Newspaper in our country for collected real-time data which are trusted source and they also provide accurate information. After collect data I store those data in csv file and apply the Data Decision Tree Algorithm and Data Naive Bayes Algorithm for prediction result of Natural disaster and also got an accuracy of implementation result.I also implemented the Statistical method to find a most affected disaster in specific place, Pie chart of disaster affecting information of specific disaster and Bar chart of top affected area by specific disaster and got a result of statistical analysis.

5.2 Limitation and conclusions

Natural disasters in the form of Floods, river erosion, cyclone, storm, thundering that claim numerous lives, cause significant damage of the properties. The effects have been much more serve in developing country such as Bangladesh compared to developed countries. Reason of study about natural disaster and research about it in context of our country to find a prediction result and track the natural disaster.

By using this system our government and Non-government organization (NGO) will take a perfect decision and make a pre plan for upcoming year. This is really helpful for them.Every research work have some limitation. So that, in this research work have one limitation. Which is skewed dataset. Skewed dataset means some of disaster data are more then with others disaster data. This is happed for real time based dataset. But this not a big issue to effecting research implementation work.

5.3 Recommendation for future works

For the further development of this research work in future, it have really a good opportunity to create a more big data set and implementation this data set using this method. It has opportunity to build an Android apps or web-based application to find an information about natural disaster. Real time data collection procedure will change by getting information form online source. So data collection procedure will get a dynamic way.

REFERENCE

[1] The Daily Prothom Alo, Available at <<https://en.prothomalo.com/>>, last accessed on 6th January 2020.

[2] The Daily bdnews24.com Available at <<http://bdnews24.com/>>, last accessed on 15th January 2020.

[3] Web scraping Available at <<https://www.datacamp.com/community/tutorials/web-scraping-using-python>>, last accessed on 16th January 2020.

[4] Jupitar notebook Available at <<https://jupyter.org/>>, last accessed on 17th January 2020.

[5] Anaconda, Available at << https://www.anaconda.com/>>, last accessed on 17th January 2020.

[6] Adventist Development and Relief Agency (ADRA) Available at<< https://www.adrabangladesh.org/>>, last accessed on 20th January 2020.

[7] Saptarsi Goswami, Sanjay Chakraborty, Sanhita Ghosh, Amlan Chakrabarti, Basabi Charkarborty A Review on application of data mining techniques to combat natural disasters. Available at 16th April 2016 p. 1-28.

[8] Brooks Nick, Adger W Neil. Country level risk measures of climate-related natural disasters and implications for adaptation to climate change. Clim Res. Tyndall Centre for Climate Change. 2013. p. 1–30.

[9] Chris Hodaap, Matt Robbins, Leff Gray and Andrew Graettinger. Damage tracker a cloud and mobile system for collecting damage information after natural disasters. 2013. P. 1-6.

[10] Datar Ashlesha, Liu Jenny, Linnemayr Sebastian, Stecher Chad. The impact of natural disasters on child health and investments in rural India. Soc Sci Med 2013. Page 1-45.

[11] Liu SB, Bouchard B, Bowden DC, Guy M, Earle P. USGS tweet earthquake dispatch (@USGSted): using twitter for earthquake detection and characterization. AGU fall meeting abstracts, vol. 1; 2012.

[12] Kradolfer Urs. SalanderMaps: a rapid overview about felt earthquakes through data mining of web-accesses. EGU general assembly conference abstracts, vol. 15; 2013.

[13] Sakaki Takeshi, Okazaki Makoto, Matsuo Yutaka. Earthquake shakes twitter users: real-time event detection by social sensors. In: Proceedings of the 19th international conference on World Wide Web. ACM; 2010.

[14] Takako Hashimoto, Chakraborty Basabi, Kuboyama Tetsuji, Shirota Yukari. Temporal awareness of needs after east japan great earthquake using latent semantic analysis. Information modelling and knowledge bases XXV, vol. 260. IOS Press; 2014.p. 200–12.

[15] Doan Son, Vo Bao-Khanh Ho, Collier Nigel. An analysis of Twitter messages in the 2011 Tohoku earthquake. In: 4th International Conference on eHealth. Electronic Healthcare. Berlin Heidelberg: Springer; 2012. p. 58–66.

[16] Panigrahi Sangram, Verma Kesari, Tripathi Priyanka, Sharma Rika. Knowledge discovery from earth science data. In: Fourth international IEEE conference on communication systems and network technologies (CSNT); 2014. p. 398–403.

[17] Pabreja Kavita, Datta Rattan K. A data warehousing and data mining approach for analysis and forecast of cloudburst events using OLAP-based data hypercube. Int J Data Anal Tech Strat 2012;4(1):57–82.

[18] Srivastava Kuldeep, Bhardwaj Rashmi. Real-time now cast of a cloudburst and a thunderstorm event with assimilation of Doppler weather radar data. Nat Hazard 2014(70):1357–83.

APPENDICES

Appendix A: Research Reflection

During my research I was faced many problems. I had to learn different type of Data mining techniques, machine learning and implementing different type of Algorithm on my research work. After then I had use various prediction algorithm and classifier algorithm then train my data set for expected out of this research work. After finishing this research work I have get a depth knowledge about machine learning and data mining techniques and implementing processing. I hope this knowledge will helpful for me in future.

Appendix B: Related Issues

Figure 4.2.2.1: Result and accuracy of disaster tracking using prediction algorithm.

In Figure 4.2.2.1 shows that prediction result of disaster and result. In this case i use a Data decision tree Algorithm and get and accuracy 0.95% out of 1.00.

MSC_Nafiur

22%	18%	9%	21%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

1	Submitted to Daffodil International University Student Paper	10%
2	Submitted to Universiti Malaysia Sabah Student Paper	5%
3	dspace.daffodilvarsity.edu.bd:8080 Internet Source	1%
4	Saptarsi Goswami, Sanjay Chakraborty, Sanhita Ghosh, Amlan Chakrabarti, Basabi Chakraborty. "A review on application of data mining techniques to combat natural disasters", Ain Shams Engineering Journal, 2018 Publication	1%
5	Submitted to Indian Institute of Technology, Bombay Student Paper	<1%
6	Submitted to University of Stellenbosch, South Africa Student Paper	<1%
7	dblp.uni-trier.de Internet Source	<1%