Sentiment Analysis of User-Generated Reviews of Women Safety Mobile Applications

BY

Afrin Jaman Bonny
ID: 181-15-10768
AND
Mehrin Jahan
ID: 181-15-10895
AND
Zannatul Ferdhoush Tuna
ID: 181-15-10918

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

Supervised By

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

Co-Supervised By
Ahmed Al Marouf
Senior Lecturer
Department of CSE
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

JANUARY 2022

APPROVAL

This Project titled **"Sentiment Analysis of User-Generated Reviews of Women Safety Mobile Applications"**, submitted by Afrin Jaman Bonny, ID No:181-15-10768, Mehrin Jahan, ID No:181-15-10895 and Zannatul Ferdhoush Tuna, ID No:181-15-10918 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 5th January, 2022.

BOARD OF EXAMINERS

Chairman

____

Dr. Touhid Bhuiyan (DTB)
Professor and Head
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner

____

Md. Riazur Rahman (RR)
Assistant Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner

____

Md. Ohidujjaman Tuhin (MOT)
Assistant Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

External Examiner

____

Shah Md. Imran
Industry Promotion Expert
LICT Project, ICT Division, Bangladesh

DECLARATION

We hereby declare that this project has been done by us under the supervision of Shah **Md Tanvir Siddiquee**, **Assistant Professor, Department 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:

____

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

Co-Supervised by:

____

Ahmed Al Marouf
Senior Lecturer
Department of CSE
Daffodil International University

Submitted by:

Afrin Jaman

____

Afrin Jaman Bonny
ID: 181-15-10768
Department of CSE
Daffodil International University

Mehrin

____

Mehrin Jahan
ID: 181-15-10895
Department of CSE
Daffodil International University

Zannatul Ferdhoush Tuna

____

Zannatul Ferdhoush Tuna
ID: 181-15-10918
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

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

We are really grateful and wish our profound indebtedness to **Shah Md Tanvir Siddiquee**, **Assistant Professor,** Department of CSE Daffodil International University, Dhaka. Deep Knowledge & keen interest of our supervisor in the field of "*Machine Learning and NLP (Natural Language Processing*" to carry out this project. His endless patience, scholarly guidance ,continual encouragement , constant and energetic supervision, constructive criticism , valuable advice, reading many inferior drafts and correcting them at all stages have made it possible to complete this project.

We would like to express our heartiest gratitude to Dr. Touhid Bhuiyan, professor and Head, Department of CSE, for his kind help to finish our project and also to other faculty members 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 discussion while completing the course work.

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

ABSTRACT

Google Play Store is basically an app store from where we get various kinds of applications for our android certified devices which makes our life a lot easier and faster through the diverse functionalities the apps contain. Numerous users are using applications as per their needs and putting their experience, thoughts of using that application via reviews in form of ratings and texts. As the safety of women is threatened, whether or not applications like women's safety apps are appreciated, can be detected through text reviews and ratings by the users. In this study, we try to analyze the polarity (positive, negative, neutral) of the sentences or text reviews that are given by the users of the women's safety app through the google play store. To detect the emotions of the users through the given text reviews and star ratings, different machine learning (ML) and deep learning methods using natural language processing (NLP) are conducted to analyze the sentiments of the review given by the users. For this study, we have collected data from the app reviews and star ratings provided by the users of the women's safety related applications whose main purpose is to provide necessary functionality that can keep women safe in any dangerous and unwanted situation. The purpose of this paper is to mine the opinion of the users and get their viewpoint about those apps of specific purpose whether it is positive, negative, or neutral. As the current user's ratings, reviews, or as a whole their viewpoint helps the new user understand the performance of the applications and insights in advance, so the mining of their opinion is helpful for both parties - developers and general users. To detect the level of the sentiment, we have applied several supervised machine learning algorithms namely Multinomial Naive Bayes (MNB), Logistic Regression (LR), Support Vector Machine (SVM), and k-nearest neighbor (K-NN) as well as unsupervised deep learning algorithm Bidirectional Encoder Representations from Transformers (BERT). Among these algorithms, the BERT has outperformed all other algorithms in terms of accuracy (86.06%).

TABLE OF CONTENTS

CONTENTS	PAGE
Board of examiners	i
Declaration	ii
Acknowledgements	iii
Abstract	iv
Table of content	v-ix
List of chapters	v-vii
List of figures	vii
List of tables	ix

LIST OF CHAPTERS	PAGE
CHAPTER 1: INTRODUCTION	**1-6**
1.1 Introduction	1-3
1.2 Motivation	3
1.3 Rationale of Study	3-4
1.4 Research Questions	4
1.5 Expected Output	5
1.6 Project Management and Finance	5

1.7 Report Layout	6
CHAPTER 2: BACKGROUND	**7-15**
2.1 Preliminaries/Terminologies	7
2.2 Related Works	7-11
2.3 Comparative Analysis and Summary	11-15
2.4 Scope of the Problem	15
2.5 Challenges	15
CHAPTER 3: RESEARCH METHODOLOGY	**16-29**
3.1 Research Subject and Instrumentation	16-18
3.2 Data Collection Procedure	18-19
3.3 Statistical Analysis	19
3.4 Proposed Methodology/Applied Mechanism	20-29
3.5 Implementation Requirements	29
CHAPTER4: EXPERIMENTAL RESULTS AND DISCUSSION	**30-36**
4.1 Experimental Setup	30
4.2 Experimental Results & Analysis	30-35

4.3 Discussion	36
CHAPTER 5: IMPACT ON SOCIETY,ENVIRONMENT AND SUSTAINABILITY	**37-39**
5.1 Impact on Society	37
5.2 Impact on Environment	38
5.3 Ethical Aspects	38
5.4 Sustainability Plan	38-39
CHAPTER 6: SUMMERY, CONCLUSION, RECOMMENDATION AND IMPLICATION FOR FUTURE RESEARCH	**40-41**
6.1 Summary of the Study	40
6.2 Conclusions	40-41
6.3 Implication for Further Study	41
REFERENCES	**42-43**

LIST OF FIGURES

FIGURES	PAGE NO
Figure: 1.1.1 Some of the Icons of the Women Safety Related Mobile Applications	2
Figure: 3.2.1 Procedure of Collecting Data Manually	18
Figure: 3.2.2 Tabular Form Our Dataset	19
Figure: 3.4.1 Proposed Methodology of Research Model	20
Figure: 3.4.2 Scatter Plot of the Sentiments	22
Figure: 3.4.3 Hierarchy of Applied Algorithms	23
Figure: 3.4.4 Input Tokens in BERT	26
Figure: 3.4.5 Architecture of BERT	26
Figure: 4.2.1 Plot of Star Review	30
Figure: 4.2.2 Plot of the Review Sentiment (Positive, Neutral, Negative)	31
Figure: 4.2.3 Word Cloud of the Reviews	32
Figure: 4.2.4 Confusion Matrix of Multinomial Naïve Bayes (MNB)	32
Figure: 4.2.5 Confusion Matrix of K-Nearest Neighbour (KNN)	33
Figure: 4.2.6 Confusion Matrix of Support Vector Machine (SVM)	33
Figure: 4.2.7 Confusion Matrix of Logistic Regression (LR)	34
Figure: 4.2.8 Confusion Matrix of BERT	34
Figure: 4.2.9 Comparison Between Applied Algorithms	35

LIST OF TABLES	
Tables	**Page No**

Table: 1.6.1 Timeline of the Research Management	5
Table: 2.3.1 Comparative Analysis of the Related Publications	11-14
Table: 3.1.1 Information of the Safety Related Applications	16-18
Table: 3.3.1 Statistical Insight of the Dataset	19
Table: 4.2.1 Evaluation metrics of applied algorithms	35

CHAPTER 1

Introduction

1.1 Introduction:

To determine the polarity of a sentence which means to find out if a sentence or writing is positive or negative or neutral we use Sentiment analysis. The sentiment analysis technique unifies machine learning algorithms along with deep learning algorithm using natural language processing to categorize and distribute any given sentence or sentences according to sentiment ratings, themes, and other factors. This strategy is used to understand a huge number of public sentiments, monitor brand reputation, and track product status on the market [1].

Sentiment analysis mainly emphasizes text polarity, which means it looks for positive, negative, and neutral opinions, but it can also detect other emotions like anger, happiness, emergency or not-so-emergency, interesting or not-so-interesting, and so on. We identified the most commonly utilized text polarity category for this study [2].

Women's safety has been a major concern for decades, and even currently, the world's most secure locations aren't secure enough for women. We are living in a century in which the world is taking significant strides toward women's empowerment, assisting women in areas such as education, acceptance of their opinions, raising awareness, and supporting in the advancement of their status. But still now women of all ages are victimized to various forms of violence at home, in the workplace, at school, on the road, and elsewhere. However, because we live in a technological age, numerous mobile application developers have released developed applications connected to safety, particularly for women's safety. According to studies, around 61 percent of women take routine efforts to prevent sexual assault or violence [3].

There are a variety of mobile applications accessible in this technological age to make our lives simpler, safer, and less stressful. Many developers have worked to develop applications that may be used to give safety to women on various digital distribution services, as women's safety is a major priority. Many apps related to women's safety are accessible on Google Play Store, which has been a popular platform to download and purchase for Android users.

Women Only Community bSafe - Never Walk Alone

Disha SOS Women Safety

Figure: 1.1.1 Some of the Icons of the Women Safety Related Mobile Applications

For our study, we employed sentiment analysis, which is a type of text analytics that assesses a user's attitude about the components of a service that the "women's safety app" provides through the reviews they have given in the google play store applications. The reviews are extremely beneficial to future app users in terms of familiarizing themselves with the app's quality and offerings.

Sentiment analysis based on aspects is a fantastic approach that focuses on the polarity of phrases as well as their feelings. New users like to read previously submitted reviews and base their decision to download a certain application on these reviews.

Sentiment detection is a technique to determine a user's opinion on certain kinds of applications. It identifies the text content's polarity [4]. In this research, we will utilize the star rating to examine the polarity (positive, negative, and neutral) of the words or text evaluations provided by users of the various women's safety related apps [5]. For this purpose, we have used four different machine learning classifiers which are Multinomial Naive Bayes (MNB), Support Vector Machine(SVM), Logistic Regression(LR), K-Nearest Neighbor (KNN), and a deep learning algorithm which is

Bidirectional Encoder Representations from Transformers (BERT). This will assist in detecting the amount of accuracy from the collection of data and will help to obtain real accuracy.

1.2 Motivation:

In terms of women's safety, our nation lags significantly behind other industrialized countries such as the United States, the United Kingdom, Australia, etc. It is really unfortunate that our country is in such a bad state when it comes to creating a safe atmosphere for women. We have to hear of a woman's awful harassment, rape, or murder on a daily basis. Many developers from many nations have created a variety of mobile applications that may be utilized in an emergency when a person, particularly a woman, feels uncomfortable and is experiencing a safety crisis. Such applications are very much welcomed and greatly required in our country.

Many developers have created mobile applications for this reason, but it is a question of research to discover how the users of these applications feel about them, which will significantly assist the developers in understanding their emotions and needs. In today's data-overloaded environment, it's impossible to tell the user's reaction by reading each review one by one. Our motivation is to determine the user’s experience in using these applications and help developers find more ways to improve their work in this domain of security applications. So, in order to determine the users' general impression, we divided them into three groups: those who are optimistic about the app’s service, those who do not believe their service is good enough to offer security, and those who are unbiased about the applications accessible.

1.3 Rationale of the Study:

This study topic analyzes the feedback received from users of this review and can be used to encourage the development of similar applications, motivating developers to enhance their work until it reaches its highest degree of satisfaction. There are various women's safety applications accessible for download and usage on digital platforms, but many individuals are ignorant of them, are unsure of the level of security they give, and wish to learn more about them. There has been a lot of sentiment analysis work done on reviews before but no studies have been conducted before using this specific women’s safety apps review. Also, we believe that this study will assist mobile application developers in learning more about the performance of their applications in real-world scenarios, how users are reacting to their developed apps, and where they can make improvements to improve the quality of their work. We believe that our research is significant for both consumers of mobile applications, particularly those looking for safety apps on the Google Play store, and developers who have already built or wish to contribute to this area in the future. As a result, we undertook this study to provide individuals, particularly women, with a new frame of reference.

1.4 Research Questions:

Research questions are very essential to find out what kind of study is conducted to improve our knowledge about the research and to find out the clear view and purpose of the research. Related questions to our study are discussed below-

Q1: Can we exploit user reviews from the Google Play store to detect users' sentiment?
By using the reviews from the Google Play store given by the users we can detect the emotion and mine the opinion of the reviewers by classifying them into positive, negative, and neutral.

Q2: Is there any automatic mechanism that can provide a summary of user reactions?
Yes, there are many automatic mechanisms that can summarize the users' reactions. We have used four supervised learning such as KNN, SVM, LR, and MNB, and an unsupervised learning BERT for our research.

Q3: What are the advantages of gathering user feedback on women's safety-related apps?
Gathering user feedback on women's safety related applications helps in understanding the real user's experience after using them, helps developers to understand real world users better and improve their service, helps people know more about these applications to create awareness among the society.

1.5 Expected Output:

Our project revolves around research. As a result, the primary objective is to publish research articles related to this research. The suggested study is a significant step forward in the field of assessing the feelings of users of women's safety mobile applications. This study will give developers a clear picture of what users think about current apps, allowing them to expand their work and try to fix issues within those apps that already exist, as well as allowing new users to learn more about those apps and prior users' experiences. This will enhance public knowledge of women's safety concerns in our society, assist general users in making informed decisions, and benefit the community as a whole.

This work intends to obtain a result where we must prove our analysis with mathematical evidence using Natural Language Processing by categorizing the text with the help of Machine Learning and Deep Learning algorithms. This study will also aid in the comprehension of several algorithms, as well as the top performing algorithm on this novel dataset.

1.6 Project Management and Finance:

There was no money spent on this study, and no software, hardware, or other equipment was purchased. This study took a while to complete because of the data collection and data processing. The Table 1.6.1 shows how much time we spent on each stage of this research-

Table: 1.6.1 Timeline of the Research Management

Stages of Research work	**Timeline for Each Stage**
Determining and finalizing the topic	1 month 15 days (approximate)
Data collection	1 month (approximate)
Literature review	15 days (approximate)
Experiment	15 days (approximate)
Experiments and validations	1 month (approximate)
Report writing	20 days(approximate)
Total	**5 months 5 days (approximate)**

1.7 Report Layout:

The following is the content of this research paper:

- In Chapter One, we have given an overview of the research, the rationale for this study as well as the expected outcomes. We have also given a quick overview of sentiment analysis and the necessity for women's safety applications in modern culture.

- Chapter two includes associated work with this research effort, including a brief discussion, a summary of the work with comparative analysis, as well as any problems or obstacles that may develop.
- Chapter three explores the workings of our research project, including the topic of our study, the data collection process, statistical analysis, our methodology, and the research's implementation requirements.
- Our experimental results and discussion are covered in Chapter four, as well as the experimental setup, outcomes, and analysis.
- The topic of social impact on our society, ethical considerations, and a sustainability strategy are discussed in Chapter five of our research project.
- The sixth chapter speculates on the summary and conclusion of our entire study, as well as the implications for future research.

CHAPTER 2
Background

2.1 Preliminaries/Terminologies:

In this era of technology where everything is within our reach, Mobile applications play a very important role in it. There are millions of mobile applications with various kinds of usage which make our life a lot easier and faster. Women safety applications is an application that provides safety services via mobile. We have collected almost 3360 reviews on this kind of safety application and applied machine learning algorithms as well as a deep learning algorithm to classify the opinion of the users of these applications. Over recent years so many studies have been done on review of different and various essentials. In those review related sentiment analysis studies different algorithms have been applied to detect and classify different types of results on different datasets.

For our study, we have applied algorithms like KNN, LR, SVM, MNB, and BERT.

2.2 Related Works:

In this section, the related works found in the literature are compared and discussed over our problem definition. No works are reported on the sentiment analysis from the women's safety mobile application reviews. The mobile application reviews make our paper novel, with no existing algorithms to compare our work with. Therefore, the related works are presented in this section.

Ali Hasan et. al. presented a paper regarding sentiment analysis on an Indonesian online marketplace application named "Shopee " in [6] where they proposed analysis on the feedback given by the users which were then classified into positive and negative sentiments to evaluate the performance of the application. The dataset consisted of 200 reviews and the method "Naive Bayes" was applied which yielded an accuracy of 96.667%.

Ronal Watrianthos et. al. have done a sentiment analysis on an Indonesian e-commerce app called "Traveloka" in [7] where they implemented the Naive Bayes algorithm by testing the Vmap value which gave the largest probability of the test data for each class.386,646 data were collected to analyze the sentiments of the reviews given by the customers. After analyzing they got greater negative sentiment (0.31020) than positive sentiment (0.16132), which basically indicates the reviewers are disappointed with the expensive ticket prices.

Atiqur Rahman et. al. used Sentiment Analysis on movie reviews to learn the strong and weak points of the movie through the review and ratings given by the viewers through collecting 2000 reviews. Machine Learning techniques such as Bernoulli Naïve Bayes (BNB), Decision Tree (DT), Support Vector Machine (SVM), Maximum Entropy (ME), and Multinomial Naïve Baye(MNB) was used to classify data into Positive and negative sentiment. Here, Parts of speech (PoS) tags were also added to get the more accurate sentiment and Multinomial Naïve Bayes (MNB) acquired higher accuracy of 88.50%. There is a scope to implement deep learning methods in this project.[8]

Sakshi Ranjan et. al. performed a sentiment classification of the google app reviews of 13 fields in [9] and identified the behavior of university students towards the app market by collecting their reviews from 6 fields of the Utkal University students. For this purpose, they collected 10,841 Google app reviews to train the model and 400 reviews from that particular university through a local survey conducted department wise to test the model. Supervised Machine Learning methods such as Naive Bayes (NB), Support Vector Machine (SVM), Logistic Regression (LR), K-Nearest Neighbors (KNN), and Random Forest (RF) was implemented and on the trigram+TF-IDF scheme, SVM secured the highest accuracy of 93.37%.

Xing-Min Lin et. al. used a Sentiment Analysis approach in [10] on the review of "Online Car hailing app" which can promote low carbon travel that saves energy and reduce pollution, based on the combination of long and short-term memory neural networks- LSTM, Bi LSTM with attention focus mechanism to predict sentiment. This paper mainly focuses on users' comments for sentiment analysis technology using deep learning methods. In this paper,among all the models such as LSTM + Mean, LSTM + Last, LSTM+Attention, and BiLSTM + Attention used, the BiLSTM + Attention model gave the highest accuracy of 98.70%.

Gopalkrishna Barkur et. al. analyzed in [11] the feelings of Indians concerning Covid-19 lockdown by extracting tweets distinctly tweeted with the hashtags #IndiaLockdown and #IndiafightsCorona from March 25th to March 28th, 2020. The analysis was done by using the software R and a Word Cloud was generated that depicts the sentiments of the tweets where a total of 24,000 tweets were considered. Further, there's a scope to filter out the fake news that spreads on social media in the future to find out the actual feelings of the people of that particular place or who are using the hashtags to tweet.

Ayushi Mitra et. al. in [12] proposed a Hybrid Based approach for classification with the combination of Machine learning and Lexicon-based approach to analyze the data of movie reviews given by the viewers. Different approaches like Naïve Bayes, SKlearn BernoulliNB, Sklearn SVC () as a first approach and Decision Tree, Random Forest, KNN as the second approach were used to find out different accuracy where the Random Forest (RN) got the highest accuracy of 80% among all other algorithms used.

Mohammad Rezwanul Huq et. al. developed a functional classifier in [13] to correctly detect the sentiment of an unknown tweet into positive and negative. Two methods were introduced which is Sentiment Classification Algorithm (SCA) based on K-Nearest Neighbor (KNN) and the other one is based on Support Vector Machine. They used quite a number of features such as Word feature, N-Gram Feature, Pattern Features, Punctuation Feature, and Key-based feature. These features were used with KNN and SVM with different versions. Among them, KNN with normalization and keyword base (5 features) outperformed acquiring the highest accuracy of 84.32%. Although a large number of tweets could have been used to develop this functional classifier and also only two approaches were used to calculate the accuracy and performance evaluation.

Lopamudra Dey et. al. discussed and compared two supervised machine learning algorithms on two data sets which are Movie reviews and Hotel reviews by taking 5000 positive and 5000 negative reviews for each in [14]. KNN and Naive Bayes were used to analyze sentiment classification and to evaluate the performance of both algorithms in respect to accuracy, precision, and recall. For the movie reviews both the algorithm performed well whereas Naive Bayes gave an accuracy of more than 80% and performed better than the KNN approach. The accuracies for the hotel reviews were lesser in both algorithms giving almost similar results.

Kudakwashe Zvarevashe et. al. in [15] modeled a framework to mine the opinion or to analyze the sentiment of the unlabeled datasets of hotel reviews given by the customers which can help the hotel managers to have deeper knowledge about their customers perspective of their services. In this paper, from about 259000 unlabeled reviews on cars and hotels, the hotel reviews of London, Beijing, and Montreal were chosen and these unlabeled details were labeled based on sentiment polarity to find the polarity(positive, negative, neutral) of the reviews. Classification Algorithms such as Naive Bayes multinomial (NBM), Sequential minimal optimization (SMO), Compliment Naïve Bayes (CNB), and Composite hypercube on iterated random projections (CHIRP) were applied where Naïve Bayes algorithm got higher precision which is 80.9% among them. While classifying some data were labeled as neutral which could have been positive or negative. Automatic labeling, feature extraction, and performance classification of the data set using deep learning are considered for future work.

Zeenia Singla et. al. conducted sentiment analysis on about 4,000,00 Mobile Phone reviews for 4500 mobile phones in [16]. These reviews and ratings were extracted from the e-commerce platform Amazon. These unstructured datasets/reviews were classified into positive and negative sentiment where the sentiment score was established using the NRC sentiment dictionary.ML algorithms like Naïve Bayes, Support Vector Machine (SVM), and Decision Tree was used to classify the model.10 Fold Cross Validation was used to evaluate the proposed model. The highest accuracy of 81.75% was achieved by the SVM model amid all other two models for a number of iterations.

Omar Sharif et. al. proposed a system classify customers sentiment on restaurants whether they are positive or negative in [17].1000 of restaurant reviews translated in Bengali from an English benchmark dataset were used as the dataset and Random forest, Decision tree, Multinomial Naive Bayes algorithm were used not only to analyze the intrinsic sentiment of the customers but also to compare the accuracy among the models they have used. Exact numbers of train and test data were used for these algorithms along with different K values of every fold to find the optimum value of K. 80.48% accuracy for 6-fold cross-validation was achieved from Multinomial Naive Bayes among all other algorithms that were used for this classification. There's a scope to use other powerful algorithms to obtain more accurate results which is a limitation of this paper.

Anang Anggono Lutfi et. al. proposed [18] sentiment analysis in the sales review on an Indonesian marketplace named “Bukalapak”.In this study the data was classified into positive and negative sentiments.Naive Bayes and Support Vector Machine with linear kernel was used to classify the model where SVM with linear kernel achieved the highest average accuracy of 93.65% [13].

Eftekhar Hossain et. al. introduce us in [4] to a machine learning-based sentiment classification framework that can determine the sentiment into positive and negative categories from Bengali book reviews containing 2000 data by exploiting the various features extraction techniques. They used trained classifier LR(Logistic Regression), MNB, RF, DT, KNN, SVM, and SGD and among them, only LR, MNB, SVM, and SGD provided acceptable cross-validation accuracy where Naive

Bayes with unigram features obtained the highest accuracy 87% and for validation and test datasets, this classifier achieved 84% accuracy.

Palak Baid et. al. analyzed the Movie reviews in [19] using various Machine Learning techniques like Naïve Bayes, K-Nearest Neighbor (k=3), and Random Forest. The data was collected from 2000 user-created movie reviews, among them, 1000 were positive reviews and 1000 were negative reviews and a relation was created. In this research, they used various classifier techniques to identify the polarity of the movie reviews collected.10 fold cross-validation was used for the validation phase and the TextDirectoryLoader component was used to import the dataset into the WEKA which is open-source software under GNU. Among all the algorithms that were applied, the Naïve Bayes classifier achieved the highest accuracy of 81.45%.In the future, other classifier algorithms or hybrid models could have been applied to increase the final accuracy.

2.3 Comparative Analysis and Summary:

We have reviewed some papers related to sentiment analysis and are comparatively relatable with our research work. The following table is about those related work and their method as well as the acquired evaluation of the method used for these research studies.

Table: 2.3.1 Comparative Analysis of the Related Publications

Serial No.	The Name of the Authors	Applied Methods	Evaluation / Accuracy	Year
1.	Dany Pratmanto, Rousyati Rousyati, Fanny Fatma Wati, Andrian Eko Widodo, Suleman Suleman. Ragil Wijianto	Naive Bayes	Naive Bayes (96.667%)	2020
2	Ronal Watrianthos, Sudi Suryadi, Deci Irmayani, Marnis Nasution, Elida F. S. Simanjorang	Naive Bayes Classifier	Polarity: Negative sentiment (0.31020), Positive sentiment (0.16132)	2019
3.	Atiqur Rahman, Md. Sharif Hossen	Bernoulli Naïve Bayes (BNB), Decision Tree (DT),Support Vector Machine (SVM),	Bernoulli Naïve Bayes (87.50%), Decision Tree (80.17%), Support Vector Machine (87.33%), Maximum Entropy (60.67%),	2019

		Maximum Entropy (ME), Multinomial Naïve Bayes (MNB)	Multinomial Naïve Bayes (88.50)	
4.	Sakshi Ranjan,Subhankar Mishra	Naive Bayes (NB), Support Vector Machine (SVM), Logistic Regression(LR), K-Nearest Neighbors (KNN), and Random Forest(RF)	Naive Bayes (80%), Support Vector Machine (93.41%), Logistic Regression (84.08%), K-Nearest Neighbors (91.01%), and Random Forest(85.11%)	2020
5.	Xing-Min Lin , Chun-Heng Ho , Lu-Ting Xia , Ruo-Yi Zhao	LSTM + Mean, LSTM + Last, LSTM+Attention, BiLSTM +Attention	LSTM + Mean (92.35%), LSTM + Last(94.39%), LSTM+Attention (97.16%), BiLSTM +Attention (98.70%)	2021
6.	Gopalkrishna Barkur, Vibha, Giridhar B. Kamath	Software R	Evaluated through word cloud 	2020
7.	Ayushi Mitra	First approach: Naïve Bayes, SKlearn BernoulliNB, Sklearn SVC () Second approach: Decision Tree, Random Forest, K-Nearest Neighbors (KNN)	First approach: Naïve Bayes (70.44%), SKlearn BernoulliNB(70.15%), Sklearn SVC ()(75.37%) Second approach: Decision Tree (52%), Random Forest(80%), K-Nearest Neighbors (71%)	2020
8.	Mohammad Rezwanul Huq,Ahmad Ali,	Support Vector Machine (SVM), K-	Algorithm in [13] (79.99%), KNN with normalization (4	2017

	Anika Rahman	nearest neighbor (KNN)	features) (80.80%), KNN with normalization and keyword base (5 features) (84.32%), SVM with (4 features) (58.79%), SVM with normalization (4 features) (58.39%), SVM with normalization and keyword base (5 features) (67.03%), SVM with normalization and keyword base (5 features) with grid search (77.97%)	
9.	Lopamudra Dey, Sanjay Chakraborty, Anuraag Biswas, Beepa Bose, Sweta Tiwari.	K-Nearest Neighbors(KNN), Naive Bayes	For movie reviews KNN achieved above 80%, for hotel reviews both achieved lower and approximate similar results.	2016
10.	Kudakwashe Zvarevashe, Oludayo O Olugbara	Naive Bayes multinomial (NBM), Sequential minimal optimization (SMO), Compliment Naïve Bayes (CNB), Composite hypercubes on iterated random projections (CHIRP)	Naive Bayes multinomial (precision: 80.9%), Sequential minimal optimization (above 75%) Compliment Naïve Bayes (precision: 80.5%), Composite hypercubes on iterated random projections (precision: 75.6%)	2018
11.	Zeenia Singla, Sukhchandan Randhawa, Sushma Jain	Naïve Bayes, Support Vector Machine (SVM), Decision Tree(DT)	Naïve Bayes (66.95%), Support Vector Machine (81.77%), Decision Tree (74.75%)	2017
12.	Omar Sharif, Mohammed Moshiul Hoque,Eftekhar Hossain	Decision Tree, Random Forest and Proposed system using Multinomial naive bayes	For K value=5: Decision Tree (70.40%), Random Forest (75.51%) and Proposed system using Multinomial naive bayes (79.59%) For K value=6: Decision Tree (69.51%), Random Forest (78.04%)	2019

<table>
<tr><td></td><td></td><td></td><td>and Proposed system using
Multinomial naive bayes
(80.48%)
For K value=7:
Decision Tree(71.42%),
Random Forest(70%)
and Proposed system using
Multinomial naive bayes
(78.57%)</td><td></td></tr>
<tr><td>13.</td><td>Anang Anggono Lutfi,
Adhistya Erna
Permanasari, Silmi
Fauziati</td><td>Naive Bayes and
Support Vector
Machine with linear
kernel</td><td>Naive Bayes (Average
accuracy:89.28%) , Support
Vector Machine with linear
kernel(Average
accuracy:93.65%)</td><td>2018</td></tr>
<tr><td>14.</td><td>Eftekhar Hossain, Omar
Sharif Mohammed
Moshiul Hoque</td><td>Logistic
Regression(LR),
K-Nearest
Neighbour (KNN),
Decision Tree(DT),
Random Forest(RF),
Multinomial Naïve
Baye(MNB),
SVM,
SGD</td><td><table>
<tr><td></td><td>Classifier</td><td>Uni-gram</td><td>Bi-gram</td><td>Tri-gram</td></tr>
<tr><td rowspan="7">Accuracy(%)</td><td>LR</td><td>83</td><td>77</td><td>66</td></tr>
<tr><td>KNN</td><td>58</td><td>54</td><td>60</td></tr>
<tr><td>DT</td><td>69</td><td>77</td><td>59</td></tr>
<tr><td>RF</td><td>73</td><td>77</td><td>68</td></tr>
<tr><td>MNB</td><td>88</td><td>78</td><td>69</td></tr>
<tr><td>SVM</td><td>81</td><td>77</td><td>63</td></tr>
<tr><td>SGD</td><td>77</td><td>72</td><td>68</td></tr>
</table></td><td>2020</td></tr>
<tr><td>15.</td><td>Palak Baid,
Apoorva Gupta,
Neelam Chaplot</td><td>Naive Bayes, K-
Nearest Neighbour
and Random Forest</td><td>Naïve Bayes (81.40%), K-
Nearest Neighbour (55.30%)
and Random Forest(78.65%)</td><td>2017</td></tr>
</table>

In our study, we have applied four supervised machine learning classifier algorithms which are Multinomial Naive Bayes(MNB), Support Vector Machine(SVM), Logistic Regression(LR), and K-Nearest Neighbor(K=3) and a unsupervised algorithm BERT. After training our model we have achieved a satisfactory result.

2.4 Scope of the Problem:

We used the novel dataset of reviews that we gathered for our research. Users have the option to leave a review of the application on the Google Play Store. Users from all sorts of backgrounds who have download access to such applications can utilize them and provide feedback on a specific application or applications to that choice. Because so many people submit reviews in their original language or transliterate them, the data for our dataset becomes noisy. As a result, our algorithm was unable to identify every word uttered as being in the English language. It was challenging for our model to relate to other text evaluations, resulting in slightly lower accuracy.

2.5 Challenges:

Because the Google Play Store has a large number of mobile applications, we had to actively search for each and every application connected to safety, particularly related to women's safety. We had to check over all of the descriptions of the safety-related applications to make sure they met our study's requirements. The evaluations sometimes included the reviewer's original language, and other times merely emojis with no text, which was avoided while manually gathering the data. The collection and organization of the entire dataset was a laborious and time-consuming operation.

As this is the first work based on this unique dataset we need a large number of dataset to get a good accuracy of the work. So, we tried to manually get the raw data that can help to extract the best emotions out from those reviews which was actually a very challenging task to accomplish.

CHAPTER 3

Research Methodology

3.1 Research Subject and Instrumentation:

"Sentiment Analysis of User-Generated Reviews of Women Safety Mobile Applications," we propose as the title of our thesis. Sentiment analysis is a popular topic for determining the opinions of users or consumers of a product or service. Using various algorithms, we attempted to depict the emotion of users of several specific mobile applications, particularly women safety-related applications. We have gathered user reviews from 32 different apps that exist in the Google Play Store for Android users. In the table below we are mentioning some details about those applications from which we have taken reviews-

Table: 3.1.1 Information of the Safety Related Applications

Name of the mobile application	Country/ Region	Downloads
Women Safety	Tamil Nadu, India	100,000+
Indian SOS Women Safety | Disha | Emergency Alert	India	10,000+
Women Safety with Security	Global	1,000+
Chilla: Women safety app with scream detection	India	10,000+
RAKSHA-Women Safety App	India	5,000+
Sister - Personal safety app	Global	50,000+
Women Safety Tips	India	1,000+
SHESafe	Global	5,000+
My Safetipin: Complete Safety App	India	50,000+

GetHomeSafe - Personal Safety	Global	10,000+
Eyewatch SOS for Women	New Delhi, India	10,000+
bSafe - Never Walk Alone	Global	1,000,000+
WanderSafe Safety App	Global	1,000+
SafeUP - Walk With Women	Global	50,000+
MSMR Women Safety App	India	5,000+
Shake2Safety - Personal Safety	Global	100,000+
Disha SOS	India	1,000,000+
Best free and safe social app for women - SHEROES	India	1,000,000+
Women Protection Wing	N/A	N/A
Khayar - For your safety	India	1,000+
Safety - Help - SOS	Global	50,000+
Raksha, A Women's Safety App	India	10,000+
Safety App for Silent Beacon	Global	10,000+
InSec (Intelligent Security) - Personal Safety App	N/A	N/A
Woman Safety Resq	India	100+
Behn - A women security app	India	N/A
Carelife- Personal Safety App	N/A	N/A
Safety App	N/A	N/A
Her Amigo (Women Safety and Consultancy)	N/A	N/A

Protection of Women from Domestic Violence Act2005	Global	10,000+
Mayday Safety	Global	5,000+
Noonlight: Feel Protected 24/7	Global	1,000,000+

3.2 Data Collection Procedure:

For our research, we manually collected our data from the Google Play Store. While manually collecting data, we ran into some issues, such as searching for all of the apps that fulfill the study's objectives from a large number of applications in the Google Play Store and having to double-check the descriptions of those applications to ensure that they are the apps that are required for this study. We collected a total of 3360 text data together with the star rating provided by users of those women's safety applications since the more data we can collect, the better we can train our model, which will offer us higher accuracy.

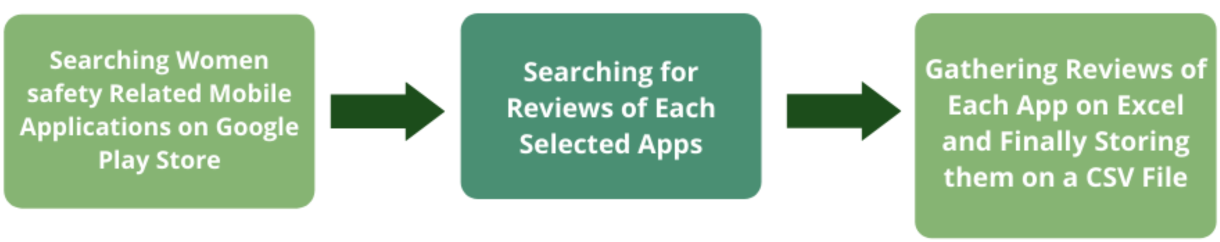

Figure: 3.2.1 Procedure of collecting data manually

After gathering data in an excel file, we converted it to a CSV file containing five columns: app name, review (text data), star(1 to 5 scale numeric rating), and the app link. A snapshot of our own gathered data is shown below:

	Serial Number	App Name	a_review	star	App Link
0	1	Women Safety	Nice app! Its working sending map location in ...	4	https://play.google.com/store/apps/details?id=...
1	2	NaN	Ads? Are you serious? You are making an emerge...	1	NaN
2	3	NaN	Sad. This app is 's not about any marriage iss...	1	NaN
3	4	NaN	Simple feature is missing "to press power butt...	2	NaN
4	5	NaN	You can only seem to add three numbers to send...	3	NaN
...	...	...	...	...	...
3355	3356	NaN	They actually care & check to see if you are s...	5	NaN
3356	3357	NaN	It helps when I walking back for somewhere and...	5	NaN
3357	3358	NaN	Just downloaded this app. God forbid I should ...	5	NaN
3358	3359	NaN	Accidentally clicked it and they responded ext...	5	NaN
3359	3360	NaN	Maybe this app was made to figure out women's ...	1	NaN

Figure: 3.2.2 Tabular Form Our Dataset

3.3 Statistical Analysis:

A total number of 3360 text data along with the star ratings was collected and the star ratings of the applications on a scale of 1 to 5 were divided into three categories: "negative," "positive," and "neutral." The ratings below 3 were labeled "negative", above 3 were labeled "positive" and only the reviews with a rating of 3 were labeled "neutral". A total of 3360 data was collected and they were divided these review data into two categories, i.e., training set and test set which were then divided into 8:2 ratio respectively. These are all the real-life opinions that users of those specific applications have provided as reviews.

Table: 3.3.1 Statistical Insight of the Dataset

Sentiment	Count	Percentage
Positive	2414	74%
Negative	198	10.4%
Neutral	748	15.5%

3.4 Applied Mechanism:

This methodology part helps to understand the overall rationality of the study. It gives a precise concept to the reader about the whole procedure of particular research and helps them in understanding the work. In this section, a detailed description of models used in the study is mentioned. For review related sentiment analysis studies, machine learning and deep learning algorithms play the most important role to find out the accurate sentiments or emotions of the reviewers or users. This part includes the whole process starting with the data collection then data preprocessing, feature selection, training the model with selected algorithms, classification of users' sentiment on the test set to evaluate and analyze the predicted results together with accuracy, precision, recall, and F-score. Also, for running such algorithms a computer with good configuration is needed. The full work process is shown through a step-by-step diagram which makes it clearer to understand the whole process.

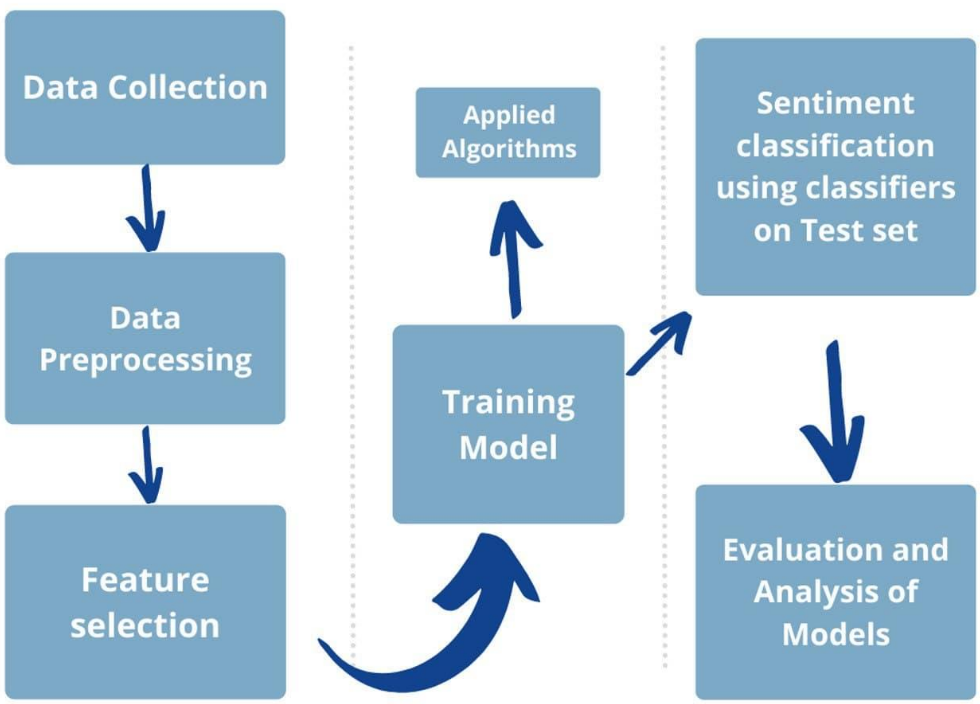

Figure: 3.4.1 Proposed Methodology of Research Model

3.4.a. Collection of Reviews: For this study, a novel and authentic dataset were used that was collected from the google play store with 32 different safety-related applications specially made to provide security to the women as well as others. A total number of 3360 text data along with the star ratings was collected and there were five columns of the dataset which consist of the app name, review(text data), star(1 to 5 scale numeric rating), and the link of the application. The star ratings of the applications on a scale of 1 to 5 were divided into three categories: "negative," "positive," and "neutral." The ratings below 3 were labeled "negative", above 3 were labeled "positive" and only the reviews with a rating of 3 were labeled "neutral". From the total 3360 gathered data 2414 data were positive, 748 data were neutral and 198 data were negative. These are all the real-life opinions that users of those specific applications have provided as reviews.

3.4.b. Data Pre-Processing: On the dataset, data cleaning and text pre-processing techniques were used to develop efficient learning models and improve overall performance. As we have collected the data manually the reviews with URL and reviews with only emojis were avoided. Reviews that were written in languages other than English weren't also considered. Because the acquired review data was not yet sufficiently structured, several preprocessing processes were required to make them well structured. The classification process will be made easier with structured data. Preprocessing was carried out in a number of phases.

Removal of white spaces: On numerous occasions, extra spaces or more than one space is left between the texts in text data thus, we have removed any extra white spaces from the data because they do not hold any value.

Removal of punctuation: In this text processing technique, we had to take care of 32 main punctuations and replace them with empty strings wherever they were present in the review dataset.

Removing mention: The mention sign was also omitted while preprocessing the dataset.

We extracted the subjectivity and polarity values of each case after acquiring the cleansed data. The total subjectivity values suggest that our dataset followed the conventional "V-shape" where the color "red" denotes negative polarity, "green" denotes positive polarity, and "blue" denotes neutral polarity.

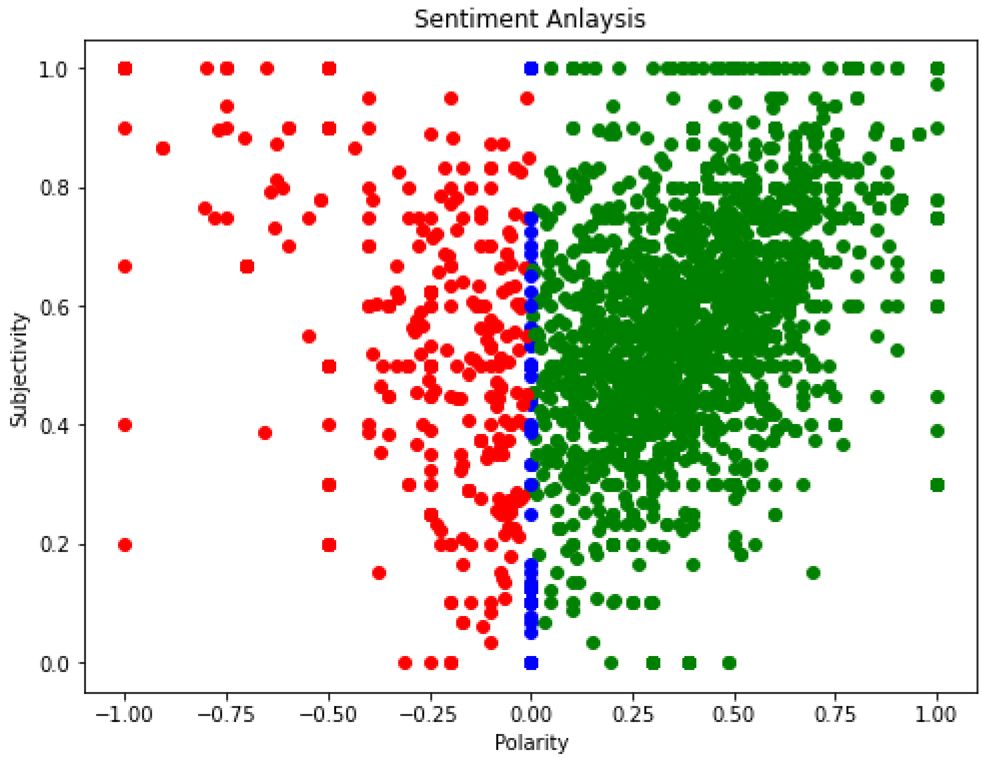

Figure: 3.4.2 Scatter Plot of the Sentiments

Text Vectorization: After removing all the noisy values from the dataset we proceed with the vectorization part where we have used CountVectorizer as it is the most basic method for quantitatively representing text data [20]. The CountVectorizer method is used when we wish to convert a given text into a vector based on the frequency of each word which would appear in the full text. This approach is adopted in this study which is a fantastic tool provided by the scikit-learn module in Python.

3.4.c. Data Processing: There exist so many machine learning and deep learning algorithms for different types of motives of work. In our study, at first we have gathered star ratings in addition to the text dataset, which we then utilized to classify the collected text reviews. We divided the reviews into three categories: positive, negative, and neutral.

We used data preprocessing methods to clean up the unstructured and noisy text data by eliminating punctuation, white space, mention(@), hashtag(#) and then used count vectorization to turn the provided text into a vector based on the frequency of each word that occurs in the entire text. After preprocessing, we have divided the women's safety app review data into two categories, i.e., training set and test set which were then divided into 8:2 ratio respectively

As we are dealing with the text reviews of some particular applications, supervised machine learning algorithms like MNB, LR, SVM, KNN and unsupervised open source deep learning algorithm BERT have been used. We have implemented these approaches or algorithms to identify the polarity of a text. These algorithms are the best fit for our research because the major goal was to see the users' opinions through their reviews utilizing the supervised and unsupervised method. A summary of the applied techniques for determining sentiment levels from user reviews is presented in this section.

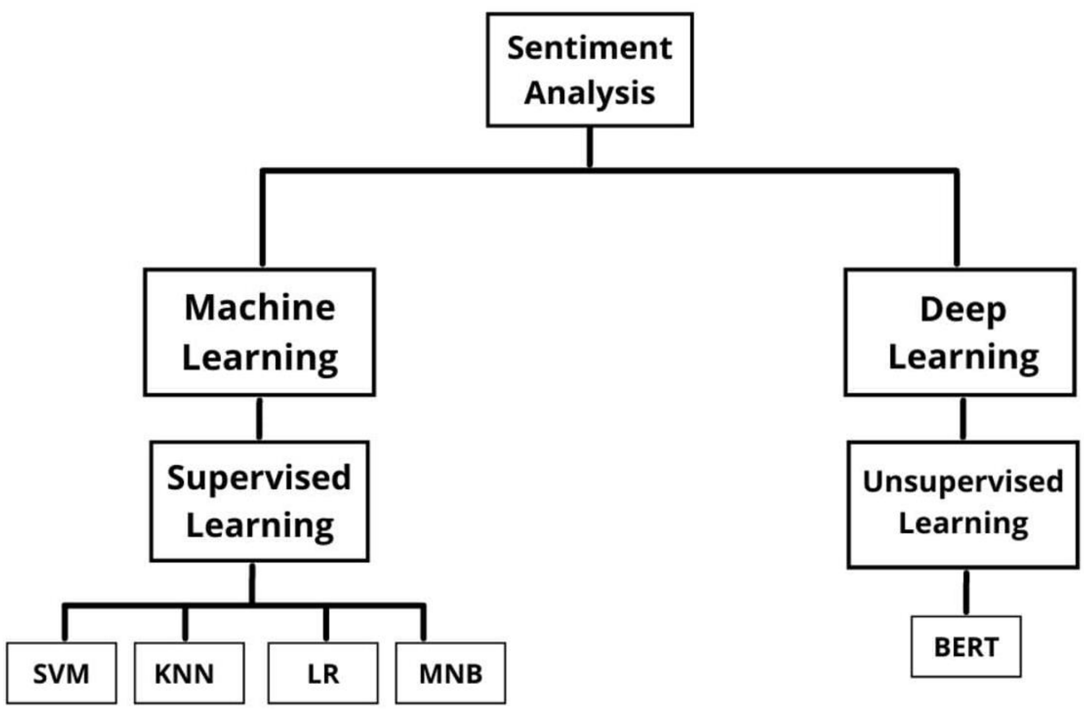

Figure: 3.4.3 Hierarchy of Applied Algorithms

3.4.c.1. Supervised Learning:

KNN: K-Nearest Neighbour (KNN) is a supervised machine learning algorithm that is very simple to understand and easy to implement. The K-NN algorithm can be used for both Regression and Classification Problems but it is widely used for classification problems. Even though it is very popular for its effective performance and efficient outputs for pattern recognition, machine learning, text categorization, data mining, object recognition, and many more, the KNN classifier is regarded as the most popular classifier for pattern recognition. This K-NN method stores all available data that is given and classifies a new data point based on its similarity to the existing data and Euclidean distance is chosen for calculating the distance [21]. This means that when fresh data comes, the K-NN algorithm can quickly classify it into the best suitable category. We can simply determine the category or class of a dataset with the help of K-NN where the larger the dataset the better the accuracy. Here we have used K=3 considering our implementation of the algorithm for our classification.

SVM: The support vector machine (SVM) approach is widely used as supervised machine learning for both classification and regression problems, however, it is most usually utilized for classification tasks because it can handle both linear and nonlinear difficulties. SVM model is a representation of distinct classes in a hyperplane in multidimensional space that is created to iteratively distinguish between the two classes and find the closest point of the lines from the classes. Moreover, the SVM algorithm minimizes the error. In this study, SVM is used for the purpose of multiclass classification and it works great with fewer datasets and can give reliable results. As it is used less computation and can give significant accuracy for the proposed model, it is preferred over other classification algorithms [22].

LR: Logistic regression (LR) is a supervised machine learning approach for predicting the probability of a target variable with a binary dependent variable or nominal, ordinal, or interval which are independent variables. When the input must be divided into two groups by a linear boundary, logistic regression is employed [23]. Multinomial Logistic Regression is useful when there is a need to categorize subjects based on the values of a set of predictor variables. As the dependent variable is not limited to two categories so it is more general. In this study there are more than two classes such as positive, negative, and neutral, as a result, the multinomial logistic regression is used [24]

MNB: Just like the other three algorithms, Multinomial Naive Bayes (MNB) is a supervised machine learning classification algorithm that is used for the analysis of categorical text data, and it is the most popular classification model. As MNB is a probabilistic learning method that is widely used for Natural Language Processing (NLP) thus, for text data analysis and with problems with multiple classes this algorithm is sought after. In Bayes theorem, each feature being categorized is independent of the others so the presence or absence of a feature doesn’t matter to other features. Even if this algorithm is not suitable to predict numeric values, for textual data it is the fastest and most suitable. For our study where we have a novel dataset that we have collected from the reviews, MNB has achieved the highest accuracy [25].

3.4.c.2. Unsupervised Learning:

BERT: Bidirectional Encoder Representations from Transformers is an unsupervised open source machine learning framework for natural language processing that uses surrounding text to assist computers to grasp the meaning of ambiguous words in a text and it can grasp the contextual meaning of a word or sub word in a text or sentence [26].

To generate a sequence of word BERT pre-trained model is used as it is a good feature representation for unsupervised learning method .In this research, we employed the cased version of Bert and performed all of the essential preprocessing, such as tokenization, padding, and attention mark, before using the Bert deep learning algorithm. To make the BERT model understand that we are doing classification and to assist the model in differentiating between the two sentences during training we must add a special token [CLS] to the start of each sentence and [SEP] to the end of those sentences.

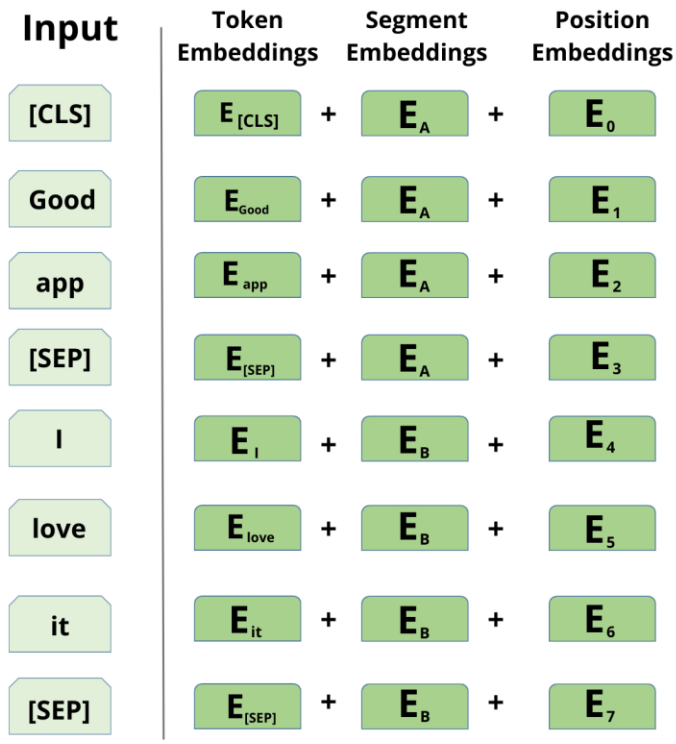

Figure: 3.4.4 Input Tokens in BERT

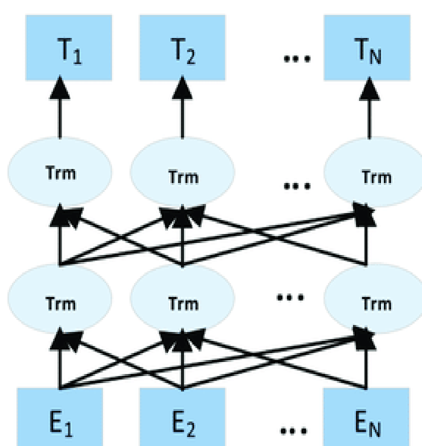

Figure: 3.4.5 Architecture of BERT [27]

As a pre-trained model, we utilized "BertTokenizer". To make evaluating our model's predictions easier, we divided the data into an 8:2 ratio for training and testing. We built our sentiment classifier using the simple “BertModel” to make utilizing this algorithm easier. We returned the raw output of the last layer since the cross-entropy loss algorithm in PyTorch requires it.

The AdamW optimizer given by Hugging Face was used to reproduce the training technique from the BERT study. To evaluate the model, we have first trained it, then save the state of the best model, as greatest validation accuracy, which can be used to evaluate how well our model predicts emotions with a deeper sense of language context. [28].

It is clear that sentiment analysis has improved significantly with the adoption of BERT, a natural language processing algorithm that surpassed all four supervised machine learning algorithms, obtaining an 86.06% greater accuracy.

3.4.d. Evaluate Performance of the Algorithms:

For classification models, performance measurements are used to evaluate how effectively the algorithms perform in a specific situation. These performance metrics include precision, recall, F1-score, and accuracy. The matrices used to evaluate the performance of these algorithms are briefly described here in this section.

Precision: The model's precision score reflects its ability to accurately forecast the positives out of all the positive predictions that have been made [29]. It is the ratio of accurately predicted positive observations to total expected positive observations. When the cost of false positives is high, precision helps. The low false-positive rate is related to the high precision.

$$\text{Precision} = \frac{True\ Positive}{True\ Positive + False\ Positive}$$

Recall: The model's capability to properly forecast positives out of real positives is represented by the recall score [29]. It actually answers the question of how many of the positives did the model predict. When the cost of false negatives is large, recall aids in the prevention of false negatives. It actually evaluates a classifier's sensitivity. The better the learning model is in identifying both positive and negative samples, the higher the recall score.

$$\text{Recall} = \frac{True\ Positive}{True\ Positive + False\ Negative}$$

F1-Score: F1 score or F-measurement is the weighted average of Precision and Recall that measures a model's accuracy on a dataset and it is also known as the harmonic mean of the model's precision and recall [29]. The F1-score provides equal weight to Precision and Recall when calculating accuracy, thus making it a viable alternative to Accuracy metrics. If the class distribution is uneven then it indicates a good F1 score.

$$\textbf{F1 Score} = 2 * \frac{Recall * Precision}{Recall + Precision}$$

Accuracy: Accuracy describes performance statistics for classification models. It mainly specifies the correlation of true positives and true negatives to all positive and negative reviews. Simply, the percentage of correct answers given by the model is called accuracy. When the datasets are symmetric and the values of false positives and false negatives are almost equal accuracy works greatly. But other than that, to evaluate the accomplishment of the model one has to consider other parameters as well [29].

$$\text{Accuracy} = \frac{True\ Positive + True\ Negative}{True\ Positive + True\ Negative + False\ Positive + False\ Negative}$$

Here, True Positive (TP) means correctly predicting positive values from real positive values. Hence, the anticipated value corresponds to the real value which means when the model anticipated a positive value, the actual result was truly positive.

True negative (TN) refers to the predicted value of the model being negative while the actual value is also negative. In simple words, the value of correctly predicting negatives from real negative occurrences is known as true negative.

False-positive (FP) defines the contradiction between the actual class and the predicted class. It predicts a class as positive while the actual class is negative.

False Negative (FN) represents the value of inaccurate negative predictions meaning that the number of predictions where the classifier incorrectly predicts the positive class as negative

3.5 Implementation Requirements:

"Sentiment Analysis of User-Generated Reviews of Women Safety Mobile Applications," we propose as the title of our thesis. Sentiment analysis is a popular topic for determining the opinions of users or consumers of a product or service. Using various machine learning and deep learning methods, we attempted to depict the emotion of users of several specific mobile applications, particularly women safety-related applications. To execute such algorithms accurately, we need a high-configuration computer, a GPU, and other relevant equipment and technologies.

The following is a summary of the necessary instrument and technology for our model-

Hardware:

- ❖ Intel Core i3 8th generation with 4GB RAM, ROM: 1TB
- ❖ Windows 10 64 Bit

Software:

- ❖ Google Collab with GPU
- ❖ Google Chrome
- ❖ Python packages(Scikit-learn, Numpy, Pandas, NLTK, Textblob, matplotlib)

CHAPTER 4

Experimental Results and Discussion

4.1 Experimental Setup:

According to our experimental need, from data collecting to evaluating our model we have arranged everything that is needed. Data collection, pre-processing of the data, applying classification algorithms to mine the opinion of the reviewers and evaluating the model was done using our own laptop with a configuration of : Corei3 8th Gen, RAM: 4 GB, ROM: 1TB, OS: Windows 10 64 Bit.

4.2 Experimental Results & Analysis:

Sentiment analysis is a method of determining the sentiment of textual data. We used different classifiers to find out the percentage of correct answers given by the model and other relevant things like precision, recall, and F-score that will accurately evaluate the model in this research to extract those sentiments from the given reviews from users of all kinds of women safety mobile applications. We labeled our text dataset with the star rating using a 1 to 5 scale numeric rating that was supplied by the users along with the text data that play a significant part in conveying users' experience and emotions before attaining our intended outcome. 593 people gave a one-star rating, 155 people gave a two-star rating, 198 people gave a three-star rating, 376 people gave a four-star rating, and a massive 2038 people gave a five-star rating.

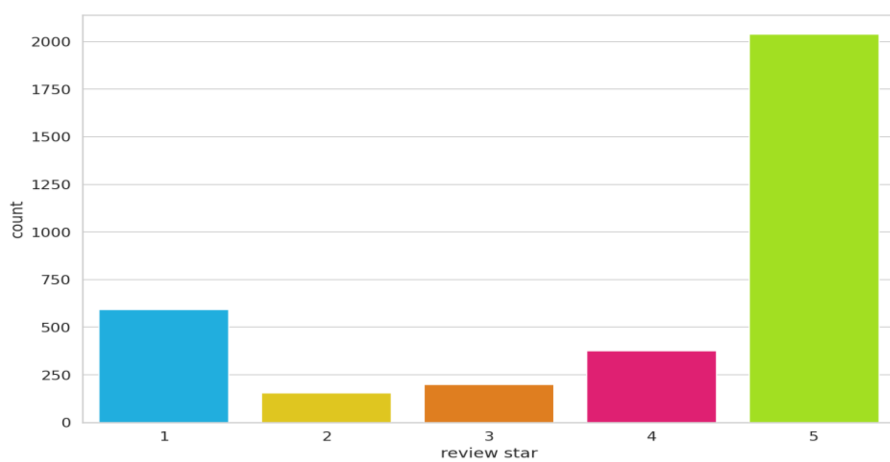

Figure: 4.2.1 Plot of Star Review

After labeling the dataset we have divided those data into three categories: "negative," "positive," and "neutral." The ratings below 3 were labeled "negative", above 3 were labeled "positive" and only the reviews with a rating of 3 were labeled "neutral" where we get 74% positive reviews,15.5% of neutral reviews, and only 10.4% of negative reviews.

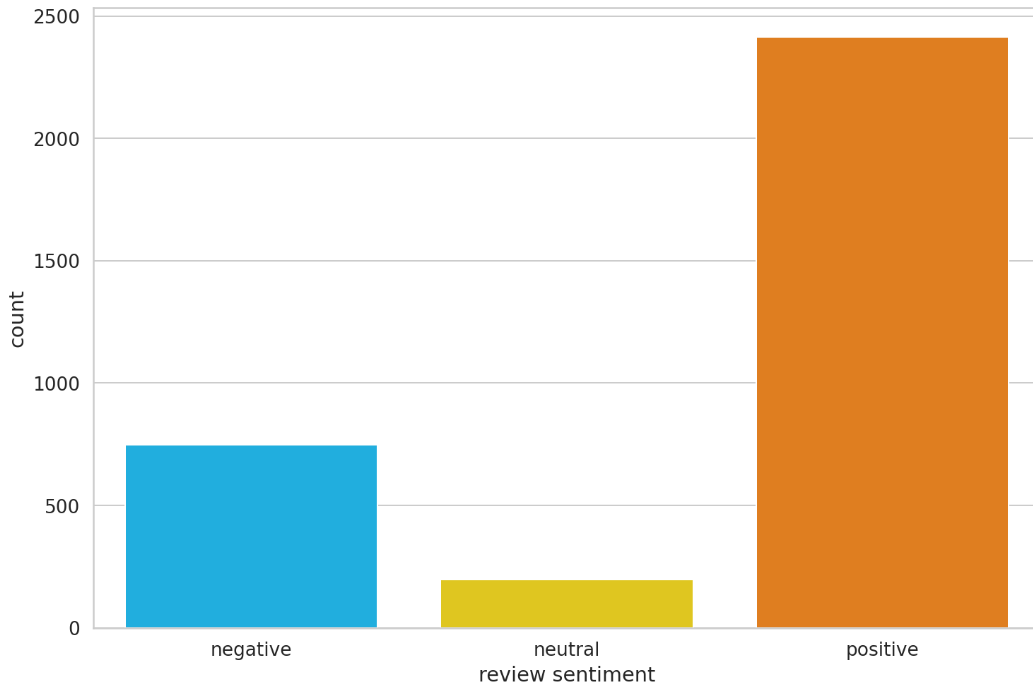

Figure: 4.2.2: Plot of the Review Sentiment (Positive, Neutral, Negative)

Using sentiment analysis data to create word clouds allows you to see how subjects are addressed in your data. For our research, we created a word cloud to display the text data, with the size of each term indicating its frequency and significance.

Figure: 4.2.3 Word Cloud of the Reviews

After that, we cleaned the data and used count vectorization to quantitatively describe it.

To mine the opinion of the reviewers as well as to find out the percentage of accuracy, precision, recall, and F-score of the model, we have used five different algorithms: MNB, KNN, SVM, LR and BERT which provided us with good results. The following is the confusion matrix for the algorithms that we have attained:

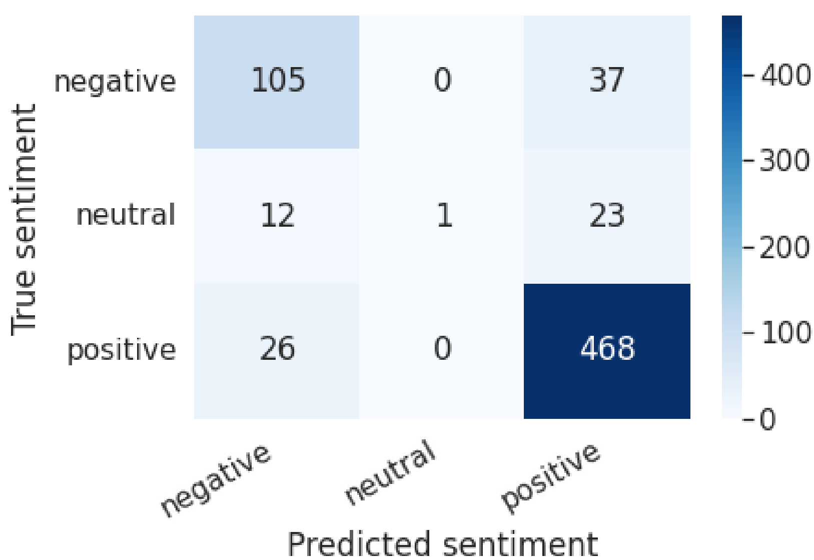

Figure: 4.2.4 Confusion Matrix of Multinomial Naïve Bayes (MNB)

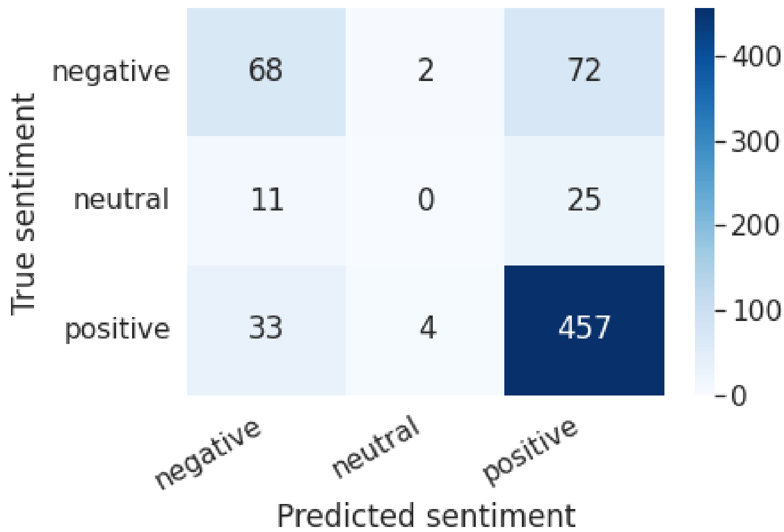

Figure: 4.2.5 Confusion Matrix of K-Nearest Neighbour (KNN)

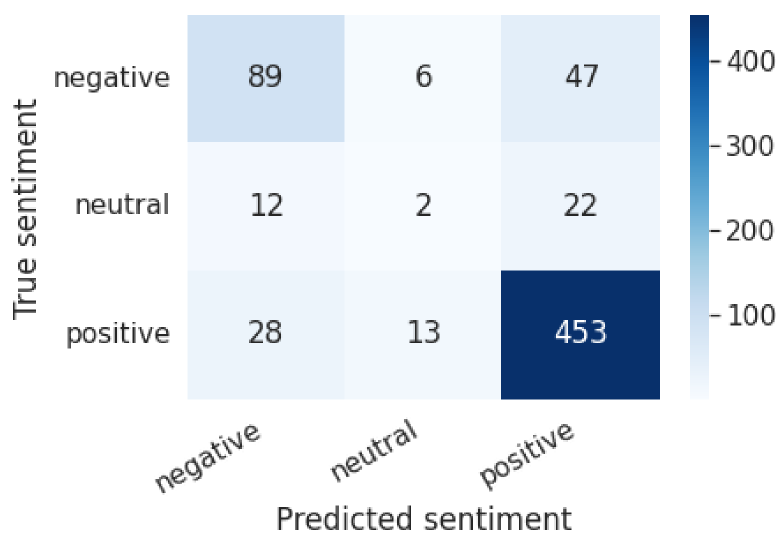

Figure: 4.2.6 Confusion Matrix of Support Vector Machine (SVM)

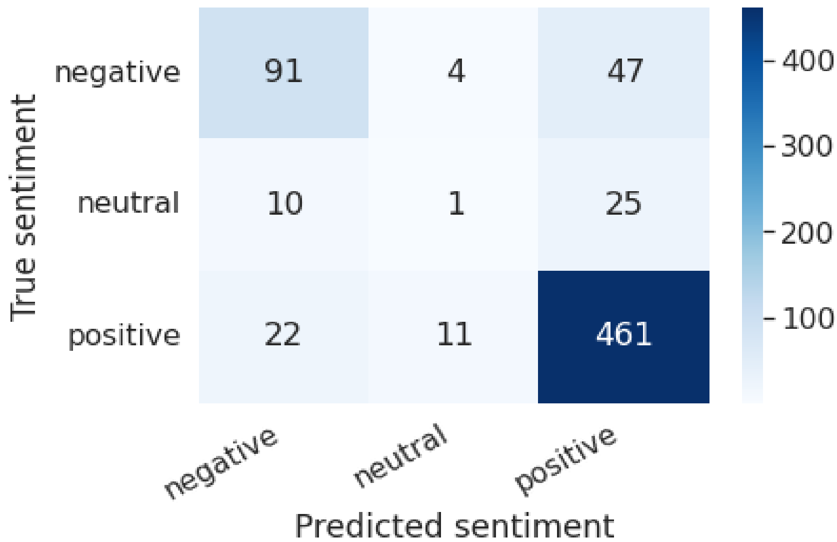

Figure: 4.2.7 Confusion Matrix of Logistic Regression (LR)

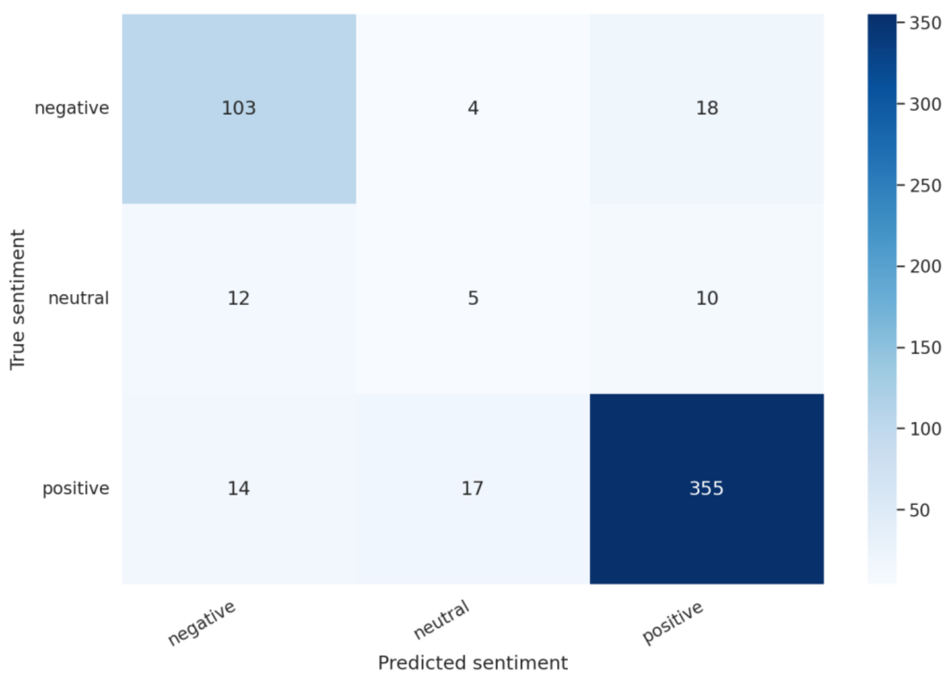

Figure: 4.2.8 Confusion Matrix of BERT

It is known that there are no machines in the world that can provide us with a 100 percent correct result; nonetheless, our model did provide us with a significantly good result. The following is the tabular representation of our model's final result:

Table: 4.2.1 Evaluation metrics of applied algorithms

Applied Algorithms	Precision	Recall	F-Measure	Accuracy (%)
KNN	0.84	0.78	0.81	78.12%
SVM	0.83	0.81	0.82	81%
LR	0.86	0.82	0.84	82.29%
MNB	0.90	0.85	0.88	85.42%
BERT	0.86	0.86	0.86	86.06%

From this table, it is clear that BERT outperformed the other four algorithms in terms of accuracy, following closely behind MNB acquired the highest precision, recall, and F-measure. KNN got the lowest accuracy.

The graphical representation of the table is given below:

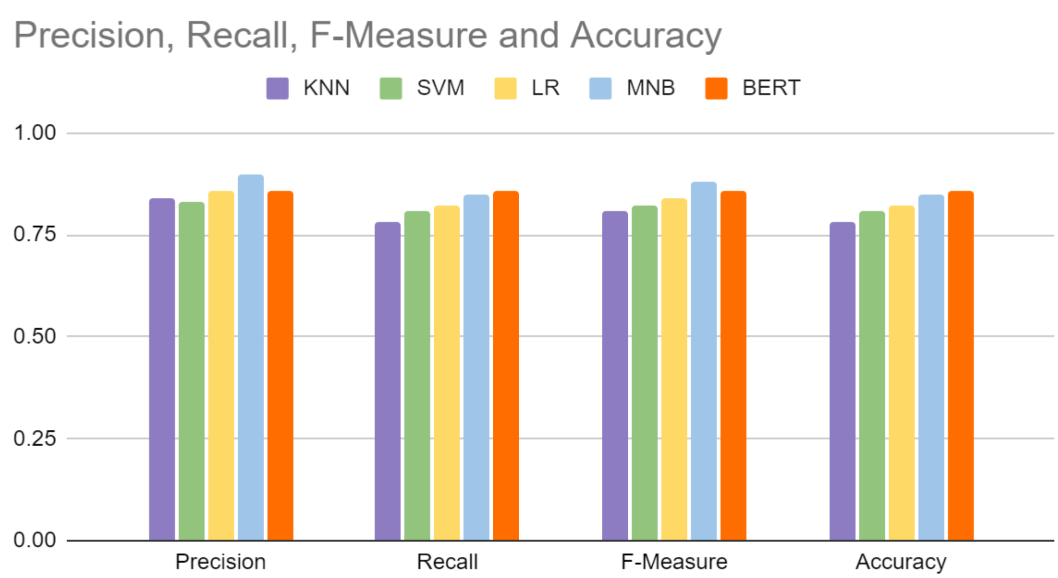

Figure: 4.2.9 Comparison between Applied Algorithms

4.3 Discussion:

We've seen a lot of work on reviews where sentiment analysis was applied before, but no models in this specific field, especially women safety app reviews, have been used. These algorithms, on the other hand, have been employed in the past on a variety of review-related studies, allowing us to learn more about the dataset's accuracy. In our study, we have applied total four machine learning algorithms where KNN, SVM, LR and MNB are supervised classifiers and BERT is an unsupervised deep learning classifier.

We have acquired a very excellent accuracy rate using several of those techniques, with the Bert algorithm outperforming every other algorithm on that unique dataset.

For this study the BERT algorithm achieved highest accuracy of 86.06% along with precision, recall and F-measure all being 0.86 where, following closely the MNB algorithm acquired an accuracy of 85.42% with precision being 0.90, recall being 0.85 and F-measure being 0.88.

KNN algorithm achieved accuracy of 78.12%, with 0.84 precision, 0.78 recall and 0.81 F-measure. SVM machine learning classifier attained 81% accuracy along with 0.83 precision, 0.81 recall and 0.82 F-score or F-measure.LR classifier got 82.29% accuracy as well as 0.86 precision, 0.82 recall and 0.84 F-measure. These values help to evaluate our model.

These models are used to determine and depict user or consumer attitudes. The dataset we utilized was unique, therefore we separated it into train and test datasets before applying the algorithms to prevent models from overfitting and to achieve an accurate evaluation from the models we have implemented.

CHAPTER 5

Impact on Society, Environment and Sustainability

5.1 Impact on Society:

This study will have a significant social impact. One of the most fundamental and unquestionable notions of a civilized society is women's protection. Every human being's safety is a top responsibility, but when it comes to women's safety issues, it becomes even more essential to tackle the difficulties. With the advancement of technology in this digital age, the idea of developing mobile applications connected to safety, notably women's safety, has arisen. We observe a lot of women's safety-related apps accessible for download on digital platforms, and it appears that the majority of them are user-friendly. Almost every application in this category is user-friendly, and some don't even require users to open them during an emergency since they can recognize a certain gesture or motion to activate and complete the mission.

However, as time passes, new users are becoming more interested in learning more about these applications and how they work before installing them. As a result, our study will assist individuals in gaining a general understanding of the applications from which we have gathered feedback from prior users. This study will have a significant impact on our society since it will broaden people's thinking and perspectives on various applications, and some may be persuaded to download their chosen app from among all of these. In addition, the sector of mobile application development is flourishing, and many individuals are pursuing professions in it. As a result, the developers will gain from this study since it will illustrate how the actual world reacts to their work, giving them room to improve. This will inspire them to create more innovative and secure yet user-friendly applications, which will undoubtedly assist our society towards becoming a more secure one in which every human being, regardless of background, caste, or gender, may wander securely at any time without trepidation. Because women's safety is a high priority in our society, our effort will aid in better understanding safety applications and assisting women in emergency circumstances. This research will add to the understanding of both users and developers in our society, as well as set the foundation for future work in this area, given there have been no previous studies on these applications.

5.2 Impact on Environment:

Our research does not impact the environment in any way, so it does not violate any law that can harm or cause a great effect on our environment.

5.3 Ethical Aspects:

We can confidently state that the research we conducted did not breach any research ethics. We avoided any kind of personal data like name, address, or other sensitive information about the reviewer while manually collecting data from the Google Play Store. As a result, the information gathered cannot be utilized against another individual and will not cause harm to anyone else. The information we have gathered is unique and authentic; it was not obtained from any website and has never been utilized before.

Because the data was acquired via the Google Play Store, every Android user has access to it, we made sure not to damage or pressurize anyone when obtaining it. This project is a way to gather user feedback on women's safety-related work, which will motivate developers to enhance it and bring it to a satisfactory level. This type of application can be a little step toward ensuring a safe and secure environment for women and other people in our society. We are confident that this research will never cause harm to anyone.

While completing any project-related tasks, our own personal computer (PC) was used. The research was carried out without the assistance of any other entity. We didn't use any other person's equipment, data, or information, and we didn't ask anyone for aid. We conducted our research while maintaining our moral integrity and adhering to all relevant institutional and governmental standards. Human dignity and privacy have been protected.

5.4 Sustainability Plan:

The principal goal of this research is to assist users of mobile applications by providing a viewpoint on how previous users have experienced using these apps and if they are worth installing and using. Our study also attempts to assist developers by providing information on how their developed apps perform in the actual world, allowing them to make improvements.

We need more reviews connected to these applications, additional applications with advanced capabilities in this sector with actual user reviews, in order to make this work progress and sustainable in the future. It will benefit more individuals in the future, whether they are general users or mobile application developers working in IT farms, by enhancing their knowledge of existing apps. As a result, our research will assist both current and future users of these applications, as well as existing and future developers, in expanding their work in the field of safety.

CHAPTER 6

Summary, Conclusion, Recommendation, and Implication for Future Research

6.1 Summary of the Study:

Our whole project relies upon user feedback for applications related to women's safety. This type of mobile application is really beneficial to everyone, but particularly to women. Our research will assist the developers of such applications in gaining a clear insight into the level of user satisfaction. We've divided the reviews into three main categories based on user feedback: positive, negative, and neutral, and we're able to determine the users' emotions and satisfaction levels in this way.

As feedback, we gathered written reviews as well as the star ratings offered by users of those applications and then labeled the textual data with star ratings. Then we sorted it into three categories: positive, negative, and neutral. To analyze the model, After collecting the data of women’s safety app review, the scikit-learn library from python was used to preprocess the dataset, then a Machine learning supervised classifier such as Multinomial Naive Bayes, Support Vector Machine, Logistic Regression, and K-Nearest Neighbor(K=3) was applied along with unsupervised deep learning algorithm BERT. BERT achieved the highest accuracy of 86.06%, outperforming the other four algorithms.

We concluded our research by completing all of the steps outlined above. We discovered that a high proportion of users of these applications have a positive opinion of them. This research will contribute to the field of safety-related concerns by educating individuals about current apps and urging developers to improve their work.

6.2 Conclusions:

Sentiment analysis is an excellent approach to comprehend other people's perspectives on something or about anything, whether good, negative, or neutral, which is easy for us humans to grasp but tough for computers. This paper heeds on sentiment analysis using Natural Language Processing (NLP) of the reviews of a safety-related app designed specifically for women that were accessible on the Google Play store.

This women's safety app review for evaluation has never been conducted to any prior study. According to our findings, this is the first study to use this dataset, which was collected straight from the Google Play store's app review section.

This research can help the developers to improve their work to reach their full potential and open scope for development. These kinds of applications are very easy to use and can actually be used as a safety procedure for women, as a result, women can have the confidence to face any danger.

This is a basic attempt to develop women's empowerment, since how can they acquire empowerment if they do not even feel comfortable going out alone.

6.3 Implication for Further Study:

Technology evolves on a daily basis, and everything evolves and improves along with it for the betterment of mankind. In the same way that every other research industry is generating and inventing new models for a variety of sectors. We are utilizing five models for our research, although there are many more contemporary and efficient models other than those five. As a result, various machine learning models, as well as deep learning models, can be applied to this research in order to improve it. In order to improve, we want to broaden the scope of our studies. In the future, we'd like to expand our dataset and use different algorithms.

Additionally, the developers are working on a number of new and complex applications, all of which have the potential to provide new user experiences through reviews. This will aid in the expansion of our dataset, allowing us to deploy more advanced algorithms. We also want to undertake a sentiment analysis of each yearly update individually so that we can see if the developers are upgrading these programs in the appropriate direction.

References:

[1] Sentiment Analysis Explained, available at <<https://www.lexalytics.com/technology/sentiment-analysis>>,last accessed on 26-11-2021 at 11.00 PM

[2] Sentiment Analysis: A Definitive Guide, available at <<https://monkeylearn.com/sentiment-analysis/>> ,last accessed on 26-11-2021 at 11.00 PM

[3] Women safety, sexual ,assault, awareness ,available at <<https://today.yougov.com/topics/lifestyle/articles-reports/2019/03/28/women-safety-sexual-assault-awareness>> ,last accessed on 26-11-2021 at 12.00 PM

[4] Eftekhar Hossain, Omar Sharif, Mohammed Moshiul Hoque, "Sentiment Polarity Detection on Bengali Book Reviews Using Multinomial Naïve Bayes", ICACIE Advances in Intelligent Systems and Computing,281-292.July 2020

[5] S. Al-Natour and O. Turetken, "A comparative assessment of sentiment analysis and star ratings for consumer reviews", International Journal of Information Management, vol. 54, p. 102132, 2020.

[6] Dany Pratmanto, Rousyati Rousyati, Fanny Fatma Wati, Andrian Eko Widodo, Suleman Suleman, Ragil Wijianto,"App Review Sentiment Analysis Shopee Application In Google Play Store Using Naive Bayes Algorithm",Journal of Physics: Conference Series,Volume 1641,24 November 2020.

[7] Ronal WatrianthosSudi SuryadiDeci IrmayaniMarnis NasutionElida F.S. Simanjorang,"Sentiment Analysis Of TravelokaApp Using Naïve Bayes Classifier Method",International Journal of Scientific and Technology Research,Volume 8, Issue 07, July 2019.

[8] Atiqur Rahman, Md. Sharif Hossen,"Sentiment Analysis on Movie Review Data Using Machine Learning Approach",International Conference on Bangla Speech and Language Processing (ICBSLP), 27-28 September, 2019.

[9] Sakshi Ranjan,Subhankar Mishra,"Comparative Sentiment Analysis of App Reviews",Computing, Communication and Networking Technologies, ICCCNT 2020 (2020).

[10] Xing Min Lin,Chun Heng Ho,Lu Ting Xia,Ruo Yi Zhao, "Sentiment analysis of low-carbon travel APP user comments based on deep learning ",Sustainable Energy Technologies and Assessments,Volume44(3):101014 (2021).

[11] Gopalkrishna Barkur,Vibha,Giridhar B. Kamath, "Sentiment analysis of nationwide lockdown due to COVID 19 outbreak:Evidence from India ",Asian J Psychiatr,2020 Jun;51:102089.

[12] Ayushi Mitra, "Sentiment Analysis Using Machine Learning Approaches (Lexicon based on movie review dataset)",Journal of Ubiquitous Computing and Communication Technologies (UCCT) Vol.02/ No.03,Pages: 145-152,(2020).

[13] Mohammad Rezwanul,Ahmad Ali,Anika Rahman, "Sentiment Analysis on Twitter Data using KNN and SVM",(IJACSA) International Journal of Advanced Computer Science and Applications,Vol. 8, No. 6, 2017.

[14] Lopamudra Dey,Sanjay Chakraborty,Anuraag Biswas,Beepa Bose,Sweta Tiwari, "Sentiment Analysis of Review Datasets using Naïve Bayes' and K-NN Classifier",International Journal of Information Engineering and Electronic Business, Volume-8, Issue-4, pp.54-62, 2016.

[15] Kudakwashe Zvarevashe,Oludayo O. Olugbara, "A Framework for Sentiment Analysis with Opinion Mining of Hotel Reviews",2018 Conference on Information Communications Technology and Society (ICTAS).

[16] Zeenia Singla, Sukhchandan Randhawa,Sushma Jain, "Sentiment analysis of customer product reviews using machine learning",2017 International Conference on Intelligent Computing and Control (I2C2),India,June 2017.

[17] Omar Sharif, Mohammed Moshiul Hoque, Eftekhar Hossain, "Sentiment Analysis of Bengali Texts on Online Restaurant Reviews Using Multinomial Na¨ıve Bayes",1st International Conference on Advances in Science, Engineering and Robotics Technology 2019 (ICASERT 2019).

[18] Lutfi, A. Permanasari and S. Fauziati, "Sentiment Analysis in the Sales Review of Indonesian Marketplace by Utilizing Support Vector Machine", Journal of Information Systems Engineering and Business Intelligence, vol. 4, no. 1, p. 57, 2018. Available: 10.20473/jisebi.4.1.57-64.

[19] Palak Baid,Apoorva Gupta,Neelam Chaplot,"Sentiment Analysis of Movie Reviews using Machine Learning Techniques",International Journal of Computer Applications (0975 – 8887) Volume 179 – No.7, December 2017.

[20] P. Johri, J. Verma and S. Paul, Applications of Machine Learning.

[21] Bayhaqy, A., Sfenrianto, S., Nainggolan, K.; Kaburuan, E. R. Sentiment analysis about e-commerce from tweets using decision tree, K-nearest neighbor, and Naïve Bayes. 2018 International Conference on Orange Technologies (ICOT). 2018

[22] How Does Support Vector Machine (SVM) Algorithm Works In Machine Learning? Analytics Steps ", Analyticssteps.com, Available online at <<https://www.analyticssteps.com/blogs/how-does-support-vector-machine-algorithm-works-machine-learning>> last accessed on 27-11-2021 at 10.00 PM

[23] A. Tyagi and N. Sharma, "Sentiment Analysis using Logistic Regression and Effective Word Score Heuristic", International Journal of Engineering & Technology, vol. 7, no. 224, p. 20, 2018.

[24] Mayur Wankhade, A Chandra Sekhara Rao, Suresh Dara, Baijnath Kaushik, "A Sentiment Analysis of Food Review using Logistic Regression", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT),, Volume 2, Issue 7, pp.251-260, September,2017.

[25] Multinomial Logistic Regression ,Ibm.com, available at <<https://www.ibm.com/docs/en/spss-statistics/version-missing?topic=regression-multinomial-logistic>> ,last accessed on 28-11-2021 at 11.00 PM

[26] J. Shen, X. Liao and Z. Tao, "Sentence-level sentiment analysis via BERT and BiGRU", 2019 International Conference on Image and Video Processing, and Artificial Intelligence, 2019.

[27] BERT Explained: State of the art language model for NLP, available at <<https://towardsdatascience.com/bert-explained-state-of-the-art-language-model-for-nlp-f8b21a9b6270>>, last accessed on 1-12-2021 at 12.00 PM

[28] E. Biswas, M. Karabulut, L. Pollock and K. Vijay-Shanker, "Achieving Reliable Sentiment Analysis in the Software Engineering Domain using BERT", 2020 IEEE International Conference on Software Maintenance and Evolution (ICSME), 2020.

[29] Data Analytics, available at <<https://vitalflux.com/accuracy-precision-recall-f1-score-python-example/>>, last accessed on 27-11-2021 at 11.00 PM

Final Defense Report- Project ID-FL20D021

ORIGINALITY REPORT

27%	20%	18%	16%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	dspace.daffodilvarsity.edu.bd:8080 Internet Source	5%
2	Submitted to Daffodil International University Student Paper	3%
3	www.jatit.org Internet Source	1%
4	arxiv.org Internet Source	1%
5	Xing-Min Lin, Chun-Heng Ho, Lu-Ting Xia, Ruo-Yi Zhao. "Sentiment analysis of low-carbon travel APP user comments based on deep learning", Sustainable Energy Technologies and Assessments, 2021 Publication	<1%
6	www.journaltocs.ac.uk Internet Source	<1%
7	"Sentimental Analysis and Deep Learning", Springer Science and Business Media LLC, 2022 Publication	<1%