A consequential ML intersection for figurative meaning detection using Bengali linguistic dataset

BY

Nahin Rowshon Dipty

ID:183-15-11806

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

Supervised By

Md.Azizul Hakim
Senior Lecturer
Department of Computer Science and Engineering
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

September 2022

APPROVAL

This Project/internship titled **"A consequential ML intersection for figurative meaning detection using Bengali linguistic dataset"**, submitted by **Nahin Rowson Dipty, ID No: 183-15-11806** 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 Bachelor of Science in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on **14th September**.

BOARD OF EXAMINERS

Dr. Touhid Bhuiyan — **Chairman**
Professor and Head
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Al Amin Biswas — **Internal Examiner**
Senior Lecturer
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

14.09.22

Mushfiqur Rahman — **Internal Examiner**
Senior Lecturer
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

14.9.22

Dr. Md Sazzadur Rahman — **External Examiner**
Associate Professor
Institute of Information Technology
Jahangirnagar University

DECLARATION

I hereby declare that, this project has been done by us under the supervision of Md.Azizul Hakim, Senior Lecturer, Dept. of CSE, Daffodil International University. We also declare that neither this project nor any part of this project has been submitted elsewhere for the award of any degree or diploma.

Supervised by:

Md.Azizul Hakim
Senior Lecture

Department of Computer Science and Engineering
Daffodil International University

Submitted by:

Nahin Rowshon Dipty

NRowson

Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

At first, I’m thankful to the almighty for showing me the path to achieve the honorable degree like B.Sc. in Computer Science & Engineering from Daffodil International University. Then I’m grateful to my parents for their support and good wishes forever.

I would like to thank my supervisor for his utmost support and guideline to successfully finishing this interesting research work on consequential Machine Learning intersection for figurative meaning detection using Bengali linguistic dataset. His help and guidance provided me with the courage to complete this research project profoundly. He helped me all related resources and information regarding to do this research from the scratch. I also thank my co-workers who supports me to shape up the dataset and other related tasks.

I am also humble towards my Co-Supervisor, for whose guidance was a massive contribution regarding the goal that was intended in achieving in this project.

ABSTRACT

Each human feeling may presently be connected to the writings we see on a day by day premise on different online stages. Such stages gives client inside the autonomy to share, connected with another clients suppositions and announcements in different terms and themes. With which case, it is basic for such settings to have had a framework that can recognize between honest to goodness feelings and threatening vibe. A prime illustration of how people employ imaginative linguistic methods in social communication is humor. In addition to exchanging information or conveying implicit meaning, humor fosters interpersonal connections among individuals who are exposed to it. It can assist people in separating themselves from tensions and assist them in finding the humorous side of issues. Additionally, it aids in controlling our emotions. Additionally, the ways in which people create humorous content offer information into their genre and character attributes. Which is why we have chosen to center our endeavors on distinguishing one of the foremost interesting sorts of all time: humor. For tasks like predicting sentiment polarity at the document level, we think an n-gram model in conjunction with latent representation will result in a more appropriate embedding. For the purpose of creating a quick and effective embedding for brief text segments, our suggested embedding mixes n-gram encoding with such a latent model. We utilized Direct Relapse, Choice Tree, Arbitrary Timberland, KNN, Multinomial Credulous Bayes, RBF SVM, and Direct SVM with in division strategy to get the most extreme exact findings. The findings which may well be fundamental in case of deciding between the cases between the conceptional intrigued for those crossing points of categorization of the animosity and veritable commenters who proposed for joke or per say humors conjunctions. The best accuracy was achieved by the Unigram was MNB (Multinumia Naiv bias) at 93%, by biagram it was same as 93% by MNB, with the triagram the feature with the MNB at 94% accuracy in toal.

TABLE OF CONTENTS

CONTENTS **PAGE**

CHAPTER

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

Introduction

1.1 Introduction

Within the sort of common dialect preparing, coming to feeling examination for request, where a major sum that's related to work done ahead of time. Such works are the foundational motivation for our work in humor location. Humor is essentially a state of intellect where somebody feels both interested by various comical references. Such as, Investigation of feeling is additionally called to the analysts [23] as opinion mining. So, within the case of humor, it may well be a joke, deride or a few kind of word which makes somebody snicker or grin. In different cases where the SA has been the device to extricate the most extreme important information from a strong sum of data's (24) Extremity Classification can be appeared to demonstrate much advantageous picking up such clarifications. Such as within the cases where deciding the emojis to urge an thought of one's deliberate behind utilizing them. Emojis can be demonstrated to have the thirst in later eras communications. Intentionally our youthful eras utilize them in case of uncovering their feelings. Modern communities now have simple, appealing means to express their opinions on a wide range of topics thanks to social media. This has led to the opening of new difficulties in data management, decision support tools, and human contact. Numerous studies have focused on coping with multilingualism, various genres and writing styles, and other qualitative communication tools (notions, feelings, attitudes, and views). However, when imagination and metaphorical language are employed in spoken and written communication, these issues become quite challenging. A human can comprehend the fundamental meaning of these expressions with ease, but a machine would need a lot more information to grasp the significance of artistic expression like irony and comedy.

1.2 Motivation

When it comes to NLP research, Natural language processing has always been a fascination towards scientists to build and work on the languages to understand and determine human behavior. The language is the only medium to understand this criteria. And that's why I wanted to work toward this research genre for the further progress. As for the humor, in many cases the humor becomes confused term with the toxicity. So, for the solemn distinction finding between toxic behaves, I become motivated to work on this peculiar topic.

1.3 Research Questions

- What is social networking?
- What is Humor comment classification?
- What is NLP?
- Why NLP is important?
- What is the effect of social networking?
- What is machine learning?
- Why machine learning is important?
- How to machine learning works?
- Which algorithm is better for Humor comment analysis?

1.4 Expected Output

The expected output was that, by using the model which is consists of many different algorithms which are to be trained and to be then tested by the dataset that contains humor and non-humor data's, the model should be specify to choose of which are humor and which are not. The algorithms used are mostly frequent while working with text data. As we know that the text data in Bengali are quite tough to deal with genuine machine learning models, in that case the dataset was preprocessed properly so that the outcome doesn't vary on the conditions which are the garbage values interfare.

1.5 Report Layout

The report has six chapters. Every chapter describes the different aspects of the " **A consequential ML intersection for figurative meaning detection using Bengali linguistic dataset**". Every chapter has different parts described in detail.

Chapter 1: Introduction

The inspiration is clarified and the proposition objective and introduction are presented.

Chapter 2: Background Studies

The applicable work is talked about and significant popular techniques are introduced corresponding related work.

Chapter 3: Research Methodology

Presents the information assortment, information pre-handling, and the element determination methodology.

Chapter 4: Design Specification

the philosophies for assessment grouping are clarified and the result discussed.

Chapter 5: Implementation

The 3-assessment plan, the precision assessment, and the investigation are introduced.

Chapter 6: Conclusion and Future Scope

The end is drawn and my commitments are portrayed.

CHAPTER 2

Background Studies

2.1 Introduction

Within the think about of [1], we have distinguished a certain feeling investigation such can be worn or finished in numerous diverse grounds such as spaces and dialects. Where in case of [2], they worn visual feeling examination, content feeling examination. Here geo-feeling disappointment encompassed by information equipment's in an adjoining geographic locale and watching sundry sentiments from see and strong information equipment's was sent for freeing the separate feeling were included to basic for the support to identical part territorial outline. Within the [3], We tried the matches of instruction work without a doubt by tolerating information over researchers content assessment and applying a lexicon-placed get to. Presently [4], we upheld LDA to remove the question of the store where the question is forward as the presence of the squabble of the various mark anticipation. We taken after a certain finest brew additionally the character of alpha, beta, fundamental credit, edge and perplexity. Presently within the [5], a few calculations in

2.2 Related Work

Within the case of [9], we expecting on an uncommon zone which is Bangladesh where people's stamp of see in Bengali dialects on social websites is by each minute. Presently [10], To commonly look at the Thai feeling of the purchaser's survey within the thing, taken a toll and conveyance ambit over utilize a multi- topographical lexicon and feeling recompense strategy we presented. Thai dialect tokenized for utilizing the longest planning with the calculation by a customer's study. It was broken descending to design its feeling. Feeling stipend strategy tried to thus reimburse the feeling to an perspective where a shopper's look characterizes the feeling with absent a Sentiment-strength calculation credit by understanding feeling extremity and quality of a book [11]. It boosts along characteristic unpredictable moving elucidation, connected to segregation and hostility. By explored tweets the is covers reveal a winning strong dissent feeling. For allotment here wore Gullible Bayes Classifier and the endeavor

2.3 Research Summary

The dataset that was used in the term of experiment, It was collected randomly from different websites and platforms from internet. As for the datasets pre-processing terms, the unnecessary data, the numeric values, the special characters were removed in case of getting the full and proper result in the term of detection. In the dataset there were in some cases where there were repeat cases of some values and texts, that were thoroughly checked and removed for the sake of performance.

2.4 Scope of the problem

This research presents and talks about examinations in supposition investigation of Twitter presents relating on U.S carrier organizations. The objective of this examination is to decide if tweets can be ordered either as showing the positive, negative, or impartial notion. Online life permit clients to communicate and share their data, thoughts and suppositions through informal communication destinations. Carrier tweets are getting well known and utilized as a dataset to check the worries of the clients. In this paper, the specialists grouped extremity through explanation utilizing classifiers KNN, Logistic Regression and Random Forest to build up a model. The specialists assembled tweets of nearby aircraft concerning their involvement with the administrations gave by neighborhood carriers in the Philippines. The analysts likewise decided the estimation of a tweet on the off chance that it is sure, nonpartisan or negative and gave the quantitative and subjective investigations, just as conclusion examination, to more readily comprehend the consequence of the trial. It would help people a step ahead of dynamicity and automation in this field of observation.

2.5 Challenges

In this paper, we have contemplated the slant examination dependent on the inputs of voyagers in regards to carrier organizations. Our proposed approach demonstrated that both element determination and over-inspecting procedures are similarly significant with respect to boosting our outcomes. The utilization of highlight choice strategies has restored the best subset of highlights and decreased the calculations expected to prepare our classifiers. Though, it has decreased the slanted appropriation of the classes found in a large portion of our littler datasets without causing overfitting. Our outcomes are convincing proof that the proposed model has high grouping precision in anticipating occasions structure the three classes toxic, severe toxic, positive, severe positive toxic comment, and normal comment. It was also a challenging task to manage Bengali text and make it processed for model training. As can be seen, a portion of the applied classifiers has beaten the others.

CHAPTER 3

Research Methodology

3.1 Introduction

Here in the section, we are going to briefly describe the procedure we went over to complete our research work. We variously wore Python Programming language to conceal the elaborated machine learning algorithms. The algorithms were selected and placed on our dataset frequency, because when it comes to Natural Language Processing, the dataset has the most crucial part in the full procedure. The algorithms we selected were Linear Regression, Decision Tree, Random Forest, Multinomial Naive Bayes, KNN, Linear SVM, RBF SVM. These are all profound machine learning techniques which had a preferred result as estimated. In our research, our goal was to access a rare rule placed algorithm. In fig 1, we have shown the process by which we contemplate the procedure visualizing a flow chart. Technically, a mappings () is applied to strings of varying lengths into the a training dataset of constant size in the basic bag-of-words interpretation of text. A common classifier, such the linear autoencoder or support vector machine, can be used in this area. Let S stand for the set of any and all finite duration word vectors from D, and let D stand for the underpinning vocabulary. To indicate a set's cardinality, we use the symbol |.| The abovementioned mapping, S RM, which maps sequences of words in S to a finite dimensional feature space, will also be assumed. The mood labels will be organized into a set called Y = 1,..., K, where K = 2 designates types of sentiment, including such "humor" or "no humor" A labelled training set containing adapt new from Y would also be referred to as (xi, yi) I = 1,..., L | xi = S, yi = Y If wj D, then a text entry string of fixed Length will indeed be represented as x = (w1,...,wN). Let stand for the corpus's n-gram vocabulary, and let j = (wj, wj+1,..., wj+n1) represent the j-th location in x. Given the description as a bag of unigrams, the function (x) naturally transfers the input x to a (sparse) vector with dimensions M = |D|. The mapping from x to a M = ||-dimensional form with || = O(|D|n) is equivalent in a bags-of-n grams (x). We acquire a "embedding" for each phrase, which is a mapping of the each phrase into a realvalued subspace, because we know that each word contains a lot of semantic information. Specifically, Every word from wj to D has an m-dimensional element incorporated in it. utilizing a lookup table LTE(), space is described as

$$LT_E(w_j) = E \times (0, \ldots.,0, 1_{at\ index\ wj}, \ldots.,0) = E_{wj} = [E_{1,WJ}, E_{2,WJ}, \ldots., E_{m,WJ}]$$

where E Rm|D| is a matrix that contains learning-to-be-done word - based variables. Here, Ewj Rm represents the word wj's lexicon embedding, while m stands for the target word's encoding dimensionality. It's crucial to remember that backpropagation is used during the education process to continuously train the variables of E.

3.2 Research Subject and Instrumentation

My recommended subject name is " Humor Detection ". It can be attained as a consequential research sense in Natural Language Processing. I have scrutinized a path in creating estimation investigation in Bangla with the determined and theoretical method. A noteworthy ML model needs a high structure pc and various instruments. The instance a concept analyzer is given beneath the fundamental required hardwares for this model.

Tools:

- Windows 10
- Python 3.8
- Numpy
- Pandas
- NLTK
- Jupyter Notebook

3.3 Data Collection

- Our collected amount of a certain dataset contains 2103 data's which build up sentences. The sentences are profoundly humor collected from numerous sources such as social media, books, television shows, Jokes etc. The dataset was divided into two different classes which are humor and non-humor. Based on class differences the dataset distribution is shown below

Table 1.0: Number of data distribution by Classes

Class Name	Amount
Humor Data	1099
Non-Humor Data	1004

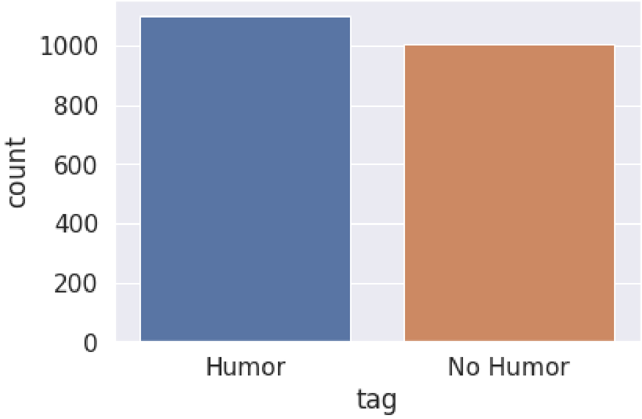


Fig 1.1: Data statistics Diagram

3.4 Dataset Distribution

After all the cleaning and clearing of all the untamed data the dataset finally reached the phase where it was meant to be distributed into classes for train and test. The total amount of data was splinted into an 80/20 ratio. The amount of train data size was 1683 and the amount of test data size was 420.

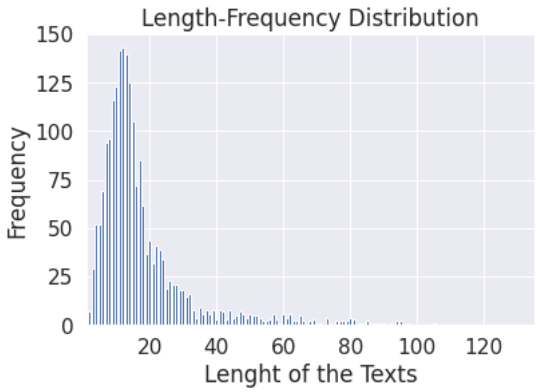


Fig 1.2: Dataset length frequency distribution

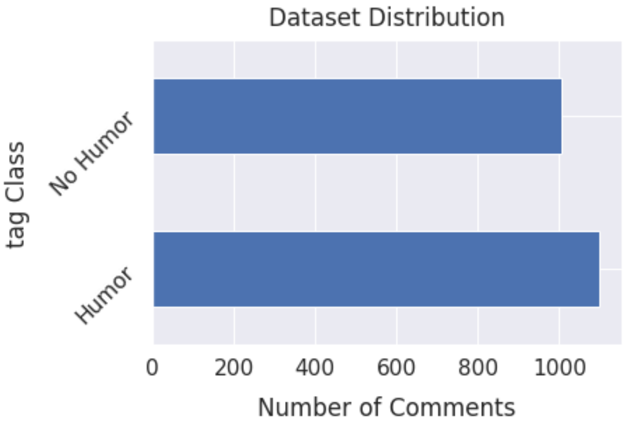


Fig 1.3: Dataset Distribution based on sentiment class

3.5 Preprocessing

As for preparing the dataset for the train case of the algorithms, we needed to refresh the whole dataset by deleting the garbage character. Which are special characters ("!", "@", "#", "$", "%", "^", "*") numerical character (1, 2, 3, 4, 5, 6, 7, 8, 9, 0) English Alphabets (A to Z both capital and small letter format), White space and duplicate character. The duplicate character could be also retained as the character a certain got reconciled repeatedly while accepting the data set. It is important that the dataset is raw in case of training the machine, so it can certainly apprehend the difference of the classes.

3.6 Stop word remove

A stop word is an ordinarily utilized word, for example, ".", ",", "’", "|" etc. that such a web engine has indeed been configured to ignore, both when arranging parts for searching and when retrieving them as a result of a pursue enquiry I wouldn't need these terms to take up space in our database or take up a lot of processing time. For this, I may effectively evacuate them by storing a list of terms that we regard to be stop words. In Python, the NLTK (Natural Language Toolkit) provides a list of stop words stored in 16 different dialects. They may be found in the nltk data index.

3.7 Tokenization

Tokenization means breaking downward the altercation regarding from the sentence, and these singular altercation are called tokens. Tokenization is important in such processes for Training the data’s to the algorithm. Such an example could be made as “হাত দিয়ে চানাচুর মেখে চামচ দিয়ে খাবার নামই বড়লোকি” in such context each of the individual parts of speech “হাত” , “দিয়ে” and others would be renowned as tokens. As for the labelling case, the dataset was apart into two different classes, which are the humor and non-humor.

3.8 Statistical Analysis

1. In the dataset total 2103 data is presented.
2. The dataset is divided into 2 classes.
3. 1683 data is used for the train.
4. 420 data is used for the test.
5. Highest accuracy achieved 96%.

3.9 Implementation Requirements

For the implementation machine learning model, I used python language as a programming language. Loading the dataset panda’s library is used and NLTK library for preprocessing. Total implementation is built in a jupyter notebook environment.

CHAPTER 4

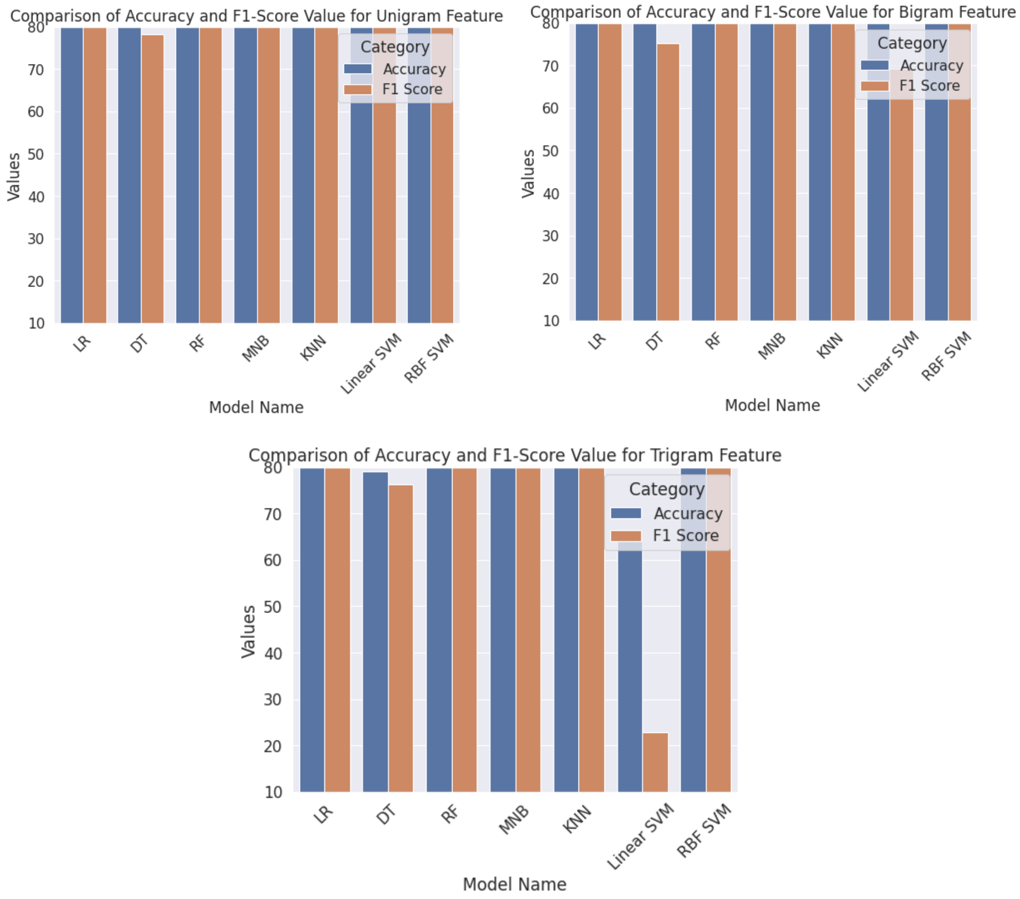
Experimental Results and Discussion

4.1 Introduction

The algorithms were able to gain a very volatile and appreciating level of result in the case of our detection of humor from non-humor. The graph below shows the Unigram, Bigram and Trigram of the humor detection by our selected algorithms. The algorithms were selected and placed on our dataset frequency, Because, when it comes to Natural Language Processing, the dataset has the most crucial part in the full procedure. The algorithms we selected were Linear Regression, Decision Tree, Random Forest, Multinomial Naive Bayes, KNN, Linear SVM, RBF SVM. These are all profound machine learning techniques which had a preferred result as estimated.

4.2 Model Performance

As for the unigram feature, The highest accuracy was achieved by MNB at 93%, The highest F1-score was achieved by MNB at 92, The highest precision score was achieved by RBF SVM at 92 and the highest recall score was also achieved by the MNB which was 93. For the Bigram performance features, the highest accuracy was achieved by the MNB at 93%, The highest F1 score was also achieved by MNB at 92, The highest precision score was achieved by the Linear SVM at 96, Highest recall score was also achieved by the MNB at 93. Now for the final Trigram feature for our work, the highest accuracy was achieved by the MNB at 94%, The highest F1 score was also achieved by MNB at 92, The highest precision score was achieved by the Linear SVM at 98, Highest recall score was also achieved by the MNB at 90.



1.3 Fig model performance

Model Name	Accuracy	Precision	Recall
Linear Regression	91.47	90.36	88.24
Decision Tree	82.46	78.57	77.65
Random Forest	87.20	88.18	80.00
MNB	93.36	90.00	92.94
KNN	85.78	84.81	78.82
Linear SVM	90.05	92.11	82.35
RBF SVM	91.00	92.31	84.71

4.4 Summary

There was a lot of texts which were unable to comprehend when it was cleaned out of special characters, as those were also responsible for the evaluation in the case of humor expression. In those cases I needed to determine different approach by re-processing them into more humor like context.

CHAPTER 5

Impact on Society, Environment and Sustainability

5.1 Impact on Society

In the modern era, every human emotion can be related to the texts we see on a daily basis on numerous internet platforms. In which motion, it is fundamental for these platforms to have a system to determine which are actual emotions and which are aggression placed. This is why we have decided to draw our work towards detecting one of the most interesting genres of all time, known as Humor. By this kind of work, in hope to build a more conclusive and variance free internet era.

5.2 Impact on Environment

The widespread dissemination of false information via social networking sites creates harm. You should be more cautious if you're an individual or corporation that provides a lot of content, perhaps with the help of internet-based life programming. It just stops for a few moments to check anything you've seen on the internet. Consider the story's beginning. Whether you're unfamiliar with it, look it up on the internet to see if it's credible. If you don't have that much time, it's best to ignore it, especially if it looks like satire, deceptive material, or deliberate publicity. You may help to reduce the spread of deceit and fake news by refusing to share faulty information.

One of the key advantages of a web-based existence, as mentioned before, is the ability to transmit information to a large number of people in a short period of time. While this may be a huge benefit in a crisis, it can also be a huge hindrance because material with no validity can be transmitted in a matter of seconds. This may lead to a lot of misinformation and a lot of hysteria. When rumors arose that the Queen had died because to the Queen missing a Christmas administration due to a common sickness, this was an example of this. This, combined with several hoaxers fabricating fake news reports, led many to believe the Queen had passed away.

5.3 Ethical Aspects

Online networking allows you to find meetings that are focused on your advantages and pastimes, making it one of the best ways to meet and interact with new people who have similar interests to you. This is fantastic for meeting new people, as well as for finding love interests and web dating, which has become more popular than a traditional face-to-face meeting as a result of online life and Tinder-like apps.

The internet is a fantastic way to convey news quickly throughout the world, with "breaking news" tweets receiving thousands of retweets in seconds. When it comes to important data, such as climate updates and missing children, this may be really helpful.

As previously said, internet networking has impacted society in a variety of beneficial ways with no additional cost, as all key web-based life phases are available for free. Consider another product or service that has ever changed your life as much as the internet, and then think about how much it costs.

5.4 Sustainability

- There are generally speaking 2.3 billion overall dynamic internet-based life clients.
- 91% of huge corporate brands have at least two internet-based life stages.
- 65% of all people feel awkward and uncomfortable when they can't access their online life profiles.

CHAPTER 6

Conclusion and Future Work

6.1 Conclusion

In the vast world where emotional intelligence is getting broader alignment with all the commotion between our daily life structure and habits, we think that this work of ours is capable of adding a new addition to the work's conception of the ongoing developmental era of BNLP. The emotion detection has always remained a hot elaborated topic for the researchers in this genre, we hope to inspire others with our work in this contribution. So that we could build a future where technology would help one's mind to have calamity.

6.2 Recommendations

Experimental analysis is the most recent pattern to comprehend the requirements of the mass open; it's a simpler and the savvy approach to see how the individuals are feeling about a specific subject of issue and the brand effect of smaller scale blogging. In this situation, we had thought about the assessment of the individuals towards the aircraft business and handled the ongoing issues of United Airlines and how the open feel about it. The examination affirmed our presumption on how powerful a way to deal with twitter assumption investigation is. The Logistic Regression and Random Forest classifier utilized in the calculation, alongside two programming for better outcomes portray plainly the assumption of the mass group and subsequently, the aircraft could without much of a stretch decipher the information and advantage from it by attempting to enhance the angles that appear to be negative or are loathed by the directed crowd. There are some recommendations for this work such as,

- Remove bias from dataset.
- Need the ratio of data categories is equally distributed.
- Use more machine learning classification algorithms.
- Create neural network for better result.
- Need parameter tuning for classification.

6.3 Implication for Further Study

The touchy development of information on the web and internet-based life locales permits organizations to utilize conclusion examination to pick up understanding into customers sentiments about their items or administrations. Toxic comment investigation in existing writing is frequently performed dependent on the restricted social media information utilizing a couple of days' worth of information. Except if social media information is as a rule constantly downloaded this impediment forestalls factually noteworthy and important outcomes being acquired. A far-reaching investigation of tweets to determine factually huge client assessment needs to address a few measures. These incorporate, at the base, 1) an adequately significant stretch of time over which tweets are gathered to guarantee representativeness as opposed to individuals' prompt response following a bit of news the occasion, 2) an adequate number of tweets that best speak to every geographic area, 3) a gauge of conceivable inclination if the tweets start from a given geographic area and 4) if the ends drawn match the prominent attitude got from other market sources.

REFERENCES

[1] Sabra, K. S., Zantout, R. N., El Abed, M. A., & Hamandi, L. (2017, September). Sentiment analysis: Arabic sentiment lexicons. In *2017 Sensors Networks Smart and Emerging Technologies (SENSET)* (pp. 1-4). IEEE.

[2]Alfarrarjeh, A., Agrawal, S., Kim, S. H., & Shahabi, C. (2017, October). Geo-spatial multimedia sentiment analysis in disasters. In *2017 IEEE International Conference on Data Science and Advanced Analytics (DSAA)* (pp. 193-202). IEEE.

[3] Aung, K. Z., & Myo, N. N. (2017, May). Sentiment analysis of students' comment using lexicon based approach. In *2017 IEEE/ACIS 16th international conference on computer and information science (ICIS)* (pp. 149-154). IEEE.

[4] Bashri, M. F., & Kusumaningrum, R. (2017, May). Sentiment analysis using Latent Dirichlet Allocation and topic polarity wordcloud visualization. In *2017 5th International Conference on Information and Communication Technology (ICoIC7)* (pp. 1-5). IEEE.

[5] Zhang, X., & Yu, Q. (2017, September). Hotel reviews sentiment analysis based on word vector clustering. In *2017 2nd IEEE International Conference on Computational Intelligence and Applications (ICCIA)* (pp. 260-264). IEEE.

[6] Ikoro, V., Sharmina, M., Malik, K., & Batista-Navarro, R. (2018, October). Analyzing sentiments expressed on Twitter by UK energy company consumers. In *2018 Fifth international conference on social networks analysis, management and security (SNAMS)* (pp. 95-98). IEEE.

[7] Vanaja, S., & Belwal, M. (2018, July). Aspect-level sentiment analysis on e-commerce data. In *2018 International Conference on Inventive Research in Computing Applications (ICIRCA)* (pp. 1275-1279). IEEE.

[8] Zvarevashe, K., & Olugbara, O. O. (2018, March). A framework for sentiment analysis with opinion mining of hotel reviews. In *2018 Conference on information communications technology and society (ICTAS)* (pp. 1-4). IEEE.

[9] Mahtab, S. A., Islam, N., & Rahaman, M. M. (2018, September). Sentiment analysis on bangladesh cricket with support vector machine. In *2018 international conference on Bangla speech and language processing (ICBSLP)* (pp. 1-4). IEEE.

[10] Porntrakoon, P., & Moemeng, C. (2018, July). Thai sentiment analysis for consumer’s review in multiple dimensions using sentiment compensation technique (SenSecomp). In *2018 15th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON)* (pp. 25-28). IEEE.

[11] Dilai, M., & Levchenko, O. (2018, September). Discourses Surrounding Feminism in Ukraine: A Sentiment Analysis of Twitter Data. In *2018 IEEE 13th International Scientific and Technical Conference on Computer Sciences and Information Technologies (CSIT)* (Vol. 2, pp. 47-50). IEEE.

[12] Naf'an, M. Z., Bimantara, A. A., Larasati, A., Risondang, E. M., & Nugraha, N. A. S. (2019). Sentiment Analysis of Cyberbullying on Instagram User Comments. *Journal of Data Science and Its Applications*, *2*(1), 38-48.

[13] Rabeya, T., Chakraborty, N. R., Ferdous, S., Dash, M., & Al Marouf, A. (2019, February). Sentiment analysis of Bangla song review-a lexicon based backtracking approach. In *2019 IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT)* (pp. 1-7). IEEE.

[14] Dhawan, S., Singh, K., & Chauhan, P. (2019, October). Sentiment analysis of Twitter data in online social network. In *2019 5th International Conference on Signal Processing, Computing and Control (ISPCC)* (pp. 255-259). IEEE.

[15] Ramanathan, V., & Meyyappan, T. (2019, January). Twitter text mining for sentiment analysis on people's feedback about oman tourism. In *2019 4th MEC International Conference on Big Data and Smart City (ICBDSC)* (pp. 1-5). IEEE.

[16] Lee, J. S., Zuba, D., & Pang, Y. (2019, April). Sentiment analysis of Chinese product reviews using gated recurrent unit. In *2019 IEEE Fifth International Conference on Big Data Computing Service and Applications (BigDataService)* (pp. 173-181). IEEE.

[17] Wongkar, M., & Angdresey, A. (2019, October). Sentiment analysis using Naive Bayes Algorithm of the data crawler: Twitter. In *2019 Fourth International Conference on Informatics and Computing (ICIC)* (pp. 1-5). IEEE.

[18] Gupta, S., Lakra, S., & Kaur, M. (2019, February). Sentiment Analysis using Partial Textual Entailment. In *2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon)* (pp. 51-55). IEEE.

[19] Gupta, S., Lakra, S., & Kaur, M. (2019, February). Sentiment Analysis using Partial Textual Entailment. In *2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon)* (pp. 51-55). IEEE.

[20] Li, Z., Li, R., & Jin, G. (2020). Sentiment analysis of danmaku videos based on Naïve Bayes and sentiment dictionary. *IEEE Access*, *8*, 75073-75084.

[21] Poornima, A., & Priya, K. S. (2020, March). A comparative sentiment analysis of sentence embedding using machine learning techniques. In *2020 6th international conference on advanced computing and communication systems (ICACCS)* (pp. 493-496). IEEE.

[22] Baskar, M., Ramkumar, J., Rathore, R., & Kabra, R. (2020). A deep learning based approach for automatic detection of bike riders with no helmet and number plate recognition. *Int J Adv Sci Technol*, *29*(4), 1844-1854.

[23] Baskar, M., Ramkumar, J., Reddy, V. V., & Reddy, G. N. Cricket Match Outcome Prediction using Machine Learning Techniques. *International Journal of Advanced Science and Technology*, *29*(4), 1863-1871.

[24] Biswas, R. (2021). Sentiment Analysis on National Education Policy Change 2020. *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, *12*(11), 1480-1488.