IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

NEGATIVE THINKING DETECTION USING SOCIAL MEDIA CHARTS

Lalitha.A
Assistant professor, Department of Computer Science and Engineering
SRM Valliammai Engineering College, Chennai, Tamil Nadu, India

Syed Muntasir SK
Department of Computer Science and Engineering
SRM Valliammai Engineering College, Chennai, Tamil Nadu, India

Surya GM
Department of Computer Science and Engineering
SRM Valliammai Engineering College, Chennai, Tamil Nadu, India

Roshan R
Department of Computer Science and Engineering
SRM Valliammai Engineering College, Chennai, Tamil Nadu, India

Abstract- Social media text analytic is the process of deriving information from text sources. Text analysis can be applied to any text-based dataset, including social media. Negative thinking is a major problem faced today by society even we are finding in social media. Negative thinking have affected the quality of life and economic growth badly. We can identify the negative thinking patterns and predict the negative thinking by detecting and analyzing the historical data. However, some negative thinking are unregistered and unsolved due to a lack of evidence. Thus, detecting negative thinking is a still challenging task. We can use social media to detect negative thinking related activities. Because social media users sometimes convey messages related to his or her surrounding environment via social media message. In this paper, we proposed a machine learning approach to detect the negative thinking and analyze it on its type. As the first step, we fetch the text messages using predefined keywords relating to the negative thinking. Then, after the Pre-processing, we applied a support vector machine- based filtering approach to eliminate the noise. And then Random forest is used for classification. Then in the final stage, we analyze and categorize the negative thinking type.

Keywords – Support Vector Machine, Negative Thinking, Evidence, Detecting, Messages, Social Media, Historical.

I. INTRODUCTION

Suicide may be a serious public health concern, with many people worldwide experiencing suicidal thoughts or attempting suicide annually. Social media platforms became an increasingly popular venue for people to precise their struggles with psychological state and suicidal ideation. Detecting suicidal content on social media can provide a chance to intervene and offer support to those in need. •Recent advances in tongue processing (NLP) and machine learning (ML) techniques have enabled the event of algorithms which will detect suicidal content in social media data. These algorithms can analyze large volumes of social media data and identify individuals who could also be in danger of suicide, allowing timely intervention and support. •Detecting suicidal content in social media data may be a complex task, because it involves understanding the nuances of language and context employed by individuals who express suicidal ideation. Researchers and data scientists have used a spread of NLP and ML techniques, including deep learning, transfer learning, and graph-based approaches, to develop accurate and efficient algorithms for detecting suicidal content in social media. •Overall, the detection of suicidal content in social media has the potential to save lots of lives and improve psychological state outcomes. By detecting and intervening with individuals in danger of suicide, we will provide timely support and resources to those in need, ultimately reducing the incidence of suicide and improving psychological state for all.

II. LITERATURE REVIEW

Web Mining To Detect Online Spread Of Terrorism:

In several regions of the world, terrorism has recently increased exponentially. Due to the massive increase in terrorist activity, it is critical to put an end to terrorism and stop its further development before it threatens human life or property. The internet has evolved into a tool for disseminating terrorism through speeches and videos as a result of technological advancement. Terrorist groups utilise the medium of the internet to hurt and defame individuals and also promote terrorist acts through web pages that force people to join terrorist organisations and conduct crimes on the behalf of those organisations. For the goal of effective system development, web mining and data mining are utilised together. Web mining even includes a variety of useful text mining techniques.

Detection of Dangerous Web Pages Based on the Analysis of Suicidal Content Using Machine Learning Algorithms:

The mission of preventing suicide is currently one of the health industry's top concerns. As a result, it's critical to spot suicide risk factors early on. The feasibility of real-time detection of websites viewed that contain suicidal statements is discussed in this article. Web pages are categorised using text analysis of the content on them. The development of the server and the browser extension might be considered separate aspects of this

endeavour. The extension gathers data on the content of the user's visited web sites and sends it to the server. On the server, the page classification procedure is carried out. The effectiveness of different machine learning algorithms for identifying suicidal webpages is compared in the work's concluding section.

A Deep Learning Approach to Fast, Format-Agnostic Detection of Malicious Web Content:

Today's Internet has a severe problem with malicious web material. In this study we offer a deep learning strategy to detecting malicious online pages. Our approach operates directly on a language-agnostic stream of tokens extracted directly from static HTML files with a straightforward regular expression, in contrast to previous work on web content detection that relied on syntactic parsing or on emulation of HTML and Javascript to extract features. This enables it to work quickly in high-frequency data environments, such as firewalls and web proxies, and prevents the sophisticated parsing and emulation code from exposing its attack surface. Our neural network analyses content at hierarchical geographical scales, in contrast to well-known methods like bag-of-words models, which disregard spatial information, enabling our model to capture localization.

III. PROPOSED MODEL

In this study, we provide a technique for identifying suicidal profiles. Then, we analyse a number of profiles from the social network by utilising all of the data that is currently available. Then, in order to differentiate between suicidal and non-suicidal profiles, we incorporate a number of criteria. Using various data mining methods and methodologies, these features can either be explicitly extracted from the user profile or implicitly inferred. Here, we concentrate on emotional characteristics and sentiment analysis, which provide clues regarding the mental health of suicidal profiles. Additionally, we employ account features to recognise users based on the shared data on their profiles.

We propose a system with the primary purpose of building a website where users may examine any webpage or any website for any sign of terrorist activity. Our website will give users the option to input the URL of the pages they wish to scan in order to achieve this. Our technology will count the words on the entire webpage when you enter the URL and compare them to words in our database. We will assign a score to each word that we keep in our database. Our technology will retrieve the scores for each word that appears in the user's web page from our database before calculating a website's overall rank.

.3.1 Random Forest–

Like its name suggests, the random forest method uses a lot of different decision trees that work together. The class with the most votes becomes the prediction made by our model. The random forest's individual trees each spit out a class prediction.

3.2 Decision Tree–

A decision tree is a graph in the shape of a tree where the nodes stand in for the places where we choose an attribute and pose a question, the edges for the replies, and the leaves for the actual output or class label.
They work with a straightforward linear decision surface when making non-linear decisions.

3.3 Support Vector Machine–

One of the most well-liked supervised learning algorithms, Support Vector Machine, or SVM, is used to solve Classification and Regression problems. However, it is largely employed in Machine Learning Classification issues.

The SVM algorithm's objective is to establish the best line or decision boundary that can divide n-dimensional space into classes, allowing us to quickly classify fresh data points in the future. A hyperplane is the name given to this optimal decision boundary. SVM selects the extreme vectors and points that aid in the creation of the hyperplane. Support vectors, which are used to represent these extreme instances, form the basis for the SVM method.

3.4 Naïve Bayes–

A group of classification algorithms built on the Bayes' Theorem are known as naive Bayes classifiers. It is a family of algorithms rather than a single method, and they are all based on the idea that every pair of features being classified is independent of the other.

3.5 Logistic Reression–

A predictive analysis is the logistic regression. To describe data and explain the relationship between one dependent binary variable and one or more independent nominal, ordinal, interval, or ratio-level variables, we employ logistic regression.

3.6 K-Nearest Neighbours–

K nearest neighbours is a straightforward method that categorises new cases based on a similarity metric and stores all of the existing cases (e.g., distance functions). KNN is a non-parametric technique that has been utilised in statistical estimates and pattern recognition since the early 1970s.

3.7 Architecture Diagram–

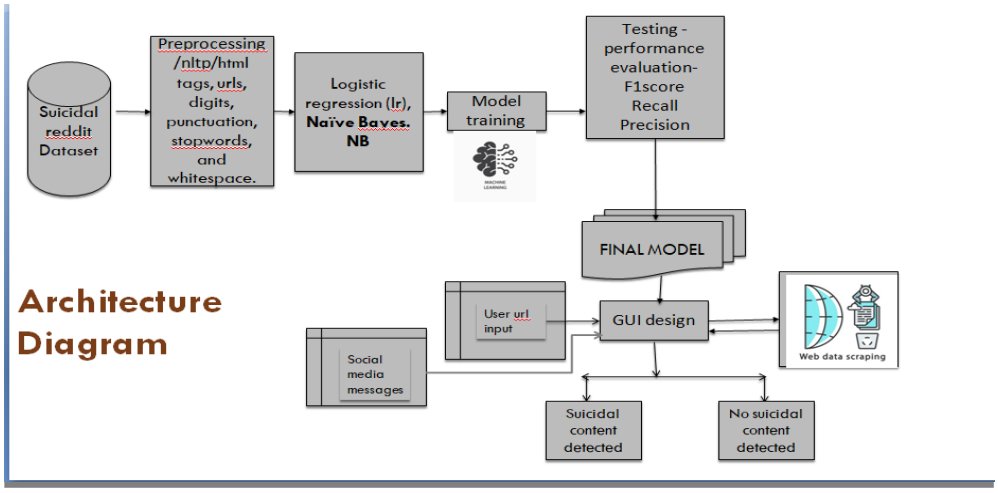

Figure.1

IV. METHODOLOGY

4.1 Data Collection–

In order to employ data analysis to discover repeating patterns, collection of data enables you to keep a record of earlier events. With machine learning techniques, you may create prediction models from these patterns that track trends and foretell future changes. Effective data collection techniques are essential to creating high-performing predictive models since predictive models are only as good as the data on which they are based. The information in the data must be accurate (trash in, garbage out) and pertinent to the work at hand. For instance, a debt default model might profit from rising petrol prices over time but not from tiger population increases. We pull the data for this module from the Kaggle dataset archives. This dataset includes details about divorce in previous years.

4.2 Data Cleaning–

A crucial stage in any machine learning effort is data cleaning. Data that may be inaccurate, incomplete, duplicated, or formatted inappropriately are removed or modified as part of this module's data cleaning process to make the data ready for analysis. You can investigate your tabular data using a variety of statistical analysis and data visualisation tools to find data cleaning activities you might wish to carry out.

4.3 Data Pre processing-

An essential phase in the data mining process is data preparation. It describes the processes of preparing data for analysis by cleansing, converting, and integrating it. The purpose of data preprocessing is to enhance the data's quality and suitability for the particular data mining operation.

4.4 Model Training & Testing–

One of the most well-liked supervised learning algorithms, Support Vector Machine, or SVM, is used to solve Classification and Regression problems. However, it is largely employed in Machine Learning Classification issues.

The SVM algorithm's objective is to establish the best line or decision boundary that can divide n-dimensional space into classes, allowing us to quickly classify fresh data points in the future. A hyperplane is the name given to this optimal decision boundary. SVM selects the extreme vectors and points that aid in the creation of the hyperplane. Support vectors, which are used to represent these extreme instances, form the basis for the SVM method.

4.5 Exploratory Data Analysis–

To further comprehend our data, we examined the most frequently used words from each subreddit using a Count Vectorizer (The words used were similar with some subtle differences).

V.EXPERIMENTAL ENVIRONMENT

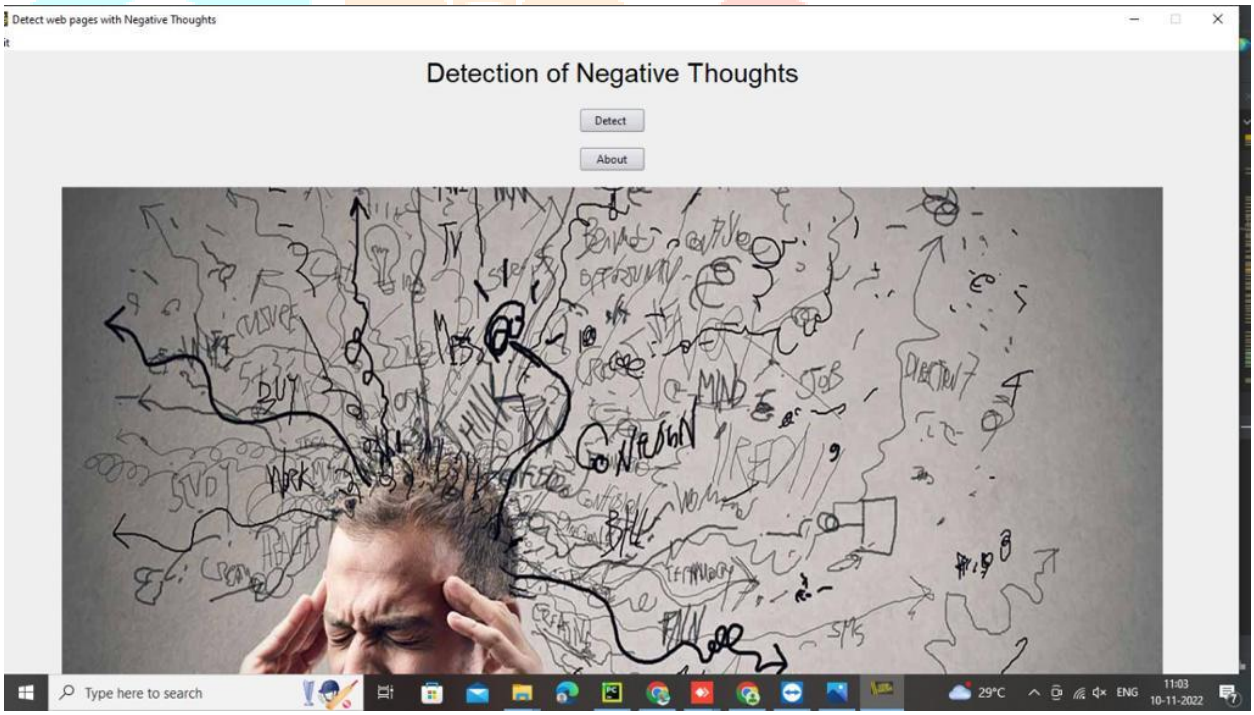

Figure.2

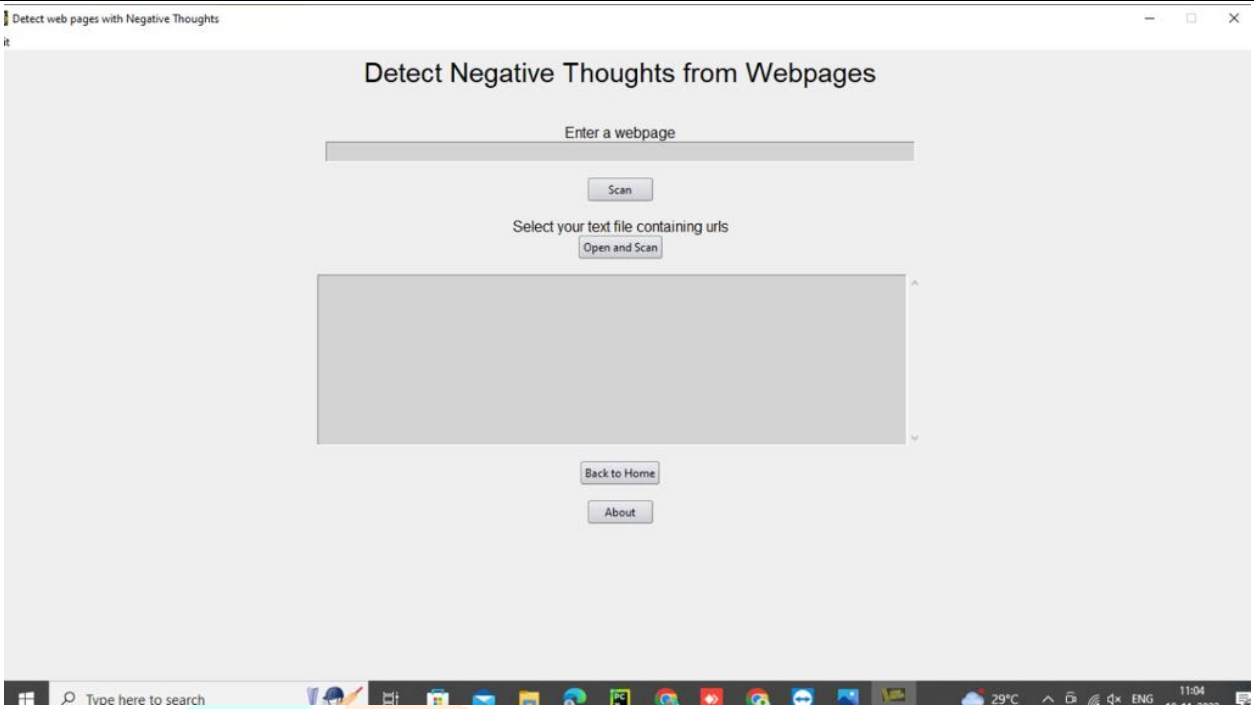

Figure.2.1

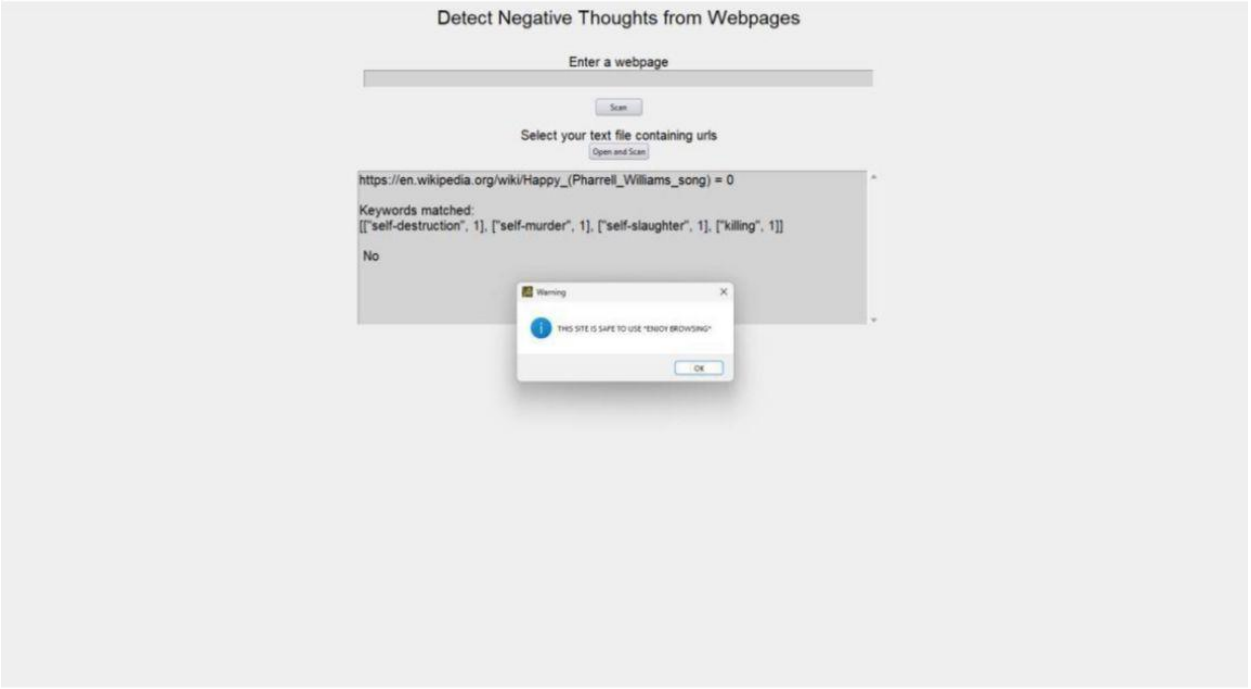

Figure.2.2

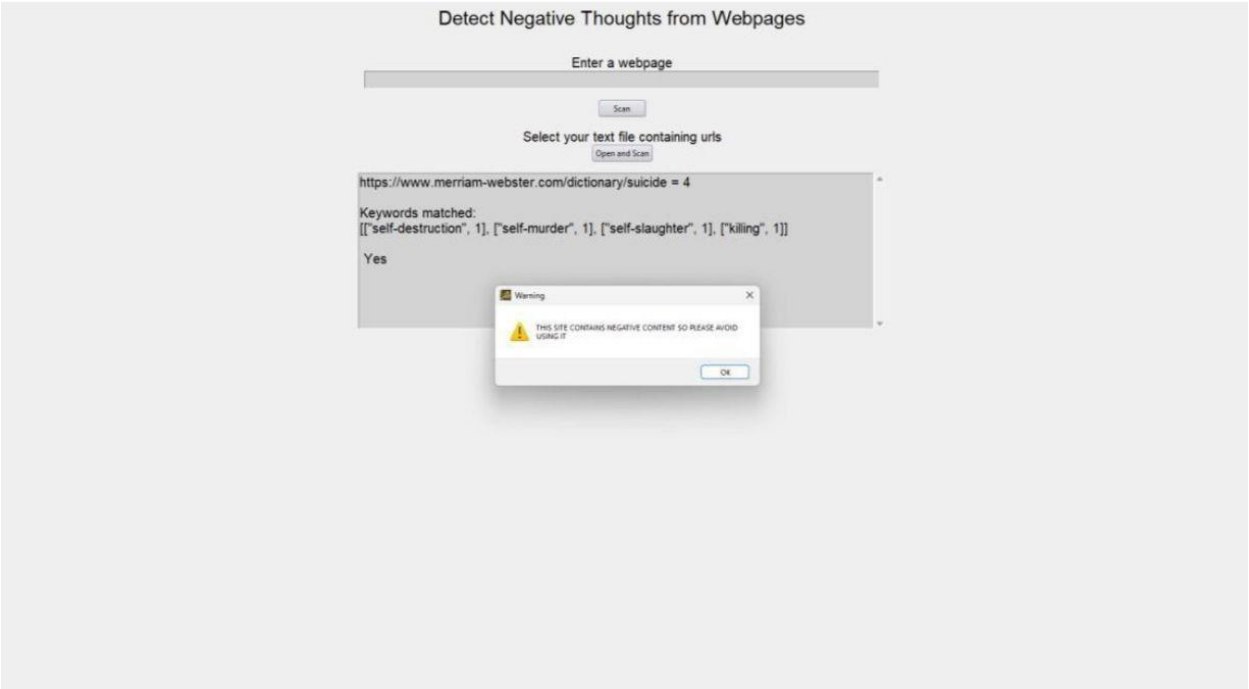

Figure.2.3

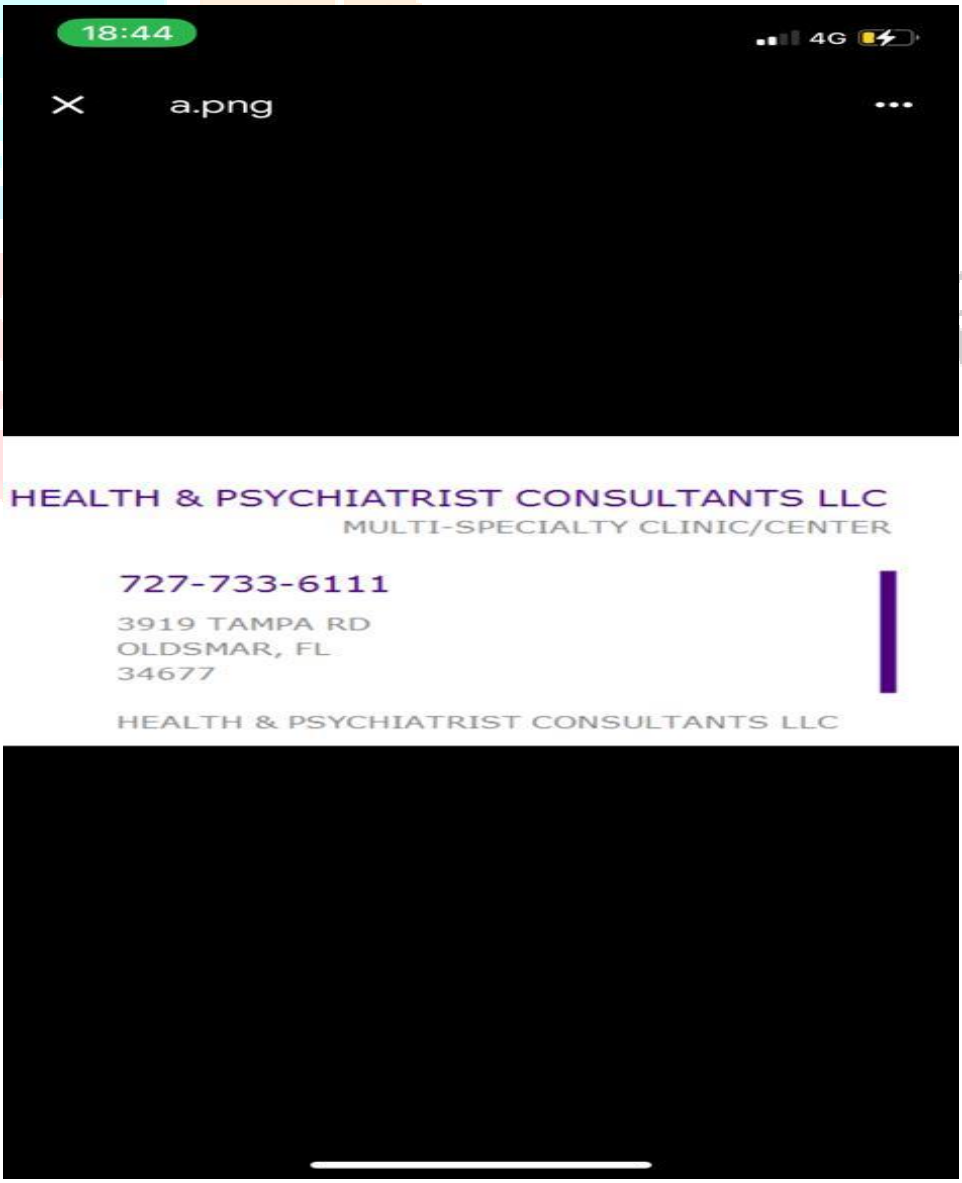

Figure.2.4

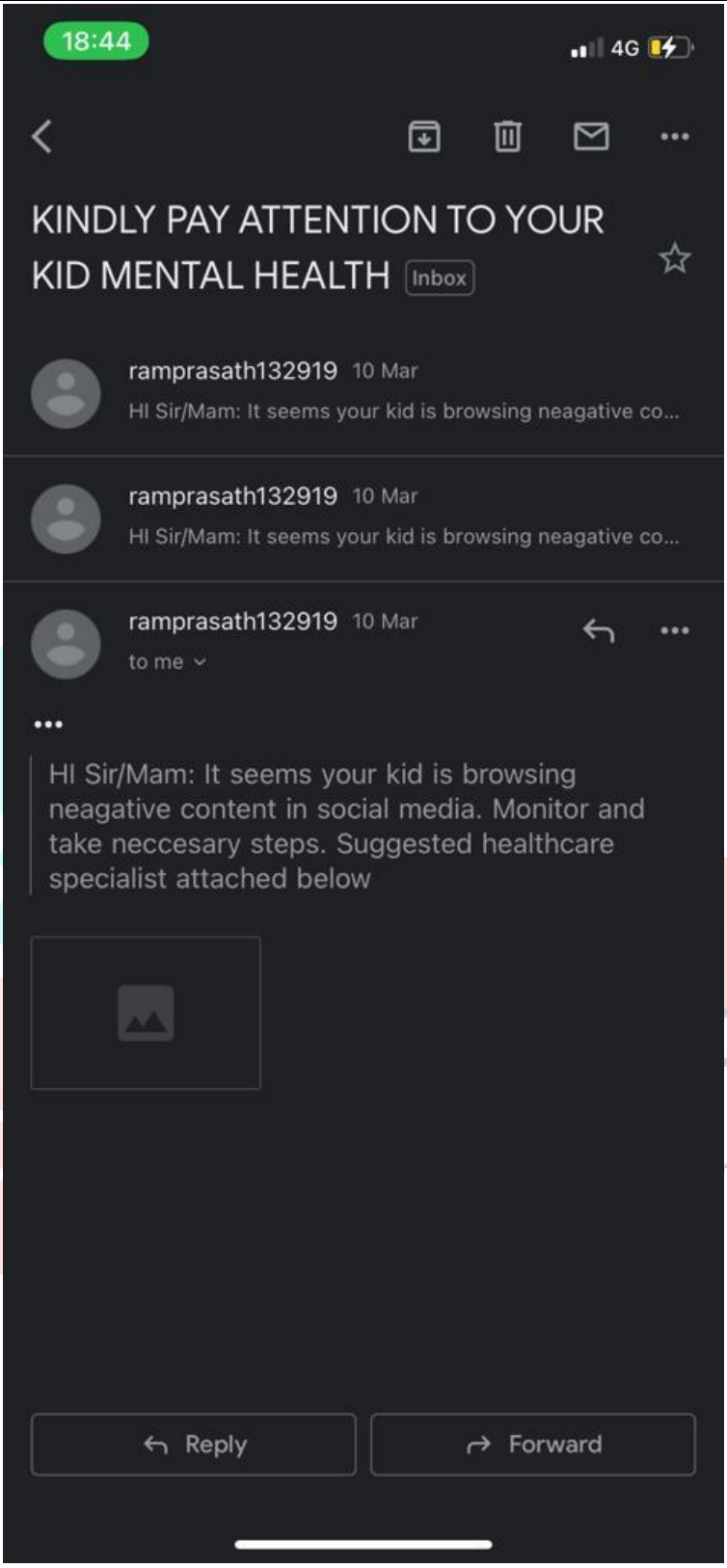
18:44
4G
KINDLY PAY ATTENTION TO YOUR KID MENTAL HEALTH
Inbox
ramprasath132919 10 Mar
HI Sir/Mam: It seems your kid is browsing neagative co...
ramprasath132919 10 Mar
HI Sir/Mam: It seems your kid is browsing neagative co...
ramprasath132919 10 Mar
to me
HI Sir/Mam: It seems your kid is browsing neagative content in social media. Monitor and take neccesary steps. Suggested healthcare specialist attached below
Reply
Forward

Figure.2

VI. DISCUSSSION

Everything is coded, from the environment to the finished system projects, primarily in web-oriented languages like Python. Since the Jupyter notebook makes it possible to create kernels for machine learning and deep learning, it is utilised as an integrated development environment. Then moving on to the coding portion, there are a number of library packages that can be used, such as Scikit Learn, Pandas, and Matplotlib. A number of algorithms are then used to test their performance by visualising the data, comparing its accuracy, and ultimately forecasting the negative thoughts. The data is further divided into training and testing derived from a method model selection with a testing size of 20% of the data we feed.

VII. CONCLUSION

In contemporary communities, cyber bullying has become a serious concern. This study provided a model for the detection of cyber bullies that made use of a number of classifiers based on the TF-IDF and Word2Vec feature extraction. Also, a number of machine learning-based text classification techniques were researched. The tests were carried out using a global Twitter dataset. The experimental results show that LR, with classification accuracy and F1 score of 90.57% and 0.9280, respectively, had the best accuracy and F1 score in our dataset. Between LR, SGD, and LGBM classifier performance, there is a little difference; SGD attained an accuracy of 90.6%, but the F1 score was lower than LR. Nonetheless, the F1 score for the LGBM classifier was 0.9271, and its accuracy was 90.55%. This indicates that L outperforms other classifiers. Also, it was discovered during the trials that L outperforms the other classifiers employed in this study in terms of prediction time and performance as data size grows. As a result, SGD performs almost as well as LR, but the error is not as small. In order to improve the detection accuracy of machine learning, feature extraction is a crucial component. We didn't look at many feature extraction methods in this research.

REFERENCES

[1] James SL, Abate D, Abate KH, Abay SM, Abbafati C, Abbasi N, et al. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories,

[2] Kessler RC, Birnbaum H, Bromet E, Hwang I, Sampson N, Shahly V. Age differences in major depression: results from the National Comorbidity Survey Replication (NCS-R). Psychol Med. (2010)

[3] Hodgetts S, Gallagher P, Stow D, Ferrier IN, O'Brien JT. The impact and measurement of social dysfunction in late-life depression: an evaluation of current methods with a focus on wearable technology. Int J Geriatr Psychiatry. (2017) 32:247-55. doi: 10.1002/gps.4632Ke Yuan, Longwei Yang, Yabing Huang, Zheng Li," Heart Disease Prediction Algorithm Based on Ensemble Learning", 7th International Conference on Dependable Systems and Their Applications (DSA), pp: 30-55, 2020.

[4] Fiske A, Wetherell JL, Gatz M. Depression in older adults. Annu Rev Clin Psychol. (2009) 5:363-89. do:

[5] Rodda J, Walker Z, Carter J. Depression in older adults. BMJ. (2011)

[6] "Suicide."https://www.who.int/news-room/fact-sheets/detail/suicide (accessed Nov. 24, 2020).Mahaveer, Puneet, Deepika, " Cardiovascular Disease Prediction Analysis using Classification Techniques", IEEE Delhi Section Conference (DELCON) , pp. 36-45, 2022.

[7] "GHO World Health Statistics data visualizations dashboard Suicide, WHO.S

[8] Facebook artificial intelligence spots suicidal users - BBC News." https://www.bbc.com/news/technology-39126027 (accessed Nov. 24, 2020)

[9] "«Rekomendacii po rasprostraneniyu v SMI informacii o sluchayah samoubiistva»."https://www.rospotrebnadzor.ru/documents/details.php? ELEMENT_ID= 6735 (accessed Nov. 24, 2020)..

[10] "O devatel'nosti Rospotrebnadzora po predotvrashcheniyu samoubijstv sredi detej i podrostkov- RSS - Oficial'nyi sat Rospotrebnadzora."http://11.rospotrebnadzor.ru/rss_all/-/asset_publisher/Kq6J/content/id/382348 (accessed Nov.24, 2020).

[11] "Chislo detskih suicidov v Rossi v 2016 godu vyroslo pochti na 60%."https://www.interfax.ru/russia/554375 (accessed Nov. 24, 2020).

[12] V. Chandler, "Google and suicides: what can we learn about the use of internet to prevent suicides?," Public Health, vol. 154, pp. 144-150, Jan. 2018, doi:10.1016/i.puhe.2017.10.016.

[13] L.Biddle et al., "Suicide and the Internet: Changes in the accessibility of suicide-related information between 2007 and 2014," J. Affect. Disord., vol. 190, pp. 370-375, Jan. 2016, doi: 10.1016/j.jad.2015.10.028.

[14] Tor Proiect Anonymity Online."https://www.torproject.org/ (accessedNov. 27, 2020).

[15] C. M. Mörch, L. P. Côté, L. Corthésy-Blondin, L.Plourde-Léveillé, L. Dargis, and B. L. Mishara, "The Darknet and suicide," J. Affect. Disord., vol. 241, pp.127-132, Dec. 2018, doi: 10.1016/j.jad.2018.08.028.