

**POLARITY PREDICTION OF NEWS HEADLINES OF BENGALI
NEWSPAPERS USING MACHINE LEARNING**

BY

APU BARUA

ID: 153-15-6557

AND

FARJANA YESMEN

ID: 153-15-6561

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

Supervised By

Md. Tarek Habib

Assistant Professor

Department of CSE

Daffodil International University

Co-supervised By

Dewan Mamun Raza

Lecturer

Department of CSE

Daffodil International University



DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

SEPTEMBER 2019

APPROVAL

This Project titled “**Polarity prediction of news headlines of Bengali newspapers using machine learning**”, submitted by Apu Barua and Farjana Yesmen, ID No: 153-15-6557 and 153-15-6561 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 B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on 14 September 2019.

BOARD OF EXAMINERS



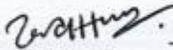
Dr. Syed Akhter Hossain
Professor and Head
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Chairman



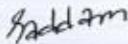
Nazmun Nessa Moon
Assistant Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner



Md. Zahid Hasan
Assistant Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner



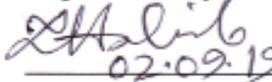
Dr. Md. Saddam Hossain
Assistant Professor
Department of Computer Science and Engineering
United International University

External Examiner

DECLARATION

We hereby declare that, this project has been done by us under the supervision of **Md. Tarek Habib, Assistant Professor, Dept. of CSE** Daffodil International University. We also declare that neither this project nor any part of this project has been submitted elsewhere for award of any degree or diploma.

Supervised by:


02-09-19

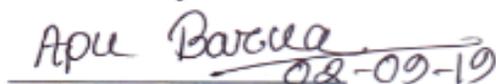
Md. Tarek Habib
Assistant Professor
Department of CSE
Daffodil International University

Co-supervised by:

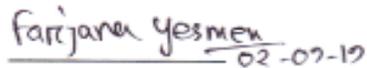

02-09-19

Dewan Mamun Raza
Lecturer
Department of CSE
Daffodil International University

Submitted by:


02-09-19

Apu Barua
ID: 153-15-6557
Department of CSE
Daffodil International University


02-09-19

Farjana Yesmen
ID: 153-15-6561
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

First we express our heartiest thanks and gratefulness to almighty God for His divine blessing makes us possible to complete the final year project.

We really grateful and wish our profound our indebtedness to **Md. Tarek Habib**, Assistant Professor, Department of CSE Daffodil International University, Dhaka. Deep Knowledge & keen interest of our supervisor in the field of Computer Vision to carry out this project. 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.

We would like to express our heartiest gratitude to **Dr. Syed Akhter Hossain**, Professor and Head, Department of CSE, for his kind help to finish our project and also to other faculty member and the staff of CSE department of Daffodil International University.

We would like to thank our entire course mate in Daffodil International University, who took part in this discuss while completing the course work.

Finally, we must acknowledge with due respect the constant support and patients of our parents.

ABSTRACT

Time is becoming more and more scarce resource keeping pace with the advancement of technology. Automatic detection of polarity of newspaper heading can save time of reading. It helps to analyze social status to reduce the number of bad news. It provides a platform for serving good news and create a positive environment. In this project work, we develop a system that detects the positivity or negativity of a newspaper headline in Bangla. It has been studied a lot earlier, but all of the research work has been done in English based on English Newspapers. It is a matter of fact that no work has been done in Bangla. We use a large number of Bangla newspaper's headlines in order to build our data set in this project work since there was no such work before in Bangla. Web crawler is used to create the Bangla data set. Then we use three classifiers based on machine learning, e.g. support vector machine (SVM), logistic regression and random forest and one classifier based on deep learning, e.g. sequential model. We thoroughly investigate the feasibility and applicability of the classifiers. We train and test the data set with them following machine learning approach. Sequential deep learning model outperforms all other classifier achieving an accuracy of 80%, SVM shows the poorest result, i.e. an accuracy of 55%.

Keywords: Polarity prediction, Data Set, SVM, Logistic Regression, Random Forest, Deep learning.

TABLE OF CONTENTS

CONTENTS	PAGE
Board of examiner	i
Declaration	ii
Acknowledgement	iii
Abstract	iv
CHAPTER	
CHAPTER 1: INTRODUCTION	1-2
1.1 Introduction	1
1.2 Motivation	1
1.3 Research Questions	1
1.4 Expected Output	1
CHAPTER 2: BACKGROUND	3-4
2.1 Introduction	3
2.2 Related Works	3
2.3 Scope the Problem	4
2.4 Challenges	4
CHAPTER 3: RESEARCH METHODOLOGY	5-6
3.1 Data Collection	5
3.2 Data Tagging	5
3.3 Model Selection	5

3.4 Workflow	6
CHAPTER 4: MACHINE LEARNING ALGORITHM	7-13
4.1 Support Vector Machine (SVM)	7
4.2 Logistic Regression	12
4.3 Random Forest	13
4.3 Deep Learning Model	13
CHAPTER 5: EXPERIMENTAL RESULTS	14-16
AND DISCUSSION	
5.1 Introduction	14
5.2 Experiment with Logistic Regression	14
5.3 Experiment with SVM	15
5.4 Experiment with Random Forest	15
5.5 Experiment with Sequential Model	15
5.6 Comparative Analysis	15
CHAPTER 6: SUMMARY, CONCLUSION, RECOMMENDATION	
AND IMPLICATION FOR FUTURE RESEARCH	17-18
6.1 Summary of the Study	17
6.2 Conclusions	17
6.3 Recommendation	17
6.4 Implication for Further Study	18
REFERENCES	19-21

LIST OF FIGURES

FIGURES	PAGE NO
Figure 3.4.1: Working procedures	6
Figure 4.1.1: Hyperplane in scatter	8
Figure 4.1.2: Creating hyperplane margin	8
Figure 4.1.3: Finding right hyperplane	9
Figure 4.1.4: Correct hyperplane	9
Figure 4.1.5: Hyperplane with outliers	10
Figure 4.1.6: Hyperplane with maximum margin	11
Figure 4.1.7: Solving with SVM	11
Figure 5.1.1: A small portion of the dataset	14

CHAPTER 1

INTRODUCTION

1.1 Introduction

Polarity prediction is the method of realizing in case a piece of writing is positive, negative or neutral news. It is also known as opinion mining, performing the attitude of the lecturer. A common use area for this technology is to find out how people feel about a particular topic. Human decision-making is influenced by opinion. Finally, we can say that, Polarity prediction is a qualitative study of ideas, feelings, attitudes, influences, emotions, etc. developed in the text.

Text = Reviews, Blogs, Discussions, Headlines, Comments, feedback.

1.2 Motivation

In this project, we are expecting to extract the sentiment led by many Bengali newspapers and trying to identify the polarity of the newspaper headlines as positive or negative. Although, this research will classify given headlines into one of three categories: positive, negative or neutral.

1.3 Research Questions

- How Polarity prediction Impacts On Our Everyday Life?
- Can we simultaneously apply machine learning and deep learning models?
- What are the challenges faced by Polarity prediction applications?
- Which approach is better for Polarity prediction?
- What are the areas of research in Polarity prediction, where I can explore my research topics?

1.4 Expected Output

In this project we have worked with Bangla Newspaper and we tried to find out the polarity in positive and negative sentences. Newspaper can be positive, negative or neutral. After exceeding the capabilities of the natural text language complete

understanding machines, the statistical analysis of comparatively simple statements can provide a surprisingly significant range of positive and negative sentiments. Then this analyzed data can be used in many sectors and other predictions.

CHAPTER 2

BACKGROUND

2.1 Introduction

Earlier, it was very difficult to do this research because the newspapers were manual. But with the gradual advancement of technology, these problems have been solved. Because now almost all the newspapers have come to the online platform. Already a lot of research papers have been done on English newspapers. Since many analyzes have been done on English newspaper, so, we have worked with Bengali newspaper. Bengali is a popular language which is in the seventh position. That is why we have felt that it is necessary to analyze the Bengali newspaper.

2.2 Related Works

This project is largely stimulated by the work of Kaur and Chopra [5]. They classify newspaper articles into three poles. Positive, Negative and Neutral. In this project, we will implement SVM [2], logistic regression [17] random forest classification and sequential model of keras.

In their research [5] used Naive Bayes classifier, and this work applied on English Language. In reality, most of the works were done on English of Polarity prediction. Work on other natural languages is increasing including Chinese [11] [12], Japanese [7] [8] [9] [10] and other languages. Very few Works that are done in Bengali are done related to the blog's [13] [14].

In this project, we select Logistic classifier because in comparison with Naïve Bayes classifier, Logistic classifier always offer the better result. Logistic classifier [17] has less error [17] than Naïve Bayes classifier.

In, logistic regression estimates the minimum measurements on the dataset, it measures the connection between categorical binary dependent variables and categorical variables. However it evaluates the probabilities of categorical variables using the logistic function.

[15] In this, news analyzes were performed on news and blogs. Their systems include sensing detection, aggregation and scoring phases. Wanner and. Al [5] presented a visual tool for comparative sensitivity analysis of large-scale news articles.

2.3 Scope the Problem

Since there is less research on Bangla News Paper, there is a much bigger opportunity to work with Bangla News Paper. According to recent data, we know that, About 210 million people speak in Bengali, Of these, 100 million people who live in Bangladesh speak There are more than 85 million Indians who speak Bengali and there are many people all over the world who speak Bengali at home. Since so many people speak in Bangla, so we think that, we should work with Bangla.

2.4 Challenges

Since there has never been such a thing to do with Bangla Newspaper, working with Bangla News Paper is challenging. In this case, building a data set is the most challenging. There are usually two ways we can collect data: one is coding and the other is manually. Manually collecting data is very boring for newspaper headlines. And if you want to collect data through coding, then will require skill on coding. So to work on Bangla, you have to face these challenges.

CHAPTER 3

RESEARCH METHODOLOGY

3.1 Data Collection

In this project we have analyzed the newspaper headlines. We have used multiple newspapers for these headlines like Prothom Alo, Samakal etc. Since no previous data set was available, we had to create a new data set for this project. We used Web Crawler to collect headlines from news portals. Through web crawler, we can easily collect the data using coding. For data collection we first had to find the archives of the news portal. Then, with the help of snipping tool, we had to find out the headline of our newspaper in the HTML tag. Then we put the HTML tags in the web crawler we created and parsed the headlines into a text file. Our headlines are stored date wise and we're converting all the text files into one text file. Then convert that text file to the CSV file.

3.2 Data Tagging

After successfully collecting the data and converting it to CSV format, our job was to tag the headlines. We need manually tagged information for our model training. We had to read every headline to tagging the data those headlines positive, negative or neutral. When the data is finished tagging, we saved the entire data set in the CSV format as our main dataset.

3.3 Model Selection

There are many types of machine learning classifiers for text classification. In the context of the search, we chose to utilize SVM, Random Forest, Logistic Regression and Sequential Model in our project. SVM provides a good accuracy for text classification. Since SVM and logistic regression are binary classifiers, they are better at classifying the polarity of a sentence. Our job is to distinguish between positive and negative which, like binary classification and SVM, logistic regression works better for it. Random Forest algorithm accept the decision tree idea further

by conducting a large number of decision trees. It is useful for building simple classification network.

3.4 Workflow

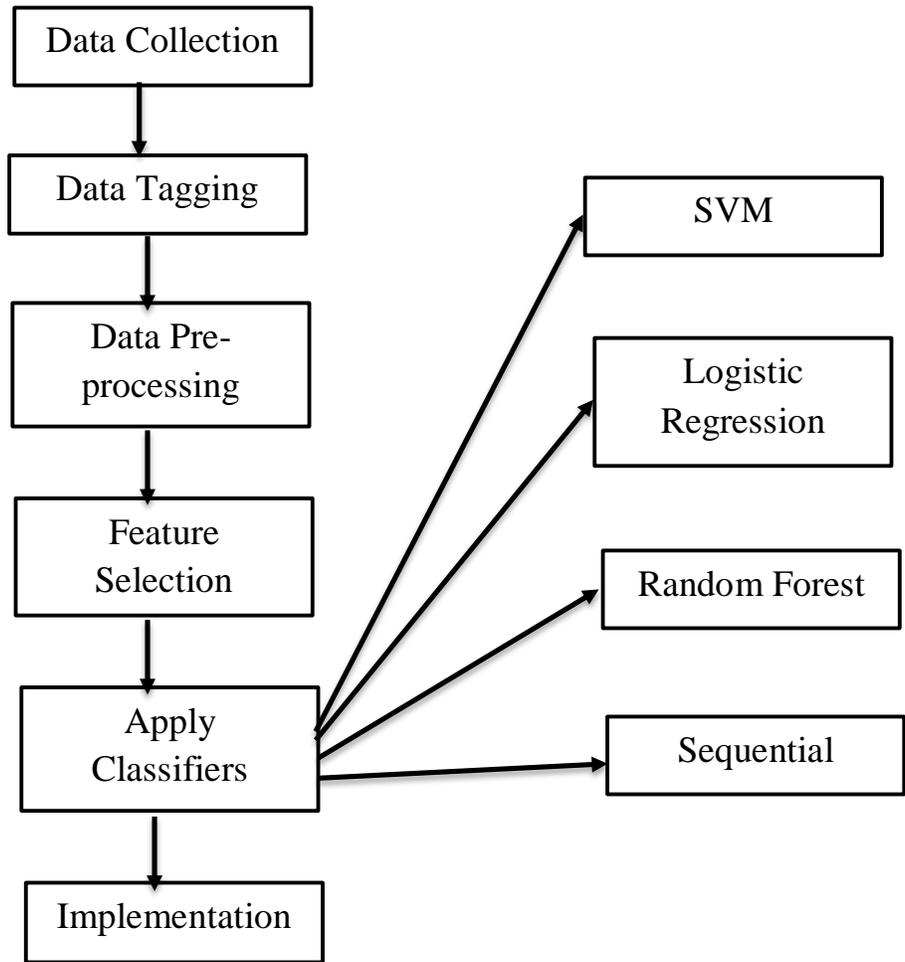


Figure 3.4.1: Working Procedures

CHAPTER 4

MACHINE LEARNING ALGORITHM

4.1 Support Vector Machine(SVM)

SVM is a supervised learning model, which analyzes data for categorization and prevention analysis. SVM builds a model using training algorithms that give new examples in one or two sections. The SVM creates a gap and separates as wide a segment as possible. New application section spaces can be spaced in the same space or space problem is not possible when the data is not labeled correctly. Then for analysis we need to follow an obsolete teaching approach, through which we can split the data into separate groups.

Example

Outlier

The outlier is an observation point that is far from the other observation. An observation that is beyond the standard values expected in the study or test, and which is often excluded from the data set.

Hyperplane

In geometry, a hyperplane is a symptom of lower levels than its surroundings. If a space is three-dimensional, its hyperplanes are 2 dimensional planes, while the space is 2-dimensional, then its hyperplanes are 1 dimensional lines. This idea can be used in any common place where the idea is the dose of a symptom is defined. Suppose we have three hyper-planes (A, B and C). Now, to classify the stars and the chakras, we need to identify the correct hyper-plane. We need to remember a thumb rule to detect the correct hyper-plane.

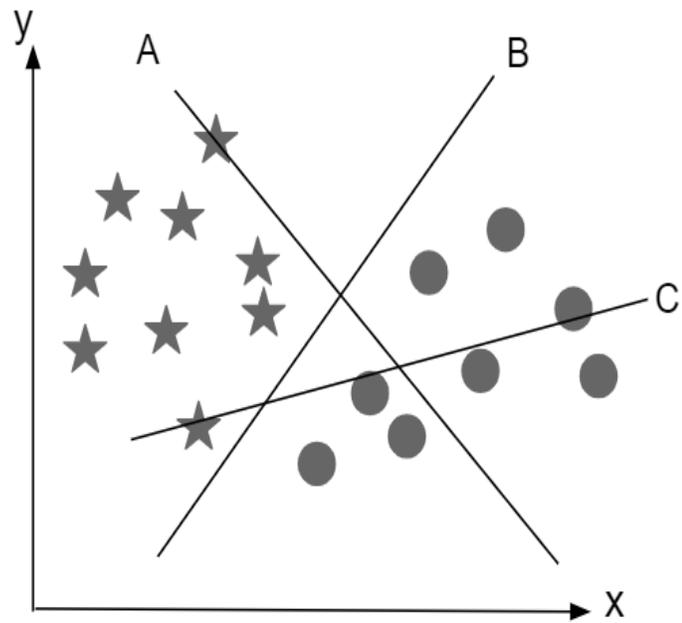


Figure 4.1.1: Hyperplane in scatter

Now, here we have three hyperplane all are in scatter possible ways.

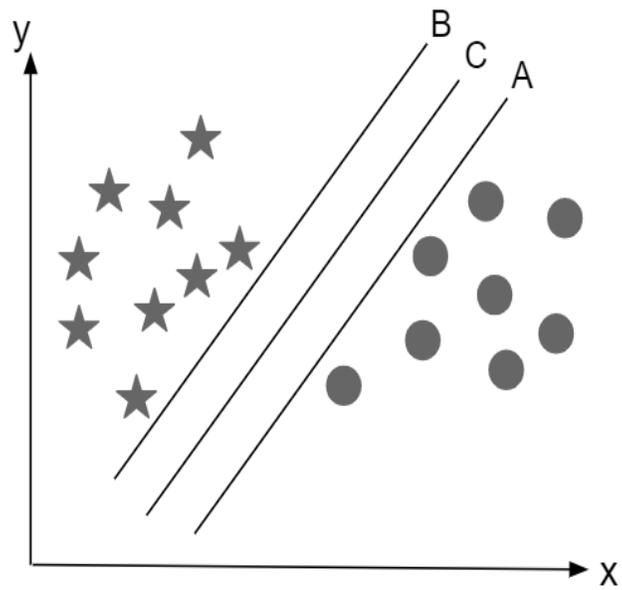


Figure 4.1.2: Creating hyperplane margin

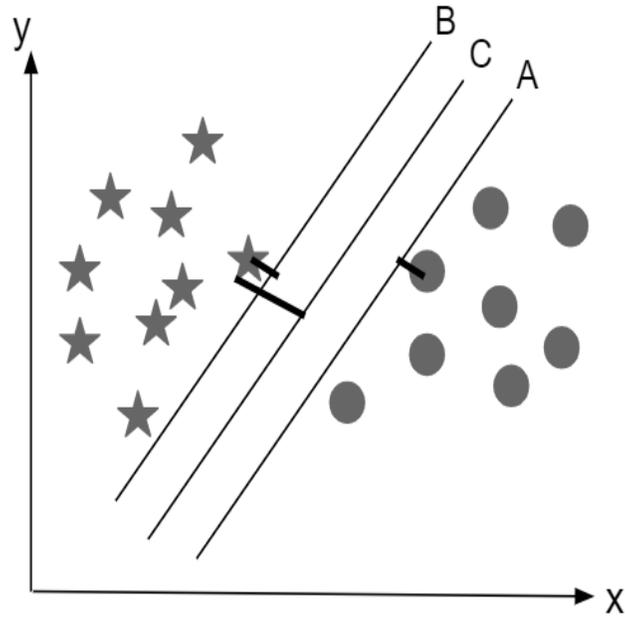


Figure 4.1.3: Finding right hyperplane

Above, we can see that the margin for hyper-plane C is very high rather than A and B. Hence, we name the right hyper-plane as C. We have selected the hyper-plane with higher margin is robustness, because if we select a hyper plane having low margin then there is a high chance that the margin will be miss-classified.

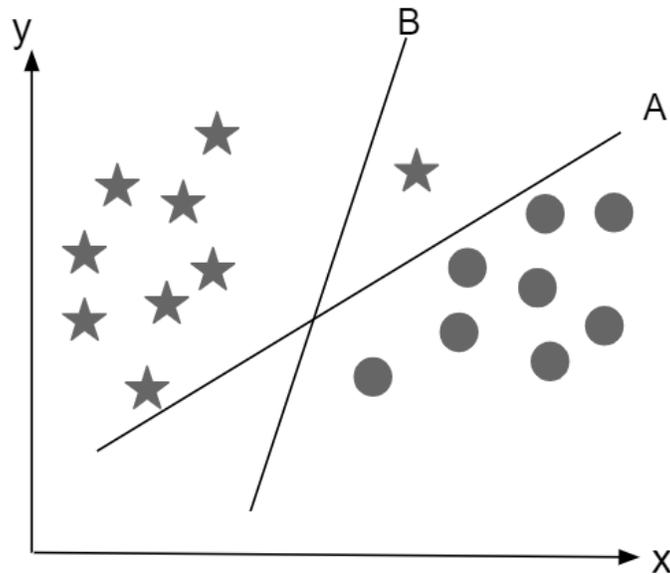


Figure 4.1.4: Correct hyperplane

Here, one might select B as the higher margin, but SVM selects the hyper plane, which classifies the classes accurately prior to maximizing margin. B has a classification error and A has categorized it all correctly. Therefore, right hyper plane is A.

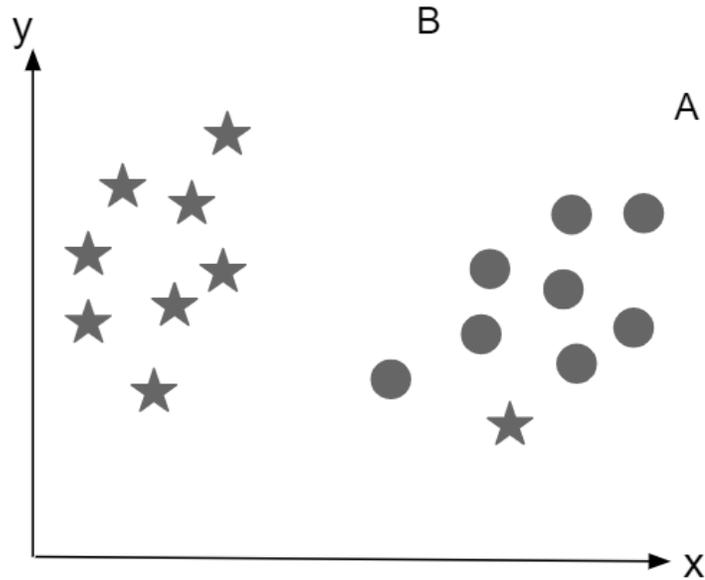


Figure 4.1.5: Hyperplane with outliers

Above, we cannot segregate two classes with a straight line, as one of the stars is in the territory of others. SVM has the feature of ignoring the exterior and looking for the highest-margin hyper plane. Therefore, it can be said that, SVM is robust to outliers.

In the scenario below, a linear hyper plane between two classes is not possible.

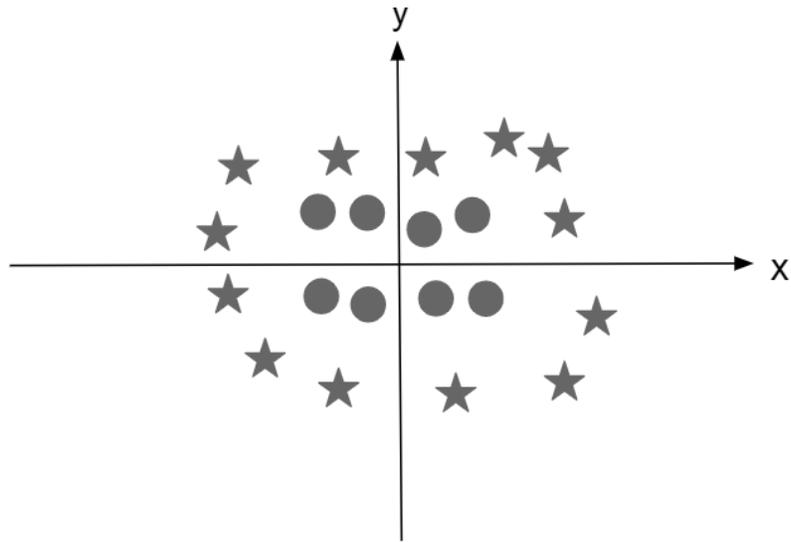


Figure 4.1.6: Hyperplane with maximum margin

This problem also can be solved by SVM. SVM introduces an additional feature $z=x^2+y^2$. If we plot the data points on axis x and z:

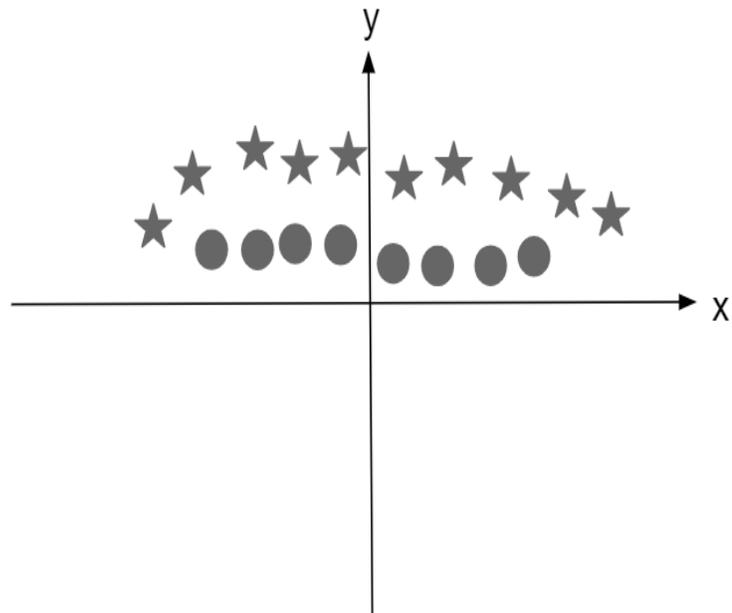


Figure 4.1.7: Solving with SVM

4.2 Logistic Regression

There are two types of regression analysis, one is linear regression and the other is logistic regression. We have used logistic regression. The logistic regression task is to find the best fitting model for the relationship between the dependent and the independent variables. Logistic regression deals with the probability to measure the relation between dependent and independent variable. Logistic regression helps for the classification like sentiment analysis. Logistic regression also called binary classification. For logistic regression analysis, we get a mathematical equation called a sigmoid or logit function. The logit function defined as:

$$\text{Logit}(z) = \frac{1}{1+e^{-z}}$$

Where, z is the usual data point of the dataset and e is the Euler's number which value is 2.71828 (Approx.). The sigmoid function will convert the input value from 0 to 1.

4.3 Random Forest:

We have used further classifier which is random forest classifier. Random forest classifier is another popular machine learning technique which used in both regression and classification. It is a kind of ensemble classifier which is using decision tree algorithm. This classifier effectively handle the large datasets and avoid the over-fitting.

4.4 Deep Learning Model

One more model we have used is sequential model, which is the deep learning model. Now, we are going to introduce sequential models with Keras. In Keras, essentially two types of models available. One is called sequential, and we use it to define sequential models which means we just stack layers one by one, sequentially. Another model is available in Keras, which is mainly used for non-sequential models and is run by the named model. Normally, these model uses the functional

API. Regardless of which model we define, the core abstraction to each Keras model is the notion of a layer. So each model is made up of a bunch of layers. In a sequential model, we stack the layers sequentially. So, each layer has unique input and output, and those inputs and outputs then also come with a unique input shape and output shape. If we want to retrieve the weights a layer has, we can simply call `get_weights` on layer and retrieve all those weights as a list of numpy arrays. If we want to set weights, that's also possible by simply calling `set_weights` with a given list of numpy arrays weights. Also each layer has its defining configuration, which we can get by calling `get_config` on a layer. To build a sequential model, we have to carry out a few steps. First one is, we instantiate a sequential model, then we add layers to it one by one. Afterwards, we need to compile a model with a mandatory loss function and a mandatory optimizer, and if we like, also with optional evaluation metrics such as accuracy. Afterwards, we fit our model to training data that we provide. And after that, it really depends what we want to do. Most of the time we evaluate our model but, sometimes we also want to maybe serialize or persist model or deploy it somewhere or start a new experiment altogether. It really depends on our use case. Essentially, this is the list of steps that we have to carry out each time, when we define a Keras model. Defining a loss function can be done in two ways. First, we can import a specific loss function from the Keras losses module. The second option is simply call the respective loss function by name. That is, we provide a string that stands for the loss function. If we are done compiling our model, we simply fit it to training features and labels and we also always have to specify a batch size, the number of epochs we want to train and optionally we can also specify validation data. Afterwards, we can evaluate our model or simply predict on new features, as we like. So, this is the high level overview of Keras sequential model.

CHAPTER 5

EXPERIMENT RESULTS AND DESCRIPTION

5.1 Introduction

We gathered and tagged 2254 headlines for our dataset. This number consists of all the negative, positive and neutral headlines. Before using the classifiers, we had to work on the dataset, so we can use it for creating the model. This was the head of the dataset before anything was changed.

1	বেঁচে গেলেন বিধ্বস্ত বিমানটির সবাই	Positive
2	বিমান চালালেন মেয়ে	Positive
3	হোমনায় ব্যবসায়ীকে পিটিয়ে হত্যা	Negative
4	রিয়াল-বার্সা দু'দলেরই হার	Negative
5	জাবালে নূরের সেই বাসের মালিক শাহাদৎ গ্রেফতার	Positive
6	নেত্রকোণায় হত্যা মামলায় এক পরিবারের ৩ জনের যাবজ্জীবন	Negative
7	রোনালদোর নতুন রুটিন	Positive
8	দোষীদের দ্রুত বিচারে দৃষ্টান্তমূলক শাস্তি দেওয়া হবে: আইনমন্ত্রী	Positive
9	এমপিওভুক্ত শিক্ষক-কর্মচারীদের জুলাইয়ের বেতন ব্যাংকে	Positive
10	উচ্চ রক্তচাপে বাড়ে ডিমেনশিয়ার ঝুঁকি	Negative
11	পলিথিন টাঙিয়ে অফিস করেন সমাজসেবা কর্মকর্তা!	Positive
12	যাত্রীর পেটে ১২ স্বর্ণবার	Negative
13	চট্টগ্রামেও রাস্তায় নেমে বিক্ষোভ শিক্ষার্থীদের	Negative
14	১০০ কোটির ক্লাবে জাহুবীর 'ধড়ক'	Positive
15	দাবি মানার আশ্বাস	Positive
16	জাবালে নূরের চালক ৭ দিনের রিমান্ডে	Negative

Figure 5.1.1: A small portion of the dataset

5.2 Experiment with logistic regression

The logistic regression model as a linear model that regression between 0 and 1 in mathematically. This model very effective for text classification like our work. We very effectively use this model and get the better accuracy which is almost 0.736.

5.3 Experiment with SVM

We used the same dataset that we used with logistic classifier to experiment with Support vector machine. Same dataset was used so that we can compare all the classifiers that would be used. Test accuracy for SVM Classifier was 0.56204379562044.

5.4 Experiment with random forest

Random forest model avoids over fitting. Random forest also give better accuracy which near by the logistic regression. This is the test accuracy 0.7267355982274741 we got in the case of random forest using this same data set.

5.5 Experiment with sequential model

The first deep learning model that we are about to develop is a simple deep neural network. This model actually works based on the layer. Our project done where we used dense layer. Dense layer means every node connected each nodes. When, we changed epochs then finally get the better result. Where test accuracy is 0.8035.

5.6 Comparative Analysis

Some of the classifier, feature vectors we've set are logistic regression, random forest and sequential model. We select these three classifiers because they give better accuracy. Then we used this classifiers in the train data set and using this data set, we measure the performance. After using all the above classifiers, we have measured performance on the data set. Based on the performance, we get accuracy on that data set. In performance analysis, we know that only accuracy can't be better for classification evolution. So, that's why we need some more metrics for classification evolution. A variety of performance metrics are used for the classification problem to evaluate the prediction. The easiest way to measure the performance of a classification problem is a confusion matrix that gives output 2 or more classes.

Table 1: Comparison

Classifiers	F1	Precision	Recall	Accuracy
Sequential	0.7910	0.8688	0.7589	0.8035
Logistic Regression	0.7256	0.8225	0.6986	0.7385
Random Forest	0.7215	0.8109	0.6897	0.7267
SVM	0.5394	0.6571	0.5396	0.5553

As we can see from the Table 1, Sequential Model has the highest F1 score, which indicates Sequential Model test was most accurate for this dataset. Sequential Model has the highest precision score too, and it indicates that the ratio of correct prediction vs search results of Sequential Model is higher than the other three classifiers.

CHAPTER 6

SUMMARY, CONCLUSION, RECOMMENDATION AND IMPLICATION FOR FUTURE RESEARCH

6.1 Summary of the study

In the summary, First of all, we collect a lot of Newspaper headlines. Then we created a data set. After that, we selected some models to train the data set. After reading a lot of research papers, finally, we selected some models. Then we train our data set with these models. We used 70% of the data as training data and 30% of the data used as test data. In our project, we have applied machine learning model as well as deep learning model. Deep Learning classifiers are more effective rather than Machine Learning Classifier and also give good accuracy.

6.2 Conclusions

In this research work, we tried to make a good model which can predict if a newspaper headline is positive or negative. We have collected a lot of newspaper headlines from various news portals. But we've only used 2254 newspaper headlines. Since we got very few neutral headlines rather than positive and negative headlines so we remove neutral headlines. Since our main concern is to check polarity between positive and negative news for this reason we delete neutral news.

6.3 Recommendations

In the case of our project, since we cannot work with very large data sets, we will have a recommendation that in the future, those who work with Bangla Newspaper should be able to take on more data. There will be another reminder that deep learning models work very well from machine learning, so without using machine learning I would say to use deep learning models.

6.4 Implication for further study

In the future, we want to work with larger data sets and Deep Learning wants to do the whole thing. One of the biggest problems in working with Bengali newspapers is the creation of data sets. Although we have collected the headlines of newspapers through coding, the But Main Problem is data tagging. Since data tagging process is the manual process so this is a boring part. That is why we will be working on a large data set in the future and after tagging all the data, we will publish the entire data set so that no one wants to do this related work and they do not have to face the problem with the data.

References:

- [1] <http://www.prothom-alo.com/> & [1] <http://www.Samakal.com/>
- [2] Mullen T. and Collier N., “Polarity prediction using support vector machines with diverse information sources”, National Institute of Informatics (NII), Japan
- [3] Gamallo P. and Garcia M., “Citius: A Naive-Bayes Strategy for Polarity prediction on English Tweets” , Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014), pages 171–175, Dublin, Ireland, August 23-24 2014.
- [4] Liu B., “Polarity prediction and Opinion Mining”, Morgan & Claypool Publishers, May 2012.
- [5] Kaur H. and Chopra V., “Polarity prediction of News Headlines using Naive Bayes Classifier”, International Journal for Science, Management and Technology
- [6] Chowdhury S. and Chowdhury W., “Polarity prediction for Bangla Microblog Posts”, BRAC University, Dhaka, Bangladesh
- [7] Kanayama H. and Nasukawa T., “Fully automatic lexicon expansion for domain-oriented Polarity prediction”. In Proceedings of the Conference on Empirical Methods in Natural Language Processing, Sydney, Australia, 2006.
- [8] Kobayashi N., Inui K., Tateishi K., and Fukushima T., “Collecting evaluative expressions for opinion extraction”. In Proceedings of IJCNLP 2004, pages 596–605, 2004.
- [9] Suzuki Y., Takamura H., and Okumura M., “Application of semi-supervised learning to evaluative expression classification”. In Proceedings of the 7th International Conference on Intelligent Text Processing and Computational Linguistics, 2006.
- [10] Takamura H., Inui T., and Okumura M., “Latent variable models for semantic orientations of phrases”. In Proceedings of the 11th Meeting of the European Chapter of the Association for Computational Linguistics, 2006.
- [11] Hu Y., Duan J., Chen X., Pei B., and Lu R., “A new method for sentiment classification in text retrieval”. In IJCNLP, pages 1–9, 2005.
- [12] Zagibalov T. and Carroll J., “Automatic seed word selection for unsupervised sentiment classification of Chinese text”. In Proceedings of the Conference on Computational Linguistics, 2008.
- [13] Das A. and Bandyopadhyay S., (2009a). “Subjectivity Detection in English and Bengali: A CRF-based Approach.”, In Proceedings of ICON 2009, December 14-17, 2009, Hyderabad.

- [14] Das D., Bandyopadhyay S., “Labeling emotion in Bengali blog corpus—a fine grained tagging at sentence level”, In Proceedings of the 8th Workshop on Asian Language Resources, pages 47–55, Beijing, China, August 2010.
- [15] Srinivasaiah M., Godbole N. and Skiena S., “Large Scale Sentiment Analysis for News and Blogs”, Stony Brook University, Stony Brook, NY 11794-4400, USA.
- [16] Wanner F., Rohrdantz C., Mansmann F., Stoffel A., Oelke D., Kristajic M., Keom D.A. Luo D., Yang J. and Atkinson M., “Large-scale Comparative Sentiment Analysis of News Articles”
- [17] Yan-Tak A., Jordan M., “On Discriminative vs. Generative classifiers: A comparison of logistic regression and naive Bayes”
- [18] Campbell, C., Jiang R., Wright W., Yan-Tak A, Ahres Y.,” Predicting Volatility in Equity Markets Using Macroeconomic News”.

Plagiarism report

ORIGINALITY REPORT

22%
SIMILARITY INDEX

13%
INTERNET SOURCES

12%
PUBLICATIONS

20%
STUDENT PAPERS

PRIMARY SOURCES

1	www.analyticsvidhya.com Internet Source	3%
2	Submitted to Daffodil International University Student Paper	3%
3	Submitted to October University for Modern Sciences and Arts (MSA) Student Paper	2%
4	Submitted to University of North Texas Student Paper	1%
5	Submitted to Coventry University Student Paper	1%
6	"Information and Communication Technology for Intelligent Systems", Springer Nature, 2019 Publication	1%
7	Mittelstädt, S., A. Stoffel, and D. A. Keim. "Methods for Compensating Contrast Effects in Information Visualization : Methods for Compensating Contrast Effects", Computer Graphics Forum, 2014.	1%
