Applicability of Machine Learning Algorithms to Analyze Despondency Comments on Social Media Using Analytical Hierarchical Process (AHP)

Damilola Alaba HALLY
LCU/PG/001538

Being a MSc Thesis Submitted to the Department of Computer Science, Faculty of Natural and Applied Sciences, Lead City University, Ibadan, Oyo State, Nigeria

In Partial Fulfillment of the Requirements for the Award of Master Degree (MSc) in Computer and Information Science

2022

Certification

This is to certify that Damilola Alaba HALLY with Matriculation Number LCU/PG/001538 carried out this research work titled" An Advanced Integration Of Machine Learning Techniques To Extract Despondency (Depression) Symptoms on Social Media" in the Department Of Computer Science ,Faculty of Natural and Applied Sciences , Lead City University Ibadan, Oyo state for the Award of Master Degree (MSc) in Computer and Information Science and this has not been previously submitted , Nigeria for the award of Master Degree in computer and information science.

…….. ……

Dr. Wumi Ajayi Date
(Supervisor)

..... ……..

Dr Wilson Sakpere Date
(Head of Department)

Dedication

This Research work is dedicated to the Almighty God for his Grace and mercy upon my life, and to my family members who have been a wholesome support to me.

Acknowledgement

I want to thank the prestigious institution Lead City University for the opportunity to learn and complete my Master's Program and to the University Library for provision of adequate materials needed.

I am grateful to the Department of Computer Science for giving me the privileged and opportunity to study and to learn, my appreciation goes to my supervisor Dr. W.S. Ajayi, I also like to acknowledge my lecturers, Prof. S.O. Akinola, Dr. W. Sakpere, Dr. A.A. Waheed, Dr. R.A. Badru for their supports and always ensuring that things are done rightly.

My appreciation goes to my family who has always supported and sponsored me through this program.

"Even though the above-mentioned institutions and persons have assisted in the process of this research work, I alone stand responsible for the errors, if any, found in the work."

Abstract

Depression is a serious mental illness that affects an individual's professional and personal life. With the development of internet usage people have started to share their experiences and challenges with mental disorder through online platforms. Social media platforms come close to being a true digitization of the human social experience. In many cases people would prefer to express themselves online rather than offline especially in completed suicide attempts around the world. This thesis objectives are to extract despondency indicative social media posts, categorize these posts and then apply an integration of machine learning techniques to generate markers in identifying depressive comments in social media. This will be examined using five algorithms; Support Vector Machine, Logistic Regression, K-Nearest Neighbor, Naves Bayes and Linear Regression. The Analytical hierarchical Performance was used to determine the best algorithm to detect depression in terms of performance metrics. This process identifies users who are at risk of depression to initiate quick intervention. The result of this thesis shows that the support Vector machine has the highest performance metrics. This signifies that the support vector Machine is the best algorithm to apply for the extraction of Despondency symptoms in social media, this will also determine the best machine learning Algorithm that has the best performance and accuracy in detecting despondency symptoms on social media. It is recommended based on the conclusions of this thesis that the support vector machine Algorithm be used in institutions social media platforms to adequately monitor despondency in individuals

Keywords: Despondency, Machine learning Algorithm, Analytical Hierarchical Performance, Support Vector Machine, Logistic Regression,

Word count: 246.

Table of Contents

Chapter Two; Literature Review

Chapter Five: Conclusion

List of Tables

List of Figures

List of Acronyms

Acronyms	Meaning
PHQ-9	Patient Health Questionnaire 9 (Mental Health Disorder Screening)
CES-D	Centre for Epidermiologic Studies Depression Scale
SNS	Social Networking Sites
KNN	K Nearest Neighbour (Machine Learning Algorithm)
LR	Logistic Regression (Machine Learning Algorithm)
DT	Decision Tree (Machine Learning Algorithm)
SVM	Support Vector Machine (Machine Learning Algorithm)
NB	Naïve Bayes (Machine Learning Algorithm)
AHP	Analytic Hierarchy Process
WEKA	Waikato Environment for Knowledge Analysis
NLP	Natural Language Processing
LIWC	Linguistic Inquiry and Word Count

Appendices

Chapter 0ne

Introduction

1.1 Background to the Study

It is evident in the world today the subtle and extreme effects of the various forms of internet explorations, Several researchers have also concluded through studies that these experiences tend to increase expressiveness in personalities through social media platforms. This gives us an interesting depth of knowledge into the behavior of persons when they are using these platforms or otherwise .The process has given researchers the ability to study the traits of different forms of personality dimensions through the data gathered. The data includes responses and interactions on social media and bringing it into perspective .There has been an immense increase in social media engagements since the availability of several internet providers in Africa ,making it affordable and easy to connect to the world community, thereby giving everyone a voice to express themselves[1]. According to report, there were 33.00 million social media users in Nigeria as at January 2021. The number of social media users in Nigeria increased by 6.3 million users between 2020 and 2021. The number of social media users in Nigeria was equivalent to 15.8% of the total population in January 2021[2]. Subsequent to this data it is possible to infer depression from social media platforms.

There were 33.00 million social media users in Nigeria as at January 2021. The number of social media users in Nigeria increased by 6.0 million (+22%) between 2020 and 2021. The number of social media users in Nigeria was equivalent to 15.8% of the total population in January 2021. Owing to this data it has become possible to infer depression from social media platforms.

Despondency is a common mental disorder affecting more than 264 milloin people worldwide[3]. This is characterized by persistent sadness and lack of interest and pleasure in previously enjoyed activities or events. World Health Organization approved that despondency is arguably one of the most detrimental diseases in the world, made so because sufferers of despondency are not adequately evaluated and proper intervention in course of treatment executed.

It has been widely concluded that despondency can permeate through the different layers of a person's existence, truncating the way they live their daily lives, activities and their relationship with people and the community as a whole[5] .Accordingly, within an estimated period of time, one sixth of the general populace encounter despondency in their lifetime. Research also proves the young people are more susceptible to despondency putting into consideration their immense involvement in social media and way of life[6].

Studies depict that one in three women experience despondency in their lives, either as an experience of sadness after having a child or experiences from a tragic event that can cause despondency symptoms[7], The intense schedule and distraction available to students makes despondency more prevalent among them which is a source of concern to world bodies on health[8], particularly in third world countries, the pressure is prevalent among students in higher learning institutions[9]. Several literatures have described the inherent complexity pertaining to identifying depression in social media platforms in which numerous researchers have highlighted key features that can be employed using machine learning techniques[10].

Learning is basically converting experience into knowledge or expertise. Immense volume of user data is generated on social media platforms, which can be used to extract valuable

tendencies and propensities. In this thesis we will be using machine learning algorithms ,particularly Analytic Hierarchical Process comprising of several machine learning algorithms, to identify keywords, hash tags that can indicate the early prognosis of depression. Machine learning is an aspect of Artificial intelligence that imitates the function of the human brain by processing data and creating patterns that will be used in decision making and predictions[11].

Support vector machines are a powerful kernel method used for classification and regression tasks. When trained optimally, it produces excellent separating differentials. The quality of this training as with other machine learning algorithms such as Logistic Regression,.Decision Tree,Naïve Bayes,Support Vector Machines and K-nearest neighbours depends not only on the given training data but also on additional learning parameters. It is proposed an automated approach be used in adjusting the learning parameters to make the optimization process more efficient. Evaluating a well-known dataset will show this approach can help produce quintessential processes that are very well tuned to their classification tasks. A large language library is essential to these applications as the system proposed to be developed should be able to identify persons that are susceptible to despondency

1.2 Statement of the Problem

Despondency affects the psychological state of mind of an individual, this is known to be greatly influenced by the constant use of the internet space, which influences the way of life of some people, although this has also been a platform for patients to discuss their experiences and symptoms online. The world health organization suggested that lack of early diagnosis is a major

cause of despondency and research needs to be done to reach persons who are vulnerable to despondency through their comments on social media.

This thesis intends to explore methods of machine learning techniques and algorithms to identify generated markers in identifying depressive statements, words and comments in social media. Extensive research will be done using SVM(support Vector Machines), Logistic Regression ,naves Bayes, Decision tree and K-nearest Neighbor algorithms to analyze and process comments and post, create patterns and make predictions if a statement is depressive, indicative depressive and non-depressive.

Due to the large volume of Data generated on social media, classical machine learning techniques and algorithms can help solve big data classification problems. Particularly, in the multidomain applications in a big data environment. The logistic regression model compares the odds of a prospective comment in those that are depressed and those that are not. The logistic regression function can also be used to calculate the probability that an individual belongs to one of the groups or the other. Naves bayes algorithm is a probabilistic classifier that can achieve high accuracy levels.

1.3 Justification of the Study

Depression symptoms are not easily diagnosed as people tend to hide their depressive symptoms and tendencies physically. Therefore social media has become an outlet for people to discuss and express their feelings, which gives an insight into their mental state. This contains data that can be analyzed and extracted using machine learning algorithms to depict individuals at risk of depression. This thesis exemplifies the innovation and contibution of artificial intelligence to the health sector by early detection of despondency and intervention in individuals.

1.4 Aim and Objectives

The aim of this project is to use machine learning algorithms and data analysis approach to extract depression comments from social media posts for identifying at-risk users.

The following are the specific objectives to be achieved in this project.

i) To extract and categorize depressive comments from social media post using Natural Language Processing (NLP) tools such as Ncapture and LIWC 2016

ii) To analyze the depressive comments using Logistic Regression, Naïve Bayes,Decision tree, Support Vector machine and K-nearest Neighbour, based on the categorization.

iii) To evaluate the performance of selected machine learning algorithms for accurate identification of at-risk user based on the analysis using the selected metrics and Analysis Hierarchical Process (AHP)

1.5 Research Question

This thesis is to select the best metrics and Analytical hierarchical process to determine the optimal classifier for early detection of at risk despondency persons using machine learning algorithms based on five performance metrics which was raised as a further work by Raza ul Mustapha et al at the international conference of information technology and new generation in April 2021,On the Research of "A Comparative Analysis of Machine Learning Algorithms For At Risk Despondency Symptoms Using Analytic Hierarchical Performance (AHP)"

1.6 Significance of the Study

Despondency creates a psychological imbalance in the mind of a person, who is an essential part of any community, and the effects are long lasting and can affect a person's ability to function and live a rewarding life. Depression symptoms are not easily diagnosed as people tend to hide their depressive tendencies and symptoms. Social media has become an outlet for people to discuss and express their feeling which gives an insight into their state of mind and how depressed individuals are .It contains vast amounts of data to analyze and extract depression symptoms in people at an early stage. This thesis exemplifies the innovation and contribution of artificial intelligence through machine learning, to help early detection and intervention in depressed individuals.

1.7 Scope of the Study

This thesis can be implemented in proper identification and intervention in at risk users of social media. It can also give proper perspective to the issue of mental health and the prevalence among youths and adults, and to know the high risk factors and signs of depression. This thesis can be implemented in various institutions to apply adequate intervention in those that are depressed.

1.8 Limitation of the Study

This thesis may be limited by the data set that will be used for the analysis and the keywords used to extract depression symptoms as this may not adequately cover all symptoms of depression.it may also be limited by ethical processes to retrieve comments from social media.

1.9 Operational Definition of Terms

- Supervised Learning- This is a process that uses structured input and output data. These data sets are trained to supervise algorithms into classifications and to predict the outcomes accurately .this can be divided into
- Classification – This process uses an algorithm to assign data to specific categories such as separating blue T- shirts from Pink. Examples of these algorithms include linear classifiers and support Vector Machines.
- Regression – This is a learning method that uses an algorithm to decipher the relationship between dependent and independent variables. This process is beneficial in predicting numerical values on different data structures examples are linear regression and logistic regression and polynomial regression.
- Unsupervised Learning – This is a method that utilizes an algorithm to label, analyze and cluster related data sets in which the algorithm will discover patterns in the data sets without any intervention from a human.
- Clustering – This is a data mining technique used to group unlabeled data based on some differences and similarities, this ia helpful in image compressions and market segmentation
- Decision Trees – This is a support tool that uses a tree like model to depict possible decisions and outcomes. It displays an algorithm that only contains conditional control statements, this is usually used indecision analysis and operations research to identify the strategy best applied to reach a goal.
- Deep Learning – Deep learning is a modeling approach which is made up of many processing layers to understand the representation of data after levels of abstraction

- AHP – Analytical hierarchical Process is a framework of quantifying a criteria and alternatives elements that can achieve the overall goal of the process initiated.
- Machine Learning Algorithms – these are mathematical model mapping and computation methods used to uncover underlying information embedded in vast amount of data by predicting output values.

Endnotes

[1] **Maxwell, E Aaron., A Timothy, Warner, & F Fang. *"Implementation of Machine-Learning Classification in Remote Sensing*: An applied review." International Journal of Remote Sensing, 2018:39,no 9 2784-2817**

[2] WHO. *Depression Fact Sheet*. 2020. Retrieved from: https://www.who.int/news-room/fact-sheets/detail/depression .

[3] Kumar, V. Uday, Alekhya Savithri, M. Jhansi Bhavani, A. Madhu Priya, K. Venkata Sai Bindu Jahnavi, & Namburu Divya Naga Lakshmi. *"Finding Psychological Instability Using Machine Learning*." **In 2020 7th International Conference on Smart Structures and Systems (ICSSS)**, IEEE, 2020, pp. 1-4.

[4] *Depression & Other Common Mental Disorders*: **Global Health Estimates World Health Organization, Geneva, Switzerland** 2017

[5] De Choudhury, M., Counts, S. & E Horvitz, May. *Social Media as a Measurement Tool of Depression in Populations*. **In Proceedings of the 5th Annual ACM Web Science Conference** 2013 pp. 47-56.

[6] Md Islam, ,A. Muhammad, A Kabir, A Ahmed, A Raihan M. Kamal, H Wang, & A Ulhaq. "*Depression Detection from Social Network Data Using Machine Learning Techniques*." **Health information science and systems** 2018: pp. 1-12, no 1.

[7] **T Ahalpara, Trishala, D Kalyani ,D Prathamesh , & N Parkar. "*Literature Review on Self-Harm and Suicidal Tendencies using Machine Learning.*A Journal of Physical Sciences,Engneering and Technology,12,no.SUP 1June 30 2020; pp. 157-165.**

[8] **A Akakandelwa, & Walubita, F Akindehin, & M Akindehin, . *Students' Social Media Use and the Perceived Impact on their Social Life:* A Case Study of the University of Zambia. The International Journal of Multi-Disciplinary Research .Online social networking practices of some Nigerian university undergraduates: Implications for counseling. British Journal of Arts and Social Sciences, 3(1), 2018, pp. 68-78.**

[9] T Tsai,H Chang ,Y Chang ,S Chang YS . *An Empirical Examination of Differences in Facebook Usage Behavior, Profile Contents and Privacy Settings.* **Personality disclosure on social network sites**: Computers in Human Behavior; 2017, pp. 76

[10] A Ahmed, O Husseini , B Prasadith , M Husseini Orabi, & D Inkpen. . *Deep Learning for Depression Detection of Twitter Users.* **In Proceedings of the Fifth Workshop on Computational Linguistics and Clinical Psychology: From Keyboard to Clinic,** 2018 pages 88– 97. Association for Computational Linguistics.

[11] N Jiang , H Yang , G Qingchuan, & J Huang **,** . *Development of Machine Learning and Several Learning Methods. Industry and Science Forum***. Machine learning and its algorithm and development analysis [J]. Information and Computer Science (Theoretical Edition)**, 15 (10) 2019: pp. 198-199.

[12] Jahan, Roshan, & Manish M Tripathi. *"Multimodal Depression Detection Using Machine Learning."* **In Artificial intelligence,Machine learning and Mental Health Pandemic pp. 53-72. Academic Press**

[13] H Yusuf , Wada, L Rajwani, E Anyam, E Karikari, M Njikizana, L Srour, M Garba & M. Khalid, *Mental Health in Nigeria: A Neglected issue in Public Health*, **Public Health in Practice,** Vol 2, 2021, 100166,ISSN 2666-5352,

Chapter Two
Literature Review

2.1 Conceptual Review

This Chapter will focus on the background study of most relevant research works related to this topic. Relevant works on identifying signs of depression from social media and tools used in prediction of depressed symptoms will be discussed. Then, we will provide insight on the research done in depression detection in line with the proposed approach using relevant research on identifying signs of depression from social media, and the use of a language processing system and algorithm techniques to bring out in depth meanings from depressive comments that depict despondency symptoms

Despondency has been acclaimed to be one of the biggest causes of people committing suicides, recorded to be 850,000 people every year[1]. . Regrettably, the majority of the instances involve teenagers aged 16 to 30. There is a problem among those experiencing from any mental disease in that they are unaware of who to consult or what form of treatment they need. The diagnostic criteria for major despondency disorder (MPD) are stated in the Diagnostic and Statistical Manual of Mental Disorders 5th edition (DSM-5): "If a person has several clinical signs for a contiguous period of two weeks and there is a marked difference from previous highly functional levels, then the person is suffering from a depressive disorder[2]

"Despondency influences a persons, thoughts,feelings and actions, and it can make it difficult to function and go about one's everyday activities." Despondency can be caused by a variety of

factors, some of which are unknown. The following are some of the more common forms of despair.

2.1.1 Major Depressive Disorder (MDD)

Major depressive disorder(MDD) is a mood disorder characterized by a number of features:

1. An extended period of depression

2. Inability to pay attention

3 Finding little enjoyment or enthusiasm in routine tasks and occurrences.

4 Having little desire to perform routine duties.

5 Feelings of apprehension and restlessness

6 extreme situations of sleep deprivation

7 Tiredness and restlessness 8 Depression and the desire to terminate one's life[3]

2.1.2 Persistent Depressive Disorder (PDD)

Dysthymia, or persistent depressive disorder, is a kind of chronic depression that manifests symptoms for longer periods of time. There are three levels of severity: mild, moderate, and severe. A person may enjoy brief periods of remission from depressive symptoms, but this relief only lasts a few months. Although the symptoms aren't as severe as those of major depressive disorder, they do tend to linger longer. Symptoms of PDD include:

1. Loss of enjoyment and interest

2. Irritability and rage

3. sadness

4. Lack of self-confidence

5. Insomnia

6. Fatigue and a lack of energy

7. Concentration issues[4]

2.1.3 Bipolar Disorder

Mania is a symptom of bipolar disorder, which is characterized by moments of unusually heightened mood. These phases can be moderate (hypomania) or severe (hypermania), causing significant impairment in a human being's life and requiring hospitalization. They can also impact a person's perception of reality. The majority of people with bipolar disorder are also depressed. Despondent persons often suffer a variety of medical and emotional symptoms, in addition to a sad mood and decreased interest in activities.

1) Fatigue, sleeplessness, and drowsiness

2) Spasms, discomfort, and conduct disorder that aren't described

3) Despair of hope and self-esteem

4) Anxiety and irritation

5) Inability to make decisions and disorganization[5]

2.1.4 Post-Traumatic Stress Disorder (PTSD)

This is a psychological disorder brought on by a terrible and life-changing event, either by personal experience or participation. Flashbacks, nightmares, and anticipatory mania, as well as panic disorder about the event, are all possible symptoms. Most persons who experience traumatic situations may have trouble adjusting and coping for a period of time, which can interfere with daily functioning. People with PTSD have powerful, unsettling thoughts and

sensations about the traumatic experience that continue long after it has occurred. They may have flashbacks or hallucinations about the occurrence, experience despair, dread, or fury, and feel disconnected or estranged from others. Persons with PTSD may reject situations or people who remind them of the terrible ordeal, and they may have intense unpleasant reactions to seemingly innocuous things like loud noises or unintentional touches[6].

2.1.5 Detecting Depression

Notable academics used crowdsourcing to find Twitter users who had been diagnosed with depression using traditional psychometric tests[7]. Others, as well. In addition, numerous strategies for predicting mental health status on social media were used. Other authors managed to identify vulnerable individuals to have depression before the start of their major depressive disorder[8], with an accuracy of 85 percent prediction rate using an anxious prediction model published at the conference of advanced Engineering science[9]. Contributors to the mentioned accuracy include reduced social activity, increased negative affect, clustered social network of the individual, and raised interpersonal engagement among individuals who have depression. Psychological assessment can also be used to detect anxiety, according to more recent research. Religious participation and radical undertones have both been identified as substantial indicators of serious depression. Determining the severity of depression might be an important element to consider when implementing therapies so that the appropriate authorities can prioritize people based on their severity of depression. Other researchers developed a regression model based on Facebook status updates to predict depression levels, with moderate accuracy surpassing the baseline. The algorithm was able to detect seasonal and temporal variations in depression rates. This concept

has contributed significantly to the usage of continuous data Other than depression itself, values rather than discrete values are used to determine the level of depression. Despite the fact that some researchers developed similar measurements to identify the social media depression index (SMDI)[10], they used a classification model to predict who would be depressed, it was approved to use nonclinical Facebook users who agreed to answer 100 personality questions and share access to their status updates. For one year, W.i.r.e.d science used 223 volunteer's Facebook posts to extract attributes from messages using AI algorithms, sentiment analysis utilizing sentiment lexicons, and the quantity of words. When compared to the baseline results generated by using simply sentiment features[11], the model produced better outcomes The authors concluded that depression can be successfully identified through the use of language in social media, and clear patterns of change in the degree of despondency and can be identified over the seasons where depressive linguistic style is more prominent in one time over the other, based on the findings and aligning with the literature in despondency and its correlation with change of seasons.

2.1.6 Social Media

These are interactive technologies that enable virtual communities and networks to create, share, and disseminate information, concepts, professional experiences, and other forms of expression[12]. While the wide variety social media services currently accessible poses issues to the context of social media, there are several common aspects that apply to all: Users typically use web-based programs on their desktops and laptops to access social media services, or they download services that provide social media features to their mobile devices (e.g., smartphones and tablets Individuals, communities, and organizations can share, discuss, participate in, and alter user-generated or self-curated content posted online when users interact with various electronic

services. Furthermore, social media is used to chronicle memories, learn and discover new things, advertise, and create friendships, as well as the development of new ideas through the creation of blogs, podcasts, films, and gaming sites . Facebook TikTok, WeChat, Instagram, Twitter, and LinkedIn are some of the most popular social networking platforms, with over 100 million registered members. YouTube, Quora, Telegram, WhatsApp, Snapchat, Pinterest, Microsoft Teams, Google Meet, and others are examples of prominent platforms that are commonly referred to as socialmediaservices. Social media use has a wide spectrum of beneficial and bad consequences, according to observers. For firms, startups, non-profit organizations, advocacy organizations, political parties, and governments, social media can boost an individual's sense of connectedness with real or virtual communities and can be an effective communication and marketing tool.

Social media, according to the Pew Research Center, began as a way to connect with friends and family, but it was quickly adopted by businesses looking to reach out to customers through a popular new communication tool. The way to interact and exchange information with everyone on the planet or with a large number of people at once is the power of social media. Media users are typically younger. Between the ages of 13 and 50, about 90% of people utilize at least one sort of technology. Observers have also seen a rise in social movements' use of social media to communicate and organize during periods of political unrest. People can exchange their thoughts, opinions, and facts through social media by forming virtual communities and networks. By its very nature, social media is a network, allowing users to exchange and receive content electronically. The content includes personal data, documents, movies, and images. Users can interact with social media through web-based software or applications on a computer, tablet, or smartphone social media[13]

Furthermore, these consumers are more educated and at ease. Because the usage of social media has made everyone a potential author, everyone may now express their ideas and beliefs, and the language is now closer to the user than to any predefined conventions. Blogs, tweets, and status updates are all written in the same way . in a casual, conversational tone—often more of a torrent of consciousness than what one might anticipate from traditional print media, which is painstakingly worked out and meticulously edited. The informal nature of social media poses significant obstacles to autonomous language processing at all levels. On the surface, simple NLP systems designed for traditional data encounter a number of obstacles. Punctuation and capitalization that is inconsistent (or absent) can make detecting sentence boundaries difficult - even for human readers. Social media is also significantly noisier than traditional print media. Ads and other unwanted, irrelevant, or distracting items abound on social networks. Even if these types of noise are ignored, most of the true, authentic content on social media might be considered either relevant or irrelevant. This is illustrated in a study of the perceived worth of tweets by users. They gathered almost 40,000 ratings from followers, with only 36% of tweets regarded as "worth reading" and 25% rated as "not worth reading." The so-called presence maintenance tweets were the least valued. In any language-processing effort targeting social media, applying pre-processing as a procedure to filter away spam and other irrelevant content, or models that are more capable of coping with noise. NLP techniques are particularly concerned about many features of social media text. The characteristics of a specific medium and how it is used can have a significant impact on what defines an effective summarizing strategy. Individual tweets on Twitter, for example, are very context-poor when in comparison to more traditional texts or platforms like Facebook because of the 140-character restriction. However, due to the practice of retweeting posts, duplication might become an issue across several tweets.

Social media can take many forms. Photo sharing, blogging, social gaming, social networks, video sharing, business networks, virtual worlds, reviews, and many other activities are examples of these activities. Governments and politicians use social media to communicate with their people and voters. Individuals utilize social media to communicate with friends and family members. Various social media platforms will be used by some people. professional possibilities, connect others around the world with similar interests, and share their thoughts, experiences, insights, and emotions Participants in these activities are members of a virtual social network. Social networking is a must-have tool for businesses. Companies utilize the platform to locate and communicate with customers, increase sales through advertising and promotion, assess consumer trends, and provide customer service and support.

Despite having similar audiences to Twitter and Instagram, Facebook is the world's most popular social networking platform. The following are the most popular social media websites as of January 2021: Facebook is one of the most popular social media platforms (2.74 billion users) YouTube is an excellent source of information (2.29 billion users),WhatsApp is a messaging service that allows people to communicate (2 billion users),Messenger for Facebook (1.3 billion users),Instagram is a social media platform where people may share (1.22 billion users),Whatsapp is a messaging application that allows you to communicate with others (1.21 billion users),TikTok is a social media platform that allows people to share their experiences (689 million users)[14].

The way we engage with one another online has altered due to social media. It has enabled us to learn about what's going on in the world in real time, to interact with one another and keep in contact with long-distance relationships, and to have access to an infinite quantity of information at our fingertips. Many people have found common interests with others online thanks to social

media, making the world seem more relatable. According to a Pew Research Center survey, using social media is linked to having more friends and a more diverse personal network, particularly in emerging nations. Friendships can start digitally for many youngsters, with two - thirds of them meeting a friend online

2.1.7 Machine Learning

Machine Learning (ML) is an artificial intelligence subfield. Arthur Samuel best summarizes it as "the branch of study that allows computers to learn without being explicitly programmed."[15] This suggests that computers could well be trained to learn by observing others. Machine Learning aims to make decisions based on unknown facts without being explicitly told how to do so supervised learning, unsupervised learning, and semi-supervised learning are the three kinds of machine learning tasks. Supervised machine learning is a sort of machine learning in which a model is "trained" on a set of labeled "training instances" before being applied to new data to make predictions. Classification problems are another term for supervised learning problems. When a model is developed on unlabeled data by detecting patterns and relationships in the data, it is known as unsupervised machine learning.

Clustering difficulties are another term for unsupervised learning problems. Semi-supervised machine learning is a type of machine learning that uses labeled examples to train the model and then uses unlabeled data to refine the class boundaries. Credit card fraud detection, character recognition, speech understanding, face characterisation, product suggestion, the health sector, consumer segmentation, form identification, and sign language interpretation are just a few examples of real-world Machine Learning applications.

Machine Learning operates by looking for patterns in data (associated or not with given classes). Data is first turned into a representation (a set of features) that a computer can understand in all of the following applications. In order to use Machine Learning methods, text must be translated into a quantitative (or discrete) representation in NLP (natural language processing). Similarly, Computer Vision works with images to turn them into a format that a computer can understand. Data collection and preprocessing, feature engineering, model training, and system testing are all part of a typical machine learning application.

2.1.8 Engineering Features from Data Set

. Feature engineering is a term used in Machine Learning to describe the process of applying domain knowledge of data to produce features that machine learning algorithms may utilize to detect patterns. A feature is a piece of data that can be used to make predictions. When creating a Machine Learning task, attributes, variables, and other phrases are utilized as features. A dataset including dependent and independent variables is provided to a Machine Learning algorithm. The dependent variable is usually a nominal value, which is the predicted outcome. When a variable has a fixed number of categories, it is called nominal. Independent variables, on the other hand, can be numerical in nature such as an integer ,real or nominal . A few Machine Learning algorithms, such as WEKA's "Nave Bayes Multinomial Text" classifier, may accept variables of type "string." In the context of text data, features might be included in the dataset or extracted from it.type

2.1.9 Preference of Classifiers

There are several different Machine Learning algorithms in use today. Support Vector Machine (SVM), Naive Bayes (NB), Random Forest (or other tree-based algorithms), Decision trees, and ensemble approaches are some of the most extensively used algorithms in Natural Language Processing jobs. There is no idea of a universal algorithm as at the time of this research. Researchers typically experiment with various algorithms and optimize them to represent the topic of interest. Finding suitable features is an important part of any Machine Learning process. The algorithms that classify these features then look for patterns. These properties are frequently derived from text in NLP tasks, Examples of Bags of Words features are those that rely on word frequency or arrangement of such words. Other functions, such as sentiment analysis, document value, tone, and readability level, are more problem specific. The goal of generating features is to extract information from text and convert it into a format that a machine learning algorithm can understand.

2.1.10 The Processing of Natural Language

The study of human language interaction with a computer is known as natural language processing[16]. It enables a computer to comprehend natural language. Artificial Intelligence encompasses it. Bags of Words characteristics that rely on word frequency or arrangement are examples of language processing strategies that allow computers to evaluate, fully comprehend words. Other features are more problem-specific, such as sentiment analysis, document value, tone, and readability level. The purpose of generating features is to extract information from text and convert it into a format that can be understood by a machine learning algorithm. Human languages to derive meaning In this field, significant progress has been accomplished, currently, developers may use NLP to organize and structure knowledge in order to execute tasks like

summarization, translation, entity recognition, relationship extraction, sentiment analysis, and segmentation of a process with a fair level of accuracy.

Understanding human language is challenging because it involves more than simply a knowledge of words[17]. It necessitates an awareness of how words are connected to form meaning. A literal or word-for-word translation isn't always accurate. Many words have many meanings, which makes understanding language even more difficult.

NLP applications are now frequently employed, even among those who are unaware of their existence. Machine translation, spam filters, and autocomplete suggestions are among the most often utilized applications (on phones and search engines),. Text summarization apps and sentiment analysis apps (which allow us to learn how people feel about a product, service, or person) are examples of alternative applications. Text classification techniques are used to categorize SMS/emails as spam or not, and books are divided into genres, to name a few of the many NLP uses. Other NLP presentations include:

• **Sentiment Analysis**: This procedure entails extracting sentiment from text, such as positive, negative, neutrality, or even sentiments like happiness, sadness, anger, disgust, and fear.

• **Text Summarization**: This comprises summarizing a single article or a collection of articles based on a common theme.

• **Textual Entailment**: Using directional relationships between text fragments to link numerous word phrases/words with the same meaning.

• **Extraction**: Extracting structured information from unstructured data, such as recognizing entities and their relationships, and decoding co-references.

• **Topic Segmentation**: Separating subjects from text because they may overlap.• Question Answering: Answering closed (specific) and open (subjective) questions. These are the basis for virtual assistants, like iPhone's Siri.

• **Speech-Parts Tagging**: Classifying words based on their parts of speech, such as nouns, verbs, adjectives, and so on.

• **Translation is the process of translating one language into another**.

The goal of developing NLP approaches is to extract valuable information from huge amounts of data available online and offline without the need for human participation. Interventions by humans are costly and time-consuming. Artificial intelligence systems can execute jobs far faster than humans, but they have yet to achieve the same level of precision. In NLP, there are two processes.

1. System Based on Rule Engineering; These applications are guided by a set of rules. Building a machine translation system, for example, requires a technically comprehensive set of rules that define allowed words, their parts of speech, and sequences of parts of speech. Furthermore, these restrictions are based on a well-defined vocabulary, and any new words may not be included in that lexicon, which would be difficult to do with the ever-changing language of social media.

2. Statistical Process in Natural Language; This comprises quantitative and probabilistic techniques to dealing with language, such as implicitly modeling language by counting words or short sequences of words, or employing massive collections of aligned parallel text to achieve machine translation rather than requiring explicit rules. We apply statistical NLP approaches in this study, specifically NLP methods based on machine learning algorithms. By evaluating a set

of instances and drawing statistical conclusions, NLP can use machine learning techniques to learn these principles.

2.1.11 Natural Language Interpretation in Social Media

Social media, which includes discussion forums, blogs, Twitter (microblogging), Facebook, and Instagram (photosharing), has recently grown in popularity. These platforms enable the production of large amounts of text in an informal setting. Abbreviations, acronyms, punctuation, and emoticons are just a few of the message trends that have emerged on social media. Existing words have acquired new meanings, such as "tagged" and "poster." Social media has its own lexicon. In order to extract relevant information, NLP applications that are deemed rule-based systems would need to incorporate new rules to take into consideration the new terms and abbreviations being developed every day, Machine Learning programs are better at adapting to changing languages. To interpret text, NLP analyzes language patterns. Tracking the sentiment in the tone of a written message (tweet, Facebook update, etc.) and tagging that text as good, negative, or neutral is one of the most appealing ways NLP provides significant intelligence. Predicting user opinions is a common activity, such as using social media to forecast the outcome of electoral campaigns or creating user profiles to anticipate their likes and dislikes. Data from facebook (a social media platform) is also used in this thesis.

2.1.12 Support Vector Machines

A classification algorithm is Support Vector Machines (SVM). Support Vector Machine is a traditional machine learning technique that can still aid in the classification of large amounts of data[18]. SVM's training technique produces a model that determines an ideal hyperplane

separating examples from the two classes using a dataset training examples for a binary class problem. Based on which side of the higher dimensional space the sample falls on, SVM provides a prediction. SVM stands for supervised machine learning. An SVM training method produces a model that predicts the category of a new example given a set of training examples, each labelled as belonging to one of the many categories. SVM is better at generalizing problems, which is the purpose of statistical learning. This theory lays out a framework for investigating the problem of gaining knowledge, generating predictions, and making judgments based on a set of data.

The support vector machine is mostly used for pattern classification, which implies it is used to categorize various types of patterns. There are two types of patterns now: linear and non-linear. Non-linear patterns are patterns that are not clearly discernible or cannot be quickly separated in low dimensions, and so must be further altered to be easily separated.

Basically, the main idea behind SVM is the construction of an optimal hyper plane, which can be used for classification For linearly separable patterns. The optimal hyper plane is a hyper plane selected from the set of hyper planes for classifying patterns that maximizes the margin of the hyper plane that is, the distance from the hyper plane to the nearest point of each patterns

.2.1.13 K-Nearest Neighbors (KNN).

This is a non-parametric method for classification and regression is K-nearest neighbors. It is one of the most widely used machine learning techniques because it is simple to understand and implement[19]. The underlying concept of this algorithm is to look for similarities in the neighborhood, assume a test data point that is comparable to them, and derive an output. With this process, we search for k neighbors and make a forecast. In this classification, the majority of

the data is applied to the k closest data points, whereas in KNN regression, the mean of the k closest data points is determined as the output, and odd numbers are chosen as k. KNN is a model that only processes data while it is running.

No prior knowledge of this algorithm is required. During the test, the k closest neighbors will participate in classification or regression. This is accomplished by employing a few hyper parameters, including the value of k and the distance function, but it necessitates a significant computation if the sample size is big, and it should be broadly selected to decide optimal feature scaling. In the the algorithm, the K value defines how many neighbors will participate. The validation error will be used to adjust the value of k. The distance function is then determined using Euclidean distance, which is the most commonly utilized in similarity functions, as well as Manhattan and Hamming distances, which are different alternatives to the Distance function.

2.1.14 Decision Tree

A decision tree is a tree-based approach for solving regression and classification issues. For obtaining the result, an inverted tree is framed from a homogeneous probability dispersed from the root node to extremely heterogeneous leaf nodes. For continuous dependent variables, regression trees are utilized, whereas for discrete dependent variables, classification trees are employed. Each node of the decision tree has a condition over a feature, which is derived from the independent variables. Based on the condition, the node chooses which node to examine next. Once the leaf node is reached, an output is predicted. The right sequence of conditions makes the tree efficient and the Information gained is used as the criteria to select the conditions in nodes. A recursive, greedy based algorithm is used to derive the tree structure[20].

- For classification and regression trees, gini index is used as the classification metric. It is a metric to calculate how well the data points are mixed together. . An output is expected after the leaf node has been reached. The tree is efficient when the conditions are in the appropriate order, and the information gathered is utilized as a criterion for selecting conditions in nodes. The tree structure is computed using a recursive, greedy-based algorithm.
- The Gini index is used as a classification metric for classification and regression trees. It's a statistic for determining how well data points are combined. At every stage of the decision tree creation, the attribute with the highest Gini index is chosen as the next criterion. The gini score is the highest when a set is unequally mixed.

At each stage of the decision tree creation, the attribute with the highest gini index is chosen as the next criterion. The gini score will be maximum when a set is unequally mixed.

Entropy and information gain are utilized to determine the next attribute in the Iterative algorithm (dichotomiser).

This eliminates the need for data preprocessing or assumptions about data distribution because the decision tree can adequately explain the prediction result. As a result, when training complex data sets, the tree might become complex, resulting in the loss of useful information when continuous variables are employed. Max depth, sample leaf, sample split, and criterion are some of the hyper parameters employed.

2.1.15 Logistic Regression

Logistic regression analysis is used to assess the relationship between many categorical or continuous predictor variables and a binary outcome. A high-quality logistic regression model

possesses predictive power, interpretability, significance, and resilience to data mistakes. Logistic regression calculates the linear output before applying a stashing function to the regression output, similar to linear regression. A frequent logistic function is the sigmoid function

Although a machine is incapable of comprehending text, it is capable of replicating it[21]. To complete this duty, it needs have a basic understanding of natural language laws. Phonology, morphology, syntax, semantics, and pragmatics, as well as ambiguity and tautology, are all aspects of language that NLP is concerned with. Working with text data is most commonly done using Bag of Words, part-of-speech tags, and syntactic dependency relations.

2.1.16 Word Counting Model

The frequency of words is used as a feature for training a classifier in this model, which represents text (a phrase or a document) as a bundle of its words, bypassing grammar and word order but maintaining multiplicity[22].

2.1.17 Components of Speech

The technique of associating each word in a sentence with the part of speech that it represents in the phrase is known as component of speech tagging. This is accomplished by rule-based systems, however classification methods have been used to solve the challenge throughout time.

.

2.1.19 Relationships of Syntactic Dependency

Syntactic dependence descriptors display syntactic syllables. Component of speech tagging is the process of matching each word in a sentence to the part of speech it represents in the phrase.

Although rule-based systems have been utilized to handle the problem in the past, categorization methods have also been used. A phrase is constructed using binary token relationships. A verb's dependents, for example, are related. These connections form a graph that looks like a tree.

2.1.19 Subject Modelling

Subject modeling is a statistical modeling technique for determining the subjects that appear in a group of documents. Given that a paper is about a specific topic, one would expect certain words to occur more or less frequently. Topic modeling techniques frequently create groups of comparable phrases as "themes." A topic model encapsulates this intuition in a mathematical framework, allowing for the examination and finding of a series of documents based on the statistics of the words in each document.

2.1.20 Naves Bayes Algorithm

Naïve Bayes is one of the data classification algorithms. This algorithm is one of the most accurate in data mining. It applies simple probabilistic classification by calculating a set of probabilities to retrieve the frequency and combination of values in a data set. The probability of certain features appearing as members in a probability sequence is obtained by calculating the frequency of each feature value in its class from the training data set. The training data set is a part of the data set used to train classification algorithms. The training process uses known values to predict unknown values.

The Naïve Bayes Classifier algorithm has been proven effective in many practical applications, including text classification[23].

2.1.21 K-fold Validation Processes

K-fold cross-validation is an effective strategy used by data scientists. It involves partitioning data to effectively use the available dataset to create a more generalized model. The main focus of doing this form of machine learning is to build a more generalized model which can perform well on projected data .A model can be built with a high level of accuracy but isn't flexible with new data. It should adequately fit the training data. Machine Learning is all about generalization which means that model's performance can fit to a criteria of data points that have never been used during the training process. This is why the data set is divided into a training set and a test set[24]

.

2.1.22 Training and Validation Data Sets

The training dataset is the set of data used to train a model and it also proves to be the largest set of data. This is the set of data that the model will uses to learn the behavior of the data set. The model will then be trained continuously to understand the behavior and patterns in it.

2.2 Review of Empirical Studies

Majority of the groundwork on identifying depression from social media attempts at making the process of diagnosing individuals instinctive. These works do not intend to replace psychologists but they attempt to provide a beneficial tool to ease a clinical burden clinical burden, these are primarily based on statistical methods by collating relevant data by quantifying activity on social media platforms, standard psychometric measures, positive or negative emotions, social segregation and social media platforms specific statistics.This may be based on the number of followers or friends on twitter, the network of users, and or user interests or mutual connections.

Noteworthy works that fall into this category are statistical models that were used to characterize differences between mothers with Postpartum Depression (PD) and mothers without PD[25]. Postpartum Depression is a mood disorder that can affect mothers after childbirth. Facebook data was collected from 165 volunteers during prenatal and postnatal periods to predict likelihood of PD. The ground veracity was formulated based on volunteer reports and Health check Questionnaire in a standard psychometric measure. The authors based their statistics on their predictions on belonging to four categories:

(i) Individual characteristics; number of social media status updates, income, age, entropy of activity.

(ii) Principal interests; number of likes on status updates, comments on Uploaded media, wall posts by friends etc.

(iii) Content characteristics; positive and negative emotion affect measures by applying the Linguistic Inquiry and Word Count tool, and the number of posts which are Questions.

(iv) The use of articles, conjunctions, verbs, adverbs and personal Pronouns etc.

However, the authors do not report the details of the most effective statistical mode used to give the highest accuracy and performance, needless to say that the model was based on data from the entire prenatal period plus one month from the postnatal period and the model successfully explains 48% of the variance in the data. The predictive variables were variables that quantified user interaction on social media from the user characteristics category and principal interest.

Self-harm leading to death is of national concern all around the world. This has also been compounded by the lack of early detection in persons as well as the large volume of data to be processed in time ,this makes this task complex[25] .Research spanning over 40 years has

unearthed an analysis of previous suicide research that shows that stereotyped methods are not adequate in detecting despondency. Worthy of mention is that these methods of prediction are only slightly better than chance, although they show promise in the area of using natural language processing to extract words that show risk to the individual in question[26].

These limitations include demographic location as to where the individual lives as well as their medical history and psychological predisposition, and this can be identified on social media, such as location applications and flagging of incorrect information and action on social media platforms. The datasets are usually substitute judgments of the resource specialists and not the actual demographic information available, these processes do not give accurate demonstration of suicide risks in social media; this also does not justify the conclusions on these data according to some researchers. It is believed that the research process using a text extraction method which is based on pre-defined meaning of such words which do not reflect the true state of an individual's mind as comments depend on context.

Language has become informal with the use of emojis and abbreviation of words, these are not the usual lexicons of natural language and cannot be processed using traditional methods. Machine learning techniques have the capacity to encapsulate the different language dimensions in social media as psychological health condition are compounded and variable in nature as well[27].This gives a broader perspective to this health condition as existing research do not put these innovative processes into consideration[28].Adequate information based on research that natural language process through machine learning algorithms can enhance predictions. The techniques have been used in other health research to predict either improvements in health conditions or depreciation based on data used. In review of over 250 health related studies[29],it

can therefore be concluded that innovative tools can be used to predict and detect despondency conditions[30].

Furthermore, these predictions are limited to only medical sources such as the data base of a hospital as discerned in the research done at Berkshire health system hospitals which was based on only counseling session, MRI ,and clinical results .This results did not depict the natural behavior of an individual ,as recent studies have shown that data is more accurate in daily exchanges of an individual with the society[31].

The internet offers an an opportunity to use vast amounts data generated through user content, this data reflects several communication trends that show psychological imbalance in some individuals, few studies also show symptoms of despondency, but not accurately in the area of suicide. Some individuals showed consent in allowing researchers use their facebook comments to predict despondency as these information is linked to their medical records. Facebook posts of consenting individuals have also been used to forecast despondency as recorded in electronic medical records of several health facilities.

The access to the Facebook statuses posted by 683 patients visiting a metropolis academic emergency department was ascertained,, 114 of whom had a diagnosis of depression in their medical records, by applying the language preceding their first documentation of a diagnosis of depression, it was identified that depressed patients with fair accuracy approximately matched the accuracy of the screening surveys which were benchmarked against medical records(Accuracy of 0.69).a higher accuracy was documented a few weeks after an initial diagnosis a set of data which aided the prediction ,Despondency symptoms can be reflected emotionally or cognitive manner and prediction can be achieved using data of about 12weeks which are significant in processing, access of depression assessment through social media of

consenting individuals is sufficiently a feasible scalable complement to existing screening and monitoring procedures. Each year, approximately 17% of the world population experiences depression, of whom only 31% receive minimally adequate treatment. Despondency may very well become a cause for dysfunction in the society in a few years especially in thriving communities, these rates are alarming and are caused by insufficient diagnosis and treatment which strongly suggest that innovative processes should be utilized to increase prediction rate[32]

In other research to exemplify the diagnosis, Natural language in social media posts were used by a department to build a model to predict the first documentation of a diagnosis of depression in the electronic medical record (EMR). This has demonstrated the practicability of using Twitter and Facebook post data activity to predict depression, postpartum depression, suicidal tendencies and post-traumatic stress disorder. The use of social media data to identify mental disorders (despondency) is the most important aspect to look into when considering detection of mental disorders through social web mining, this makes it possible to use social media content to identify mental illness in a given population. A major contributor for this task is attributed to De Choudhury who has investigated to a large extent that social media could be used as a source of information to understand mental illness among individuals as well as within a population. It is well documented that among the 69 percent of American adults who are online, 66 percent use Facebook, followed by 20 percent who use LinkedIn, and 16 percent who use Twitter. This creates a unique flow of sensors to allow the identification of psychological and social behavioral patterns in data[33], giving a thorough representation of a society of social media users inside multiple social media platforms. The researcher further said that the language used and the emotions expressed in the postings could suggest depression. In other words, postings that convey feelings of worthlessness, remorse, powerlessness, and self-hatred can be regarded

depression indications. Second and third person pronouns was also an indicator of depression awareness. Absence from social activities and changes in social network links, in contrast to linguistic traits, are also indicators of depression.

Multimedia content is also significant in the context of sharing more images, for example, could be an indicator of someone who is inversely connected with depression. Aside from the amount of images, the content of those photos, such as facial expressions, has been discovered as a significant indicator[34]. A link between Instagram filters and depressed users was discovered[35], which is similar to this research effort. One of the most commonly used filters among depressed participants was one that converts color photographs to black and white. .

Researchers have emphasized that it is advantageous to identify mental illness such as depressive disorder through social media posts and to develop tools that can be used by health care sector and the user to take proactive measures to obtain intervention treatments to avoid deterioration[36]. Needless to say it is imperative to raise awareness on the conscientiousness of accessing social media content to identify indicators of mental health disorders with the intent to alert health care authorities or relevant parties such as family members on adequate intervention and treatment. Other notable research has been done to distinguish suicide related risk factors from Twitter conversations

Mental illness is a major cause of global disability[37]. It has an impact on a person's cognitive and language mode, mood, daily activities, and work ability. Medical resources are used in large quantities around the world to deal with despondency. Anxiety, bipolar disorder, depression, borderline personality disorder, obsessive-compulsive disorder (OCD), post-traumatic stress disorder (PTSD), schizophrenia, and other disorders are among them. Depressed mood,

exhaustion, sorrow, sleep problems, lack of focus, no interest in activities, feelings of guilt, diminished energy, reduced motivation to work, and suicidal tendencies are all indicators of major depressive disorder (MDD)[38].In a related analysis of despondency , Symptom diagnosis is an important first step in the treatment of mental disease. Because depression recovery rate has a high correlation with the depressive period[39], prompt identification and prevention may be the preferred approach for Major depressive disorder. Despair is a complex phenomenon that is influenced by a variety of factors, some of which are not always readily available. This knowledge is difficult to obtain directly and explicitly from the topic. As a result, much work needs to be done to identify and diagnose depression in the general population using various evaluation tools such as asking patients to fill out depression questionnaires.[40]. Patients are assessed for despondency using this method, and their scores are compared to a specified chart. The Beck Depression Inventory (BDI)[41], the Center for Epidemiological Studies Depression (CES-D)[42], the Patient Health Questionnaire 9-item depression module (PHQ-9)[43], and the Hamilton Rating Scale for Depression (HRSD)[44] are some of the questionnaires used to assess depression.

Although these series of questions assessment tools are a good first step in identifying and diagnosing despondency in its early stages, their fundamental disadvantage is bias, as well as people with depression's low interest in filling out questionnaires. This makes it difficult for care providers to gain a complete understanding of an individual with depression using only questionnaire-based self-reported data, which is grossly inadequate.

However, there are no distinct ways for assessing MDD in medical science. Comprehensive study has been performed to better understand how persons with despondency manage their social connections, as well as how to recognize depressive symptoms early on.

Recently, social networking sites (SNSs) have been utilized to identify MDD as a social sensor. SNSs offer a flexible platform for sharing information efficiently. Individuals can readily communicate their thoughts and opinions on social media platforms, allowing academics to look into a variety of psychological issues and human behaviors. Twitter and Facebook are two of these sites Individual everyday life activities, user behavior, emotions, and feelings can all be detected utilizing these platforms. Recent researches have looked at textual posts, social media usage, and interactions on SNSs to see if they may detect people who are depressed. SNSs are utilized as a social sensor to infer behaviors and illnesses in individuals using behavioral markers, which are implicitly biased by an individual's decision on what to report. Social networking sites accurately and in real time record everyday actions and events, making them less vulnerable to memory bias. Individuals' thinking, activities, communication, sociability, and emotion can all be captured through social media. . User profiles and postings, for example, may reveal sentiments of self-hatred, worthlessness, guilt, and helplessness, all of which are common indicators of serious depression. In comparison to people who are not depressed, depressed people utilize more negative words/terms as well as the first pronouns in frequent comments. In retrospect, the link between social media communication and sadness is a bit vague. Predicting depression signals only based on language features of posts is a big accomplishment. However, other factors such as communication frequency, inflow and outflow, and textual comments on social media can accurately predict mental illness To this purpose, it is critical to comprehend the role of social media in combating and dealing with depression.

It is also critical to comprehend the significance of interactions on SNSs, as well as how to evaluate succinct information and features on social media, notably Facebook. To describe the consequences of depressive episodes on an individual's behavior, advanced research is required,

which should comprise a comprehensive collection of features such as user information, statistics, groups, status updates, activities, and the like. These attributes are used to extract relevant data such as user temperament, depression references, polysubstance abuse references, support and advice references, self-discourse, racial disparities, and radicalization for the purpose of detecting radicalization of individuals having despondency. More so social media has such a significant and profound impact on a person's well-being, these characteristics have both a good and negative relationship with depression.

The majority of studies have found a correlation between the two features and usage. For example, frequency of status updates, whether favorable or unfavorable, has been linked to despondency. Rumination can be exacerbated by negative behavioral facebook posts or negative interactions with others. Less depressed people, on the other hand, were found to post optimistic status updates frequently. Furthermore, research shows that persons suffering from depression post more negative than good messages in, additional studies demonstrate that for more depressed people, Facebook activities grow more prevalent and crucial. These findings show that the behavior might be due to loneliness in online activities. However, researchers have showed that depressed people prefer to use social media to exacerbate their feelings, get social support and mitigate uncertainty[45]. However, more research into social networking sites is required to clarify these discrepancies:

A study suggested a conceptual framework that uses Facebook users' data to assess MDD, and it used user demographics, behaviors, and information provided by the SMPP application to improve depression assessment methodology[46]. For major depression identification analysis, the approach includes both observable and unobserved variables. This reflects the values of the observed variable, such as the number of status updates, likes, friend network size, and others.

Because various user attributes can determine expressed depressive symptoms, this information could help anticipate depression. Because of the widespread use of social media to diagnose depression, the study presented a framework for investigating SNS as a screening method for despondency in individuals. To create a successful screening method, we must first determine Facebook's predicted link with depressive traits. The SMPP tool uses Facebook features to detect depression-related signals (depression, help-seeking behavior, and SUD). A study of the relationship between Facebook attributes and CES-D consideration of the various crucial traits that can effectively separate persons either with or without depression were identified as a result of the investigation.

Research also reported that, while depression diagnosis utilizing social network data has taken a firm grasp around the world, there are still some dimensions that have yet to thoroughly explored[47]. The goal of the study was to use machine learning approaches to conduct despondency analysis on Facebook data acquired from an online public source in order to assess the effect of depression detection. Needless to say, part of the above-mentioned research has looked at emotional processes, temporal processes, and linguistic style in order to detect depression .As a result, there are few distinct investigations in a research publication that have utilized machine learning approach SVM, KNN, Decision Tree, and Ensemble separately[48]. There were also well-known research that merged all of these strategies on the same set of data to look into the differences in technique-based results.

RapidMiner was used to develop this Depression Prediction Model[49]. This model combines a variety of procedures to evaluate both SVM and Nave Bayes classifiers. Two sets of data and seven key operators make up the model. The first dataset would be the training dataset, which contains depressed and non-depressed postings that were manually trained. The second dataset

contains patient SNS postings, which were altered for each participant to test the model's prediction. The first operator chose characteristics and determinants that contributed to the training dataset's performance; some were maintained, while others were deleted. The second and third operators are the Nominal to Text operators, which change the type of chosen categorical attributes to text and map all attribute values to matching string values. For both training and validation datasets made use of this. The major goal of the study was to investigate if there was a significant relationship between social media usage and mental illness. SNS activity were thought to be able to detect mental illness in its early stages.

A psychiatrist cannot obtain complete information from a depressed patient using typical questioning approaches, but the SNS-based system can address this obstacle by introducing a self-reporting process. We acquire a closer observation and viewpoint of a depressed individual's natural behavior and way of thinking based on the user's social activities, and better identify the mental behavioral levels. The UGC from the patient's Twitter and/or Facebook posts was collected using a web application, which was then examined using numerous text analysis APIs. Following that, the patient was assigned to one of four levels of depression: minor, mild, moderate, or severe. Using RapidMiner to evaluate classifiers, the SVM and the Nave Bayes Classifier, a model for predicting depression was created. People with untreated depression or people who are unaware of their mental illness for a variety of reasons can be taken over on a machine learning architecture with the use of social media as a single platform that contextualizes everything on a publicly exposed platform, thanks to the advent of new medical advanced methods and technologies.

To buttress further, if an advertisement to buy a product shown on a Facebook ads, the meta tags and other constraints will recommend the next advertisement to be similar from the previous one,

this is a recommender system, because a pattern from the user, gets observed and analyzed by the machine learning algorithms working on a the background system, also for an online-shopping website, if the user mostly browse in the account at late hours and buys more products at night, then the advertisements will also learn a pattern of user time engagement to show more of the new items similar to the previous purchased products, but only during the night time because of various factors in addition to previously purchases. For example, vast numbers of people are linked to the internet every second, resulting in endless discussions, data transfers, and downloads[50]. This information is used in a variety of data analyses aimed at providing relevant information for a variety of industries, including telecommunications, commerce, medical, government intelligence, and other international and technical sectors.[51] Additionally, there are a variety of internet sites where people can upload, research, and explore new features and methods that influence their behavior. Each of these platforms can be used to perform various aspects of user analytics in a range of technological fields. Product category has as a result, a technique can be built that uses a combination of data analytics to forecast various insights among social media users. .

This technique uses a configuration that breaks down the structure of a user's posts, comments, tweets, location, and trip traces, as well as other publicly available data, which can be found largely on social media sites like Instagram, Snapchat, Facebook, and Twitter. Unfavorable future events can be avoided, and positive consequences can be identified. For example, a data collection by a user on social media with locations, sad emoticons, and depressive remarks, as well as other connected information, can be evaluated by an educational institution to obtain various statistics that can assist in providing appropriate support and intervention.. Similarly, a pattern of user activity can be used to forecast various potential outcomes including positive,

negative, and neutral interest in any hot issue by constructing a sentiment analysis model. This study focuses on the development of an architecture model and procedure for forecasting depression, a serious health concern among social media users. Furthermore, the first step necessitates the selection of an acceptable data source to target the amount of publicly available real-time user data. However, due to its open source qualities and increased compatibility of variables such as time, speed, and connectivity, Twitter shows to be the finest source for data collecting among all social media platforms[52]..

2.3 Conceptual Framework of Processes

The social media posts are extracted using N-capture[53], this progresses to the preprocessing stage which includes data cleaning and categorization using natural language processing and psycholinguistic assessment which will be applied to the data set.

The performance phase will be done by applying the 10 fold cross validation and the hold out method where the machine learning algorithms namely logistic Regression ,Naïve bayes, Decision tree, Support vector Machine, and K nearest neighbor will be trained to identify post which indicate despondency[54]. This performance will also be analyzed using the WEKA 3.8.5(Waikato Environment for Knowledge Analysis)[55] and will be run through five performance metrics namely sensitivity ,specificity ,accuracy ,Precision and F-measure. This process will depict the selection of the optimal Algorithm using the Analytical Hierarchy Process[56].

2.5 Summary of Gaps in Literature Reviewed

The various processes used in extracting and processing social media posts have been researched by different authors to show the correlation between an individual's post on social media and their mental health, However a few studies have examined the use of Natural language processing in classifying text using support vector machines to detect depression using post in Reddit social media platform.This process had an accuracy of 80%. Further research was done with the multinomial Naïve Bayes theorem to detect depression on social media using the case study of Twitter,this was used to categorize depression level and the degree of depression[59].In recent research Despondency was detected using an ensemble classifier using Quality of life scales reflecting on an individual's access to healthcare as a classifier for Despondency. Initially a model was created using Questionnaires given to social media users to ask questions relating to their mental health[60],this process was flawed based on the inaccuracy of information and personal bias of the individual, it was therefore realistic to extract data from social media where the individual had a community to express their ideologies and challenges.

This thesis fulfills the future recommendations of a research done by of creating a multiclass model and an optimal classifier to measure performance based on five metrics and using several algorithms on the same dataset to extract despondency symptoms on social media, and also to indicate the individuals at risk of depression through their social media post[61].

Endnotes

[1] Gao, Shuang, DVince . Calhoun, & J Sui. *"Machine learning in major depression: From classification to treatment outcome prediction."***CNS neuroscience and therapeutics** 24, no. 11 2018: 1037-1052.

[2]. Farland, Gabriella. "Morningside Hospital*: A Historical Case Study of the Diagnosis and Treatment of Depression in Mid-20th Century American Psychiatry."* **University of Oregon Libraries,Scholars Bank** 2019.

[3]. Kennis, Mitzy, L Gerritsen, V Marije Dalen, A Williams, P Cuijpers, & C Bockting. *"Prospective biomarkers of major depressive disorder: a systematic review and meta-analysis."***Molecular Psychiatry 25**, no. 2 2020: pp. 321-338.

[4] **Hatton,M Christopher .,W Lewis. Paton, D McMillan, J Cussens, S Gilbody, & P. Tiffin. *"Predicting persistent depressive symptoms in older adults: a machine learning approach to personalised mental healthcare.*" Journal of Affective Disorders 246, 2019: 857-860.**

[5] **Carvalho, Andre F., J Firth, & E Vieta. *"Bipolar disorder.* "New England journal of medicine 383, no. 1 2020: pp. 58-66.**

[6] Watkins, E Laura.,R Kelsey,. Sprang, & O. Rothbaum. *"Treating PTSD: A review of evidence-based psychotherapy interventions."***Frontiers in behavioural neuroscience** 12 2018: pp. 258

[7] AlSagri,S Hatoon & M Ykhlef. *"Machine learning-based approach for depression detection in twitter using content and activity features."***IEICE Transactions on information systems** 103, no. 8 2020: pp. 1825-1832.

[8] Kulkarni, Akshay Bhavani Kumar. *"Early Detection of Depression." PhD dissertation.* **Department of computer Science, Univesidad Nacional de Colombia** 2018.

[9] **Kesharwani, P Kumar. *"A Critical Analysis on Effects of Social Media on School Students.*"International journal of Reviews and Research in social sciences 9, no. 2 2021: pp. 65-70**.

[10] Hegde, Shridhar, G. Santosh, M. Shivakumar, & R. Srihari. *"User Interest Prediction based on Social Network Profile with Machine Learning."* **In 2021 6th international conference for convergence in Technology (I2CT*)*,** 2021, pp. 1-6. IEEE.

[11] **Dhir, Amandeep, Y Yossatorn, P Kaur, & S Chen. *"Online social media fatigue and psychological wellbeing—A study of compulsive use, fear of missing out, fatigue, anxiety and depression.* "International Journal of information Management 40 2018: pp. 141-152**.

[12] **R Thapa, & S. Subedi. *"Social media and depression*." Journal of psychiatrists Association of Nepal, no. 2 2018: pp. 1-4.**

[13] Lin,Y Liu,E Jaime,. Sidani, A Shensa, A Radovic, E Miller,B Jason . Colditz, L. Hoffman, M Leila, Giles, & A. Primack. *"Association between social media use and depression among US young adults."* **Depression and anxiety vol** 33, no. 4 2016: pp. 323-331.

[14]**Shensa, Ariel,E Jaime . Sidani, M Amanda , G. Escobar-Viera, & A. Primack. *"Social media use and depression and anxiety symptoms: A cluster analysis.*" American journal of health behaviour'42, no. 2 2018: pp. 116-128.**

[15] Jalali, Alireza, J Sherbino, J Frank, & S Sutherland. *"Social media and medical education: exploring the potential of Twitter as a learning tool."* **International review of psychiatry** 27, no. 2 2015: pp. 140-146.

[16] Mirjalili, Seyedali, H Faris, & I Aljarah.**Evolutionary machine learning techniques Singapore: Springer**, 2019

[17] . D Tredici, Marco, D Marcheggiani, S Schulte Walde, & R Fernández. *"You shall know a user by the company it keeps:* **Dynamic representations for social media users in nlp**." arXiv preprint arXiv:1909.00412 2019.

[18] P Lavanya, & E. Sasikala. *"Deep learning techniques on text classification using Natural language processing (NLP) in social healthcare network:* **A comprehensive survey." In 2021 international conference on signal processing and communication** IEEE, 2021, pp. 603-609.

[19] Cervantes, Jair, F Garcia-Lamont, L Rodríguez-Mazahua, & A Lopez. "*A comprehensive survey on support vector machine classification: Applications, challenges and trends."***Neurocomputing** 408, 2020: pp. 189-215.

[20] Islam, Md Rafiqul, A Raihan M. Kamal, N Sultana, R Islam, & M Ali Moni. *"Detecting depression using k-nearest neighbors (knn) classification technique."* **In 2018 international conference on computer, communication, chemical, material and electronic engineering (IC4ME2),** IEEE, 2018, pp. 1-4.

[21] Islam, Md, M Ashad Kabir, A Ahmed, A Raihan M. Kamal, H Wang, & A Ulhaq. *"Depression detection from social network data using machine learning techniques*." **Health information science and systems** 6, no. 1 (2018): pp. 1-12.

[22] Jiang, Haihua, B Hu, Z Liu, G Wang, L Zhang, X Li, & H Kang. *"Detecting depression using an ensemble logistic regression model based on multiple speech features."* **Computational and mathematical methods in medicine** 2018

[23] Riekert, Martin, M Riekert, & A Klein. *"Simple Baseline Machine Learning Text Classifiers for Small Datasets."* **SN Computer Science** 2, no. 3 2021: pp. 1-16.

[24] Walkowiak, Tomasz, S Datko, & H Maciejewski. *"Bag-of-words, bag-of-topics and word-to-vec based subject classification of text documents -a comparative study."* **In International Conference on Dependability and Complex Systems,** Springer, Cham, 2018, pp. 526-535.

[25] Hasan, Md Rakibul, M Maliha, & M. Arifuzzaman. *"Sentiment Analysis with NLP on Twitter Data."* **International Conference on Computer, Communication, Chemical, Materials and Electronic Engineering** (IC4ME2), IEEE, 2019, pp. 1-4.

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

[27] **Bakken, Suzanne. *"Predictive models: important problems and innovative methods."* Journal of the American Medical Informatics Association 29, no. 1 2022: 1-2.**

[28] **L Glaz, Aziliz, Y Haralambous, D Kim-Dufor, P Lenca, R Billot, T C. Ryan & Jonathan Marsh.** ***"Machine learning and natural language processing in mental health: Systematic review."*** **Journal of medical internet research 23, no. 5 2021: e15708.**

[29] Tate,E Ashley.,C Ryan . McCabe, H Larsson, S Lundström, P Lichtenstein, & R Halkola. *"Predicting mental health problems in adolescence using machine learning techniques."* **plos one** 15, no. 4 2020: e0230389.

[30] **Wong, H Yan, H Yi Mo,N Marc. Potenza,N Mung, M Chan, W Man Lau, T Kwan Chui, A Pakpour, & C Lin.** ***"Relationships between severity of internet gaming disorder, severity of problematic social media use, sleep quality and psychological distress.*****"International journal of environmental research and public health 17, no. 6, 2020, pp. 1879.**

[31] Bhavya, &S Anitha, Pillai. *"Prediction models in healthcare using deep learning*." **In International Conference on soft computing and pattern Recognition**. Springer, Cham, 2019. pp. 195-204

[32] Mumtaz, Wajid, & Abdul Qayyum. ***"A deep learning framework for automatic diagnosis of unipolar depression."*****International journal of medical informatics 132, 2019, 103983**.

[33] Chancellor, Stevie, & D Choudhury. "*Methods in predictive techniques for mental health status on social media: a critical review*."**NPJ Digital medicine** 3, no. 1, 2020

[34] **. Ridout, Brad, & A. Campbell.** ***"The use of social networking sites in mental health interventions for young people: systematic review."*** **Journal of medical internet 20, no. 12, 2018, e12244.**

[35] **Kim, Jiin,A Zara. Uddin, Y. Lee, F Nasri, H Gill, M Subramanieapillai & R. Lee .** ***"A systematic review of the validity of screening depression through Facebook, Twitter, Instagram, and Snapchat."*** **Journal of Affective Disorder 286, 2021, pp. 360-369.**

[36] **Laacke, Sebastian, R Mueller, G Schomerus, & S Salloch.** ***"Artificial intelligence, social media and depression. A new concept of health-related digital autonomy.*****"The American Journal of Bioethics 21, no. 7, 2021 pp. 4-20.**

[37] **Xie, Zhigang, R Tanner,L Catherine. Striley, & M. Marlow.** ***"Association of functional disability with mental health services use and perceived unmet needs for mental health care among adults with serious mental illness."*****Journal of affective disorders 299, 2022 pp. 449-455**

[38] . Kennis, Mitzy, L Gerritsen, M. Van-Dalen, A. Williams, P. Cuijpers, & C Bockting. *"Prospective biomarkers of major depressive disorder***: a systematic review and meta-analysis. "Molecular Psychiatry** 25, no. 2, 2020: pp. 321-338.

[39] **Zhang, Fuquan, S Rao, H Cao, X Zhang, Q Wang, X Yong & J Sun.** ***"Genetic evidence suggests posttraumatic stress disorder as a subtype of major depressive disorder."*** **The Journal of clinical investigation 132, no. 3, 2022.**

[40] **Costantini, Luigi, C Pasquarella, A. Odone, M Colucci, A. Costanza, G Serafini, A Aguglia. *"Screening for depression in primary care with Patient Health Questionnaire-9 PHQ-9: A systematic review*." Journal of Affective Disorder 279, 2021: pp. 473-483**.
[41] V Glischinski, Michael, R Brachel, & G Hirschfeld. *"How depressed is "depressed"? A systematic review and diagnostic meta-analysis of optimal cut points for the Beck Depression Inventory revised (BDI-II).*"**Quality of Life Research** 28, no. 5 2019: 1111-1118.

[42] **James, Caryl, M Powell, A Seixas, A Bateman, S Pengpid, & K Peltzer. *"Exploring the psychometric properties of the CES-D-10 and its practicality in detecting depressive symptomatology in 27 low-and middle-income countries."* International Journal of Psychology 55, no. 3 2020, pp. 435-445.**

[43] Levis, Brooke, A. Benedetti, & D. Thombs. *"Accuracy of Patient Health Questionnaire-9 (PHQ-9) for screening to detect major depression: individual participant data meta-analysis*"365, 2019

[44] Bech, Per. *"The Rating scales in depression: limitations and pitfalls*."**Dialogues in Clinical Neurosciece 2022.**

[45] Ríssola, A. Esteban., A. Bahrainian, & F Crestani. *"A dataset for research on depression in social media."* **In proceedings of the 28th ACM conference on user modeling and personalization,** 2020, pp. 338-342.

[46] Yasin, Sana, S Hussain, S Aslan, I Raza, M Muzammel, & A. Othmani. *"EEG based Major Depressive disorder and Bipolar disorder detection using Neural Networks: A review."* **Computer Methods and Programs in Biomedicine** 2021, 106007, pp 202.

[47] Hussain, Zaheer, E Wegmann, H Yang, & C Montag. *"Social networks use disorder and associations with depression and anxiety symptoms: A systematic review of recent research in China.*" **Frontiers in psychology** 11 2020: 211.

[48] Islam, M, Muhammad A Kabir, A Ahmed, A Raihan M. Kamal, H Wang, & A Ulhaq. *"Depression detection from social network data using machine learning techniques*." **Health information science and systems** 6, no. 1 2018: pp. 1-12.

[49] Patinge,A Sushant., & K Vijaya. Shandilya. *A survey on mental health Prediction using social media network* No. 3521**. EasyChair**, 2020.

[50] **Arunadevi., S. Ramya, & M. Raja. *"A study of classification algorithms using Rapidminer.*"International Journal of pure and applied Mathematics 119, no. 12 2018: 15977-15988**

[51] M Valdilo,V Curcin*" Modelling depression Symptoms from social media through instance hearing"* **ReserachGate** May 2019

[52] K Cornn*"Identifying Depression on Social Media"* **Department of Statistics, Stanford University CS22N CS229A.** Jan 2021

[53] Singh, Shivani, & Sandhya Tarar. *"Mental illness diagnosis from social network data using effective machine learning technique."* **In Artificial Intelligence and Speech Technology,** CRC Press, 2021, pp. 231-238

[54] Berrar, Daniel. *"Cross-Validation*." **Data science Laboratory, Tokyo institute of Technology 2-12-1-S3,** 2019: pp. 542-545.

[55] **Ibrahim, Fatma A., & Omar A. Shiba. *"Data Mining: WEKA Software (an Overview*)." Journal of Pure & Applied Sciences 18, no. 3, 2019.**

[56] Naser, M. Z. & Amir H. Alavi. *"Error metrics and performance fitness indicators for artificial intelligence and machine learning in engineering and sciences*." **Architecture, Structures and Construction** 2021: pp. 1-19.

[57] Boer, Maartje, Gonneke, Stevens WJM, Catrin F, Margaretha E. de Looze, & Regina JJM van den E. *"Social media use intensity, social media use problems, and mental health among adolescents: Investigating directionality and mediating processes."* **Computers in Human Behavior 116,** 2021: 106645**.**

[58] Tadesse, Michael M., Hongfei Lin, Bo Xu, & Liang Yang. *"Detection of depression-related posts in reddit social media forum."* **IEEE Access 7,** 2019: 44883-44893

[59] Tsui, Fuchiang R., Lingyun Shi, Victor R, Neal D, Candice B, Satish I, Colin G. Walsh,& Brent A. *"Natural language processing and machine learning of electronic health records for prediction of first-time suicide attempts.*" **JAMIA open 4, no. 1,** 2021: ooab011.

[60] Hatton, Christopher M., Lewis W. Paton, Dean M, James C, Simon Gilbody, & Paul A. *"Predicting persistent depressive symptoms in older adults: a machine learning approach to personalised mental healthcare."* Journal of affective disorders 246 2019: pp. 857-860.

[61] Mustafa, Raza Ul, Noman A, Fahad S.A, Javed F, Basit S, & Alexander G. *"A multiclass depression detection in social media based on sentiment analysis."* **In 17th International Conference on Information Technology–New Generations ITNG 2020, Springer, Cham,** 2020, pp. 659-662.

Chapter Three

Methodology

3.1 Research Approach

This chapter presents the details of the use of machine learning and data analysis approach to extract depression comments from social media posts for identifying at-risk user. It explains the various approaches, tools and algorithms that were used in achieving the stated objectives of this research.

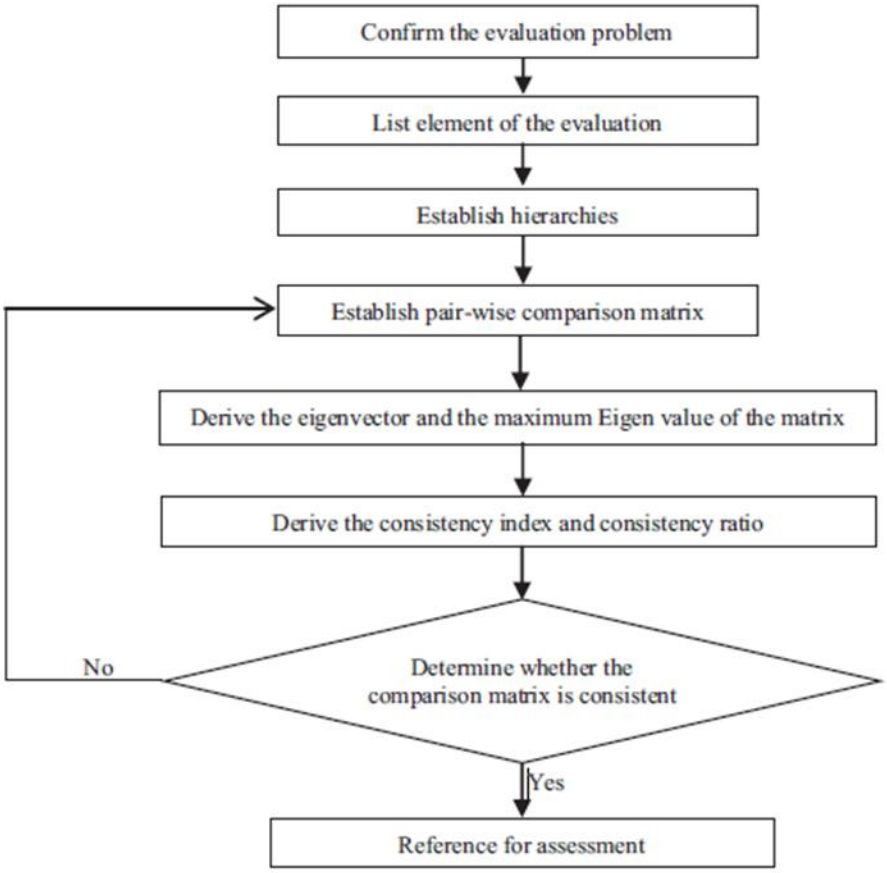


Figure 3.0: Flowchart of the decision process of AHP Business Performance Management Singapore 2022)[1].

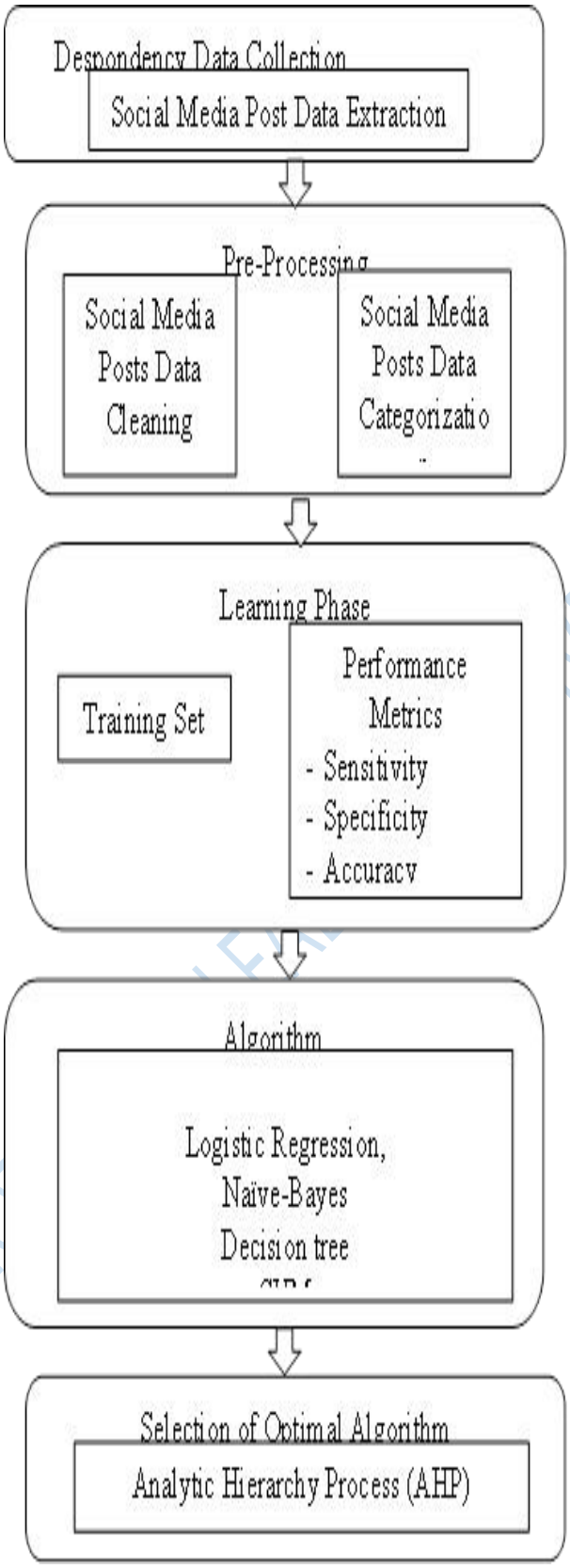


Figure 3.1: Framework for identifying the best machine learning algorithm for detecting despondency social media posts early (Researcher Hally D,2022).

3.2 System Design

The concept of this methodology necessitates the use of Agile and waterfall project processes. The Agile process involves cyclic and collaborative process in which this project goes through a series of cycles through the duration of the project. This is observed from the framework, that the data set will be run through the five machine learning algorithms using the 10 fold cross validation process, also the data set will also be applied on the algorithms using the 80;20 hold out process, each algorithm has a cycle. Subsequently this project will undergo the waterfall process after each segment of the project has been established, the waterfall process will then proceed after each cycle has been completed to validate the optimal classifier based on the findings.

3.3 Data Collection

7,145 social media posts were extracted using N capture from publicly available data from Github. This social media post is from a bipolar, despondency and anxiety facebook page containing users comments. This Data Set was obtained from a Depression and Anxiety Facebook page by Islam, to help us overcome the challenge of social media posts in detecting if a post is termed Depressive or not. This gives us the ability to discover information in a more proficient way and to avoid the road blocks of ethical approval of data.

3.4 Research Methods

The tools used for this thesis include

i) N capture – this is used to extract data and arrange it to the specification of the researcher. This processing tool is a web extension that enabled the researcher to gather web content. It can also be used to capture pages from Facebook,Twitter and Youtube as Data sets,

ii)Meaning Extraction Helper – This is an application software for text analysis with the ability to process text line by line This will be used the clean the data.

iii) Psycholinguistic measurements that extract psycholinguistic features in the data, 23 features were applied on the data set. This was used as the Ground truth label. From the 23 features the data was categorised into Depressed, non- depressed and indicative depression

iv) Performance evaluation will be applied on the categorised data on the five machine learning algorithms using the 10 fold cross validation and the 80;20 hold out process.

v) The AHP will be calculated using the performance metrics to excerpt the optimal classifier

3.5 Data Analysis with Objectives

3.5.1 Categorization of Social Media Posts for Early Detection of Despondency

To achieve the first objective, previous literatures were reviewed on the extraction of depressive comments from social media post using Natural Language Processing and later categorize in determining the appropriate data for the early detection of despondency. This comments of the user will be used in data collection as the users are diagnosed with despondency. Followed by data cleaning and categorization for early detection of despondency.

3.5.2. Social Media Posts Extraction

The extraction of the user's post, is the data collection stage. This explains the method of data collection. The social media posts extraction will be based on Facebook user's comment for despondency exploration and detection as outlined in the scope of the study. From facebook presents an opposition as it contains information regarding despondency symptoms in the comments. Ncapture is an innovative tool which deconstructs and rearranges data from different sources[1].It also aligns the contents into categories and further examined using the LIWC software. After the collection of the raw data from Facebook, it will be cleaned from any inconsistency and later analyzed by using LIWC Software[2]

.

3.5.3 Social Media Posts Cleaning

The process of removing punctuation marks,html tags etc is termed the data cleaning process. These are removed to enhance the needed content towards processing with the machine learning algorithm The procedure of data cleaning in will be performed by the removal of stop words, URL, emojis and white spaces . This was performed using the NLT tool kit.

3.5.4 Social Media Posts Categorization

The posts will be built with ground truth label information on whether the social media post is despondency indicative in order to categorize them. Psycholinguistic assessment will also be utilized to extract elements that will be used to describe and categorize social media posts. This will be accomplished by employing the psycholinguistic vocabulary package LIWC, developed by psychological analysts, to recognize the many emotive, etymological, and intellectual

components in written or vocal correspondence. The LIWC2015 scales will then be applied to the datasets to obtain numerical values. The dataset will be used to generate a list of users who are depressed or not depressed.

3.5.5 Machine Learning Algorithms Performance Evaluation

The performance of the machine learning algorithms will be evaluated using seven supervised machine learning methods to meet the second objective. Logistic Regression, Nave-Bayes, SVM, Decision Tree, and K-Nearest Neighbors are some of the techniques used. The research task will be classified by input and output, and the selection of these machine learning methods are based on the research task. The input (despondency categories) is a set of labelled data, so it's a supervised learning issue, and the model's output is a class with a known number of predicted classes, so it's a classification problem. As a result, the above-mentioned machine learning methods for categorization and prediction were used. Training and testing of the randomly partitioned partitions of the seven chosen supervised machine learning algorithms will be implemented out on the classified despondency dataset applying WEKA 3.8.5 for algorithm performance evaluation. The analysis will be performed using the Waikato Environment for Knowledge Analysis (WEKA) with a holdout of 80:20 percent of the dataset and 10-fold cross validation. Five measures of performance (metrics) will be employed. Sensitivity, specificity, accuracy, precision, and F-measure will be used as performance measures (metrics).

Sensitivity is the fraction of correctly identified positive instances, while specificity is the proportion of correctly categorized negative instances. Precision refers to the number of defective occurrences that have been classified and are therefore defective. The harmonic mean

of recall and precision is known as the F-measure, and it is often utilized for information retrieval.

The value of the sensitivity, specificity, accuracy, precision and F-measure are as shown below:

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad \text{---------- Equation 3.1}$$

$$\text{Sensitivity} = \frac{TP}{TP+FN} \quad \text{---------- Equation 3.2}$$

$$\text{Specificity} = \frac{TN}{FP+TN} \quad \text{---------- Equation 3.3}$$

$$\text{Precision} = \frac{TP}{TP+FP} \quad \text{---------- Equation 3.4}$$

$$\text{F-measure} = \frac{2\ x\ Precision\ x\ Recall}{Precision + Recall} \quad \text{---------- Equation 3.5}$$

True negative and positive are represented by TN and TP, while false negative and positive are represented by FN and FP..

3.5.6 Selection of Optimal Classifier using Analytic Hierarchy Process (AHP)

To achieve the third objective, Analytic Hierarchy Process (AHP) will be used to determine the optimal classifier for early detection of at-risk user of despondency based on the five performance (metrics) that will be employed (sensitivity, specificity, accuracy, precision and F-measure).

Endnotes

[1]P. Bazeley, K Jackson. *Qualitative data analysis with NVivo***. London: Sage**; 2013.

[2] H Yahmady & S Alabri. *Using NVivo for data analysis in qualitative research. Int Interdiscip* **J Educ**. 2013; 2 (2): pp. 181–6.

[3] W Bandara*, Using Nvivo as a research management tool: a case narrative. In: Quality and impact of qualitative research:* proceedings of **3rd international conference on qualitative research in IT & IT in Qualitative Research. 2006.**

[4] Y Tausczik & J Pennebaker. *The psychological meaning of words: LIWC and computerized text analysis methods***. J Lang Soc Psychol.** 2010: pp. 24–54

.[5] JW Pennebaker & M Francis, *Booth RJ. Linguistic inquiry and word count: LIWC 2001, vol. 71. Mahway:* **Lawrence Erlbaum Associates**; 2001. p. 2001.

[6] Y Tausczik &J. Pennebaker . *The psychological meaning of words: LIWC and computerized text analysis methods.* **J Lang Soc Psychol**. 2010: pp. 24–54

.[7] Khanmohammadi & Rezaeiahari. *AHP Based Classification Algorithm Selection for Clinical Decision Support System Development***. Procedia Computer Science** 36, 2014, pp. 328 – 334

Chapter four

Results and Discussion of Findings

The in depth breakdown and proposal carried out in the previous chapter culminated in the discussion of results in this chapter. Section 4.1 is focused on the categorization of social media post for early detection of despondency using Natural Language Processing (NLP). Section 4.2 shows the performance evaluation outputs of the seven machine learning algorithms used for early detection of despondency based on the selected metrics in addition to the presentation of various graphical interfaces. Section 4.3 highlights the determination of the optimal classifier for early detection of despondency based on the selected metrics using Analytical Hierarchical Process (AHP) while section 4.4 is focused on the discussion of evaluation results for the use machine learning techniques in analysing comments from social media posts for early detection of despondency.

4.1 Categorization of Social Media Posts for Early Detection of Despondency

An in-depth analysis of previous literatures were reviewed on the extraction of depressive comments from social media post using Natural Language Processing and later categorize in determining the appropriate data for the early detection of despondency

.

4.1.1 Extraction of Social Media Posts

We used readily accessible Facebook data from Github for this investigation. The dataset was obtained from https://github.com/ranju 12345 /Depression-Anxiety-Facebook-page-Comments-Text by Islam . The Ncapture was used to extract the social media post (from the bipolar, depression, and anxiety Facebook page) containing user comments in order to overcome the challenges of social media postings, particularly Facebook, in confirming whether the posts contain depression-related material. This was developed with the intent of structuring, breaking down, and discovering knowledge in unstructured data such as social media posts. Furthermore, it allows the capability to discover knowledge more effectively.

4.1.2 Cleaning of Social Media Posts

After the extraction of the social media post, it was analysed using the Meaning Extraction Helper application software for the text analysis strategy which also have the ability to process text line by line . The social media comment was cleaned of inconsistencies before being analyzed using Meaning Extraction Helper, an open source application program. The social media post has 21 columns, with 13 columns representing linguistic style (prepositions, conjunctions, articles, auxiliary verbs, personal pronouns, verbs, impersonal pronouns, negation, and so on), 5 columns representing emotional data (negative, anger, positive, sad, and anxiety), 3 columns representing temporal process (future, present, and past), and each column representing individual data about depressive behaviour. (See Figs. 4.1 and 4.2)

4.2 Categorization of Social Media Posts for Depressive Comments

Two steps was conducted to better categorize and determine the depressive comments according to the priorities for despondency symptoms, namely ground truth building and despondency symptoms behaviour.

4.2.1 Ground Truth Building

To construct the dataset for data analysis approach from the extracted depression comments, a ground truth label was built in view of the psycholinguistic measurement from the social media posts, a psycholinguistic vocabulary package constructed by psychological analysts to recognize the distinct emotive, intellectual, and etymological elements of a user's verbal or written conversation. This procedure generates almost 70 distinct components with varying levels of psycholinguistic characteristics. In context of the psycholinguistic qualities, 23 of the 70 components were considered and classified into three numerical categories depending on the ground truth.

4.2.2 Despondency Symptoms Behaviour

The twenty three (23) factors were categorized into three sets of despondency symptoms namely, Zero (0) category (non-depressive post), One (1) category (indicative depressive post) and Two (2) category (depressive post). The categorization of the despondency symptoms to determine the depressive behaviour of social media post was presented based on the set of the psycholinguistic attributes namely emotional process, linguistic style and temporal process. From the 7145 social media posts, 27% obtained Zero (0) category for non-depressive post, 32% obtained One (1) category for indicative depressive post and 41% obtained Two category (2) for

depressive post as illustrated in Table 4.1.

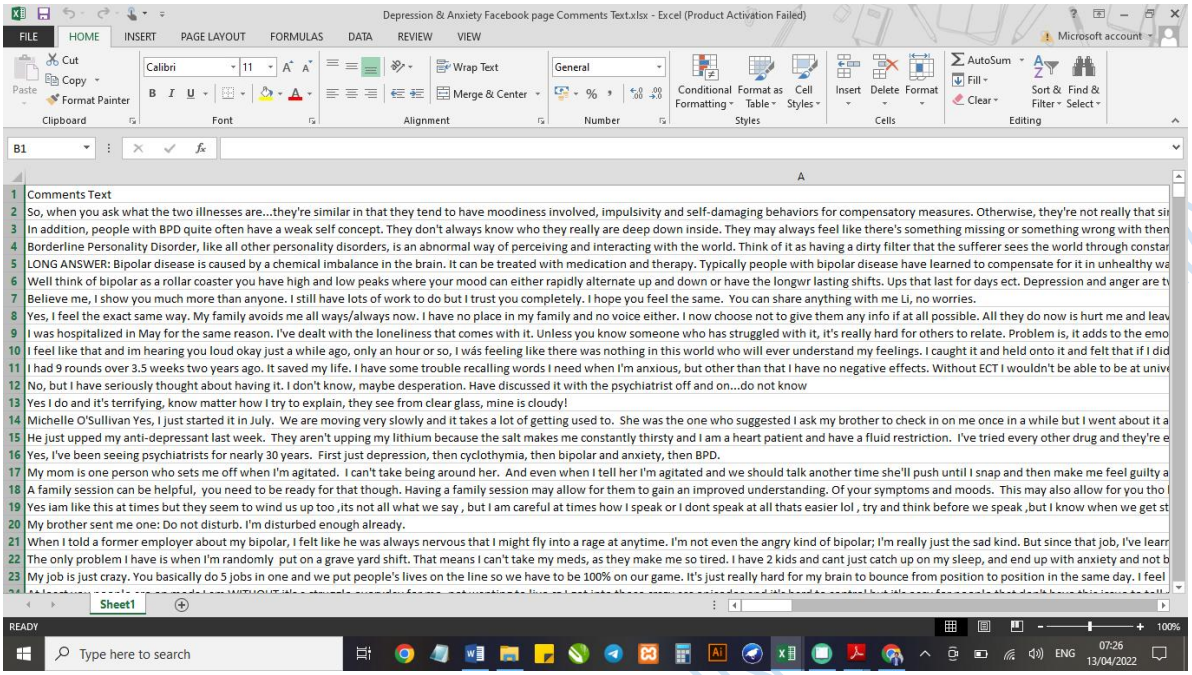


Figure 4.1: Social Media Post (https://github.com/ranju 12345 /Depression-Anxiety-Facebook-page-Comments-Text by Islam , 2018)

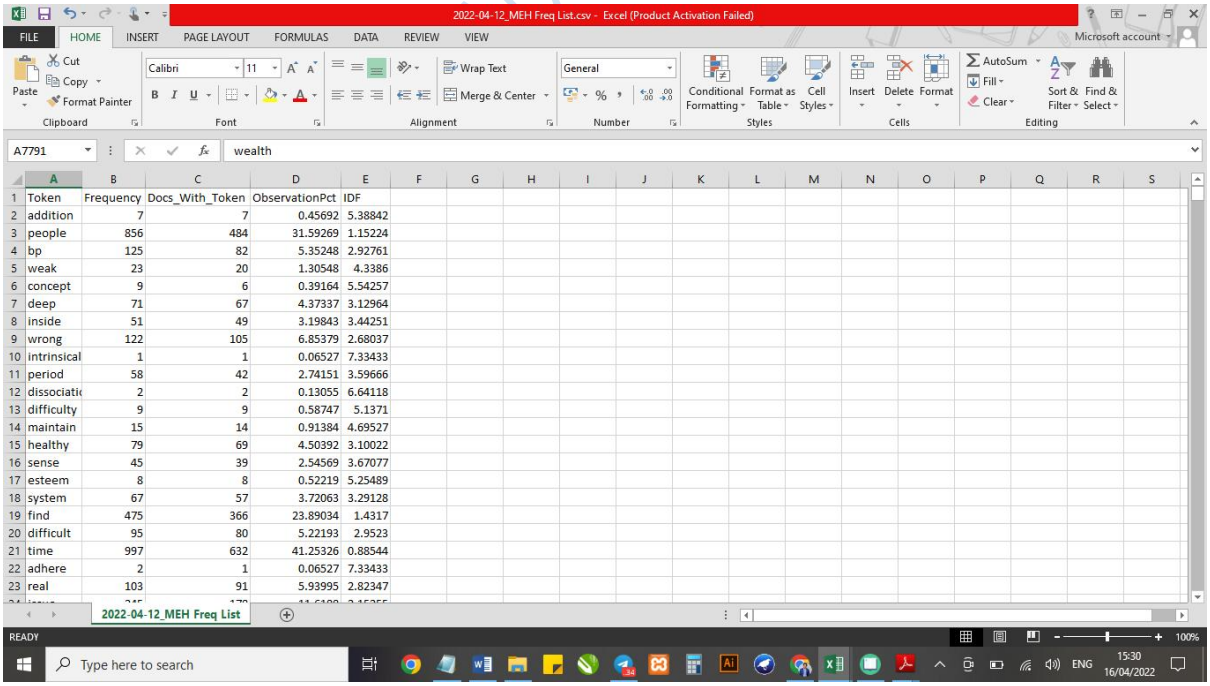


Figure 4.2: Characteristics of social media post Researcher,.Hally D, 2022)

Table 4.1: Despondency Symptoms Distribution of Social Media Posts

Social Media Posts	Quantity
Total number of posts	7145
Non Depressive posts	1929
Indicative Depressive Posts	2286
Depressive Posts	2930

Source ResearcherD Hally,2022

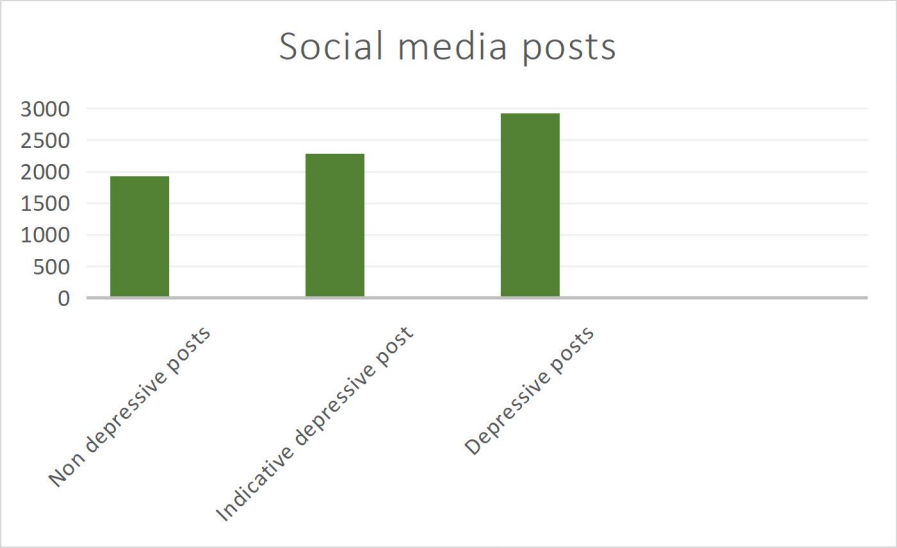


Figure 4.3: Social Media post Despondency Symptoms Distribution

4.3 Performance Evaluation of the Machine Learning Algorithms

The five machine learning algorithm in this study which represent the five classifier in this research are Logistic Regression, Decision Tree, Naïve-Bayes, Support Vector Machine, and K-Nearest Neighbors (KNN).These algorithms were used because they are classification Algorithms and the social media post are classified into 3 categories. The models for machine learning algorithms were built in WEKA and evaluated with 80% hold out and 10-fold cross validation techniques for algorithm testing and training on the social media posts extracted.

4.3.1 Performance evaluation of Algorithm using Hold Out and Cross Validation Methods

A total of 7145 instances (row of social posts) with 29 attributes (psycholinguistic attributes) were used as independent variables and 1 class variable as dependent variable and evaluated in WEKA on 10-F C-V in building the model on the machine learning algorithms. On the contrary, for hold-out method, a total of 7145 instances of social posts with 29 attributes were used as independent variables and 1 class variable as dependent variable. Out of which 5,716 instances were used as training data and 1,429 instances were used as test data in the percentage-split ration 80:20 in which the performance evaluation was carried out in WEKA.

4.3.2 Logistic Regression Algorithm performance

It was observed as shown in Table 4.2 using 10-F C-V method that TP and TN were correctly predicted for despondency symptoms type as portrayed across the main diagonal of the confusion matrix.

Table 4.2 Confusion Matrix for Logistic Regression Using Cross Validation Method

Despondency Symptoms Type	**Instances**
Indicative Depressive Post	6757
Depressive Post	383
Non-Depressive Post	5
Total Instances	**7145**

(Source: Researcher ,D Hally, 2022)

It was also observed as shown in Table 4.3 using hold-out method that TP and TN were correctly predicted for despondency symptoms type as showed across the main diagonal of the confusion matrix.

Table 4.3 Confusion matrix for Logistic Regression Using Hold-out Method

Despondency Symptoms Type	Instances
Indicative Depressive Post	1366
Depressive Post	62
Non-Depressive Post	1
Total Instances	**1429**

(Source: Researcher,D Hally, 2022)

4.3.3 Naïve-Bayes Algorithm performance

It was observed as shown in Table 4.3 using 10-F C-V method that TP and TN were correctly predicted for despondency symptoms type as portrayed across the main diagonal of the confusion matrix.

Table 4.3 Confusion Matrix for Naïve-Bayes using Cross Validation Method

Despondency Symptoms Type	Instances
Indicative Depressive Post	6757
Depressive Post	383
Non-Depressive Post	5
Total Instances	**7145**

(Source: Researcher D .Hally , 2022)

TP and TN were also successfully predicted for despondency symptoms type, as exhibited across the major diagonal of the confusion matrix, as shown in Table 4.4 using the hold-out approach

Table 4.4 Confusion Matrix for Naïve-Bayes using Hold-Out Method

Despondency Symptoms Type	Instances
Indicative Depressive Post	1366
Depressive Post	62
Non-Depressive Post	1
Total Instances	**1429**

4.3.4 Decision Tree Algorithm performance

TP and TN were successfully predicted for depression symptoms type as displayed across the major diagonal of the confusion matrix, as illustrated in Table 4.5 using the 10-F C-V approach.

Table 4.5 Confusion Matrix for Decision Tree Using Cross Validation Method

Despondency Symptoms Type	Instances
Indicative Depressive Post	6757
Depressive Post	383
Non-Depressive Post	5
Total Instances	**7145**

(Source: D.Hally, 2022)

TP and TN were successfully predicted for depression symptoms type as exhibited across the major diagonal of the confusion matrix, as illustrated in Table 4.6 using the hold-out approach.

Table 4.6 Confusion Matrix for Decision Tree Using Hold-Out Method

Despondency Symptoms Type	Instances
Indicative Depressive Post	1366
Depressive Post	62
Non-Depressive Post	1
Total Instances	**1429**

(Source:Researcher D.Hally, 2022)

4.3.5 K-Nearest Neighbors Algorithm Performance

TP and TN were successfully predicted for depression symptoms type as displayed across the major diagonal of the confusion matrix, as illustrated in Table 4.5 using the 10-F C-V approach
.

Table 4.7 Confusion Matrix for K-Nearest Neighbors Using Cross Validation Method

Despondency Symptoms Type	Instances
Indicative Depressive Post	6757
Depressive Post	383
Non-Depressive Post	5
Total Instances	**7145**

(Source: Researcher D.Hally 2022)

TP and TN were successfully predicted for depression symptoms type as exhibited across the major diagonal of the confusion matrix, as illustrated in Table 4.8 using the hold-out approach**.**

Table 4.8 Confusion Matrix for K-Nearest Neighbors Using Hold-Out Method

Despondency Symptoms Type	**Instances**
Indicative Depressive Post	1366
Depressive Post	62
Non-Depressive Post	1
Total Instances	**1429**

(Source: Researcher D.Hally, 2022)

4.3.6 Support Vector Machines Algorithm Performance

TP and TN were successfully predicted for depression symptoms type as displayed across the major diagonal of the confusion matrix, as shown in Table 4.9 using the 10-F C-V approach.

Table 4.9 Confusion Matrix for Support Vector Machines Using Cross Validation Method

Despondency Symptoms Type	**Instances**
Indicative Depressive Post	6757
Depressive Post	383
Non-Depressive Post	5
Total Instances	**7145**

(Source: Researcher,D.Hally, 2022)

TP and TN were also successfully predicted for despondency symptoms type as exhibited across the major diagonal of the confusion matrix using the hold-out method, as shown in Table 4.10.

Table 4.10 Confusion Matrix for Support Vector Machines Using Hold-Out method

Despondency Symptoms Type	**Instances**
Indicative Depressive Post	1366
Depressive Post	62
Non-Depressive Post	1
Total Instances	**1429**

(Source:Researcher D.Hally, 2022)

4.4 Determination of Optimal Despondency Symptoms Classifier Using Analytic Hierarchy Process (AHP)

The result of the machine learning algorithm performance using AHP was used in the choice of determining the optimal despondency symptoms classifier for predictive analytic model.

4.5 The Summary of the Performance of the Algorithms in Cross Validation and Hold Out Processes.

The execution of the five machine learning algorithms namely Logistic Regression, Decision Tree, Naïve-Bayes, Support Vector Machine, and K-Nearest Neighbors (KNN) using the metrics accuracy, sensitivity, specificity, precision and F-measure in 10-F C-V method and hold-out method are shown in Table 4.11 and Table 4.12 respectively. In respect to the performance Logistic Regression, Decision Tree and Support Vector Machine performed brilliantly well in relative to one metric or the other.

Table 4.11 Result Summary of Cross Validation Method for MLAs

MLAs	Accuracy (%)	TP_Rate (Sensitivity) (%)	TN_Rate (Specificity) (%)	Precision (%)	F-measure (%)
Logistic Regression	96.4	96.4	94.1	96.6	95.4
Decision Tree	95.2	95.2	93.7	94.3	86.3
Naïve-Bayes	77.3	77.3	74.7	84.3	71.1
Support Vector Machine	99.5	99.5	99.4	99.5	99.3
K-Nearest Neighbors	89.5	89.5	89.7	89.1	89.3

(Source:Reseacher D.Hally 2022)

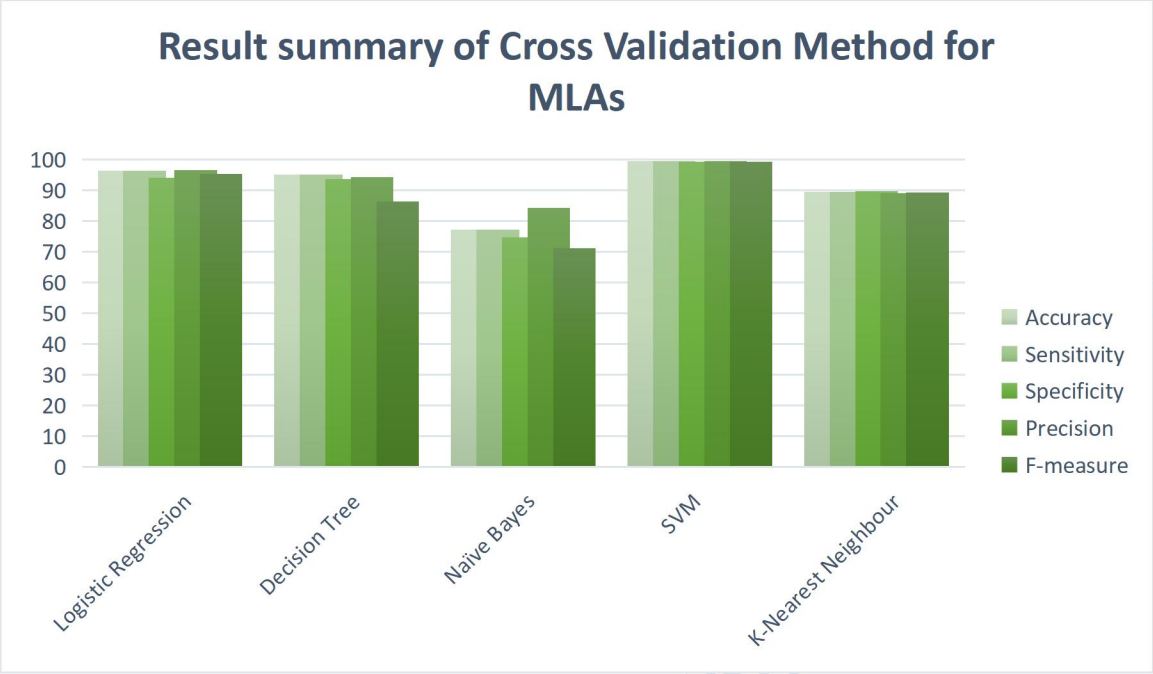


Figure 4.4: Result Summary of Cross Validation Method for MLAs

The performance analysis based on cross validation method, conducted on the algorithms as shown in Table 4.11. It is obvious that Support Vector Machine was positioned at top level in respect of classification accuracy, sensitivity, specificity, precision and F-measure with 99.5%, 99.5%, 99.4%, 99.5% and 99.3% respectively. Logistic Regression, Decision Tree, Naïve-Bayes and K-Nearest Neighbors fell in the latter category respectively.

Table 4.12 Result Summary of Holdout Method for MLAs

MLAs	Accuracy (%)	TP_Rate (Sensitivity) (%)	TN_Rate (Specificity) (%)	Precision (%)	F-measure (%)
Logistic Regression	95.3	95.3	92.7	93.8	94.6
Decision Tree	72.8	73.0	73.2	72.0	77.2
Naïve-Bayes	77.0	77.0	73.7	84.7	70.6

Support Vector Machine	98.4	98.4	98.3	98.4	99.2
K-Nearest Neighbors	83.6	84.0	96.9	74.0	84.7

(Source:Researcher, D.Hally 2022)

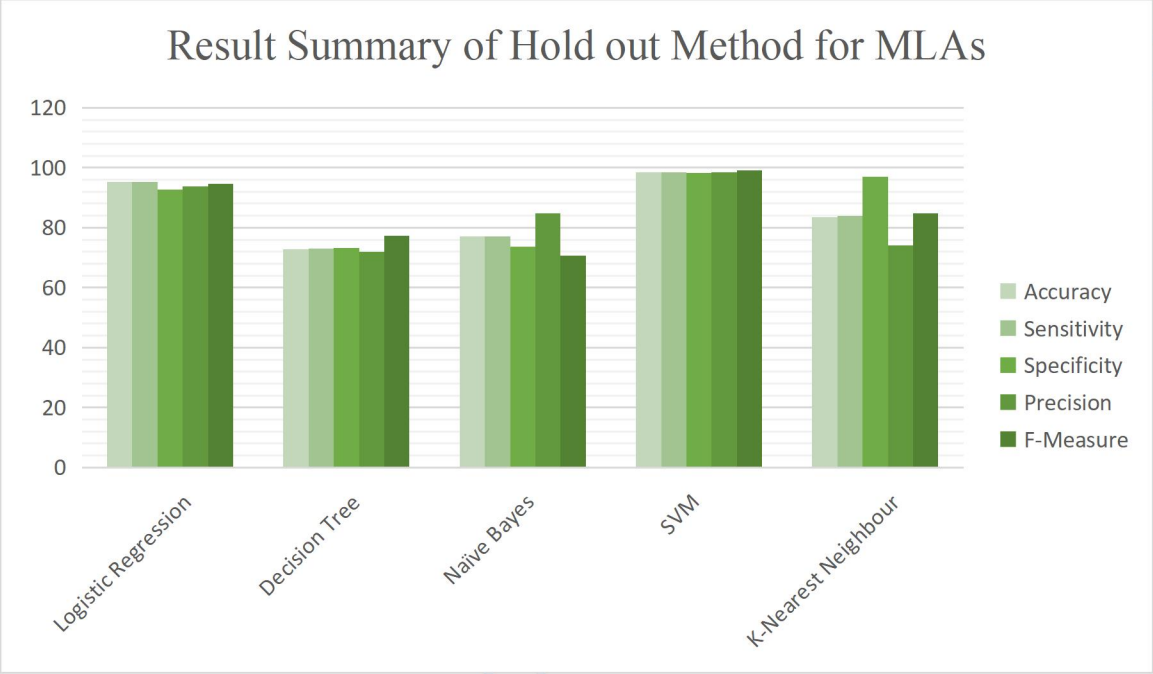


Figure 4.5: Result Summary of Hold out Method for MLAs

The performance analysis based on hold-out method, steered on the algorithms as shown in Table 4.12. It is obvious that Support Vector Machine was positioned at top level in respect of classification accuracy, sensitivity, specificity, precision and F-measure with 98.4%, 98.4%, 98.3%, 98.4% and 99.2% respectively. Logistic Regression, Decision Tree, Naïve-Bayes and K-Nearest Neighbors fell in the latter category respectively.

4.6 Analytic Hierarchy Process (AHP) Model

An AHP model was created using the results of the algorithm performance evaluation to rank the best performance measures that were used in selecting the best classifier for despondency symptoms. This

method allows experts' understanding to be incorporated into the choosing of classification algorithms, which has hardly been achieved previously. Pair-wise comparisons for qualities are stored in a matrix and priority weights are determined using the expert's knowledge. Additionally, the local consistency ratio of the qualities is calculated to verify that subjective judgments are consistent.

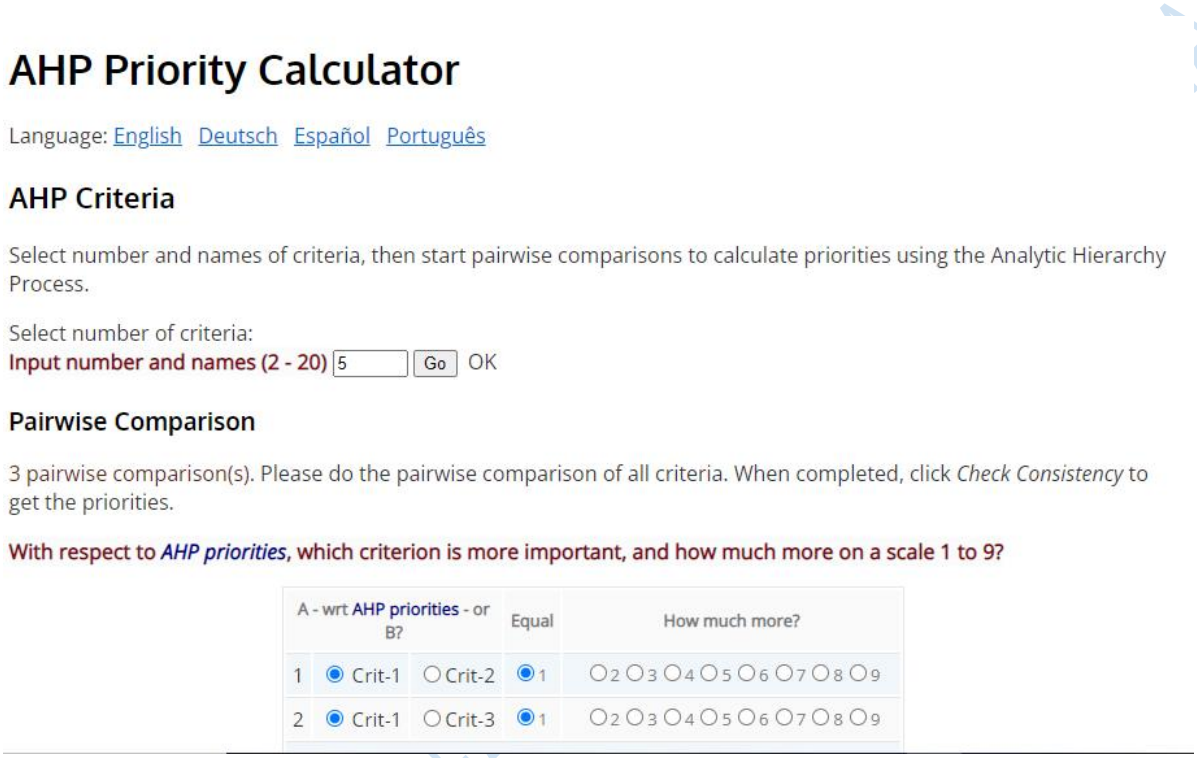


Figure 4.6: AHP-OS (Business Performance Management Singapore,bpms 2022)

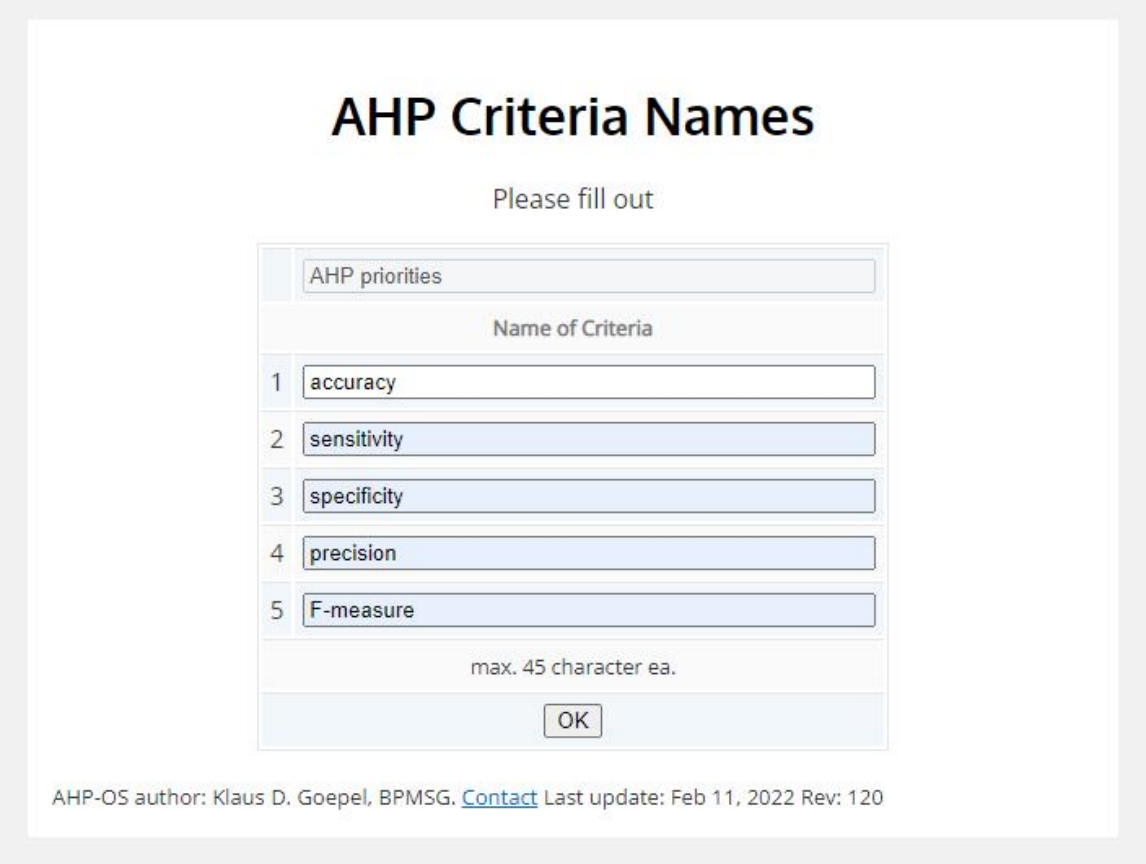


Figure 4.7: AHP Criteria Names (Business Performance Management Singapore bpms 2022)

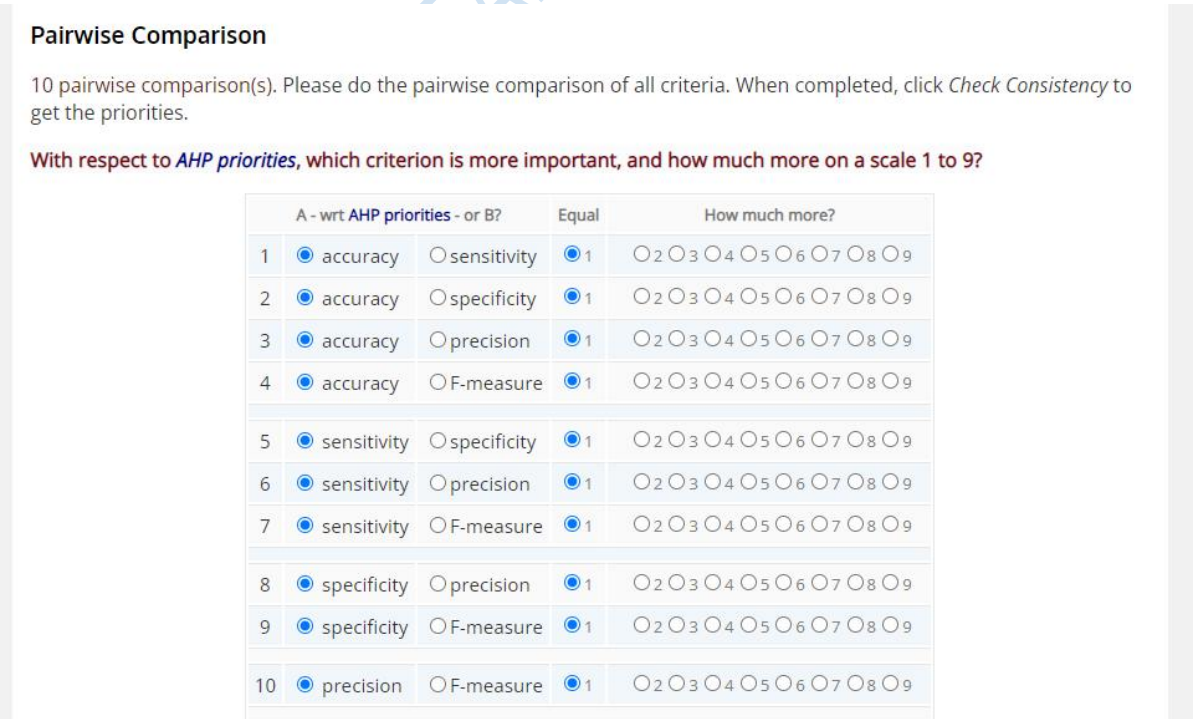
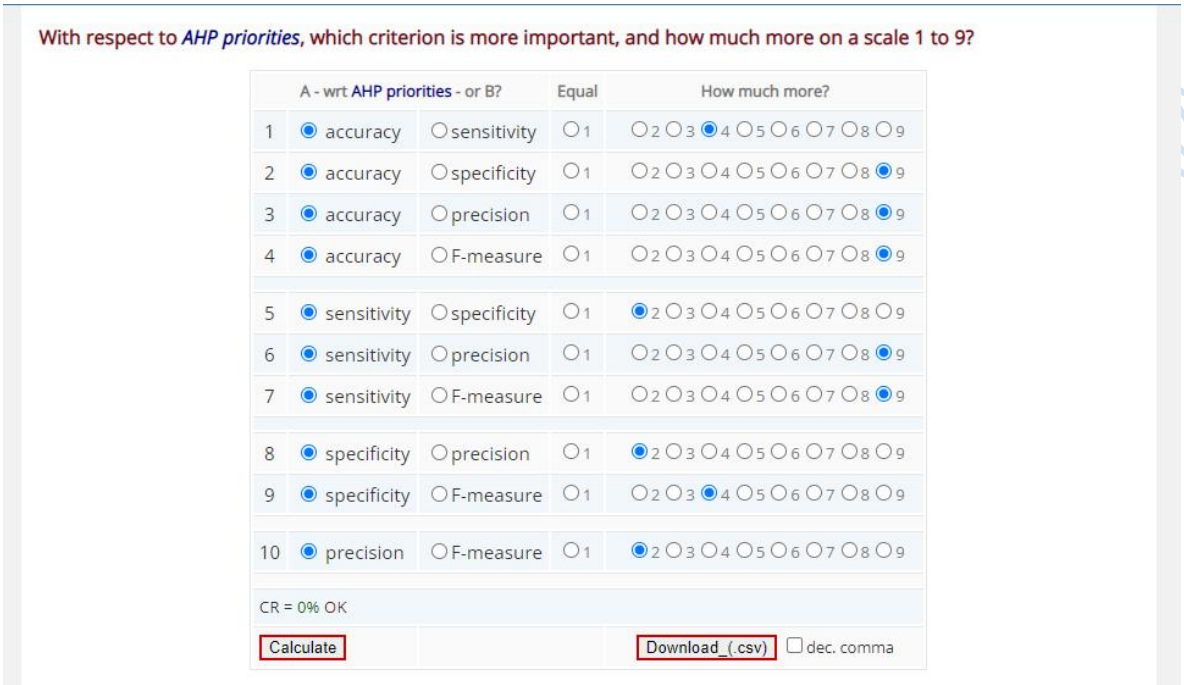


Figure 4.8: AHP Pairwise Comparison (Application of Multi-criteria Decision Making Theories in Healthcare and Biomedical Engineering.2021)



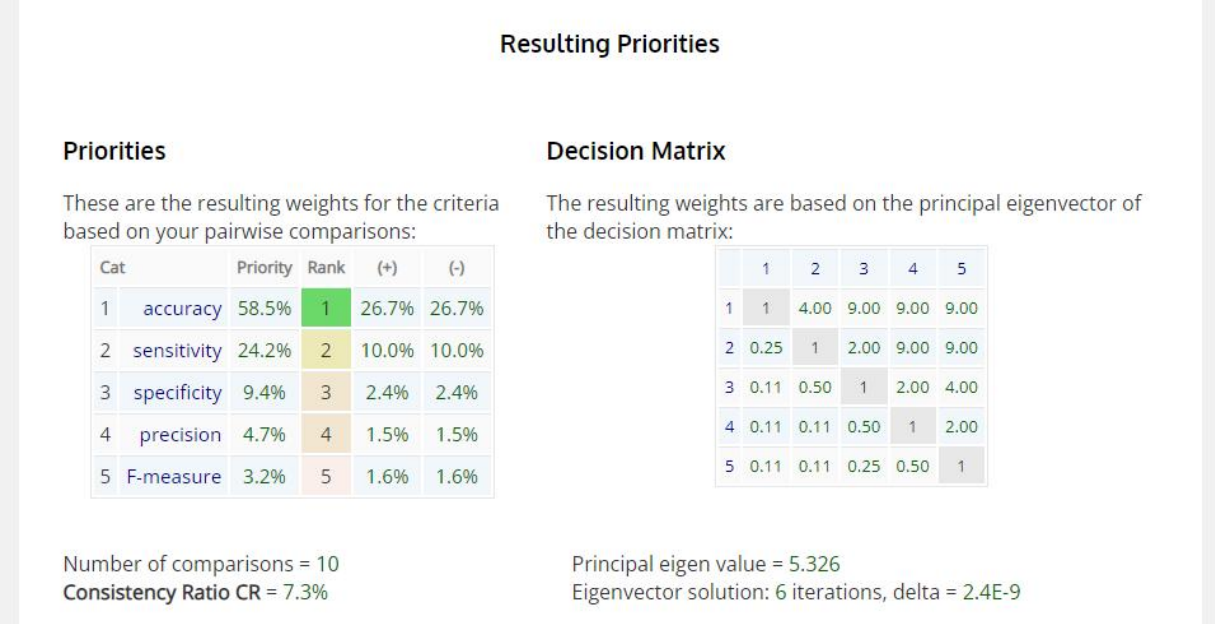
With respect to *AHP priorities*, which criterion is more important, and how much more on a scale 1 to 9?

	A - wrt AHP priorities - or B?		Equal	How much more?
1	◉ accuracy	○ sensitivity	○ 1	○ 2 ○ 3 ◉ 4 ○ 5 ○ 6 ○ 7 ○ 8 ○ 9
2	◉ accuracy	○ specificity	○ 1	○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 ○ 8 ◉ 9
3	◉ accuracy	○ precision	○ 1	○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 ○ 8 ◉ 9
4	◉ accuracy	○ F-measure	○ 1	○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 ○ 8 ◉ 9
5	◉ sensitivity	○ specificity	○ 1	◉ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 ○ 8 ○ 9
6	◉ sensitivity	○ precision	○ 1	○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 ○ 8 ◉ 9
7	◉ sensitivity	○ F-measure	○ 1	○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 ○ 8 ◉ 9
8	◉ specificity	○ precision	○ 1	◉ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 ○ 8 ○ 9
9	◉ specificity	○ F-measure	○ 1	○ 2 ○ 3 ◉ 4 ○ 5 ○ 6 ○ 7 ○ 8 ○ 9
10	◉ precision	○ F-measure	○ 1	◉ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 ○ 8 ○ 9

CR = 0% OK

Calculate Download_(.csv) ☐ dec. comma

Figure 4.9: AHP Pairwise Comparison Results (Application of Multi-criteria Decision Making Theories in Healthcare and Biomedical Engineering.2021)



Resulting Priorities

Priorities

These are the resulting weights for the criteria based on your pairwise comparisons:

Cat		Priority	Rank	(+)	(-)
1	accuracy	58.5%	1	26.7%	26.7%
2	sensitivity	24.2%	2	10.0%	10.0%
3	specificity	9.4%	3	2.4%	2.4%
4	precision	4.7%	4	1.5%	1.5%
5	F-measure	3.2%	5	1.6%	1.6%

Number of comparisons = 10
Consistency Ratio CR = 7.3%

Decision Matrix

The resulting weights are based on the principal eigenvector of the decision matrix:

	1	2	3	4	5
1	1	4.00	9.00	9.00	9.00
2	0.25	1	2.00	9.00	9.00
3	0.11	0.50	1	2.00	4.00
4	0.11	0.11	0.50	1	2.00
5	0.11	0.11	0.25	0.50	1

Principal eigen value = 5.326
Eigenvector solution: 6 iterations, delta = 2.4E-9

Figure 4.10: AHP Resulting Priorities and Decision Matrix (Application of Multi-criteria Decision Making Theories in Healthcare and Biomedical Engineering.2021)

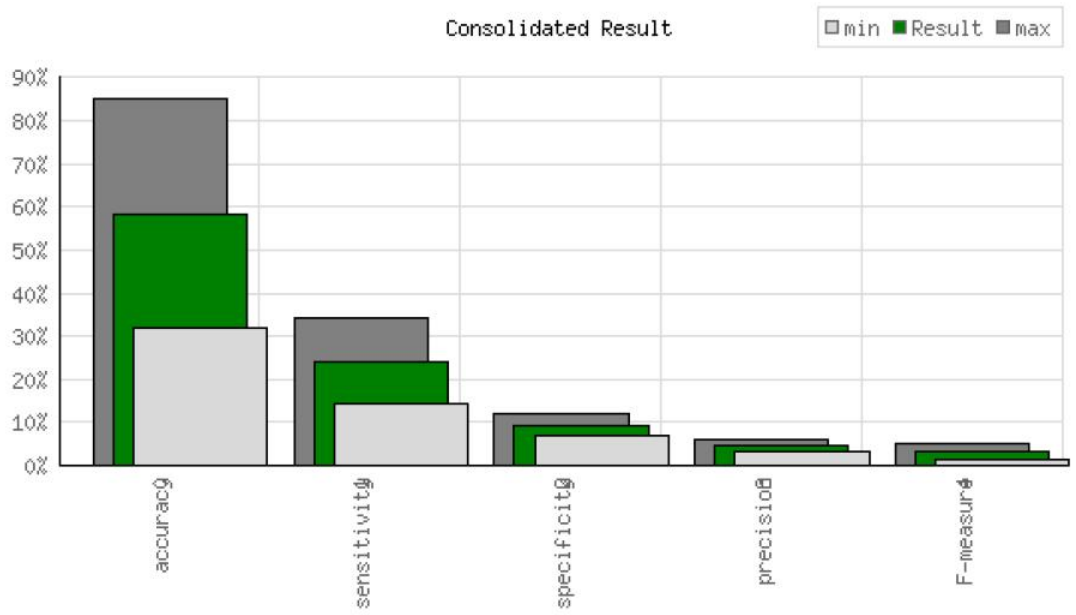


Figure 4.11: AHP Priorities Graph (D.Hally, 2022)

4.2.2.3 Optimal Despondency Symptoms Classifier Using Analytic Hierarchy Process (AHP)

The five metrics were ranked using an AHP model in selecting the best classifier for the despondency symptoms.The result indicated that accuracy is the best metrics that can be used in the selection of the optimal despondency symptoms classifier with 58.5% ranking. Therefore, the five machine learning algorithms performance were verified in relation to the obtainable standard of accuracy percentages and comparison were made on their performances. The consistency ratio (CR) of the metrics was determined as 0.073 as a result of the AHP model, which is lower than Satty's stated empirical limit (0,1). As a result, the subjective judgment can be determined to be consistent. The ranked classification techniques for five different metrics are shown in Figure 4.8 as the final results of the AHP model, as seen in Figures 4.3 to 4.8. The results of the evaluation showed that SVM outperformed other algorithms in all metrics for both hold-out and 10F cross

validation methods followed Logistics Regression which has a close range in all metrics for both hold-out and 10F cross validation methods. This indicated that SVM was the best classifier for despondency symptoms when dealing with multi-class classification for predictive analytics of despondency.

Therefore, SVM machine learning algorithm was recommended for developing despondency symptoms predictive model.

4. 8 Discussion of Findings

Result from the ranking analysis of the performance metrics using AHP model shows that accuracy is the best performance metric for the selection of the optimal class classifier for the multi-classification task on despondency symptoms. This indicates that SVM model several capabilities such as the generalization of the problem, which is the goal in statistical learning. Also, SVM has the ability to easily account for the different types of patterns in the modelling process for the despondency attributes as a multi-classification task. This shows that SVM training algorithm builds a model that predicts the category of the despondency symptoms from social media posts such as non-depressive post, indicative depressive post and depressive post.

Chapter Five

Conclusion

5.1 Summary of Findings

The psycholinguistic library can adequately extract features to determine if a comment or post is depressive, not depressive, or shows an indication of depression.

Support vector machines has the ability to account for different patterns in a multiclassification task, therefore the SVM can build a model that can predict the categories of despondency with a higher accuracy than other Algorithms tested.

Accuracy is the optimal classifier for the selection of the optimal classifier in a multiclassification task on despondency symptoms.

5.2 Conclusion

The objectives of this study was to select the best metrics using the Analytical Hierarchy process(AHP),which was used to determine the optimal classifier for early detection of at risk users of despondency based on five performance metrics2 .The design is executed in four stages which includes the pre-processing stage which entails the extraction of social media posts, data cleaning and data categorization, subsequently the learning phase is processed using a training set on five performance metrics which includes sensitivity, specificity, Accuracy ,precision and f-measure in two validation processes, the 10 fold cross validation and the 80:20 hold out method. Furthermore the validation processes were executed on five algorithms to determine the most optimal algorithm and the optimal classifier based on the performance metrics.

5.3 Recommendations

One of the most pressing issues in the world over is Mental illness, as it is termed a "pandemic"1.The process of monitoring the social media posts of individuals on various platforms such as twitter and Facebook, has been beneficial several measures have been taken to curtail hate speech, intent to cause violence and wrong information. Consequently this deals with a larger audience as compared to indivualistic opinions and response on social media, especially in an institution such as workplace and academic institutions. This thesis can enable quick interventions in providing adequate care before an individual gets to full scale depression.

. 1) The optimal algorithm can be used to build a model that can predict despondency symptoms from social media posts as it has been determined to have attributes for multi-classification task.

2) The algorithm can be made to process emojis as word replacement by incorporating the library of emojis in the psycholinguistic library in the NLPtool kit

5.4 Contribution to Knowledge

In view of the current realities of the world and the ever changing dynamics of society, multitude of persons extending to a wide range of age groups tend to express their views and opinions on social media for relevance and as change agents, this thesis provides a pathway to determine the meaning of words posted on social media in relation to the mental state of individuals. Specifically this thesis focused on at-risk(indicative depression) users as a call to action and a

tool for quick intervention in a mental illness pandemic, Likewise this will also enable institutions to monitor the mental state of persons by their social media posts.

5.5 Suggested Areas for Further Research

The results of this thesis can be implemented in future works by

1. Creating a model to be embedded into a local social media platform to run through comments in real time for despondency indicators using the optimal classifier and algorithm The optimal algorithm can be used to build a model that can predict despondency symptoms from social media posts as it has been determined to have attributes for multi-classification task.

2. To connect the system above to a tollfree help and counselling centre to administer quick intervention procedures to an individual having despondency symptoms.

3. The algorithm can be made to process emojis as word replacement by incorporating the library of emojis in the psycholinguistic library in the NLPtool kit

Biblography

Conference Proceedings

Ahmed H. O., Prasadith B., Mahmoud H. O., & Diana I. *Deep learning for depression detection of Twitter users.* **In Proceedings of the Fifth Workshop on Computational Linguistics and Clinical Psychology**: From Keyboard to Clinic, New Orleans, LA. Association for Computational Linguistics, 2018, pages 88– 97.

Bhavya, S & Anitha S P. "*Prediction models in healthcare using deep learning." In* **International Conference on soft computing and pattern Recognition**, Springer, Cham, 2019. pp. 195-204

Cervantes J., Garcia-L.F, Rodríguez M, & Lopez A. "*A comprehensive survey on support vector machine classification: Applications, challenges and trends."***Neurocomputing** 2020 408: 189-215.

De C., Counts M., S. & Hovitz E,"*Social media as a measurement tool of depression in populations.* **In proceedings of the 5th Annual ACM Web Science Conference,**2015 pp 47-56.ACM

Faruque M., Faisal D., & Iqbal H. S. *"Performance analysis of machine learning techniques to predict diabetes mellitus."* **In 2019 International Conference on Electrical, Computer and Communication Engineering (ECCE),** IEEE, 2019. pp. 1-4.

Hegde S., Santosh G, Shivakumar M, & Srihari R. "*User Interest Prediction based on Social Network Profile with Machine Learning."* **In 2021 6th international conference for convergence in Technology (I2CT),** 2021, pp. 1-6. IEEE.

Islam R. M., Raihan M. Kamal S. N, Islam R, & Ali M. A. "*Detecting depression using k-nearest neighbors (knn) classification technique.*" **In 2018 international conference on computer, communication, chemical, material and electronic engineering**. (IC4ME2), IEEE, 2018, pp. 1-4.

Kumar, U. V., Savithri. A., Jhansi M., Priya M., Bindu J, & Divya N. L. *"Finding Psychological Instability Using Machine Learning."* **In 2020 7th International Conference on Smart Structures and Systems (ICSSS),** IEEE, 2020 pp. 1-4.

Lavanya P. & Sasikala E. "*Deep learning techniques on text classification using Natural language processing (NLP) in social healthcare network:* A comprehensive survey*." In 2021* **international conference on signal processing and communication**, IEEE, 2021, pp. 603-609

Maheshwari A., N., & Daniel A. D. "*A comparative study of machine learning techniques for aviation applications."* **In *2018 Aviation Technology, Integration, and Operations Conference*,** 2018 p. 3980.

Mohammed S. A., Sadeq D., Salah A, & Gunter S. "*Analysis of breast cancer detection using different machine learning techniques*." **In International Conference on Data Mining and Big Data,** Springer, Singapore, 2020, pp. 108-117.

Mustafa R., Noman A, Fahad S.A, Javed F, Basit S, & Alexander G. "*A multiclass depression detection in social media based on sentiment analysis."* **In 17th International Conference on Information Technology–New Generations ITNG 2020***,* Springer, Cham, 2020. pp. 659-662.

Obulesu O., Mahendra M, & ThrilokReddy M. "*Machine learning techniques and tools***: A survey." In 2018 International Conference on Inventive Research in Computing Applications (ICIRCA),** IEEE, 2018 pp. 605-611.

Raj S., & Sarfaraz M. "*Analysis and detection of autism spectrum disorder using machine learning techniques.*" **Procedia Computer Science** 2020: vol 167 994-1004.

Ray S. "*A quick review of machine learning algorithms."* **In 2019 International conference on machine learning, big data, cloud and parallel computing (COMITCon)**, IEEE, 2019, pp. 35-39.

Ríssola A., Esteban O., Bahrainian A, & Crestani F. *"A dataset for research on depression in social media."* In **proceedings of the 28th ACM conference on user modeling and personalization**, 2020, pp. 338-342.

Singh S., & Sandhya T. "*Mental illness diagnosis from social network data using effective machine learning technique*." **In Artificial Intelligence and Speech Technology, CRC Press,** 2021. pp. 231-238.

Walkowiak T., Datko S., & Maciejewski H. "*Bag-of-words, bag-of-topics and word-to-vec based subject classification of text documents -a comparative study***." In International Conference on Dependability and Complex Systems***,*** Springer, Cham, 2018 ; pp526-535

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

Journals

Akakandelwa A, & Walubita,; *Students' Social Media Use and its Perceived Impact on Students' Social Media Use and the Perceived Impact on their Social Life:* A Case Study of the University of Zambia. **The International Journal of Multi-Disciplinary Research**, 2018

Akindehin, F. & Akindehin M. *Online social networking practices of some Nigerian university undergraduates:* Implications for counseling**. British Journal of Arts and Social Sciences**, 3(1), 2011, pp. 68-78

Albahri A. S., Rula A., Hamid Z. T., Al-qays A. A., Zaidan B. B., Zaidan A., Albahri, A. H. & AlAmoodi. "*Role of biological data mining and machine learning techniques in detecting and diagnosing the novel coronavirus (COVID-19)*: a systematic review." **Journal of medical systems**2020:no 7,vol 44 pp1-11.

Arunadevi J. S., Ramya M & Ramesh R.," *A study of classification Algorithm using Rapidmine*r" **International Journal of Pure and Applied Mathematics**,2018 no 2,pp119; 15977-15988

Bakken S. "*Predictive models: important problems and innovative methods."* **Journal of the American Medical Informatics Association** 2022:no 1,vol 29 1-2.

Basit W., Elbadawi M., Simon G., & Abdul "*Advanced machine-learning techniques in drug discovery."* Drug Discovery Today 2021,no 3,vol 26: 769-777.

Bech P. "*Rating scales in depression: limitations and pitfalls.*" **Dialogues in Clinical Neuroscience** 2022.

Berrar D. "*Cross-Validation."Data science Laboratory*, **Tokyo institute of Technology** 2019 pp 2-12,1-S3: 542-545.

Boer M., Gonneke, Stevens W., Catrin F., Margaretha E., Looze, & Regina J. *"Social media use intensity, social media use problems.* **Investigating directionality and mediating processes*." Computers in Human Behavior*** 2021,vol 16,: 106645.

Carvalho, A. F., Firth J, & Vieta E. "*Bipolar disorder.* **"New England journal of medicine** 2020 383, no. 1 :pp 58-66.

Charbuty, B. & Adnan A. "*Classification based on decision tree algorithm for machine learning."* **Journal of Applied Science and Technology Trends** 2021 no. 01 :pp 20-28

Chancellor S., & Choudhury D. "*Methods in predictive techniques for mental health status on social media*: **A Critical Review." NPJ Digital medicine** 2020 pp3, no. 1.

Chandrakala P, Sumithra A, Saranya A, & Bagavathi L. "*Influence of Data Mining Techniques in Healthcare Research.*" **Turkish journal of computer and mathematics Education TURCOMAT** 2021,pp 12, no. 14: 1303-1311.

Coppersmith G., Ryan L, Patrick C, & Alex F. "*Natural language processing of social media as screening for suicide risk.*" **Biomedics informatics insight** 2018: 1178222618792860

Costantini L., Pasquarella C, Odone A, Colucci M, Costanz A, Serafini G, Aguglia A . *"Screening for depression in primary care with Patient Health Questionnaire-9* (PHQ-9): A systematic review.*"* **Journal of affective disorder** 2021: pp473-483,vol 279.

Dhir A., Yossatorn Y., Kaur P, & Chen S. "*Online social media fatigue and psychological wellbeing*— **International Journal of information Management** 2018: pp141-152,vol 40.

Fortgang R. G., & Matthew K. N. "*Ringing the alarm on suicide prevention:* **Psychiatry** 84, 2021: pp192-195,no 2.

Gao S., DVince C., & Sui J. "*Machine learning in major depression: From classification to treatment outcome prediction.”***CNS neuroscience and therapeutics** 2018,vol24, no. 11 : 1037-1052.

Hasan R., Maliha M., & Arifuzzaman A. "*."* *International Conference on Computer, Communication, Chemical, Materials and* **Sentiment Analysis with NLP on Twitter Data Electronic Engineering 2019** (IC4ME2), pp. 1-4. IEEE..

Hatton M., Lewis W. P, Dean M, James C, Simon Gilbody, & Paul A. "*Predicting persistent depressive symptoms in older adults: a machine learning approach to personalised mental healthcare."* **Journal of affective disorders** 2019 pp**.** 246: 857-860.

Henrique B. M, Vinicius A. S., & Herbert K. *"Literature review: Machine learning techniques applied to financial market prediction."* **Expert Systems with Applications,**2019 vol 124 : 226-251.

Houssein E. H., Marwa M. E., Abdelmgeid A. A., & Ponnuthurai N. S. *"Deep and machine learning techniques for medical imaging-based breast cancer:* **A comprehensive review." Expert Systems with Applications** 2021,167: 114161.

Hussain Z., Wegmann E, Yang H, & Montag C. *"Social networks use disorder and associations with depression and anxiety symptoms:* **A systematic review of recent research in China.” Frontiers in psychology** 2020,vol 11: 211.

Ibrahim I., & Adnan A. "*The role of machine learning algorithms for diagnosing diseases."* **Journal of Applied Science and Technology Trends** 2021, no. 01,pp 10-19.

Ibrahim F. A., & Omar A. S. "*Data Mining: WEKA Software (an Overview)."* **Journal of Pure & Applied Sciences** 2019 vol 18, no. 3.

Islam M. D., Muhammad A, Ashir A., Abu R. M., Kamal H. W., & Anwaar U. "*Depression detection from social network data using machine learning techniques***." Health information science and systems** 2018 vol 6, no. 1 2018: 1-12

James C., Powell M, Seixas A, Bateman A, Pengpid S, & Peltzer K. *"Exploring the psychometric properties of the CES-D-10 and its practicality in detecting depressive symptomatology in 27 low-and middle-income countries."* **International Journal of Psychology** 2020, 55, no. 3 435-445

Khan A., Baharum B., Lam H L, & Khairullah K. "*A review of machine learning algorithms for text-documents classification."* **Journal of advances in information technology 1**, 2010 no. 1:pp 4-20.

Kim J. A Zara U., Lee Y, Nasri F H, Gill H, Subramanieapillai, M,Lee R . "*A systematic review of the validity of screening depression through Facebook, Twitter, Instagram, and Snapchat."* **Journal of Affective Disorder** 286; 2021 360-369.

Kwekha R, Ameer S, Heamn N. Abduljabbar, & Bilal A. "*Coronavirus disease (COVID-19) cases analysis using machine-learning applications***." Applied Nanoscience** 2021: 1-13

Laacke S., Mueller R, Schomerus G, & Salloch S. "*Artificial intelligence, social media and depression. A new concept of health-related digital autonomy*." **The American Journal of Bioethics** 2021 vol 21, no. 7 :pp 4-20.

Le G., Aziliz Y. H, Deok-H., Kim-D., Philippe L, Romain B, Taylor C. R & Jonathan M. *"Machine learning and natural language processing in mental health*". **Journal of medical internet research** 2021 **,** vol 23, no. 5: e15708.

Levis B., Benedetti A, & Thombs D. "*Accuracy of Patient Health Questionnaire-9 (PHQ-9) for screening to detect major depression*:**Individual Participant Data Meta-Analysis**" 2019, pp365

Mahesh B. *"Machine learning algorithms-a review***." International Journal of Science and Research (IJSR).** 2020:vol 9 pp 381-386

Maxwell A. E., Timothy A. W., & Fang F. *"Implementation of machine-learning classification in remote sensing:* **An applied review." International Journal of Remote Sensing** 2018 vol *39*, no. 9: 2784-2817

Mohan S., Chandrasegar T, & Gautam S. "*Effective heart disease prediction using hybrid machine learning techniques."* **IEEE access** 2019,vol 7: 81542-81554.

Mumtaz W., & Abdul Q. "*A deep learning framework for automatic diagnosis of unipolar depression.*" **International journal of medical informatics** 132 2019 103983.

Ridout B., & Campbell A. "*The use of social networking sites in mental health interventions for young people:* systematic review.*"* **Journal of medical internet** 20, no. 12 2018 e12244.

Riekert M., Riekert M, & Klein A. "*Simple Baseline Machine Learning Text Classifiers for Small Datasets."* **SN Computer Science**, 2021:no 3 pp1-16.

Sevakula R. K, Wan-Tai M. Au-Yeung, Jagmeet P. Singh, E. Kevin H., Eric M. I., & Antonis A. A. *"State-of-the-art machine learning techniques aiming to improve patient outcomes pertaining to the cardiovascular system*."**Journal of the American Heart Association**,2020 vol 9 no. 4: e013924.

Shatte A., Delyse B. R, Hutchinson M, Matthew F., & Samantha T. "*Social media markers to identify fathers at risk of postpartum depression*." 2019.

Shensa A., Jaime, Sidani E, Amanda M, Escobar V, & Primack E. "*Social media use and depression and anxiety symptoms: A cluster analysis."* **American journal of health behaviour vol**42, no. 2 2018: 116-128

Sun H., Henry V. B, & Honglan H. "*Machine learning applications for building structural design and performance assessment:* state-of-the-art review." **Journal of Building Engineering** 2021 vol 33: 101816.

Tadesse M., Hongfei L., Bo X., & Liang Y. "*Detection of depression-related posts in reddit social media forum.*" **IEEE Access** 2019 vol7: 44883-44893

Tate A. E, Ryan C, McCabe, Larsson H, Lundström S, Lichtenstein P, & Halkola R. "*Predicting mental health problems in adolescence using machine learning techniques."* **Plos one** 2020 vol 15, no. 4: e0230389.

Thapa R, & Subedi S. *"Social media and depression."* **Journal of psychiatrists Association of Nepal**, 2018: no 2 pp1-4.

Valdilo M, Curcin V" *Modelling depression Symptoms from social media through multiple instance learning"* **Research Gate** May 2019.

Watkins, L. E, Kelsey R. S., & Rothbaum. O. *"Treating PTSD: A review of evidence-based psychotherapy interventions.* **"Frontiers in behavioural neuroscience** 2018 vol 12 :pp 258

Wong Y. H, Yi M., Marc H. N. Potenza, Mung N, Chan M, Man L W T, Kwan C T, Pakpour A, & Lin C. "*Relationships between severity of internet gaming disorder,and severity of problematic social media use.* **International journal of environmental research and public health** 2020 vol 17, no. 6 ; 1879.

Xie Z., Tanner R, Catherine L. S., & Marlow M. "*Association of functional disability with mental health services use and perceived unmet needs"* **Journal of affective disorders** 2022 vol 299,:pp 449-455

Textbooks

Abdar M., Wojciech K., U. Rajendra Acharya, Ru-San Tan, Vladimir M. & Paweł P. "*A new machine learning technique for an accurate diagnosis of coronary artery disease*." **Computer Methods and Programs in Biomedicine** 179 2019 pp179: 104992

AlSagri, H. S & Ykhlef M. "*Machine learning-based approach for depression detection.* **IEICE Transactions on information systems** 103, no. 8 2020: 1825-1832.

Ayodele, Taiwo O. "*Types of machine learning algorithms*." **New Advances in Machine Learning** 3, 2010,3 pp. 19-48.

Cornn K" *Identifying Depression on Social Media*" **Department of Statistics, Stanford University CS22N CS229A**. Jan 2021

Farland G. "Morningside Hospital: *A Historical Case Study of the Diagnosis and Treatment of Depression.* **University of Oregon Libraries, Scholars** Bank 2019.

Fortgang R. G., & Matthew K. N. "*Ringing the alarm on suicide prevention,* **Psychiatry,** 2021:84,no 2 192-195.

Gao S., DVince C., & Sui J. "*Machine learning in major depression"***CNS neuroscience and therapeutics** 2018:24,no11 1037-1052.

Shikka J,Kavita P,Princi J&,Kah Phooi S. *"Multimodal depression detection using machine learning."* **In Artificial intelligence, Machine learning and Mental Health in Pandemics Academic Press**. 2022 pp. 53-72.ISBN 9780323915557

Jalali A., Sherbino J, Frank J, & Sutherland S. *"Social media and medical education: Twitter as a learning tool."* **International review of psychiatry** 27, no. 2 2015: 140-146.

Jiang Y. H, Gu Q., & Huang J. *Machine learning ,its algorithm and development analysis* [J]. **Information and Computer Science (Theoretical Edition)**, *2019*

Jiang, H., Hu B, Liu Z, Wang G, Zhang L, Li X, & Kang H. *"Detecting depression using an ensemble logistic regression model..* **"Computational and mathematical methods in medicine** 2018

Kadhim A. I. *"Survey on supervised machine learning techniques for automatic text classification."* **Artificial Intelligence Review** 2019 *52*, no.1:pp 273-292.

Kennis M., Gerritsen L, Marije V. D., Williams A, Cuijpers P & Bockting C. "*Prospective biomarkers of major depressive disorder:* **"Molecular Psychiatry** 25, no. 2 2020: 321-338.

Kulkarni, Akshay B K. "*Early Detection of Depression.*" *PhD* **dissertation. Department of computer Science, Univesidad Nacional de Colombia** 2018.

Li Z. *Development of machine learning and several learning methods.* **Industry and Science Forum,** 2018, 15 (10): 198-199.

Lin Y., Liu E. Jaime, Shensa A, Radovic A, Miller E, Jason B, Colditz, . Hoffman L, Leila M G., & Primack A. *"Association between social media use and depression among US young adults.*" **Depression and anxiety** 2016,vol 33,no 4:pp 323-331.

Naser M. Z. & Amir H. A. "*Error metrics and performance fitness indicators for artificial intelligence and machine learning.* **Architecture, Structures and Construction,** 2021:pp 1-19.

Ngiam K. Y, & Wei K. *"Big data and machine learning algorithms for health-care delivery."* **The Lancet Oncology** 2019:pp 20,no 5 e262-e273.

Patinge S., & Vijaya K. *A survey on mental health Prediction using social media network.* **EasyChair**, 2020.no 3521

Rasheed F., & Abdul W. *"Learning style detection in E-learning systems using machine learning techniques.*" **Expert Systems with Applications**, 2021:pp174; 114774

Tsai T. H., Chang H. T. Chang Y. C., Chang Y.S.; *Personality disclosure on social network sites,.* **Computers in Human Behavior**, November 2017, pp 469-482.vol 76

Tsui F. R., Lingyun S., Victor R, Neal D, Candice B, Satish I, Colin G. & Brent A. *"Natural language processing and machine learning of electronic health records."* **JAMIA** 2021:open 4,no 1 ooab011

.

Xu, H., Jian Z., Panagiotis G., Danial J, & Mahmood M. "*Supervised machine learning techniques to the prediction of tunnel boring machine.*" **Applied sciences** , 2019:no 18 vol 9 ;3715.

Yasin S., Hussain S, Aslan S, Raza I, Muzammel M, & Othmani A. "*EEG based Major Depressive disorder and Bipolar disorder detection using Neural Networks*: **A review." Computer Methods and Programs in Biomedicine** 2021, pp 202: 106007.

Yusuf H W, Linu R, Emmanuel A, Evelyn K, Mitchelle N, Lilian S, Garba M. *K,Mental health in Nigeria:* **A Neglected issue in Public Health, Public Health in Practice**. 2021,volume 2 100166,ISSN 2666-5352,

Zhang F., Rao S, Cao H, Zhang X, Wang Q, Yong X , Sun J. "*Genetic evidence suggests posttraumatic stress disorder as a subtype of major depressive disorder."* **The Journal of clinical investigation 2022** pp 132, no. 3

.

Zhang J., Zhong Y., Peng C., & Stefano N. "Emotion recognition using multi-modal data and machine learning techniques*:* **A tutorial and review." Information Fusion.** 2020:pp103-126 vol 59.

Appendix I

Run information For Decision Tree Cross Validation =

== Run information ===

Scheme: weka.classifiers.trees.REPTree -M 2 -V 0.001 -N 3 -S 1 -L -1 -I 0.0
Relation: Data for crossvalidation
Instances: 7145
Attributes: 4
temporal process
linguistic style
emotional process
despondency symptoms
Test mode: 10-fold cross-validation

=== Classifier model (full training set) ===

REPTree
==========

emotional process < 0.01 : Depressive Post (256/0) [127/0]
emotional process >= 0.01
| temporal process < 0.01 : Non-Depressive Post (3/0) [2/0]
| temporal process >= 0.01 : Indicative Depressive Post (4504/0) [2253/0]

Size of the tree : 5

Time taken to build model: 0.02 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances	7145	100 %
Incorrectly Classified Instances	0	0 %
Kappa statistic	1	
Mean absolute error	0	
Root mean squared error	0	
Relative absolute error	0	%
Root relative squared error	0	%
Total Number of Instances	7145	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	Indicative Depressive Post
	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	Depressive Post
	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	Non-Depressive Post
Weighted Avg.	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	

=== Confusion Matrix ===

a	b	c	<-- classified as
6757	0	0	a = Indicative Depressive Post
0	383	0	b = Depressive Post
0	0	5	c = Non-Depressive Post

=== Run information ===

Scheme: weka.classifiers.trees.REPTree -M 2 -V 0.001 -N 3 -S 1 -L -1 -I 0.0
Relation: Data for crossvalidation
Instances: 7145
Attributes: 4
temporal process
linguistic style
emotional process
despondency symptoms
Test mode: split 80.0% train, remainder test

=== Classifier model (full training set) ===

REPTree
==========

emotional process < 0.01 : Depressive Post (256/0) [127/0]
emotional process >= 0.01
| temporal process < 0.01 : Non-Depressive Post (3/0) [2/0]
| temporal process >= 0.01 : Indicative Depressive Post (4504/0) [2253/0]

Size of the tree : 5

Time taken to build model: 0.01 seconds

=== Evaluation on test split ===

Time taken to test model on test split: 0 seconds

=== Summary ===

Correctly Classified Instances	1429	100 %
Incorrectly Classified Instances	0	0 %
Kappa statistic	1	
Mean absolute error	0	
Root mean squared error	0	
Relative absolute error	0 %	
Root relative squared error	0 %	
Total Number of Instances	1429	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	Indicative Depressive Post
	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	Depressive Post
	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	Non-Depressive Post
Weighted Avg.	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	

=== Confusion Matrix ===

a	b	c	<-- classified as
1366	0	0	a = Indicative Depressive Post
0	62	0	b = Depressive Post
0	0	1	c = Non-Depressive Post

=== Run information ===

Scheme: weka.classifiers.lazy.IBk -K 1 -W 0 -A
"weka.core.neighboursearch.LinearNNSearch -A \"weka.core.EuclideanDistance -R first-last\""
Relation: Data for crossvalidation
Instances: 7145
Attributes: 4
temporal process
linguistic style
emotional process
despondency symptoms
Test mode: 10-fold cross-validation

=== Classifier model (full training set) ===

IB1 instance-based classifier
using 1 nearest neighbour(s) for classification

Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances	7145	100 %
Incorrectly Classified Instances	0	0 %
Kappa statistic	1	
Mean absolute error	0	
Root mean squared error	0	
Relative absolute error	0.0464 %	
Root relative squared error	0.0235 %	
Total Number of Instances	7145	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	Indicative Depressive Post
	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	Depressive Post
	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	Non-Depressive Post
Weighted Avg.	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	

=== Confusion Matrix ===

a	b	c	<-- classified as
6757	0	0	a = Indicative Depressive Post
0	383	0	b = Depressive Post
0	0	5	c = Non-Depressive Post

=== Run information ===

Scheme: weka.classifiers.lazy.IBk -K 1 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A \"weka.core.EuclideanDistance -R first-last\""
Relation: Data for crossvalidation
Instances: 7145
Attributes: 4
temporal process
linguistic style
emotional process
despondency symptoms
Test mode: split 80.0% train, remainder test

=== Classifier model (full training set) ===

IB1 instance-based classifier
using 1 nearest neighbour(s) for classification

Time taken to build model: 0 seconds

=== Evaluation on test split ===

Time taken to test model on test split: 0.28 seconds

=== Summary ===

Correctly Classified Instances	1429	100 %
Incorrectly Classified Instances	0	0 %
Kappa statistic	1	
Mean absolute error	0	
Root mean squared error	0.0001	
Relative absolute error	0.0639 %	
Root relative squared error	0.0332 %	
Total Number of Instances	1429	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	Indicative Depressive Post
	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	Depressive Post
	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	Non-Depressive Post
Weighted Avg.	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	

=== Confusion Matrix ===

a	b	c	<-- classified as
1366	0	0	a = Indicative Depressive Post
0	62	0	b = Depressive Post
0	0	1	c = Non-Depressive Post

=== Run information ===

Scheme: weka.classifiers.functions.Logistic -R 1.0E-8 -M -1 -num-decimal-places 4
Relation: DataFOR USE
Instances: 1429
Attributes: 4
temporal process
linguistic style
emotional process
despondency symptoms
Test mode: 10-fold cross-validation

=== Classifier model (full training set) ===

Logistic Regression with ridge parameter of 1.0E-8
Coefficients...

	Class	
Variable	Indicative Depressive Post	Depressive Post
temporal process	1481.7617	1382.2255
linguistic style	-339.176	-426.6349
emotional process	-95.4954	-3277.3768
Intercept	34.186	96.4722

Odds Ratios...

	Class	
Variable	Indicative Depressive Post	Depressive Post
temporal process	Infinity	Infinity
linguistic style	0	0
emotional process	0	0

Time taken to build model: 0.07 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances	1428	99.93 %
Incorrectly Classified Instances	1	0.07 %
Kappa statistic	0.9933	
Mean absolute error	0.0006	

Root mean squared error	0.0222
Relative absolute error	0.8027 %
Root relative squared error	11.7982 %
Total Number of Instances	1429

=== Detailed Accuracy By Class ===

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
1.000	0.013	0.999	1.000	1.000	0.993	0.993	0.999	Indicative Depressive Post
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	Depressive Post
0.000	0.000	?	0.000	?	?	0.641	0.002	Non-Depressive Post
Weighted Avg. 0.999	0.012	?	0.999	?	?	0.993	0.999	

=== Confusion Matrix ===

a	b	c	<-- classified as
1349	0	0	a = Indicative Depressive Post
0	79	0	b = Depressive Post
1	0	0	c = Non-Depressive Post

=== Run information ===

Scheme: weka.classifiers.functions.Logistic -R 1.0E-8 -M -1 -num-decimal-places 4
Relation: Data for crossvalidation
Instances: 7145
Attributes: 4
temporal process
linguistic style
emotional process
despondency symptoms
Test mode: 10-fold cross-validation

=== Classifier model (full training set) ===

Logistic Regression with ridge parameter of 1.0E-8
Coefficients...

	Class	
Variable	Indicative Depressive Post	Depressive Post
temporal process	1867.1589	1778.5783
linguistic style	-130.2484	-80.1366
emotional process	-75.2756	-1914.3391
Intercept	10.6166	40.8483

Odds Ratios...

	Class	
Variable	Indicative Depressive Post	Depressive Post
temporal process	Infinity	Infinity
linguistic style	0	0
emotional process	0	0

Time taken to build model: 0.2 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances 7145 100 %
Incorrectly Classified Instances 0 0 %
Kappa statistic 1
Mean absolute error 0
Root mean squared error 0
Relative absolute error 0 %
Root relative squared error 0.0007 %
Total Number of Instances 7145

=== Detailed Accuracy By Class ===

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	Indicative Depressive Post
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	Depressive Post
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	Non-Depressive Post
Weighted Avg. 1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	

=== Confusion Matrix ===

a	b	c	<-- classified as
6757	0	0	a = Indicative Depressive Post
0	383	0	b = Depressive Post
0	0	5	c = Non-Depressive Post

=== Run information ===

Scheme: weka.classifiers.functions.Logistic -R 1.0E-8 -M -1 -num-decimal-places 4
Relation: Data for crossvalidation
Instances: 7145
Attributes: 4
temporal process
linguistic style
emotional process
despondency symptoms
Test mode: split 80.0% train, remainder test

=== Classifier model (full training set) ===

Logistic Regression with ridge parameter of 1.0E-8
Coefficients...

	Class	
Variable	Indicative Depressive Post	Depressive Post
temporal process	1867.1589	1778.5783
linguistic style	-130.2484	-80.1366
emotional process	-75.2756	-1914.3391
Intercept	10.6166	40.8483

Odds Ratios...

	Class	
Variable	Indicative Depressive Post	Depressive Post
temporal process	Infinity	Infinity
linguistic style	0	0
emotional process	0	0

Time taken to build model: 0.16 seconds

=== Evaluation on test split ===

Time taken to test model on test split: 0 seconds

=== Summary ===

Correctly Classified Instances	1429	100 %
Incorrectly Classified Instances	0	0 %
Kappa statistic	1	
Mean absolute error	0	
Root mean squared error	0	
Relative absolute error	0 %	
Root relative squared error	0.0002 %	
Total Number of Instances	1429	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	Indicative Depressive Post
	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	Depressive Post
	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	Non-Depressive Post
Weighted Avg.	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	

=== Confusion Matrix ===

a	b	c	<-- classified as
1366	0	0	a = Indicative Depressive Post
0	62	0	b = Depressive Post
0	0	1	c = Non-Depressive Post

=== Run information ===

Scheme: weka.classifiers.bayes.NaiveBayes
Relation: Data for crossvalidation
Instances: 7145
Attributes: 4
temporal process
linguistic style
emotional process
despondency symptoms
Test mode: 10-fold cross-validation

=== Classifier model (full training set) ===

Naive Bayes Classifier

Attribute	Class Indicative Depressive Post (0.95)	Depressive Post (0.05)	Non-Depressive Post (0)
temporal process			
mean	0.1327	0.1063	0
std. dev.	0.0716	0.0433	0.0038
weight sum	6757	383	5
precision	0.0229	0.0229	0.0229
linguistic style			
mean	0.0999	0.0682	0.0918
std. dev.	0.0532	0.0368	0.0031
weight sum	6757	383	5
precision	0.0184	0.0184	0.0184

emotional process

mean	0.1146	0	0.1398
std. dev.	0.0665	0.0047	0.0047
weight sum	6757	383	5
precision	0.028	0.028	0.028

Time taken to build model: 0.02 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances	7116	99.5941 %
Incorrectly Classified Instances	29	0.4059 %
Kappa statistic	0.9591	
Mean absolute error	0.0108	
Root mean squared error	0.0643	
Relative absolute error	15.6935 %	
Root relative squared error	34.7241 %	
Total Number of Instances	7145	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	1.000	0.075	0.996	1.000	0.998	0.960	1.000	1.000	Indicative Depressive Post
	0.924	0.000	1.000	0.924	0.961	0.959	1.000	1.000	Depressive Post
	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	Non-Depressive Post
Weighted Avg.	0.996	0.071	0.996	0.996	0.996	0.960	1.000	1.000	

=== Confusion Matrix ===

a	b	c	<-- classified as
6757	0	0	a = Indicative Depressive Post
29	354	0	b = Depressive Post
0	0	5	c = Non-Depressive Post

=== Run information ===

Scheme: weka.classifiers.bayes.NaiveBayes
Relation: Data for crossvalidation
Instances: 7145
Attributes: 4
temporal process
linguistic style
emotional process
despondency symptoms
Test mode: split 80.0% train, remainder test

=== Classifier model (full training set) ===

Naive Bayes Classifier

Attribute	Class Indicative Depressive Post (0.95)	Depressive Post (0.05)	Non-Depressive Post (0)
temporal process			
mean	0.1327	0.1063	0
std. dev.	0.0716	0.0433	0.0038
weight sum	6757	383	5
precision	0.0229	0.0229	0.0229
linguistic style			
mean	0.0999	0.0682	0.0918
std. dev.	0.0532	0.0368	0.0031
weight sum	6757	383	5
precision	0.0184	0.0184	0.0184
emotional process			
mean	0.1146	0	0.1398
std. dev.	0.0665	0.0047	0.0047
weight sum	6757	383	5
precision	0.028	0.028	0.028

Time taken to build model: 0 seconds

=== Evaluation on test split ===

Time taken to test model on test split: 0 seconds

=== Summary ===

Correctly Classified Instances	1425	99.7201 %
Incorrectly Classified Instances	4	0.2799 %
Kappa statistic	0.9658	
Mean absolute error	0.0083	
Root mean squared error	0.055	
Relative absolute error	12.9624 %	
Root relative squared error	32.7532 %	
Total Number of Instances	1429	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	1.000	0.063	0.997	1.000	0.999	0.966	1.000	1.000	Indicative Depressive Post
	0.935	0.000	1.000	0.935	0.967	0.966	1.000	1.000	Depressive Post
	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	Non-Depressive Post
Weighted Avg.	0.997	0.061	0.997	0.997	0.997	0.966	1.000	1.000	

=== Confusion Matrix ===

a	b	c	<-- classified as
1366	0	0	a = Indicative Depressive Post
4	58	0	b = Depressive Post
0	0	1	c = Non-Depressive Post

=== Run information ===

Scheme: weka.classifiers.functions.SMO -C 1.0 -L 0.001 -P 1.0E-12 -N 0 -V -1 -W 1 -K "weka.classifiers.functions.supportVector.PolyKernel -E 1.0 -C 250007" -calibrator "weka.classifiers.functions.Logistic -R 1.0E-8 -M -1 -num-decimal-places 4"
Relation: Data for crossvalidation
Instances: 7145
Attributes: 4
temporal process
linguistic style
emotional process
despondency symptoms
Test mode: 10-fold cross-validation

=== Classifier model (full training set) ===

SMO

Kernel used:
Linear Kernel: K(x,y) = <x,y>

Classifier for classes: Indicative Depressive Post, Depressive Post

BinarySMO

Machine linear: showing attribute weights, not support vectors.

-0.001 * (normalized) temporal process
+ -0.0008 * (normalized) linguistic style
+ -22.5294 * (normalized) emotional process
+ 0.3257

Number of kernel evaluations: 19682 (48.147% cached)

Classifier for classes: Indicative Depressive Post, Non-Depressive Post

BinarySMO

Machine linear: showing attribute weights, not support vectors.

-0.1786 * (normalized) temporal process
+ 0.0041 * (normalized) linguistic style
+ 0.0021 * (normalized) emotional process
- 0.9946

Number of kernel evaluations: 5773 (55.718% cached)

Classifier for classes: Depressive Post, Non-Depressive Post

BinarySMO

Machine linear: showing attribute weights, not support vectors.

```
      -0.2809 * (normalized) temporal process
+      0.0686 * (normalized) linguistic style
+      1.4706 * (normalized) emotional process
-      0.995
```

Number of kernel evaluations: 117 (46.083% cached)

Time taken to build model: 0.1 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances	7140	99.93 %
Incorrectly Classified Instances	5	0.07 %
Kappa statistic	0.9931	
Mean absolute error	0.2225	
Root mean squared error	0.2727	
Relative absolute error	323.9275 %	
Root relative squared error	147.3461 %	
Total Number of Instances	7145	

=== Detailed Accuracy By Class ===

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
1.000	0.013	0.999	1.000	1.000	0.993	0.994	0.999	Indicative Depressive Post
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	Depressive Post
0.000	0.000	?	0.000	?	?	0.499	0.001	Non-Depressive Post
Weighted Avg. 0.999	0.012	?	0.999	?	?	0.994	0.999	

=== Confusion Matrix ===

```
   a    b    c   <-- classified as
6757    0    0 |    a = Indicative Depressive Post
   0  383    0 |    b = Depressive Post
```

```
  5    0    0 |    c = Non-Depressive Post
```

```
=== Run information ===

Scheme:       weka.classifiers.functions.SMO -C 1.0 -L 0.001 -P 1.0E-12 -N 0 -V -1 -W 1 -K
"weka.classifiers.functions.supportVector.PolyKernel -E 1.0 -C 250007" -calibrator
"weka.classifiers.functions.Logistic -R 1.0E-8 -M -1 -num-decimal-places 4"
Relation:     Data for crossvalidation
Instances:    7145
Attributes:   4
              temporal process
              linguistic style
              emotional process
              despondency symptoms
Test mode:    split 80.0% train, remainder test

=== Classifier model (full training set) ===

SMO

Kernel used:
  Linear Kernel: K(x,y) = <x,y>

Classifier for classes: Indicative Depressive Post, Depressive Post

BinarySMO

Machine linear: showing attribute weights, not support vectors.

        -0.001  * (normalized) temporal process
+       -0.0008 * (normalized) linguistic style
+      -22.5294 * (normalized) emotional process
+       0.3257
```

Number of kernel evaluations: 19682 (48.147% cached)

Classifier for classes: Indicative Depressive Post, Non-Depressive Post

BinarySMO

Machine linear: showing attribute weights, not support vectors.

```
      -0.1786 * (normalized) temporal process
+      0.0041 * (normalized) linguistic style
+      0.0021 * (normalized) emotional process
-      0.9946
```

Number of kernel evaluations: 5773 (55.718% cached)

Classifier for classes: Depressive Post, Non-Depressive Post

BinarySMO

Machine linear: showing attribute weights, not support vectors.

```
      -0.2809 * (normalized) temporal process
+      0.0686 * (normalized) linguistic style
+      1.4706 * (normalized) emotional process
-      0.995
```

Number of kernel evaluations: 117 (46.083% cached)

Time taken to build model: 0.02 seconds

=== Evaluation on test split ===

Time taken to test model on test split: 0 seconds

=== Summary ===

Correctly Classified Instances	1428	99.93 %
Incorrectly Classified Instances	1	0.07 %
Kappa statistic	0.9916	
Mean absolute error	0.2225	
Root mean squared error	0.2727	
Relative absolute error	346.6237 %	
Root relative squared error	162.3304 %	
Total Number of Instances	1429	

=== Detailed Accuracy By Class ===

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
1.000	0.016	0.999	1.000	1.000	0.992	0.992	0.999	Indicative Depressive Post
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	Depressive Post
0.000	0.000	?	0.000	?	?	0.499	0.001	Non-Depressive Post
Weighted Avg. 0.999	0.015	?	0.999	?	?	0.992	0.999	

=== Confusion Matrix ===

a	b	c	<-- classified as
1366	0	0	a = Indicative Depressive Post
0	62	0	b = Depressive Post
1	0	0	c = Non-Depressive Post

Appendix II

Token				
addition	7	7	0.45692	5.38842
people	856	484	31.59269	1.15224
bp	125	82	5.35248	2.92761
weak	23	20	1.30548	4.3386
concept	9	6	0.39164	5.54257
deep	71	67	4.37337	3.12964
inside	51	49	3.19843	3.44251
wrong	122	105	6.85379	2.68037
intrinsically	1	1	0.06527	7.33433
period	58	42	2.74151	3.59666
dissociation	2	2	0.13055	6.64118
difficulty	9	9	0.58747	5.1371
maintain	15	14	0.91384	4.69527
healthy	79	69	4.50392	3.10022
sense	45	39	2.54569	3.67077
esteem	8	8	0.52219	5.25489
system	67	57	3.72063	3.29128
find	475	366	23.89034	1.4317
difficult	95	80	5.22193	2.9523
time	997	632	41.25326	0.88544
adhere	2	1	0.06527	7.33433
real	103	91	5.93995	2.82347
issue	245	178	11.6188	2.15255
trust	120	99	6.46214	2.73921

black	23	16	1.04439	4.56174
white	18	14	0.91384	4.69527
decision	62	52	3.39426	3.38309
leave	144	114	7.44125	2.59813
terrible	24	23	1.50131	4.19884
evil	12	11	0.71802	4.93643
mistake	20	14	0.91384	4.69527
life	762	468	30.5483	1.18586
normal	132	97	6.33159	2.75962
person	299	223	14.55614	1.92716
see	453	335	21.86684	1.5202
oftentime	1	1	0.06527	7.33433
internalize	1	1	0.06527	7.33433
personal	31	29	1.89295	3.96703
failure	31	25	1.63185	4.11545
happen	192	167	10.90078	2.21634
awful	52	45	2.93734	3.52767
mean	167	146	9.53003	2.35072
fluid	7	6	0.39164	5.54257
mind	231	176	11.48825	2.16385
sufferer	34	21	1.37076	4.28981
single	39	36	2.34987	3.75081
symptom	144	113	7.37598	2.60694
diagnosis	112	90	5.87467	2.83452
pretty	72	67	4.37337	3.12964
start	368	289	18.86423	1.6679
common	33	29	1.89295	3.96703
face	83	67	4.37337	3.12964
intense	30	25	1.63185	4.11545
therapy	216	171	11.16188	2.19267
change	251	199	12.98956	2.04102
mature	4	4	0.2611	5.94803
realistic	4	4	0.2611	5.94803
viewpoint	1	1	0.06527	7.33433
rebuild	6	6	0.39164	5.54257
aberrant	1	1	0.06527	7.33433
build	26	24	1.56658	4.15628
ability	36	34	2.21932	3.80797
learn	206	156	10.18277	2.28447
cope	133	111	7.24543	2.6248
mechanism	12	12	0.78329	4.84942
dbt	27	18	1.17493	4.44396
dialectic	2	2	0.13055	6.64118

behavioral	7	5	0.32637	5.72489
develop	21	20	1.30548	4.3386
initially	8	8	0.52219	5.25489
show	103	89	5.8094	2.84569
positive	103	82	5.35248	2.92761
result	31	30	1.95822	3.93313
technique	24	20	1.30548	4.3386
live	389	269	17.55875	1.73962
fairly	15	15	0.97911	4.62628
bipolar	587	402	26.24021	1.33788
comorbid	1	1	0.06527	7.33433
meaning	23	20	1.30548	4.3386
necessarily	10	10	0.65274	5.03174
genetic	13	10	0.65274	5.03174
link	16	14	0.91384	4.69527
begin	73	66	4.30809	3.14467
suffering	68	56	3.65535	3.30898
untreated	5	5	0.32637	5.72489
tend	31	23	1.50131	4.19884
extreme	75	67	4.37337	3.12964
emotion	126	98	6.39687	2.74936
moody	4	4	0.2611	5.94803
hypersensitive	2	2	0.13055	6.64118
criticism	5	4	0.2611	5.94803
untrust	1	1	0.06527	7.33433
self-damage	4	3	0.19582	6.23572
high	153	121	7.89817	2.53854
low	132	108	7.04961	2.6522
behavior	65	51	3.32898	3.4025
love	510	322	21.01828	1.55978
chaos	8	8	0.52219	5.25489
brain	140	110	7.18016	2.63385
work	812	507	33.09399	1.10582
correctly	13	12	0.78329	4.84942
due	77	70	4.56919	3.08583
chemical	48	35	2.2846	3.77898
imbalance	29	24	1.56658	4.15628
perception	3	3	0.19582	6.23572
base	23	22	1.43603	4.24329
distrust	1	1	0.06527	7.33433
fear	97	80	5.22193	2.9523
abandon	10	8	0.52219	5.25489
reject	10	10	0.65274	5.03174

problem	183	154	10.05222	2.29738
relationship	133	99	6.46214	2.73921
friendship	10	9	0.58747	5.1371
family	323	237	15.46997	1.86627
coworker	4	4	0.2611	5.94803
end	218	182	11.8799	2.13032
hospital	138	107	6.98433	2.6615
suicidal	123	102	6.65796	2.70936
ideation	17	15	0.97911	4.62628
mania	115	65	4.24282	3.15994
depress	798	535	34.92167	1.05206
uncommon	8	8	0.52219	5.25489
proper	27	22	1.43603	4.24329
correct	21	19	1.24021	4.38989
treatment	101	81	5.28721	2.93988
plan	77	59	3.85117	3.25679
kick	18	18	1.17493	4.44396
early	46	44	2.87206	3.55014
misdiagnosis	2	2	0.13055	6.64118
lead	50	42	2.74151	3.59666
poor	17	16	1.04439	4.56174
management	10	10	0.65274	5.03174
relapse	9	9	0.58747	5.1371
rollar	2	2	0.13055	6.64118
coaster	12	11	0.71802	4.93643
peak	3	3	0.19582	6.23572
mood	210	165	10.77023	2.22838
rapidly	7	6	0.39164	5.54257
alternate	2	2	0.13055	6.64118
longwr	1	1	0.06527	7.33433
shift	14	11	0.71802	4.93643
day	782	493	32.18016	1.13382
ect	26	21	1.37076	4.28981
anger	46	37	2.41514	3.72341
borderline	39	35	2.2846	3.77898
personality	48	41	2.67624	3.62076
tilt	1	1	0.06527	7.33433
wirl	1	1	0.06527	7.33433
secerap	1	1	0.06527	7.33433
direction	16	15	0.97911	4.62628
impulsivness	1	1	0.06527	7.33433
major	59	58	3.7859	3.27389
sound	119	106	6.91906	2.67089

judgement	8	8	0.52219	5.25489
mention	41	40	2.61097	3.64545
reation	1	1	0.06527	7.33433
situation	144	125	8.15927	2.50602
commonly	3	3	0.19582	6.23572
deem	4	4	0.2611	5.94803
inappropriate	7	7	0.45692	5.38842
suffer	150	128	8.35509	2.4823
stable	45	43	2.80679	3.57313
struggle	291	245	15.99217	1.83307
long	251	204	13.31593	2.01621
answer	84	69	4.50392	3.10022
disease	58	43	2.80679	3.57313
treat	116	94	6.13577	2.79103
medication	483	300	19.58225	1.63055
typically	8	7	0.45692	5.38842
compensate	2	2	0.13055	6.64118
unhealthy	7	6	0.39164	5.54257
illegal	7	7	0.45692	5.38842
drug	132	106	6.91906	2.67089
eat	95	79	5.15666	2.96488
disorder	320	214	13.96867	1.96835
gamble	2	2	0.13055	6.64118
promiscuity	1	1	0.06527	7.33433
alcoholism	1	1	0.06527	7.33433
harm	82	69	4.50392	3.10022
assist	6	5	0.32637	5.72489
overcome	26	23	1.50131	4.19884
state	106	89	5.8094	2.84569
damage	44	39	2.54569	3.67077
lability	1	1	0.06527	7.33433
response	14	13	0.84856	4.76938
potentially	5	5	0.32637	5.72489
hypomania	21	13	0.84856	4.76938
severe	124	108	7.04961	2.6522
energy	71	57	3.72063	3.29128
happy	158	128	8.35509	2.4823
reason	104	95	6.20104	2.78045
grandiose	3	3	0.19582	6.23572
idea	77	72	4.69974	3.05766
themself	3	2	0.13055	6.64118
sleep	251	170	11.09661	2.19853
sex	77	43	2.80679	3.57313

engage	14	12	0.78329	4.84942
risky	5	5	0.32637	5.72489
consider	52	48	3.13316	3.46313
consequence	7	6	0.39164	5.54257
incapable	9	9	0.58747	5.1371
care	337	258	16.84073	1.78137
self-harm	4	4	0.2611	5.94803
hard	449	330	21.54047	1.53524
house	73	63	4.11227	3.19119
bed	110	91	5.93995	2.82347
depend	29	28	1.82768	4.00212
type	101	69	4.50392	3.10022
fall	61	55	3.59008	3.327
spectrum	12	11	0.71802	4.93643
determine	13	12	0.78329	4.84942
serious	34	30	1.95822	3.93313
occur	19	13	0.84856	4.76938
episode	111	91	5.93995	2.82347
constantly	56	52	3.39426	3.38309
counterbalance	1	1	0.06527	7.33433
offset	3	3	0.19582	6.23572
function	53	44	2.87206	3.55014
huge	54	51	3.32898	3.4025
abnormal	5	5	0.32637	5.72489
perceive	10	7	0.45692	5.38842
interact	6	6	0.39164	5.54257
dirty	4	4	0.2611	5.94803
filter	3	3	0.19582	6.23572
incorrectly	2	1	0.06527	7.33433
react	32	27	1.7624	4.03849
inappropriately	1	1	0.06527	7.33433
room	45	40	2.61097	3.64545
deviation	1	1	0.06527	7.33433
regulation	2	2	0.13055	6.64118
unexpected	2	2	0.13055	6.64118
reactive	1	1	0.06527	7.33433
childhood	16	16	1.04439	4.56174
experience	212	177	11.55352	2.15818
parent	93	65	4.24282	3.15994
caregiver	6	4	0.2611	5.94803
teacher	13	12	0.78329	4.84942
peer	10	10	0.65274	5.03174
propensity	1	1	0.06527	7.33433

short-lived	1	1	0.06527	7.33433
lover	7	2	0.13055	6.64118
cling	3	2	0.13055	6.64118
minute	50	43	2.80679	3.57313
push	114	89	5.8094	2.84569
infidelity	2	1	0.06527	7.33433
abandonment	8	6	0.39164	5.54257
rejection	1	1	0.06527	7.33433
idealize	2	2	0.13055	6.64118
villainize	1	1	0.06527	7.33433
stand	45	38	2.48042	3.69674
adore	3	3	0.19582	6.23572
eventually	47	43	2.80679	3.57313
fed	6	5	0.32637	5.72489
lack	31	30	1.95822	3.93313
reinforce	2	2	0.13055	6.64118
belief	11	11	0.71802	4.93643
left	83	76	4.96084	3.0036
alone	240	198	12.92428	2.04606
cycle	46	36	2.34987	3.75081
afterward	2	2	0.13055	6.64118
note	27	24	1.56658	4.15628
color	21	16	1.04439	4.56174
regard	17	16	1.04439	4.56174
seed	4	4	0.2611	5.94803
household	3	3	0.19582	6.23572
seemingly	4	4	0.2611	5.94803
dote	1	1	0.06527	7.33433
child	216	146	9.53003	2.35072
obvious	5	5	0.32637	5.72489
inconsistent	3	2	0.13055	6.64118
inattentive	2	2	0.13055	6.64118
cold	22	21	1.37076	4.28981
neglectful	1	1	0.06527	7.33433
absentee	1	1	0.06527	7.33433
big	108	93	6.0705	2.80173
prove	17	16	1.04439	4.56174
untrustworthy	1	1	0.06527	7.33433
absent	3	3	0.19582	6.23572
affection	3	3	0.19582	6.23572
punishment	2	2	0.13055	6.64118
handle	64	60	3.91645	3.23998
control	211	157	10.24804	2.27808

feeling		127	100	6.52742	2.72916
expect		40	37	2.41514	3.72341
deal		195	165	10.77023	2.22838
expectation		9	7	0.45692	5.38842
unrealistic		3	2	0.13055	6.64118
generally		14	11	0.71802	4.93643
physically		31	29	1.89295	3.96703
violent		11	9	0.58747	5.1371
abusive		22	20	1.30548	4.3386
devaluation		2	2	0.13055	6.64118
partner		75	50	3.26371	3.42231
occurrence		1	1	0.06527	7.33433
said		196	162	10.57441	2.24673
quick		51	47	3.06789	3.48418
impulsive		11	11	0.71802	4.93643
confuse		27	26	1.69713	4.07623
reactivity		1	1	0.06527	7.33433
	3.5	2	2	0.13055	6.64118
week		225	173	11.29243	2.18104
year		689	468	30.5483	1.18586
save		32	31	2.0235	3.90034
trouble		36	35	2.2846	3.77898
recall		5	5	0.32637	5.72489
word		58	51	3.32898	3.4025
anxious		46	46	3.00261	3.50569
negative		65	56	3.65535	3.30898
effect		207	149	9.72585	2.33038
university		8	8	0.52219	5.25489
cognitive		12	11	0.71802	4.93643
badly		18	18	1.17493	4.44396
reverse		4	4	0.2611	5.94803
downhill		1	1	0.06527	7.33433
research		70	60	3.91645	3.23998
success		11	11	0.71802	4.93643
rate		9	9	0.58747	5.1371
patient		43	43	2.80679	3.57313
significant		19	19	1.24021	4.38989
improvement		12	12	0.78329	4.84942
ultimately		14	14	0.91384	4.69527
decide		65	57	3.72063	3.29128
session		28	25	1.63185	4.11545
helpful		55	47	3.06789	3.48418
ready		44	38	2.48042	3.69674

allow	57	51	3.32898	3.4025
gain	82	54	3.5248	3.34535
improve	22	21	1.37076	4.28981
understanding	70	63	4.11227	3.19119
better	472	355	23.17232	1.46221
understand	386	278	18.14621	1.70671
relate	51	46	3.00261	3.50569
weigh	11	11	0.71802	4.93643
mi	4	4	0.2611	5.94803
ight	1	1	0.06527	7.33433
insight	14	13	0.84856	4.76938
sand	2	2	0.13055	6.64118
awareness	6	6	0.39164	5.54257
discourage	9	9	0.58747	5.1371
action	30	28	1.82768	4.00212
responsible	12	12	0.78329	4.84942
happiness	26	22	1.43603	4.24329
appreciate	23	22	1.43603	4.24329
luck	144	143	9.3342	2.37148
bitch	8	8	0.52219	5.25489
ton	24	21	1.37076	4.28981
talk	450	336	21.93211	1.51722
shake	21	20	1.30548	4.3386
guilt	31	30	1.95822	3.93313
miserable	21	20	1.30548	4.3386
skip	3	3	0.19582	6.23572
med	714	414	27.0235	1.30846
plenty	14	14	0.91384	4.69527
dose	110	75	4.89556	3.01684
realize	118	104	6.78851	2.68994
saver	4	4	0.2611	5.94803
free	80	70	4.56919	3.08583
limit	14	12	0.78329	4.84942
interaction	8	8	0.52219	5.25489
stimulus	2	2	0.13055	6.64118
dwell	11	10	0.65274	5.03174
beat	34	32	2.08877	3.86859
def	10	8	0.52219	5.25489
body	130	110	7.18016	2.63385
fact	65	56	3.65535	3.30898
yiur	1	1	0.06527	7.33433
dr	198	124	8.09399	2.51405
adjust	23	21	1.37076	4.28981

lofe		1	1	0.06527	7.33433
tie		10	10	0.65274	5.03174
chain		5	5	0.32637	5.72489
thibg		1	1	0.06527	7.33433
	7-May	1	1	0.06527	7.33433
hr		6	6	0.39164	5.54257
night		117	96	6.26632	2.76998
nothing		166	136	8.87728	2.42167
daily		96	85	5.5483	2.89168
task		17	16	1.04439	4.56174
fewl		1	1	0.06527	7.33433
side		206	154	10.05222	2.29738
stay		218	180	11.74935	2.14137
stress		166	129	8.42037	2.47452
remember		151	128	8.35509	2.4823
enjoyable		3	3	0.19582	6.23572
motion		3	3	0.19582	6.23572
withdrawal		45	31	2.0235	3.90034
damn		32	29	1.89295	3.96703
pill		80	60	3.91645	3.23998
total		71	67	4.37337	3.12964
call		209	159	10.37859	2.26543
complicate		13	11	0.71802	4.93643
easily		23	18	1.17493	4.44396
annoy		13	12	0.78329	4.84942
create		29	27	1.7624	4.03849
pharmaceutical		11	10	0.65274	5.03174
company		19	17	1.10966	4.50112
man		70	55	3.59008	3.327
friend		323	224	14.62141	1.92268
slight		13	12	0.78329	4.84942
bit		100	86	5.61358	2.87998
alter		3	3	0.19582	6.23572
ego		3	3	0.19582	6.23572
name		42	34	2.21932	3.80797
evilalina		3	1	0.06527	7.33433
funny		14	13	0.84856	4.76938
sad		59	53	3.45953	3.36404
aftermath		4	4	0.2611	5.94803
irritate		10	9	0.58747	5.1371
transform		1	1	0.06527	7.33433
main		16	16	1.04439	4.56174
choose		56	47	3.06789	3.48418

medicate	41	35	2.2846	3.77898
medical	65	48	3.13316	3.46313
marijuana	10	10	0.65274	5.03174
knock	13	12	0.78329	4.84942
rarely	12	12	0.78329	4.84942
presence	6	5	0.32637	5.72489
seroquel	55	41	2.67624	3.62076
effexor	29	24	1.56658	4.15628
antidepressant	40	34	2.21932	3.80797
buspar	17	15	0.97911	4.62628
anxiety	744	533	34.79112	1.05581
combination	38	33	2.15405	3.83782
desire	12	11	0.71802	4.93643
amount	39	37	2.41514	3.72341
excessive	4	4	0.2611	5.94803
eliminate	4	4	0.2611	5.94803
sure	227	190	12.40209	2.08731
stop	293	235	15.33943	1.87474
guidance	9	8	0.52219	5.25489
hope	298	244	15.92689	1.83716
peace	55	50	3.26371	3.42231
fair	21	19	1.24021	4.38989
resort	8	8	0.52219	5.25489
dedication	2	1	0.06527	7.33433
hardship	3	3	0.19582	6.23572
throw	42	39	2.54569	3.67077
mental	459	317	20.69191	1.57543
illness	419	277	18.08094	1.71031
mix	45	40	2.61097	3.64545
max	3	3	0.19582	6.23572
husband	132	102	6.65796	2.70936
lose	72	65	4.24282	3.15994
dedicate	1	1	0.06527	7.33433
predicament	2	2	0.13055	6.64118
prevail	1	1	0.06527	7.33433
give	485	377	24.60836	1.40208
survive	44	38	2.48042	3.69674
storm	5	5	0.32637	5.72489
easy	170	146	9.53003	2.35072
worth	113	104	6.78851	2.68994
firstly	5	5	0.32637	5.72489
entitle	5	5	0.32637	5.72489
vent	13	11	0.71802	4.93643

imagine	32	30	1.95822	3.93313
exhaust	35	34	2.21932	3.80797
arangement	2	1	0.06527	7.33433
animal	18	15	0.97911	4.62628
psychologist	38	34	2.21932	3.80797
consellor	1	1	0.06527	7.33433
affect	66	58	3.7859	3.27389
doctor	372	265	17.29765	1.7546
counsellor	14	14	0.91384	4.69527
together	74	62	4.047	3.20719
great	153	137	8.94256	2.41435
chance	40	36	2.34987	3.75081
drain	7	7	0.45692	5.38842
drop	23	22	1.43603	4.24329
group	110	88	5.74413	2.85699
pts	130	117	7.63708	2.57216
hear	137	109	7.11488	2.64298
traumatic	16	13	0.84856	4.76938
teach	21	18	1.17493	4.44396
skill	33	28	1.82768	4.00212
renew	2	2	0.13055	6.64118
forget	61	50	3.26371	3.42231
excuse	22	21	1.37076	4.28981
wrongly	5	5	0.32637	5.72489
flustered	1	1	0.06527	7.33433
outburst	7	6	0.39164	5.54257
small	100	77	5.02611	2.99052
admit	47	41	2.67624	3.62076
apologize	15	14	0.91384	4.69527
vigilant	1	1	0.06527	7.33433
benefit	40	34	2.21932	3.80797
strength	57	48	3.13316	3.46313
courage	11	10	0.65274	5.03174
describe	32	29	1.89295	3.96703
TRUE	50	48	3.13316	3.46313
honest	125	111	7.24543	2.6248
trigger	83	67	4.37337	3.12964
monster	6	6	0.39164	5.54257
head	153	132	8.61619	2.45153
marry	36	28	1.82768	4.00212
calm	64	58	3.7859	3.27389
worse	148	131	8.55091	2.45913
ridiculous	5	5	0.32637	5.72489

shitty	5	5	0.32637	5.72489
health	227	172	11.22715	2.18683
equip	3	3	0.19582	6.23572
acceptance	12	11	0.71802	4.93643
blame	40	35	2.2846	3.77898
loud	19	16	1.04439	4.56174
hour	95	75	4.89556	3.01684
◆	9	8	0.52219	5.25489
caught	8	6	0.39164	5.54257
held	9	9	0.58747	5.1371
matter	114	93	6.0705	2.80173
hopeless	11	11	0.71802	4.93643
die	104	79	5.15666	2.96488
fright	15	15	0.97911	4.62628
death	42	38	2.48042	3.69674
heal	42	38	2.48042	3.69674
rest	63	58	3.7859	3.27389
open	74	65	4.24282	3.15994
accept	75	62	4.047	3.20719
write	147	109	7.11488	2.64298
psychiatrist	127	102	6.65796	2.70936
psy	1	1	0.06527	7.33433
assess	3	3	0.19582	6.23572
follow-up	1	1	0.06527	7.33433
share	70	64	4.17755	3.17545
diagnose	201	164	10.70496	2.23446
primary	12	10	0.65274	5.03174
physician	10	9	0.58747	5.1371
qualify	6	6	0.39164	5.54257
sincerely	3	3	0.19582	6.23572
prescribe	73	63	4.11227	3.19119
condition	54	44	2.87206	3.55014
therapist	148	112	7.3107	2.61583
input	4	4	0.2611	5.94803
seeing	69	63	4.11227	3.19119
counselor	27	25	1.63185	4.11545
guess	64	56	3.65535	3.30898
attitude	5	4	0.2611	5.94803
told	176	140	9.13838	2.39269
belittle	4	4	0.2611	5.94803
pain	200	142	9.26893	2.3785
term	69	56	3.65535	3.30898
break-up	8	8	0.52219	5.25489

unfaithful	1	1	0.06527	7.33433
untruthful	1	1	0.06527	7.33433
accuse	8	7	0.45692	5.38842
kid	150	105	6.85379	2.68037
spoke	12	12	0.78329	4.84942
lady	12	12	0.78329	4.84942
pic	1	1	0.06527	7.33433
venue	1	1	0.06527	7.33433
fb	4	4	0.2611	5.94803
account	5	5	0.32637	5.72489
meeting	24	20	1.30548	4.3386
thought	391	276	18.01567	1.71393
drove	6	6	0.39164	5.54257
social	67	62	4.047	3.20719
paranoia	11	11	0.71802	4.93643
judgemental	6	6	0.39164	5.54257
compassionate	8	8	0.52219	5.25489
empathic	3	3	0.19582	6.23572
current	14	14	0.91384	4.69527
similar	49	47	3.06789	3.48418
terrify	24	22	1.43603	4.24329
bring	60	53	3.45953	3.36404
assessment	7	6	0.39164	5.54257
basically	30	28	1.82768	4.00212
history	19	19	1.24021	4.38989
esp	5	5	0.32637	5.72489
ease	21	18	1.17493	4.44396
figure	45	40	2.61097	3.64545
accurate	2	2	0.13055	6.64118
urge	21	18	1.17493	4.44396
scary	58	51	3.32898	3.4025
appointment	50	39	2.54569	3.67077
fine	32	29	1.89295	3.96703
hypersexuality	1	1	0.06527	7.33433
pregabalin	1	1	0.06527	7.33433
prescription	17	17	1.10966	4.50112
edge	17	15	0.97911	4.62628
phase	19	15	0.97911	4.62628
normalish	1	1	0.06527	7.33433
drive	70	53	3.45953	3.36404
class	36	28	1.82768	4.00212
absurd	2	2	0.13055	6.64118
stage	19	18	1.17493	4.44396

angry	59	52	3.39426	3.38309
upset	64	59	3.85117	3.25679
watch	60	53	3.45953	3.36404
porn	2	2	0.13055	6.64118
rough	15	12	0.78329	4.84942
abuse	56	52	3.39426	3.38309
bedroom	9	9	0.58747	5.1371
accord	6	6	0.39164	5.54257
separate	13	13	0.84856	4.76938
relieve	17	15	0.97911	4.62628
tense	3	3	0.19582	6.23572
add	66	65	4.24282	3.15994
finally	88	78	5.09138	2.97762
premenstrual	3	1	0.06527	7.33433
dysphoria	3	1	0.06527	7.33433
polycystic	1	1	0.06527	7.33433
ovarian	1	1	0.06527	7.33433
syndrome	13	13	0.84856	4.76938
reproductive	3	2	0.13055	6.64118
pco	3	2	0.13055	6.64118
female	7	7	0.45692	5.38842
look	263	223	14.55614	1.92716
information	37	33	2.15405	3.83782
urllink	15	11	0.71802	4.93643
woman	43	36	2.34987	3.75081
onset	6	5	0.32637	5.72489
adult	38	33	2.15405	3.83782
seek	71	60	3.91645	3.23998
numb	18	15	0.97911	4.62628
hate	104	90	5.87467	2.83452
stranger	12	11	0.71802	4.93643
excite	12	11	0.71802	4.93643
moment	102	85	5.5483	2.89168
empty	10	10	0.65274	5.03174
complete	87	85	5.5483	2.89168
quality	19	17	1.10966	4.50112
everyday	66	61	3.98172	3.22346
hyper	6	5	0.32637	5.72489
expand	2	2	0.13055	6.64118
censor	1	1	0.06527	7.33433
indulge	2	2	0.13055	6.64118
safely	4	4	0.2611	5.94803
crave	14	12	0.78329	4.84942

view	24	20	1.30548	4.3386
society	34	20	1.30548	4.3386
respond	24	23	1.50131	4.19884
ashamed	55	44	2.87206	3.55014
screw	12	12	0.78329	4.84942
careful	26	24	1.56658	4.15628
choice	47	40	2.61097	3.64545
practice	31	29	1.89295	3.96703
safety	10	9	0.58747	5.1371
form	50	40	2.61097	3.64545
brave	19	18	1.17493	4.44396
raise	25	24	1.56658	4.15628
post	84	72	4.69974	3.05766
comment	45	38	2.48042	3.69674
support	282	218	14.22977	1.94983
destructive	13	11	0.71802	4.93643
shame	30	26	1.69713	4.07623
discuss	23	18	1.17493	4.44396
appall	1	1	0.06527	7.33433
subject	8	7	0.45692	5.38842
negatively	4	4	0.2611	5.94803
scroll	1	1	0.06527	7.33433
past	111	98	6.39687	2.74936
relevant	1	1	0.06527	7.33433
disgust	8	7	0.45692	5.38842
dangerous	38	34	2.21932	3.80797
girlfriend	13	11	0.71802	4.93643
faithful	2	2	0.13055	6.64118
doubt	34	33	2.15405	3.83782
literally	40	38	2.48042	3.69674
conduct	5	5	0.32637	5.72489
guilty	29	26	1.69713	4.07623
reckless	11	8	0.52219	5.25489
motorcycle	3	3	0.19582	6.23572
moire	1	1	0.06527	7.33433
alive	24	21	1.37076	4.28981
catch	19	19	1.24021	4.38989
st	7	7	0.45692	5.38842
fun	29	23	1.50131	4.19884
kill	93	81	5.28721	2.93988
#NAME?	1	1	0.06527	7.33433
insatiable	2	2	0.13055	6.64118
touch	36	34	2.21932	3.80797

fetish	1	1	0.06527	7.33433
point	141	122	7.96345	2.53031
page	42	33	2.15405	3.83782
defeat	11	9	0.58747	5.1371
purpose	27	22	1.43603	4.24329
pay	47	36	2.34987	3.75081
judge	61	50	3.26371	3.42231
worry	128	114	7.44125	2.59813
manic	115	92	6.00522	2.81254
christian	13	12	0.78329	4.84942
amaze	45	43	2.80679	3.57313
human	34	34	2.21932	3.80797
full	65	59	3.85117	3.25679
distort	2	2	0.13055	6.64118
advice	87	78	5.09138	2.97762
god	203	111	7.24543	2.6248
bless	59	57	3.72063	3.29128
vulnerable	11	11	0.71802	4.93643
toxic	16	13	0.84856	4.76938
supportive	39	35	2.2846	3.77898
witness	7	7	0.45692	5.38842
loudly	2	2	0.13055	6.64118
invite	4	4	0.2611	5.94803
whisper	2	2	0.13055	6.64118
bold	1	1	0.06527	7.33433
explain	107	88	5.74413	2.85699
police	9	6	0.39164	5.54257
refrain	2	2	0.13055	6.64118
report	18	13	0.84856	4.76938
hit	40	36	2.34987	3.75081
arise	1	1	0.06527	7.33433
jail	10	9	0.58747	5.1371
press	2	2	0.13055	6.64118
charge	9	9	0.58747	5.1371
numerous	8	8	0.52219	5.25489
short	51	48	3.13316	3.46313
concise	1	1	0.06527	7.33433
sentence	9	9	0.58747	5.1371
grab	9	7	0.45692	5.38842
stuff	72	61	3.98172	3.22346
discussion	6	6	0.39164	5.54257
business	5	5	0.32637	5.72489
card	18	17	1.10966	4.50112

contact	34	29	1.89295	3.96703
firefighter	1	1	0.06527	7.33433
best	345	281	18.34204	1.69597
horrific	7	6	0.39164	5.54257
admin	2	2	0.13055	6.64118
remove	19	18	1.17493	4.44396
yesterday	16	15	0.97911	4.62628
scare	84	76	4.96084	3.0036
young	79	71	4.63446	3.07165
right	445	321	20.953	1.56289
self-preservation	2	2	0.13055	6.64118
smart	6	6	0.39164	5.54257
lucky	34	30	1.95822	3.93313
corner	10	10	0.65274	5.03174
environment	16	15	0.97911	4.62628
move	111	100	6.52742	2.72916
hang	49	45	2.93734	3.52767
entire	24	23	1.50131	4.19884
spasm	2	2	0.13055	6.64118
cramp	4	4	0.2611	5.94803
version	7	7	0.45692	5.38842
blue	6	6	0.39164	5.54257
ball	5	5	0.32637	5.72489
rage	13	12	0.78329	4.84942
boyfriend	43	37	2.41514	3.72341
nice	41	35	2.2846	3.77898
ache	7	7	0.45692	5.38842
pressure	67	54	3.5248	3.34535
age	62	56	3.65535	3.30898
mixture	4	4	0.2611	5.94803
sensible	1	1	0.06527	7.33433
repeat	15	14	0.91384	4.69527
break	58	55	3.59008	3.327
step	121	89	5.8094	2.84569
faith	43	36	2.34987	3.75081
interest	29	28	1.82768	4.00212
worthwhile	4	2	0.13055	6.64118
achieve	19	16	1.04439	4.56174
set	48	43	2.80679	3.57313
lift	14	14	0.91384	4.69527
foot	15	13	0.84856	4.76938
proof	10	10	0.65274	5.03174
spent	33	32	2.08877	3.86859

exact	14	14	0.91384	4.69527
tire	71	65	4.24282	3.15994
mom	96	74	4.83029	3.03026
runaway	1	1	0.06527	7.33433
prepare	14	13	0.84856	4.76938
encouragement	9	9	0.58747	5.1371
message	55	45	2.93734	3.52767
exercise	65	55	3.59008	3.327
walk	128	106	6.91906	2.67089
bike	5	5	0.32637	5.72489
ride	18	18	1.17493	4.44396
injury	8	8	0.52219	5.25489
thankfully	5	5	0.32637	5.72489
today	69	65	4.24282	3.15994
freaking	6	5	0.32637	5.72489
deffinately	1	1	0.06527	7.33433
recommend	42	37	2.41514	3.72341
insite	1	1	0.06527	7.33433
brine	1	1	0.06527	7.33433
top	21	21	1.37076	4.28981
promise	31	28	1.82768	4.00212
option	53	46	3.00261	3.50569
boss	16	13	0.84856	4.76938
key	25	23	1.50131	4.19884
listen	87	79	5.15666	2.96488
grow	45	37	2.41514	3.72341
heart	120	98	6.39687	2.74936
pray	76	55	3.59008	3.327
suppose	30	29	1.89295	3.96703
rely	16	16	1.04439	4.56174
case	70	65	4.24282	3.15994
thstvway	1	1	0.06527	7.33433
pyhsical	1	1	0.06527	7.33433
deserve	44	38	2.48042	3.69674
law	8	7	0.45692	5.38842
billet	1	1	0.06527	7.33433
son	129	85	5.5483	2.89168
bc	42	27	1.7624	4.03849
mission	2	2	0.13055	6.64118
clean	42	33	2.15405	3.83782
journey	30	25	1.63185	4.11545
prayer	40	29	1.89295	3.96703
2week	1	1	0.06527	7.33433

cop	12	10	0.65274	5.03174
legally	3	2	0.13055	6.64118
involve	29	24	1.56658	4.15628
ranaway	1	1	0.06527	7.33433
court	6	6	0.39164	5.54257
brought	17	15	0.97911	4.62628
victim	12	12	0.78329	4.84942
vial	2	2	0.13055	6.64118
threat	13	11	0.71802	4.93643
serve	13	13	0.84856	4.76938
sea	8	7	0.45692	5.38842
swear	7	7	0.45692	5.38842
laugh	19	19	1.24021	4.38989
stupid	31	29	1.89295	3.96703
continue	61	58	3.7859	3.27389
job	149	100	6.52742	2.72916
mentaly	1	1	0.06527	7.33433
forbid	5	5	0.32637	5.72489
staff	9	8	0.52219	5.25489
fantastic	5	5	0.32637	5.72489
phone	62	53	3.45953	3.36404
present	19	16	1.04439	4.56174
process	52	47	3.06789	3.48418
sever	9	8	0.52219	5.25489
outside	46	41	2.67624	3.62076
provide	25	19	1.24021	4.38989
safe	92	71	4.63446	3.07165
space	21	21	1.37076	4.28981
food	43	34	2.21932	3.80797
decent	9	9	0.58747	5.1371
nurse	35	30	1.95822	3.93313
worker	17	14	0.91384	4.69527
wear	35	31	2.0235	3.90034
street	11	10	0.65274	5.03174
clothes	11	11	0.71802	4.93643
gown	1	1	0.06527	7.33433
lace	1	1	0.06527	7.33433
string	3	2	0.13055	6.64118
hoody	1	1	0.06527	7.33433
belt	2	2	0.13055	6.64118
scarve	1	1	0.06527	7.33433
fresh	10	10	0.65274	5.03174
air	9	9	0.58747	5.1371

kalissa	1	1	0.06527	7.33433
read	118	98	6.39687	2.74936
differ	7	7	0.45692	5.38842
program	16	16	1.04439	4.56174
chat	18	18	1.17493	4.44396
affirmation	2	2	0.13055	6.64118
nasty	10	9	0.58747	5.1371
uncomfortable	15	15	0.97911	4.62628
conscious	7	7	0.45692	5.38842
effort	26	23	1.50131	4.19884
attention	31	29	1.89295	3.96703
everytime	5	5	0.32637	5.72489
force	63	53	3.45953	3.36404
awhile	15	15	0.97911	4.62628
news	11	11	0.71802	4.93643
preggo	1	1	0.06527	7.33433
zoloft	34	24	1.56658	4.15628
prozac	14	13	0.84856	4.76938
town	14	12	0.78329	4.84942
beg	18	16	1.04439	4.56174
cry	86	65	4.24282	3.15994
hospitalize	27	23	1.50131	4.19884
pregnancy	41	29	1.89295	3.96703
birth	30	28	1.82768	4.00212
month	197	147	9.5953	2.3439
worst	67	61	3.98172	3.22346
horrify	3	3	0.19582	6.23572
nightmare	29	28	1.82768	4.00212
wake	70	61	3.98172	3.22346
misery	9	8	0.52219	5.25489
enemy	20	17	1.10966	4.50112
warning	12	12	0.78329	4.84942
hellish	2	2	0.13055	6.64118
sounding	1	1	0.06527	7.33433
thick	6	6	0.39164	5.54257
neil	1	1	0.06527	7.33433
sandy	1	1	0.06527	7.33433
door	23	19	1.24021	4.38989
carer	7	5	0.32637	5.72489
feed	18	15	0.97911	4.62628
cooking	6	6	0.39164	5.54257
cook	6	5	0.32637	5.72489
danger	9	9	0.58747	5.1371

tin	1	1	0.06527	7.33433
soup	3	3	0.19582	6.23572
microwave	1	1	0.06527	7.33433
heat	4	3	0.19582	6.23572
speak	94	83	5.41775	2.91549
portray	1	1	0.06527	7.33433
maths	1	1	0.06527	7.33433
budget	2	2	0.13055	6.64118
bill	13	13	0.84856	4.76938
shop	21	19	1.24021	4.38989
tbh	4	4	0.2611	5.94803
arrive	5	5	0.32637	5.72489
bought	6	5	0.32637	5.72489
rope	3	3	0.19582	6.23572
slowly	31	29	1.89295	3.96703
plunge	2	2	0.13055	6.64118
debt	13	8	0.52219	5.25489
esa	1	1	0.06527	7.33433
dla	1	1	0.06527	7.33433
detract	1	1	0.06527	7.33433
original	7	7	0.45692	5.38842
agree	43	41	2.67624	3.62076
stubborn	5	5	0.32637	5.72489
stupidity	2	2	0.13055	6.64118
mam	4	2	0.13055	6.64118
relax	41	36	2.34987	3.75081
bore	7	7	0.45692	5.38842
memory	40	31	2.0235	3.90034
pull	28	26	1.69713	4.07623
psych	64	52	3.39426	3.38309
crutch	8	7	0.45692	5.38842
assume	20	18	1.17493	4.44396
pregnant	59	43	2.80679	3.57313
ii	19	13	0.84856	4.76938
generalize	15	15	0.97911	4.62628
opinion	42	35	2.2846	3.77898
baby	87	62	4.047	3.20719
approve	9	7	0.45692	5.38842
brutally	4	4	0.2611	5.94803
abortion	4	4	0.2611	5.94803
story	42	39	2.54569	3.67077
level	50	44	2.87206	3.55014
manner	6	6	0.39164	5.54257

counseling		38	31	2.0235	3.90034
facility		8	8	0.52219	5.25489
list		47	40	2.61097	3.64545
sick		85	65	4.24282	3.15994
quit		30	27	1.7624	4.03849
beautiful		45	39	2.54569	3.67077
daughter		85	63	4.11227	3.19119
eye		34	33	2.15405	3.83782
tunnel		10	7	0.45692	5.38842
hold		87	76	4.96084	3.0036
journal		33	30	1.95822	3.93313
alfabet		1	1	0.06527	7.33433
professional		61	57	3.72063	3.29128
squeezer		1	1	0.06527	7.33433
chemically		4	4	0.2611	5.94803
balance		29	27	1.7624	4.03849
electro		3	3	0.19582	6.23572
shock		20	18	1.17493	4.44396
erase		3	3	0.19582	6.23572
	51	7	7	0.45692	5.38842
	56	2	2	0.13055	6.64118
half		43	35	2.2846	3.77898
palpable		1	1	0.06527	7.33433
rispedol		1	1	0.06527	7.33433
	28	7	7	0.45692	5.38842
xanex		2	2	0.13055	6.64118
blow		17	17	1.10966	4.50112
panic		136	115	7.50653	2.5894
attack		142	116	7.5718	2.58074
	7.5	2	2	0.13055	6.64118
norco		1	1	0.06527	7.33433
reahmatoid		1	1	0.06527	7.33433
arthritis		4	4	0.2611	5.94803
moral		2	2	0.13055	6.64118
carelessness		1	1	0.06527	7.33433
accidentally		1	1	0.06527	7.33433
overdose		12	11	0.71802	4.93643
idiot		8	8	0.52219	5.25489
doc		67	53	3.45953	3.36404
reccom		1	1	0.06527	7.33433
provider		9	5	0.32637	5.72489
highlight		4	2	0.13055	6.64118
event		25	20	1.30548	4.3386

mad		27	21	1.37076	4.28981
emmence		1	1	0.06527	7.33433
grief		9	5	0.32637	5.72489
physical		61	53	3.45953	3.36404
torment		5	4	0.2611	5.94803
suicideal		1	1	0.06527	7.33433
homicidal		3	3	0.19582	6.23572
disapper		1	1	0.06527	7.33433
forgiveness		3	3	0.19582	6.23572
label		14	11	0.71802	4.93643
closure		2	2	0.13055	6.64118
close		123	109	7.11488	2.64298
rule		13	12	0.78329	4.84942
born		35	30	1.95822	3.93313
folk		11	9	0.58747	5.1371
script		10	10	0.65274	5.03174
money		36	35	2.2846	3.77898
paid		15	13	0.84856	4.76938
pretend		20	18	1.17493	4.44396
truth		49	40	2.61097	3.64545
adh		44	36	2.34987	3.75081
	2001	2	2	0.13055	6.64118
antianxiety		1	1	0.06527	7.33433
severely		11	11	0.71802	4.93643
chest		17	13	0.84856	4.76938
burden		25	23	1.50131	4.19884
socialize		5	5	0.32637	5.72489
temper		5	5	0.32637	5.72489
suicide		160	102	6.65796	2.70936
hurt		168	125	8.15927	2.50602
differently		32	31	2.0235	3.90034
recognize		22	22	1.43603	4.24329
drift		2	2	0.13055	6.64118
advise		17	16	1.04439	4.56174
reasonable		7	5	0.32637	5.72489
gap		5	5	0.32637	5.72489
epidural		2	2	0.13055	6.64118
avoid		44	40	2.61097	3.64545
organise		2	2	0.13055	6.64118
breastfeed		9	7	0.45692	5.38842
interfere		5	5	0.32637	5.72489
quetiapine		8	8	0.52219	5.25489
important		96	83	5.41775	2.91549

ur		54	31	2.0235	3.90034
pattern		20	17	1.10966	4.50112
manageable		15	15	0.97911	4.62628
risk		56	42	2.74151	3.59666
confidence		14	14	0.91384	4.69527
hide		34	30	1.95822	3.93313
insecurity		8	8	0.52219	5.25489
fight		158	120	7.8329	2.54684
inner		14	13	0.84856	4.76938
battle		63	54	3.5248	3.34535
scream		22	19	1.24021	4.38989
cuz		31	24	1.56658	4.15628
worthless		16	15	0.97911	4.62628
ugly		9	9	0.58747	5.1371
lonley		1	1	0.06527	7.33433
broken		21	21	1.37076	4.28981
sadness		15	15	0.97911	4.62628
sorrow		5	5	0.32637	5.72489
despair		12	11	0.71802	4.93643
ppl		49	35	2.2846	3.77898
mask		6	6	0.39164	5.54257
fake		15	14	0.91384	4.69527
smile		35	32	2.08877	3.86859
bi		42	30	1.95822	3.93313
polar		43	32	2.08877	3.86859
afraid		78	75	4.89556	3.01684
assign		2	2	0.13055	6.64118
place		163	135	8.81201	2.42905
anonymity		1	1	0.06527	7.33433
perfect		28	26	1.69713	4.07623
activity		18	18	1.17493	4.44396
obsess		9	8	0.52219	5.25489
pass		60	55	3.59008	3.327
content		9	9	0.58747	5.1371
available		29	25	1.63185	4.11545
psychiatric		13	13	0.84856	4.76938
reservation		1	1	0.06527	7.33433
wait		67	60	3.91645	3.23998
late		35	29	1.89295	3.96703
	2013	4	4	0.2611	5.94803
psychotic		30	29	1.89295	3.96703
legal		25	16	1.04439	4.56174
bust		2	2	0.13055	6.64118

tooth	8	7	0.45692	5.38842
ruin	20	19	1.24021	4.38989
simply	22	21	1.37076	4.28981
refuse	51	41	2.67624	3.62076
medicine	89	61	3.98172	3.22346
send	33	32	2.08877	3.86859
vibe	2	2	0.13055	6.64118
instantaneously	1	1	0.06527	7.33433
conquer	11	11	0.71802	4.93643
strong	143	119	7.76762	2.55521
minimum	4	4	0.2611	5.94803
effective	26	25	1.63185	4.11545
position	10	10	0.65274	5.03174
reduce	23	21	1.37076	4.28981
misunderstand	4	4	0.2611	5.94803
tablet	28	20	1.30548	4.3386
land	3	3	0.19582	6.23572
clinic	18	17	1.10966	4.50112
establish	3	3	0.19582	6.23572
lesser	3	3	0.19582	6.23572
degree	15	15	0.97911	4.62628
deny	16	15	0.97911	4.62628
shortcoming	1	1	0.06527	7.33433
concern	39	36	2.34987	3.75081
cure	30	28	1.82768	4.00212
plague	6	5	0.32637	5.72489
fibromyalgia	10	9	0.58747	5.1371
thyroid	14	11	0.71802	4.93643
cancer	28	24	1.56658	4.15628
diabetic	13	11	0.71802	4.93643
posttraumatic	2	2	0.13055	6.64118
adjustment	9	9	0.58747	5.1371
feature	5	5	0.32637	5.72489
antisocial	2	2	0.13055	6.64118
hypertensive	2	2	0.13055	6.64118
glaucoma	2	2	0.13055	6.64118
blind	6	6	0.39164	5.54257
ear	16	13	0.84856	4.76938
blood	49	38	2.48042	3.69674
count	29	23	1.50131	4.19884
metal	2	2	0.13055	6.64118
track	27	24	1.56658	4.15628
alcohol	34	30	1.95822	3.93313

dependent	7	7	0.45692	5.38842
product	12	10	0.65274	5.03174
mess	42	38	2.48042	3.69674
escape	37	27	1.7624	4.03849
crazy	64	55	3.59008	3.327
dark	73	63	4.11227	3.19119
corrupt	3	3	0.19582	6.23572
destroy	18	16	1.04439	4.56174
return	26	25	1.63185	4.11545
doom	3	3	0.19582	6.23572
design	6	6	0.39164	5.54257
favor	8	8	0.52219	5.25489
brake	4	4	0.2611	5.94803
childlessness	2	2	0.13055	6.64118
sober	6	5	0.32637	5.72489
disown	2	2	0.13055	6.64118
mother	92	74	4.83029	3.03026
fit	27	25	1.63185	4.11545
approval	7	6	0.39164	5.54257
disapproval	2	2	0.13055	6.64118
intention	8	8	0.52219	5.25489
suggest	69	66	4.30809	3.14467
possibly	25	23	1.50131	4.19884
couple	60	55	3.59008	3.327
counsel	17	17	1.10966	4.50112
educate	26	23	1.50131	4.19884
natural	40	35	2.2846	3.77898
specialist	19	15	0.97911	4.62628
area	35	32	2.08877	3.86859
mediation	2	2	0.13055	6.64118
mindfulness	18	16	1.04439	4.56174
imbalace	1	1	0.06527	7.33433
adverse	10	9	0.58747	5.1371
irregularly	1	1	0.06527	7.33433
magically	1	1	0.06527	7.33433
gonna	25	20	1.30548	4.3386
theybwill	1	1	0.06527	7.33433
jave	1	1	0.06527	7.33433
focus	85	71	4.63446	3.07165
valuable	14	11	0.71802	4.93643
mentally	49	44	2.87206	3.55014
war	12	12	0.78329	4.84942
personally	53	49	3.19843	3.44251

irresponsible	3	3	0.19582	6.23572
grown	14	13	0.84856	4.76938
disciple	2	2	0.13055	6.64118
slump	2	2	0.13055	6.64118
spare	4	4	0.2611	5.94803
unnecessarily	2	2	0.13055	6.64118
swallow	3	3	0.19582	6.23572
chronic	37	35	2.2846	3.77898
houston	1	1	0.06527	7.33433
tx	3	3	0.19582	6.23572
hurricane	2	1	0.06527	7.33433
harvey	1	1	0.06527	7.33433
office	8	8	0.52219	5.25489
noon	3	3	0.19582	6.23572
pharmacy	15	12	0.78329	4.84942
access	17	16	1.04439	4.56174
apt	3	3	0.19582	6.23572
disable	8	8	0.52219	5.25489
vet	5	4	0.2611	5.94803
joke	10	9	0.58747	5.1371
nutty	2	2	0.13055	6.64118
detrimental	2	2	0.13055	6.64118
turkey	10	10	0.65274	5.03174
seizure	11	10	0.65274	5.03174
array	2	2	0.13055	6.64118
permanent	17	15	0.97911	4.62628
zombify	1	1	0.06527	7.33433
sluggish	1	1	0.06527	7.33433
physch	1	1	0.06527	7.33433
sweety	4	2	0.13055	6.64118
reality	22	22	1.43603	4.24329
horrendous	5	5	0.32637	5.72489
horrible	92	85	5.5483	2.89168
train	34	24	1.56658	4.15628
wreck	4	4	0.2611	5.94803
thankful	10	10	0.65274	5.03174
prefer	13	11	0.71802	4.93643
method	14	14	0.91384	4.69527
manage	55	48	3.13316	3.46313
order	34	31	2.0235	3.90034
consult	4	4	0.2611	5.94803
pharmacist	8	8	0.52219	5.25489
selfish	64	34	2.21932	3.80797

thanks	28	26	1.69713	4.07623
curtis	1	1	0.06527	7.33433
encourage	29	29	1.89295	3.96703
glad	31	29	1.89295	3.96703
fail	26	23	1.50131	4.19884
attempt	45	42	2.74151	3.59666
probelly	1	1	0.06527	7.33433
fantasize	3	3	0.19582	6.23572
jealous	4	2	0.13055	6.64118
succeed	14	13	0.84856	4.76938
commit	34	34	2.21932	3.80797
becuz	1	1	0.06527	7.33433
lifetime	7	7	0.45692	5.38842
unfair.it	1	1	0.06527	7.33433
productive	20	19	1.24021	4.38989
pact	3	3	0.19582	6.23572
reach	75	63	4.11227	3.19119
reliance	2	2	0.13055	6.64118
complex	19	17	1.10966	4.50112
combat	5	5	0.32637	5.72489
life.it	1	1	0.06527	7.33433
wreck.so	1	1	0.06527	7.33433
core	3	3	0.19582	6.23572
regret	12	11	0.71802	4.93643
unmedicate	3	3	0.19582	6.23572
months.take	1	1	0.06527	7.33433
grateful	16	15	0.97911	4.62628
opportunity	4	4	0.2611	5.94803
fix	59	52	3.39426	3.38309
longsuffering	1	1	0.06527	7.33433
me.be	1	1	0.06527	7.33433
grown-up	1	1	0.06527	7.33433
medicine.it	1	1	0.06527	7.33433
belong	4	3	0.19582	6.23572
site	14	11	0.71802	4.93643
respect	26	24	1.56658	4.15628
planet	5	5	0.32637	5.72489
bare	3	3	0.19582	6.23572
decency	2	2	0.13055	6.64118
bother	19	19	1.24021	4.38989
mindset	8	8	0.52219	5.25489
compassion	11	11	0.71802	4.93643
handsome	2	2	0.13055	6.64118

hem	2	1	0.06527	7.33433
breakdown	18	18	1.17493	4.44396
couse	1	1	0.06527	7.33433
cheat	24	16	1.04439	4.56174
havimg	1	1	0.06527	7.33433
doent	1	1	0.06527	7.33433
endup	1	1	0.06527	7.33433
exacly	1	1	0.06527	7.33433
darke	1	1	0.06527	7.33433
notice	46	42	2.74151	3.59666
drink	60	36	2.34987	3.75081
neighbor	6	5	0.32637	5.72489
wellbutrin	15	11	0.71802	4.93643
devote	1	1	0.06527	7.33433
pervasive	2	2	0.13055	6.64118
loop	3	3	0.19582	6.23572
shut	26	24	1.56658	4.15628
concentrate	10	9	0.58747	5.1371
lost	104	88	5.74413	2.85699
pound	11	9	0.58747	5.1371
weight	95	60	3.91645	3.23998
jittery	1	1	0.06527	7.33433
twitchy	1	1	0.06527	7.33433
restless	11	10	0.65274	5.03174
morning	52	44	2.87206	3.55014
rear	2	2	0.13055	6.64118
exacerbate	6	5	0.32637	5.72489
tiny	12	12	0.78329	4.84942
increase	41	38	2.48042	3.69674
bothersome	1	1	0.06527	7.33433
switch	23	23	1.50131	4.19884
divorce	16	14	0.91384	4.69527
narcissistic	5	4	0.2611	5.94803
pathological	1	1	0.06527	7.33433
lie	51	45	2.93734	3.52767
deadbeat	1	1	0.06527	7.33433
dad	44	36	2.34987	3.75081
attract	5	5	0.32637	5.72489
split	9	8	0.52219	5.25489
crash	14	13	0.84856	4.76938
unable	23	22	1.43603	4.24329
rn	4	4	0.2611	5.94803
creep	6	6	0.39164	5.54257

wife		37	31	2.0235	3.90034
	35	6	6	0.39164	5.54257
ly		9	8	0.52219	5.25489
gold		4	4	0.2611	5.94803
hesitant		6	6	0.39164	5.54257
marriage		21	19	1.24021	4.38989
lock		14	12	0.78329	4.84942
code		1	1	0.06527	7.33433
random		2	2	0.13055	6.64118
bullshit		12	10	0.65274	5.03174
	46	3	3	0.19582	6.23572
spend		32	29	1.89295	3.96703
dam		3	3	0.19582	6.23572
24/7		6	6	0.39164	5.54257
loti		1	1	0.06527	7.33433
butt		4	4	0.2611	5.94803
lol		42	37	2.41514	3.72341
compatible		1	1	0.06527	7.33433
paranoid		26	24	1.56658	4.15628
aggravate		2	2	0.13055	6.64118
question		95	78	5.09138	2.97762
strange		10	9	0.58747	5.1371
member		54	43	2.80679	3.57313
betray		4	4	0.2611	5.94803
remark		3	3	0.19582	6.23572
xxx		22	22	1.43603	4.24329
basis		13	13	0.84856	4.76938
envious		2	1	0.06527	7.33433
super		21	21	1.37076	4.28981
road		31	31	2.0235	3.90034
girl		53	44	2.87206	3.55014
warrior		10	8	0.52219	5.25489
fathom		2	2	0.13055	6.64118
chin		2	2	0.13055	6.64118
hug		53	46	3.00261	3.50569
dosage		35	30	1.95822	3.93313
trial		29	28	1.82768	4.00212
error		24	23	1.50131	4.19884
perscribe		2	2	0.13055	6.64118
definantly		1	1	0.06527	7.33433
recomend		4	4	0.2611	5.94803
expirence		2	2	0.13055	6.64118
	2-Jan	3	3	0.19582	6.23572

diffrence	1	1	0.06527	7.33433
dibilitate	1	1	0.06527	7.33433
hun	12	11	0.71802	4.93643
drama.ive	1	1	0.06527	7.33433
meds.but	1	1	0.06527	7.33433
responsibility	24	22	1.43603	4.24329
dangerously	2	2	0.13055	6.64118
nowday	1	1	0.06527	7.33433
yourself.you	2	2	0.13055	6.64118
ones.do	1	1	0.06527	7.33433
did.please	1	1	0.06527	7.33433
insecure	4	4	0.2611	5.94803
unprepared	1	1	0.06527	7.33433
oc	34	34	2.21932	3.80797
trait	6	6	0.39164	5.54257
seasonal	3	3	0.19582	6.23572
affective	8	8	0.52219	5.25489
overwhelme	1	1	0.06527	7.33433
diabetes	24	17	1.10966	4.50112
technically	2	2	0.13055	6.64118
affair	5	5	0.32637	5.72489
resentful	2	1	0.06527	7.33433
brick	1	1	0.06527	7.33433
impossible-to-answer	1	1	0.06527	7.33433
thatwill	3	3	0.19582	6.23572
insane	10	10	0.65274	5.03174
betrayal	2	2	0.13055	6.64118
magnitude	1	1	0.06527	7.33433
fortunately	3	3	0.19582	6.23572
squeeze	6	5	0.32637	5.72489
toy	1	1	0.06527	7.33433
couldhave	3	3	0.19582	6.23572
wouldhave	13	6	0.39164	5.54257
under	53	49	3.19843	3.44251
circumstance	14	12	0.78329	4.84942
resentment	4	4	0.2611	5.94803
creature	2	2	0.13055	6.64118
pleasure	5	5	0.32637	5.72489
loneliness	5	5	0.32637	5.72489
wire	4	4	0.2611	5.94803
express	26	24	1.56658	4.15628
women's	1	1	0.06527	7.33433
outlet	7	7	0.45692	5.38842

justify		3	3	0.19582	6.23572
imply		3	3	0.19582	6.23572
leg		20	14	0.91384	4.69527
cut		68	56	3.65535	3.30898
underneath		2	2	0.13055	6.64118
heartache		5	5	0.32637	5.72489
	38	2	2	0.13055	6.64118
bi-polar		16	15	0.97911	4.62628
wise		12	10	0.65274	5.03174
unstable		13	12	0.78329	4.84942
downward		1	1	0.06527	7.33433
spiral		21	20	1.30548	4.3386
understandable		5	5	0.32637	5.72489
suggestion		25	23	1.50131	4.19884
stabilization		3	3	0.19582	6.23572
center		17	16	1.04439	4.56174
average		8	8	0.52219	5.25489
	5-Mar	2	2	0.13055	6.64118
priority		7	7	0.45692	5.38842
music		46	38	2.48042	3.69674
cord		1	1	0.06527	7.33433
asphyxiation		1	1	0.06527	7.33433
fend		1	1	0.06527	7.33433
meal		8	7	0.45692	5.38842
genuine		4	4	0.2611	5.94803
laughter		2	2	0.13055	6.64118
frank		1	1	0.06527	7.33433
table		7	7	0.45692	5.38842
voice		39	35	2.2846	3.77898
addiction		26	20	1.30548	4.3386
taboo		2	2	0.13055	6.64118
grace		6	6	0.39164	5.54257
comfortable		32	27	1.7624	4.03849
assimilate		1	1	0.06527	7.33433
dread		7	7	0.45692	5.38842
mall		3	3	0.19582	6.23572
grocery		8	8	0.52219	5.25489
ward		14	14	0.91384	4.69527
duck		4	4	0.2611	5.94803
longer		44	41	2.67624	3.62076
terrific		2	2	0.13055	6.64118
secret		6	6	0.39164	5.54257
building		15	12	0.78329	4.84942

	120	1	1	0.06527	7.33433
re-pay		1	1	0.06527	7.33433
regular		25	24	1.56658	4.15628
scratch		5	5	0.32637	5.72489
surface		5	4	0.2611	5.94803
separation		3	3	0.19582	6.23572
prob		3	3	0.19582	6.23572
havemt		1	1	0.06527	7.33433
herb		6	6	0.39164	5.54257
tincture		1	1	0.06527	7.33433
diff		3	3	0.19582	6.23572
wanna		29	19	1.24021	4.38989
takin		3	2	0.13055	6.64118
jst		2	2	0.13055	6.64118
cus		2	2	0.13055	6.64118
crap		24	24	1.56658	4.15628
ssri		11	7	0.45692	5.38842
fck		1	1	0.06527	7.33433
fckin		1	1	0.06527	7.33433
uf		1	1	0.06527	7.33433
crappy		4	3	0.19582	6.23572
pissed		4	4	0.2611	5.94803
row		7	6	0.39164	5.54257
slow		39	35	2.2846	3.77898
dani		1	1	0.06527	7.33433
jule		2	2	0.13055	6.64118
letelly		1	1	0.06527	7.33433
mindful		6	6	0.39164	5.54257
habit		20	18	1.17493	4.44396
tomorrow		28	24	1.56658	4.15628
roam		1	1	0.06527	7.33433
consume		6	6	0.39164	5.54257
devastation		1	1	0.06527	7.33433
empathy		6	6	0.39164	5.54257
sympathy		8	7	0.45692	5.38842
stuck		26	26	1.69713	4.07623
redirect		3	2	0.13055	6.64118
wonderful		41	37	2.41514	3.72341
faher		1	1	0.06527	7.33433
van		2	2	0.13055	6.64118
zuiden		1	1	0.06527	7.33433
irrisponsponsible		1	1	0.06527	7.33433
mislead		1	1	0.06527	7.33433

cruel		9	9	0.58747	5.1371
psychological		8	7	0.45692	5.38842
disappear		12	11	0.71802	4.93643
norm		4	4	0.2611	5.94803
catorgorize		1	1	0.06527	7.33433
	37	5	5	0.32637	5.72489
insurance		33	27	1.7624	4.03849
lavender		4	4	0.2611	5.94803
eucalyptus		1	1	0.06527	7.33433
oil		29	23	1.50131	4.19884
burn		20	19	1.24021	4.38989
diffuser		1	1	0.06527	7.33433
simple		39	36	2.34987	3.75081
burner		2	2	0.13055	6.64118
calmly		4	4	0.2611	5.94803
breathe		53	45	2.93734	3.52767
300mg		5	5	0.32637	5.72489
pluse		1	1	0.06527	7.33433
mg		57	40	2.61097	3.64545
valium		16	15	0.97911	4.62628
neulactil		1	1	0.06527	7.33433
smoke		13	13	0.84856	4.76938
canabis		2	1	0.06527	7.33433
cartiac		1	1	0.06527	7.33433
arrest		2	2	0.13055	6.64118
middle		19	17	1.10966	4.50112
highway		2	2	0.13055	6.64118
lorazapam		1	1	0.06527	7.33433
greatly		6	6	0.39164	5.54257
influence		3	3	0.19582	6.23572
deff		1	1	0.06527	7.33433
exwife		2	1	0.06527	7.33433
lithium		81	60	3.91645	3.23998
topamax		7	5	0.32637	5.72489
depakote		12	10	0.65274	5.03174
clonzapam		2	2	0.13055	6.64118
challenge		25	25	1.63185	4.11545
pristiq		7	5	0.32637	5.72489
lexapro		11	9	0.58747	5.1371
cymbalta		19	15	0.97911	4.62628
amatriptyline		1	1	0.06527	7.33433
countless		9	9	0.58747	5.1371
clutter		2	2	0.13055	6.64118

irrational	11	9	0.58747	5.1371
shrink	4	4	0.2611	5.94803
lit	1	1	0.06527	7.33433
portion	4	4	0.2611	5.94803
favorite	8	8	0.52219	5.25489
insist	9	9	0.58747	5.1371
sudden	16	15	0.97911	4.62628
afflict	2	2	0.13055	6.64118
klonopin	14	11	0.71802	4.93643
gabapentin	11	10	0.65274	5.03174
:(	13	10	0.65274	5.03174
sadly	15	15	0.97911	4.62628
religious	10	8	0.52219	5.25489
spiritual	15	9	0.58747	5.1371
distinguish	3	3	0.19582	6.23572
unforgivable	1	1	0.06527	7.33433
thank	52	51	3.32898	3.4025
liver	24	14	0.91384	4.69527
kidney	14	11	0.71802	4.93643
ny	1	1	0.06527	7.33433
power	42	34	2.21932	3.80797
sake	9	8	0.52219	5.25489
technic	1	1	0.06527	7.33433
swing	40	37	2.41514	3.72341
active	17	14	0.91384	4.69527
follow	24	20	1.30548	4.3386
surroundings	3	3	0.19582	6.23572
forthought	1	1	0.06527	7.33433
reaction	35	30	1.95822	3.93313
side-effect	3	3	0.19582	6.23572
combo	37	34	2.21932	3.80797
moon	5	4	0.2611	5.94803
instance	6	6	0.39164	5.54257
hydroxyzine	4	4	0.2611	5.94803
tolerate	12	12	0.78329	4.84942
fully	20	18	1.17493	4.44396
floor	12	10	0.65274	5.03174
klonipin	7	3	0.19582	6.23572
dissolvable	1	1	0.06527	7.33433
melt	3	3	0.19582	6.23572
tongue	3	3	0.19582	6.23572
basket	1	1	0.06527	7.33433
linger	3	3	0.19582	6.23572

guideline	2	2	0.13055	6.64118
gabatine	1	1	0.06527	7.33433
me.just	1	1	0.06527	7.33433
alike	2	2	0.13055	6.64118
painkiller	2	2	0.13055	6.64118
lash	5	5	0.32637	5.72489
electric	7	6	0.39164	5.54257
asap	10	10	0.65274	5.03174
disability	43	34	2.21932	3.80797
outcome	3	3	0.19582	6.23572
bath	15	13	0.84856	4.76938
soothe	4	4	0.2611	5.94803
symphony	1	1	0.06527	7.33433
relaxation	3	2	0.13055	6.64118
nose	5	5	0.32637	5.72489
mouth	18	17	1.10966	4.50112
atmosphere	2	2	0.13055	6.64118
flashback	8	8	0.52219	5.25489
melatonin	5	5	0.32637	5.72489
lay	14	12	0.78329	4.84942
min	7	7	0.45692	5.38842
don	18	10	0.65274	5.03174
delusional	4	4	0.2611	5.94803
obsessive	10	10	0.65274	5.03174
drowsiness	3	3	0.19582	6.23572
operate	4	4	0.2611	5.94803
vehicle	4	4	0.2611	5.94803
drowsy	5	5	0.32637	5.72489
agoraphobic	4	4	0.2611	5.94803
car	22	20	1.30548	4.3386
trazadone	10	8	0.52219	5.25489
bupropion	1	1	0.06527	7.33433
addict	39	28	1.82768	4.00212
beach	4	4	0.2611	5.94803
moreno	1	1	0.06527	7.33433
valley	1	1	0.06527	7.33433
nope	3	3	0.19582	6.23572
altogether	3	3	0.19582	6.23572
refocus	2	2	0.13055	6.64118
suck	51	45	2.93734	3.52767
path	17	14	0.91384	4.69527
kindly	1	1	0.06527	7.33433
status	6	6	0.39164	5.54257

yell	27	19	1.24021	4.38989
jump	14	13	0.84856	4.76938
fry	3	3	0.19582	6.23572
pan	3	3	0.19582	6.23572
numbness	3	3	0.19582	6.23572
guide	10	9	0.58747	5.1371
clear	31	30	1.95822	3.93313
hook	5	4	0.2611	5.94803
comfort	25	21	1.37076	4.28981
rocky	1	1	0.06527	7.33433
saving	5	5	0.32637	5.72489
feedback	4	4	0.2611	5.94803
gyn	1	1	0.06527	7.33433
hormone	13	10	0.65274	5.03174
test	49	35	2.2846	3.77898
unrelated	1	1	0.06527	7.33433
document	2	2	0.13055	6.64118
pick	37	34	2.21932	3.80797
alternative	12	12	0.78329	4.84942
stabilizer	51	45	2.93734	3.52767
positively	2	2	0.13055	6.64118
statistics	1	1	0.06527	7.33433
nervous	22	22	1.43603	4.24329
diet	39	32	2.08877	3.86859
essential	9	9	0.58747	5.1371
psychosis	31	27	1.7624	4.03849
general	27	25	1.63185	4.11545
delusion	6	6	0.39164	5.54257
dealt	17	16	1.04439	4.56174
threaten	19	19	1.24021	4.38989
nut	8	8	0.52219	5.25489
science	9	8	0.52219	5.25489
scramble	1	1	0.06527	7.33433
supplement	19	14	0.91384	4.69527
element	3	3	0.19582	6.23572
naturally	5	5	0.32637	5.72489
muscle	13	12	0.78329	4.84942
tension	9	9	0.58747	5.1371
constipation	5	3	0.19582	6.23572
loosen	1	1	0.06527	7.33433
digestive	2	2	0.13055	6.64118
tract	1	1	0.06527	7.33433
milk	4	4	0.2611	5.94803

magnesia		1	1	0.06527	7.33433
hydroxide		1	1	0.06527	7.33433
draw		18	18	1.17493	4.44396
water		35	25	1.63185	4.11545
intestine		1	1	0.06527	7.33433
soften		1	1	0.06527	7.33433
fece		1	1	0.06527	7.33433
regime		8	6	0.39164	5.54257
conditioning		1	1	0.06527	7.33433
workout		5	5	0.32637	5.72489
hockey		1	1	0.06527	7.33433
alleviate		5	5	0.32637	5.72489
soreness		1	1	0.06527	7.33433
epsom		1	1	0.06527	7.33433
salt		5	5	0.32637	5.72489
ice		13	9	0.58747	5.1371
inflammation		4	4	0.2611	5.94803
pack		5	5	0.32637	5.72489
localize		1	1	0.06527	7.33433
rda		1	1	0.06527	7.33433
400-420		1	1	0.06527	7.33433
	300	5	5	0.32637	5.72489
glycinate		2	1	0.06527	7.33433
citrate		2	1	0.06527	7.33433
oxide		2	1	0.06527	7.33433
minor		6	5	0.32637	5.72489
contraindication		1	1	0.06527	7.33433
xanax		40	29	1.89295	3.96703
magnesium		6	4	0.2611	5.94803
rare		3	3	0.19582	6.23572
clinically		1	1	0.06527	7.33433
irrelevant		4	4	0.2611	5.94803
gut		12	11	0.71802	4.93643
vitamin		22	18	1.17493	4.44396
regimen		5	5	0.32637	5.72489
insanely		2	2	0.13055	6.64118
staple		2	2	0.13055	6.64118
excrete		1	1	0.06527	7.33433
asleep		22	17	1.10966	4.50112
disturb		8	8	0.52219	5.25489
bet		11	11	0.71802	4.93643
routine		22	17	1.10966	4.50112
youtube		7	7	0.45692	5.38842

insomnia		32	30	1.95822	3.93313
play		48	42	2.74151	3.59666
stabiliser		6	5	0.32637	5.72489
	31	3	3	0.19582	6.23572
apathetic		2	2	0.13055	6.64118
logic		7	7	0.45692	5.38842
deeply		9	9	0.58747	5.1371
lessen		13	11	0.71802	4.93643
guy		28	25	1.63185	4.11545
worthy		12	12	0.78329	4.84942
gist		2	2	0.13055	6.64118
cheap		9	9	0.58747	5.1371
celestial		1	1	0.06527	7.33433
seasoning		1	1	0.06527	7.33433
tea		12	10	0.65274	5.03174
exaggerate		6	6	0.39164	5.54257
restoril		2	1	0.06527	7.33433
slept		7	6	0.39164	5.54257
warm		8	8	0.52219	5.25489
migraine		14	11	0.71802	4.93643
nausea		13	12	0.78329	4.84942
guarantee		7	7	0.45692	5.38842
	99.9	1	1	0.06527	7.33433
cost		15	13	0.84856	4.76938
	3.98	1	1	0.06527	7.33433
box		10	9	0.58747	5.1371
paranioa		1	1	0.06527	7.33433
refresh		5	5	0.32637	5.72489
midnight		2	2	0.13055	6.64118
snacker		1	1	0.06527	7.33433
heavy		10	10	0.65274	5.03174
curve		1	1	0.06527	7.33433
snack		5	5	0.32637	5.72489
replace		10	10	0.65274	5.03174
bedtime		10	8	0.52219	5.25489
nuisance		1	1	0.06527	7.33433
	45	14	13	0.84856	4.76938
licence		1	1	0.06527	7.33433
	21	13	13	0.84856	4.76938
	32	6	6	0.39164	5.54257
route		7	6	0.39164	5.54257
traffic		3	3	0.19582	6.23572
minimal		1	1	0.06527	7.33433

aka		3	3	0.19582	6.23572
agoraphobia		15	15	0.97911	4.62628
btw		6	6	0.39164	5.54257
population		4	4	0.2611	5.94803
	6,000	1	1	0.06527	7.33433
5mine		1	1	0.06527	7.33433
fast		34	31	2.0235	3.90034
clonopin		1	1	0.06527	7.33433
magic		1	1	0.06527	7.33433
passenger		1	1	0.06527	7.33433
supervision		5	5	0.32637	5.72489
cbt		10	10	0.65274	5.03174
fruit		11	10	0.65274	5.03174
smoothy		2	2	0.13055	6.64118
antioxidant		2	2	0.13055	6.64118
frit		1	1	0.06527	7.33433
protein		6	5	0.32637	5.72489
powder		2	2	0.13055	6.64118
	3-Feb	1	1	0.06527	7.33433
red		14	11	0.71802	4.93643
meat		5	5	0.32637	5.72489
veggy		7	6	0.39164	5.54257
ion		1	1	0.06527	7.33433
ext		1	1	0.06527	7.33433
outdoors		2	2	0.13055	6.64118
meditation		46	31	2.0235	3.90034
successful		6	6	0.39164	5.54257
mesd		2	1	0.06527	7.33433
disagree		3	3	0.19582	6.23572
block		9	9	0.58747	5.1371
edgy		1	1	0.06527	7.33433
crawl		5	5	0.32637	5.72489
skin		15	13	0.84856	4.76938
lower		25	22	1.43603	4.24329
taperr		1	1	0.06527	7.33433
abruptly		2	2	0.13055	6.64118
taper		6	5	0.32637	5.72489
vivi		1	1	0.06527	7.33433
ann		1	1	0.06527	7.33433
mepham		1	1	0.06527	7.33433
reply.my		1	1	0.06527	7.33433
am.my		1	1	0.06527	7.33433
	55	3	3	0.19582	6.23572

cpts	3	3	0.19582	6.23572
future	41	40	2.61097	3.64545
hell.my	1	1	0.06527	7.33433
1st	7	6	0.39164	5.54257
intern	2	2	0.13055	6.64118
source	7	6	0.39164	5.54257
away.guess	1	1	0.06527	7.33433
up.just	1	1	0.06527	7.33433
sad.god	1	1	0.06527	7.33433
receive	20	20	1.30548	4.3386
special	35	32	2.08877	3.86859
survivor	15	11	0.71802	4.93643
embrace	12	10	0.65274	5.03174
smell	6	6	0.39164	5.54257
flower	11	6	0.39164	5.54257
credit	8	8	0.52219	5.25489
sister	40	34	2.21932	3.80797
brother	44	29	1.89295	3.96703
meant	26	24	1.56658	4.15628
click	9	8	0.52219	5.25489
lovable	2	2	0.13055	6.64118
hoffman	1	1	0.06527	7.33433
institute	2	2	0.13055	6.64118
pl	5	5	0.32637	5.72489
wean	8	8	0.52219	5.25489
aim	3	3	0.19582	6.23572
goal	32	27	1.7624	4.03849
principle	2	2	0.13055	6.64118
uk	13	11	0.71802	4.93643
unsure	8	8	0.52219	5.25489
practise	2	2	0.13055	6.64118
city	4	3	0.19582	6.23572
college	28	17	1.10966	4.50112
independence	2	2	0.13055	6.64118
freak	9	8	0.52219	5.25489
school	82	57	3.72063	3.29128
panicky	3	3	0.19582	6.23572
exhaustive	1	1	0.06527	7.33433
hormonal	8	5	0.32637	5.72489
contraceptive	3	3	0.19582	6.23572
emergency	17	14	0.91384	4.69527
contraception	1	1	0.06527	7.33433
implant	1	1	0.06527	7.33433

cream	2	2	0.13055	6.64118
patch	3	2	0.13055	6.64118
intra-uterine	1	1	0.06527	7.33433
device	3	3	0.19582	6.23572
amitriptyline	2	2	0.13055	6.64118
clomipramine	1	1	0.06527	7.33433
moclobemide	1	1	0.06527	7.33433
citalopram	5	4	0.2611	5.94803
fluoxetine	4	3	0.19582	6.23572
paroxetine	1	1	0.06527	7.33433
sertraline	10	9	0.58747	5.1371
duloxetine	1	1	0.06527	7.33433
venlafaxine	7	5	0.32637	5.72489
replacement	2	2	0.13055	6.64118
hrt	4	2	0.13055	6.64118
gel	2	2	0.13055	6.64118
vaginal	1	1	0.06527	7.33433
ring	8	8	0.52219	5.25489
thin	7	7	0.45692	5.38842
anticoagulant	1	1	0.06527	7.33433
warfarin	1	1	0.06527	7.33433
acenocoumarol	1	1	0.06527	7.33433
epilepsy	11	10	0.65274	5.03174
carbamazepine	2	2	0.13055	6.64118
phenobarbitone	1	1	0.06527	7.33433
phenytoin	1	1	0.06527	7.33433
primidone	1	1	0.06527	7.33433
sodium	6	4	0.2611	5.94803
valproate	3	2	0.13055	6.64118
immunosuppressant	1	1	0.06527	7.33433
ciclosporin	1	1	0.06527	7.33433
tacrolimus	1	1	0.06527	7.33433
hiv	4	2	0.13055	6.64118
infection	4	3	0.19582	6.23572
amprenavir	1	1	0.06527	7.33433
atazanavir	1	1	0.06527	7.33433
darunavir	1	1	0.06527	7.33433
fosamprenavir	1	1	0.06527	7.33433
indinavir	1	1	0.06527	7.33433
lopinavir	1	1	0.06527	7.33433
nelfinavir	1	1	0.06527	7.33433
ritonavir	1	1	0.06527	7.33433
saquinavir	1	1	0.06527	7.33433

tipranavir	1	1	0.06527	7.33433
efavirenz	1	1	0.06527	7.33433
nevirapine	1	1	0.06527	7.33433
delaviridine	1	1	0.06527	7.33433
cholesterol	3	3	0.19582	6.23572
simvastatin	1	1	0.06527	7.33433
atorvastatin	1	1	0.06527	7.33433
irinotecan	1	1	0.06527	7.33433
dasatinib	1	1	0.06527	7.33433
erlotinib	1	1	0.06527	7.33433
imatinib	1	1	0.06527	7.33433
sorafenib	1	1	0.06527	7.33433
sunitinib	1	1	0.06527	7.33433
etoposide	1	1	0.06527	7.33433
mitotane	1	1	0.06527	7.33433
digoxin	1	1	0.06527	7.33433
ivabradine	1	1	0.06527	7.33433
amiodarone	1	1	0.06527	7.33433
almotriptan	1	1	0.06527	7.33433
eletriptan	1	1	0.06527	7.33433
frovatriptan	1	1	0.06527	7.33433
naratriptan	1	1	0.06527	7.33433
rizatriptan	1	1	0.06527	7.33433
sumatriptan	1	1	0.06527	7.33433
zolmitriptan	1	1	0.06527	7.33433
amlodipine	1	1	0.06527	7.33433
nifedipine	1	1	0.06527	7.33433
felodipine	1	1	0.06527	7.33433
verapamil	1	1	0.06527	7.33433
regulate	11	11	0.71802	4.93643
thyroxine	2	2	0.13055	6.64118
john	5	5	0.32637	5.72489
ort	1	1	0.06527	7.33433
fentanyl	1	1	0.06527	7.33433
propofol	1	1	0.06527	7.33433
sevoflurane	1	1	0.06527	7.33433
midazolam	1	1	0.06527	7.33433
anaesthetic	1	1	0.06527	7.33433
pre-operative	1	1	0.06527	7.33433
tramadol	1	1	0.06527	7.33433
analgesic	1	1	0.06527	7.33433
erythromycin	1	1	0.06527	7.33433
clarithromycin	1	1	0.06527	7.33433

telithromycin	1	1	0.06527	7.33433
antibiotic	2	2	0.13055	6.64118
itraconazole	1	1	0.06527	7.33433
voriconazole	1	1	0.06527	7.33433
antifungal	1	1	0.06527	7.33433
artemether	1	1	0.06527	7.33433
lumefantrine	1	1	0.06527	7.33433
antimalarial	1	1	0.06527	7.33433
rasagiline	1	1	0.06527	7.33433
parkinson	1	1	0.06527	7.33433
aripiprazole	3	3	0.19582	6.23572
antipsychotic	11	11	0.71802	4.93643
buspirone	3	3	0.19582	6.23572
anxiolytic	2	2	0.13055	6.64118
aprepitant	1	1	0.06527	7.33433
post-operative	1	1	0.06527	7.33433
vomit	5	5	0.32637	5.72489
butobarbital	1	1	0.06527	7.33433
phenobarbital	1	1	0.06527	7.33433
barbiturate	1	1	0.06527	7.33433
methyl	1	1	0.06527	7.33433
phenidate	1	1	0.06527	7.33433
central	2	2	0.13055	6.64118
stimulant	8	5	0.32637	5.72489
exemestane	1	1	0.06527	7.33433
antagonist	2	2	0.13055	6.64118
breast	5	5	0.32637	5.72489
eplerenone	1	1	0.06527	7.33433
diuretic	4	2	0.13055	6.64118
lansoprazole	1	1	0.06527	7.33433
omeprazole	1	1	0.06527	7.33433
proton	1	1	0.06527	7.33433
pump	4	4	0.2611	5.94803
inhibitor	3	3	0.19582	6.23572
heartburn	1	1	0.06527	7.33433
theophylline	1	1	0.06527	7.33433
bronchodilator	1	1	0.06527	7.33433
gliclazide	1	1	0.06527	7.33433
fortunate	6	6	0.39164	5.54257
reassure	10	10	0.65274	5.03174
breath	19	17	1.10966	4.50112
remind	33	29	1.89295	3.96703
sickness	12	9	0.58747	5.1371

FALSE	2	2	0.13055	6.64118
fill	35	28	1.82768	4.00212
manipulative	3	3	0.19582	6.23572
self-center	1	1	0.06527	7.33433
ground	27	23	1.50131	4.19884
drag	12	11	0.71802	4.93643
briefly	1	1	0.06527	7.33433
usual	5	5	0.32637	5.72489
hypocrite	1	1	0.06527	7.33433
liar	3	3	0.19582	6.23572
insignificant	6	6	0.39164	5.54257
connect	18	15	0.97911	4.62628
wall	6	6	0.39164	5.54257
hurriedly	1	1	0.06527	7.33433
rush	8	8	0.52219	5.25489
apartment	5	5	0.32637	5.72489
text	25	17	1.10966	4.50112
reply	10	9	0.58747	5.1371
emptiness	7	6	0.39164	5.54257
tough	36	34	2.21932	3.80797
tight	7	6	0.39164	5.54257
coz	7	4	0.2611	5.94803
ignorant	9	8	0.52219	5.25489
awkward	4	4	0.2611	5.94803
ignore	17	16	1.04439	4.56174
communicate	17	16	1.04439	4.56174
affectionate	1	1	0.06527	7.33433
kiss	4	4	0.2611	5.94803
assure	8	6	0.39164	5.54257
suffocate	2	2	0.13055	6.64118
guitar	2	2	0.13055	6.64118
pulse	3	2	0.13055	6.64118
monitor	14	12	0.78329	4.84942
stabilize	15	15	0.97911	4.62628
embarrassment	3	3	0.19582	6.23572
weekly	9	9	0.58747	5.1371
phycotheripe	1	1	0.06527	7.33433
pat	3	3	0.19582	6.23572
unit	12	9	0.58747	5.1371
noisy	1	1	0.06527	7.33433
nosy	1	1	0.06527	7.33433
light	38	33	2.15405	3.83782
hand	73	61	3.98172	3.22346

meet	23	20	1.30548	4.3386
harshly	1	1	0.06527	7.33433
gp	38	30	1.95822	3.93313
offer	24	24	1.56658	4.15628
difference	42	38	2.48042	3.69674
motivate	9	9	0.58747	5.1371
binge	5	5	0.32637	5.72489
online	28	25	1.63185	4.11545
flat	9	8	0.52219	5.25489
dress	3	3	0.19582	6.23572
afford	17	15	0.97911	4.62628
silly	11	11	0.71802	4.93643
flunk	1	1	0.06527	7.33433
slightly	3	3	0.19582	6.23572
heather	4	4	0.2611	5.94803
party	13	12	0.78329	4.84942
overwhelm	49	47	3.06789	3.48418
acid	6	6	0.39164	5.54257
reflux	3	3	0.19582	6.23572
anxity	3	1	0.06527	7.33433
groggy	4	4	0.2611	5.94803
income	7	6	0.39164	5.54257
financially	3	3	0.19582	6.23572
weather	9	8	0.52219	5.25489
lye	1	1	0.06527	7.33433
sit	67	57	3.72063	3.29128
poorly	7	6	0.39164	5.54257
distress	10	9	0.58747	5.1371
useless.there	1	1	0.06527	7.33433
. .	5	5	0.32637	5.72489
ya	12	11	0.71802	4.93643
song	7	5	0.32637	5.72489
empowerment	1	1	0.06527	7.33433
continuous	2	2	0.13055	6.64118
fault	39	34	2.21932	3.80797
ridiculously	2	2	0.13055	6.64118
tendency	7	7	0.45692	5.38842
detail	15	15	0.97911	4.62628
honey	8	7	0.45692	5.38842
large	11	9	0.58747	5.1371
glass	6	4	0.2611	5.94803
doct	1	1	0.06527	7.33433
fragile	5	5	0.32637	5.72489

peaceful		2	2	0.13055	6.64118
country		17	13	0.84856	4.76938
usa		7	6	0.39164	5.54257
sadden		4	4	0.2611	5.94803
validate		6	6	0.39164	5.54257
sue		2	2	0.13055	6.64118
radio		2	2	0.13055	6.64118
tv		12	10	0.65274	5.03174
endure		10	10	0.65274	5.03174
barefoot		1	1	0.06527	7.33433
winter		8	7	0.45692	5.38842
spring		4	4	0.2611	5.94803
coat		2	2	0.13055	6.64118
cigarette		2	2	0.13055	6.64118
sidewalk		1	1	0.06527	7.33433
	54	2	2	0.13055	6.64118
feb		8	8	0.52219	5.25489
	2010	3	3	0.19582	6.23572
	58	2	2	0.13055	6.64118
number		27	25	1.63185	4.11545
absolutely		33	31	2.0235	3.90034
multiple		20	17	1.10966	4.50112
moderate		5	5	0.32637	5.72489
aloof		1	1	0.06527	7.33433
turn		96	88	5.74413	2.85699
copy		7	6	0.39164	5.54257
paste		3	3	0.19582	6.23572
exist		20	18	1.17493	4.44396
cyclic		1	1	0.06527	7.33433
nature		12	12	0.78329	4.84942
stigmatization		1	1	0.06527	7.33433
embarrass		20	18	1.17493	4.44396
characterize		2	1	0.06527	7.33433
individual		26	24	1.56658	4.15628
talkative		1	1	0.06527	7.33433
outgoing		6	6	0.39164	5.54257
consumption		2	1	0.06527	7.33433
co-worker		1	1	0.06527	7.33433
authority		2	2	0.13055	6.64118
extra		18	16	1.04439	4.56174
motivation		16	14	0.91384	4.69527
helpless		11	11	0.71802	4.93643
socially		3	3	0.19582	6.23572

impair	3	3	0.19582	6.23572
diagnostic	5	3	0.19582	6.23572
scale	7	6	0.39164	5.54257
hypomanic	7	6	0.39164	5.54257
define	18	16	1.04439	4.56174
incredibly	11	10	0.65274	5.03174
mode	5	5	0.32637	5.72489
prolong	3	3	0.19582	6.23572
hopelessness	6	5	0.32637	5.72489
disinterest	1	1	0.06527	7.33433
fatigue	14	13	0.84856	4.76938
agitate	8	8	0.52219	5.25489
unfocus	2	2	0.13055	6.64118
indecisive	1	1	0.06527	7.33433
accomplish	10	10	0.65274	5.03174
enjoy	46	42	2.74151	3.59666
creativity	4	4	0.2611	5.94803
intensity	1	1	0.06527	7.33433
likewise	1	1	0.06527	7.33433
overtime	2	2	0.13055	6.64118
lastly	5	5	0.32637	5.72489
license	5	4	0.2611	5.94803
attest	2	2	0.13055	6.64118
stigma	17	14	0.91384	4.69527
assault	2	2	0.13055	6.64118
murder	1	1	0.06527	7.33433
misuse	1	1	0.06527	7.33433
mechanical	1	1	0.06527	7.33433
derogatory	1	1	0.06527	7.33433
inaccurate	2	2	0.13055	6.64118
association	4	4	0.2611	5.94803
client	7	5	0.32637	5.72489
official	4	3	0.19582	6.23572
cast	5	5	0.32637	5.72489
component	5	5	0.32637	5.72489
fluctuate	1	1	0.06527	7.33433
smith	1	1	0.06527	7.33433
m.	1	1	0.06527	7.33433
supervisor	6	4	0.2611	5.94803
brokenness	1	1	0.06527	7.33433
pastor	10	8	0.52219	5.25489
crack	10	4	0.2611	5.94803
pot	21	9	0.58747	5.1371

bearer		7	1	0.06527	7.33433
india		1	1	0.06527	7.33433
hung		4	3	0.19582	6.23572
pole		3	3	0.19582	6.23572
carry		15	15	0.97911	4.62628
neck		5	3	0.19582	6.23572
deliver		7	5	0.32637	5.72489
stream		5	2	0.13055	6.64118
master		9	4	0.2611	5.94803
accomplishment		2	2	0.13055	6.64118
imperfection		1	1	0.06527	7.33433
bitter		4	3	0.19582	6.23572
load		12	11	0.71802	4.93643
leak		2	1	0.06527	7.33433
flaw		7	6	0.39164	5.54257
hill		4	4	0.2611	5.94803
sun		7	6	0.39164	5.54257
wild		2	2	0.13055	6.64118
cheer		1	1	0.06527	7.33433
trail		6	5	0.32637	5.72489
advantage		10	9	0.58747	5.1371
plant		8	6	0.39164	5.54257
decorate		1	1	0.06527	7.33433
beauty		3	3	0.19582	6.23572
meltdown		6	5	0.32637	5.72489
service		34	29	1.89295	3.96703
agency		5	5	0.32637	5.72489
record		17	16	1.04439	4.56174
suit		4	3	0.19582	6.23572
fantastically		1	1	0.06527	7.33433
50mg		6	5	0.32637	5.72489
100mg		6	5	0.32637	5.72489
noticeable		2	2	0.13055	6.64118
loss		36	29	1.89295	3.96703
appetite		16	15	0.97911	4.62628
suppressant		1	1	0.06527	7.33433
topomax		4	3	0.19582	6.23572
	2-Jan	4	4	0.2611	5.94803
2x		10	6	0.39164	5.54257
pair		2	2	0.13055	6.64118
adivan		1	1	0.06527	7.33433
clinical		16	13	0.84856	4.76938
schizophrenia		21	16	1.04439	4.56174

convince		16	16	1.04439	4.56174
visit		36	32	2.08877	3.86859
conspiracy		1	1	0.06527	7.33433
sister-in-law		1	1	0.06527	7.33433
lucidity		1	1	0.06527	7.33433
compliant		2	2	0.13055	6.64118
full-blown		2	2	0.13055	6.64118
denial		5	5	0.32637	5.72489
	23	4	4	0.2611	5.94803
haha		9	8	0.52219	5.25489
frustrate		37	35	2.2846	3.77898
cent		6	6	0.39164	5.54257
mobile		2	2	0.13055	6.64118
self-care		2	2	0.13055	6.64118
wash		17	14	0.91384	4.69527
laura		1	1	0.06527	7.33433
spell		13	12	0.78329	4.84942
sequence		1	1	0.06527	7.33433
sum		2	2	0.13055	6.64118
impede		2	2	0.13055	6.64118
observe		3	3	0.19582	6.23572
speech		6	5	0.32637	5.72489
fluent		1	1	0.06527	7.33433
kempt		1	1	0.06527	7.33433
tidy		2	2	0.13055	6.64118
default		2	2	0.13055	6.64118
presentable		1	1	0.06527	7.33433
greet		1	1	0.06527	7.33433
plastic		2	2	0.13055	6.64118
friendly		6	6	0.39164	5.54257
sign		50	43	2.80679	3.57313
language		6	6	0.39164	5.54257
october		6	6	0.39164	5.54257
deadly		1	1	0.06527	7.33433
overload		5	5	0.32637	5.72489
plate		4	4	0.2611	5.94803
resource		24	22	1.43603	4.24329
heavenly		3	3	0.19582	6.23572
father		29	25	1.63185	4.11545
devastate		10	9	0.58747	5.1371
tragic		1	1	0.06527	7.33433
boy		40	30	1.95822	3.93313
tear		22	21	1.37076	4.28981

recover		25	24	1.56658	4.15628
upstairs		2	2	0.13055	6.64118
stomp		1	1	0.06527	7.33433
slam		3	3	0.19582	6.23572
bass		1	1	0.06527	7.33433
stereo		3	3	0.19582	6.23572
prolly		1	1	0.06527	7.33433
passive		1	1	0.06527	7.33433
aggressive		7	6	0.39164	5.54257
spot		6	6	0.39164	5.54257
park		5	5	0.32637	5.72489
	49	2	2	0.13055	6.64118
librium		1	1	0.06527	7.33433
zanax		2	2	0.13055	6.64118
clonaxapam		1	1	0.06527	7.33433
chantix		1	1	0.06527	7.33433
abilify		13	13	0.84856	4.76938
cocktail		14	12	0.78329	4.84942
familiar		11	11	0.71802	4.93643
crisis		20	18	1.17493	4.44396
sayin		2	2	0.13055	6.64118
specifically		7	7	0.45692	5.38842
awesome		13	12	0.78329	4.84942
thrive		3	3	0.19582	6.23572
implement		3	3	0.19582	6.23572
joy		19	18	1.17493	4.44396
inspiration		3	3	0.19582	6.23572
impact		18	18	1.17493	4.44396
forever		21	20	1.30548	4.3386
idk		17	15	0.97911	4.62628
landlord		1	1	0.06527	7.33433
manager		3	3	0.19582	6.23572
muffle		1	1	0.06527	7.33433
hinge		1	1	0.06527	7.33433
supposedly		4	4	0.2611	5.94803
bear		8	8	0.52219	5.25489
decondition		1	1	0.06527	7.33433
tolerance		5	3	0.19582	6.23572
flexibility		1	1	0.06527	7.33433
noise		11	10	0.65274	5.03174
ironically		3	3	0.19582	6.23572
bang		2	2	0.13055	6.64118
variety		4	4	0.2611	5.94803

bark	3	3	0.19582	6.23572
quiet	12	10	0.65274	5.03174
resolution	5	5	0.32637	5.72489
self-medicate	2	2	0.13055	6.64118
ounce	1	1	0.06527	7.33433
piano	3	3	0.19582	6.23572
doable	1	1	0.06527	7.33433
confide	8	7	0.45692	5.38842
rise	3	3	0.19582	6.23572
lance	1	1	0.06527	7.33433
boil	3	3	0.19582	6.23572
poison	9	8	0.52219	5.25489
bash	3	3	0.19582	6.23572
identity	5	5	0.32637	5.72489
squash	1	1	0.06527	7.33433
finish	17	14	0.91384	4.69527
project	10	9	0.58747	5.1371
burst	3	3	0.19582	6.23572
knit	5	4	0.2611	5.94803
buy	17	15	0.97911	4.62628
yarn	4	3	0.19582	6.23572
polish	2	2	0.13055	6.64118
chuckle	1	1	0.06527	7.33433
slide	3	3	0.19582	6.23572
picture	11	8	0.52219	5.25489
giant	1	1	0.06527	7.33433
caffeine	9	7	0.45692	5.38842
shower	21	16	1.04439	4.56174
google	15	13	0.84856	4.76938
image	8	7	0.45692	5.38842
cute	4	4	0.2611	5.94803
inbox	6	6	0.39164	5.54257
conversation	18	18	1.17493	4.44396
useless	8	8	0.52219	5.25489
ass	16	13	0.84856	4.76938
grandson	3	3	0.19582	6.23572
ramble	2	2	0.13055	6.64118
restraint	2	2	0.13055	6.64118
shot	18	15	0.97911	4.62628
taught	17	15	0.97911	4.62628
pure	3	3	0.19582	6.23572
frankly	3	3	0.19582	6.23572
haven	1	1	0.06527	7.33433

secretly		2	2	0.13055	6.64118
schizoeffect		2	1	0.06527	7.33433
anexity		1	1	0.06527	7.33433
advertise		1	1	0.06527	7.33433
kno		6	4	0.2611	5.94803
ether		4	2	0.13055	6.64118
epilim		4	3	0.19582	6.23572
vision		7	5	0.32637	5.72489
ritalin		5	4	0.2611	5.94803
refill		9	7	0.45692	5.38842
apathy		1	1	0.06527	7.33433
massively		5	4	0.2611	5.94803
private		15	14	0.91384	4.69527
benzo		20	14	0.91384	4.69527
agorophobia		1	1	0.06527	7.33433
regularly		9	9	0.58747	5.1371
	8-Jun	3	3	0.19582	6.23572
muck		1	1	0.06527	7.33433
spill		1	1	0.06527	7.33433
lighten		3	3	0.19582	6.23572
verse		2	2	0.13055	6.64118
bible		30	18	1.17493	4.44396
christ		8	7	0.45692	5.38842
strengthen		3	3	0.19582	6.23572
courtship		1	1	0.06527	7.33433
unwell		3	3	0.19582	6.23572
predict		4	4	0.2611	5.94803
normality		2	2	0.13055	6.64118
disclosure		4	4	0.2611	5.94803
	29	3	3	0.19582	6.23572
media		10	10	0.65274	5.03174
compare		9	7	0.45692	5.38842
trash		4	4	0.2611	5.94803
eod		1	1	0.06527	7.33433
cloud		5	4	0.2611	5.94803
remission		3	3	0.19582	6.23572
hereditary		4	4	0.2611	5.94803
existence		8	8	0.52219	5.25489
babie		1	1	0.06527	7.33433
firm		3	3	0.19582	6.23572
respectfully		1	1	0.06527	7.33433
laid		5	5	0.32637	5.72489
spouse		8	6	0.39164	5.54257

hv	9	1	0.06527	7.33433
ft	1	1	0.06527	7.33433
chiyoung	1	1	0.06527	7.33433
equipment	1	1	0.06527	7.33433
loan	2	2	0.13055	6.64118
introduce	5	4	0.2611	5.94803
career	12	8	0.52219	5.25489
2yr	4	4	0.2611	5.94803
destinctionsto	1	1	0.06527	7.33433
leisure	1	1	0.06527	7.33433
pool	2	2	0.13055	6.64118
fitness	5	4	0.2611	5.94803
elevate	5	3	0.19582	6.23572
boone	1	1	0.06527	7.33433
wud	1	1	0.06527	7.33433
postromatic	1	1	0.06527	7.33433
ambition	2	2	0.13055	6.64118
gifted	1	1	0.06527	7.33433
trauma	18	17	1.10966	4.50112

Biodata.

Personal Data;

Surname; HALLY

Other names; Damilola, A

Date of Birth; 13th June, 1985

Gender; Female

Local Government Area; Obokun

State of Origin ; Osun State

Nationality; Nigerian

Marital Status; Married

Religion; Christianity

Next of kin; Hally Victor Ayodeji

Contact

Residential Address; 26,Mesi-ogo Estate, off studio road,Akobo,Oyo-state

Postal Address: same as above

Email Address: wisebrands04@gmail.com

Mobile Number: 07068506402

Institutions Attended with Dates and Qualification

Lead City University, Toll gate, Ibadan, Oyo state 2020 till date(M.sc Computer and information science)

Olabisi Onabanjo University,Ago-Iwoye 2004 – 2008(B.sc Computer Science and Statistics)

Olabisi Onabanjo University Ago iwoye 2002 – 2004 (Diploma in Data Processing)

Providence Heights Secondary School,Fagba Agege,Lagos state. 1996- 2001

(Senior school Certificate).

Membership of Professional Bodies

Institute of strategic Management,Nigeria (Member)

Fate Foundation ,Lagos (Member)

Work Experience with Dates

Propertymart Real Estate Company,Lagos,Nigeria 2009-2011

Administative Officer

Skilledge Consults Nigeria Limited,Lagos,Nigeria	2012-2018
Administrative Manager.	
Thara Xanders Nigeria limited,Ibadan,Oyo-state	2019 – Date
Regional Manager;Product Distribution	

Academic Papers (published thesis)

Damilola Hally, Wumi Ajayi, Ayuba Atuman, Lawal olalekan, Folahan Jiboku "Big data Testing,challenges and best Practices" Technology reports of Kansai University,volume 63,issue 6,June 2021,ISSN04532198.

Referees

1) Dr Wumi Ajayi
   Department of Computer Science
   Lead City University,Ibadan,Oyo state
   080 896 3630.
2) The Rev H.S.A Osindeinde
   The Methodist Church, Nigeria
   080 3842 9765
3) Dr A. A Waheed
   Department of Computer Science
   Lead City university,Toll gate,Ibadan,oyo state
   070 3119 9441.

……. ……..

Signature Date

University Compliance Certification

This is to certify that the Thesis of Hally Damilola Alaba with Matriculation Number LCU/PG/001538,in the department of computer science, Faculty of Basic and Applied sciences,

Lead City University, Ibadan is in Full compliance with the approved University format and style

…… ……..

Signature Date