SENTIMENT ANALYSIS FOR THE DETECTION OF DEPRESSIVE USERS ON SOCIAL NETWORKS

A thesis submitted by:

Christopher Wigell

B.ESS. B.PSY.

For the award of

Master of Research

2024

ABSTRACT

Depression affects approximately 121 million people worldwide and has as significant an influence on patients as it does their carers. Until recently, psychologists used paper based psychological batteries for screening. New developments in Natural Language Processing however have enabled real-time analysis, offering increases in scalability and accessibility. We built on current knowledge by developing several Centralised Neural Network models around a pre-labelled Reddit dataset, first comparing performance. Our best model then parsed a Reddit forum (r/Depression) labelling depressed/control posts, evaluating them for changes over time whilst using a Term Frequency-Inverse Document Frequency (TF-IDF) tool to screen them for a weighted keyword list. We found neither word count nor the number of posts significantly affected our model's performance, 67% of initial forum posts had depressed labels and as time progressed there was a decrease in depressed labels (per post) which was significant between users. Our TF-IDF tool also demonstrated a new way of looking at keywords, presenting us with a list most relevant to each category, whilst we additionally developed a free research tool for release into the public. Our study was able to yield support for its use within online forums at a very low cost; justifying further exploration into the use of AI tools for the screening of depression and other mood disorders.

CERTIFICATION OF THESIS

I Christopher Wigell declare that the Thesis entitled '*Sentiment analysis for the detection of depressive users on social networks*' is not more than 100,000 words in length including quotes and exclusive of tables, figures, appendices, bibliography, references and footnotes. The thesis contains no material that has been submitted previously, in whole or in part, for the award of any other academic degree or diploma. Except where otherwise indicated, this thesis is my own work.

Date: **11 September 2024**

Endorsed by:

Professor Xiaohui Tao

Principal Supervisor

Dr Grace Wang

Associate Supervisor

Dr Ji Zhang

Associate Supervisor

Student and supervisors' signatures of endorsement are held at the University.

ACKNOWLEDGEMENTS

I would like to thank my supervisors Professor Xiaohui Tao, Dr Grace Wang & Dr Ji Zhang for their input, guidance and constructive feedback across this project.

This research has been supported by the Australian Government Research Training Program Scholarship.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

AI	Artificial Intelligence
BDI	Beck’s Depression Inventory
CBT	Cognitive Behavioural Therapy
DSM-5	Diagnostics and Statistics Manual 5
DSS	Decision Support Systems
LLM	Large Language Model
NLP	Natural Language Processing
PHQ	Patient Health Questionnaire
RSDD	Reddit Self-reported Depression Diagnosis
TF-IDF	Term Frequency-Inverse Document Frequency
USQ	University of Southern Queensland

CHAPTER 1: INTRODUCTION

Social networks and communications data, with the advent of machine learning, have enhanced our knowledge in human behaviour due to their ubiquity and investigative value. Despite the field of psychology contributing much to human research, it’s lack of more precise, scalable tools and mounting ethical/bureaucratic barriers have stymied progress (Stevens, 2017). This is only further compounded by it’s dependence on raters for screening, with inter-rater reliability heavily influenced by test choice, mood, cultural and lexical bias (Reynolds & Suzuki 2012; Fiske 2020). Consequently, we engaged our research with automated screening processes in mind to reduce these barriers.

In various areas of mental health, several computer applications have emerged to aid in the identification and management of depression. Screening apps such as PHQ-9 and BDI having continued in a digital form, serve as preliminary self-assessment/screening tools to recognize potential symptoms. Teletherapy platforms offer online counselling sessions with licensed professionals, providing remote access to mental health support for those dealing with depression. Mood tracking apps such as Daylio, eMoods and Moodpath allow users to monitor their emotional states over time, potentially revealing patterns indicative of depressive symptoms. AI-driven chatbots such as Woebot and Wysa offer conversational interfaces, enabling real time communication with users seeking more information and CBT, while wearables such as Apple Watch, Fitbit and Garmin Vivosmart offer physiologically correlated variable tracking including heart rate and sleep patterns. This data-driven approach not only enhances the efficiency of diagnosis and treatment but also empowers mental health professionals to offer more targeted support and provide a non-intrusive means of assessing emotional states. When integrated correctly, remote monitoring and timely interventions are possible across all fields, bridging geographical gaps and ensuring continuity of care.

It is essential to recognize that in their current state, these applications complement and do not replace professional medical advice. If someone experiences symptoms of depression, consulting a qualified healthcare professional for a thorough evaluation and guidance is always crucial. The integration of technology in mental health care however underscores the potential for innovative solutions in supporting individuals facing depression. At the time of this paper's publication, experimental research has advanced in the use of AI algorithms to parse data inputs (sentiment analysis) such as speech patterns and social media activity, for valanced intricacies associated with signs of depression. Although numerous applications exist, initiatives such as those applied via media for governmental lockdowns, consumer and mobile advertising identifiers, the identification of emojis as language-independent variables and the development of an ontological approach towards mental health have demonstrated it's potential use and current capacity for healthcare (Gabarron et al., 2019; Kang et al., 2017; Kralj Novak et al., 2015).

The significance of our research lies in it's potential to build on current screening methods for depression, yielding more academic support and new novel insights. By developing tools or methodologies that expedite the screening process, we are aiming to reduce the distance between symptom recognition and intervention. This will not only enhance the reach of mental health screening tools but align with a broader public health objective of reducing the societal burden associated with untreated depression. In doing so, the research not only addresses the critical need for early intervention in mental health but also contributes to a comprehensive and proactive approach to managing and preventing the adverse consequences linked to untreated depressive disorders.

Although the methods proposed within this study could be redressed for other psychological diagnoses, we sought to investigate the practicality of using an NLP algorithm to passively screen depressed users earlier than conventional methods. The end goal of this study is to contribute to the current academic framework that supports the development of tools that enable psychologists to treat depression (and other related mood disorders) more effectively than what is currently being used. Systematically, we seek to do this by:

1. Creating and training a model based on a pre-labelled dataset:

2. Evaluating factors of influence.

3. Labelling a dataset blind to this model in a separate forum.

4. Identifying weighted keywords relevant to the depressed and control groups.

5. Evaluating whether engaging in forum discussions over time influences post labels.

1.1 Psychology As A Social Construct

Psychology as a field uses psycho-social constructs to define human behaviour through the factorisation of culture, communication and physiology. Although psychological theories and concepts would benefit from universality, such as those consistent with mathematics and pure science, the clear and demonstrated behavioural differentiation observed within all psychological studies are deeply embedded within cultural, historical and social contexts that are difficult to control for (Baker, 2007; Thomas et al., 2020). The ways in which mental health, emotions and cognitive processes are conceptualized are not fixed or absolute but are constructed and reinforced by the societies in which they exist, as well as physiological/genetic predispositions (Baker, 2007). This recognition in the past, has prompted critical examination of how cultural norms, values and power dynamics contribute to the construction of psychological knowledge, but fundamental to this study, support the importance of explicit definitions in the modelling of algorithms.

In the realm of clinical psychology, for example, the diagnostic criteria for mental disorders can vary across cultures, reflecting cultural attitudes towards behaviour and mental health (Alarcón, 2009). Moreover, the emphasis on individualism or collectivism within different societies can shape therapeutic approaches and expectations (Stead, 2004). The social constructionist perspective also highlights how psychological research itself may be influenced by cultural biases, as the selection of research topics, methods and interpretations can be unconsciously influenced by the cultural context in which the research is conducted (Bargh & Morsella, 2008; Stead, 2004).

Furthermore, the social construction of psychology extends to the understanding of normality and deviance. What is considered normal behaviour in one culture may be viewed as deviant in another (Kelly et al., 2017; Nalah et al., 2013). This cultural relativity challenges the universality of psychological concepts and underscores the importance of considering diverse perspectives in both research and clinical practice (Nalah et al., 2013). In essence, recognizing psychology as a social construct invites a more inclusive and culturally sensitive approach, acknowledging the diversity of human experiences and challenging the assumption that psychological principles are universally applicable without regard to cultural and societal contexts.

Consequently, understanding and accepting how these concepts interact with language helps to illustrate how disorders can be defined and consequently, how they can be objectified within psychological batteries. To objectively screen for a disorder (or factors consistent with the definition of a series of behaviours) using logic, therefore, is entirely dependent on the definition and accuracy of the psychological literature. AI will therefore, only ever be as effective as the definitions it is modelled on.

1.2 Psychological Testing In Focus

Clinical psychologists are currently depended on for both the effective and official diagnosis; and often the subsequent treatment of numerous mental health disorders within a framework set out by the diagnostic and statistical manual of mental health. In the field of mental health, the accuracy of psychological testing, a cornerstone of clinical practice, is contingent upon various interrelated factors. Standardization, a crucial aspect, ensures consistent and uniform administration and scoring procedures, fostering reliability and facilitating the comparability of results. Reliability, encompassing test-retest and inter-rater reliability, establishes the test's consistency over time and across different raters, further enhancing the dependability of the obtained data. Concurrently, validity, examining the extent to which a test measures what it purports to measure, involves considerations such as content, criterion-related and construct validity. These psychometric properties collectively form the foundation of robust psychological assessments, instilling confidence in the accuracy of the results.

As indicated in 1.1, cultural considerations present another pivotal factor influencing the precision of psychological testing, which is dependent on both the practitioner's capacity to control their own cognitive biases, whilst recognising those in others (Yager et al., 2021). As defined by Yager et al. (2021), "cognitive and affective biases, stemming in part from intuitive, fast-thinking processes, can contribute to illogical thinking, affect medical decision making and adversely affect the conduct of psychotherapy". Controlling for the impact of these biases on test performance and interpretation is essential for clinicians, as without this, it presents a major risk of misdiagnosis and punitive outcomes dependent on overarching laws and social norms (Blumenthal-Barby & Krieger, 2015).

These ethical considerations, including informed consent and confidentiality, underscore the importance of maintaining trust between practitioners and those undergoing assessments. Moreover, this clinical judgment fundamentally determines the requirement for human interaction for diagnosis. Practitioners can navigate the interaction of test results with other relevant information to formulate a comprehensive understanding of an individual's mental health. Without this holistic approach to assessment that involves synthesizing data gleaned from psychological testing, with insights gained from clinical interviews, observations and collateral reports, diagnosis is unlikely to deliver a patient centred response.

Continued training and professional development therefore constitute the final cornerstone for ensuring accuracy in psychological testing within clinical practice. Lifelong learning is essential for clinicians to stay abreast of advancements in testing methodologies, changes in diagnostic criteria and the evolving landscape of cultural competence. This commitment to ongoing education empowers clinicians to refine their skills, ultimately contributing to the precision and relevance of psychological assessments in the dynamic field of mental health. Although risks within this area are as diverse as the aforementioned, cognitive load could be argued as the major inhibitor of decision-making processes (Fox et al., 2016; Leppink, 2017). As psychologists are required to recall and evaluate a breadth of information throughout a clinical consultation, any effort to decrease the number of variables involved should reduce the risk of error.

Consequently, we posit that with the progress of technology, relying on individual practitioners without the introduction of newer aids is not the most effective model for ongoing psychological assessments and present a case for one tool that can improve this area of practice. We ultimately suggest within the scope of our study that the improvement of the screening process is consistent with the principles of hospital triage, with clients being sent to practitioners based on hierarchical risk. To achieve this, we build an argument in support of natural language processing as the tool to fulfill this function.

1.3 Defining Depression

Depression, identified as a mood disorder characterized by persistent feelings of sadness and diminished interest, is classified by the American Psychiatric Association's Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5) under several categories. These include Disruptive Mood Dysregulation Disorder, Major Depressive Disorder, Persistent Depressive Disorder (Dysthymia), Premenstrual Dysphoric Disorder and Depressive Disorder Due to Another Medical Condition. All these depressive disorders share common features such as persistent sadness, emptiness, or irritable mood, accompanied by cognitive and somatic changes that significantly impact an individual's functional capacity (Chand & Arif, 2023; Tolentino & Schmidt, 2018).

These categories are best summarised under the following (Chand & Arif, 2023):

- Disruptive Mood Dysregulation Disorder (DMDD): Designated for children and adolescents, addressing severe irritability and frequent temper outbursts disproportionate to the situation.
- Major Depressive Disorder (MDD): Involves at least one major depressive episode marked by a pervasive low mood and a range of cognitive and physical symptoms.
- Persistent Depressive Disorder (Dysthymia): Chronic, persistent low mood lasting for an extended period.

- Premenstrual Dysphoric Disorder (PMDD): Linked to the menstrual cycle, manifesting in severe mood disturbances during the luteal phase.
- Depressive Disorder Due To Another Medical Condition: Secondary to a medical condition with a temporal relationship discernible from the condition itself and the concurrent diagnosis.

These depressive disorders share a common thread of impairing an individual's capacity to function, affecting both cognitive and somatic domains. It is crucial to recognize the comprehensive nature of these classifications, acknowledging the somatic and cognitive changes associated with depression. Understanding the diverse manifestations of depressive disorders enables interprofessional teams to collaboratively evaluate and manage depression, ensuring well-coordinated care that enhances patient outcomes.

Moreover, the impact of depressive disorders extends beyond individual suffering to encompass broader societal implications, including economic burdens and healthcare utilization. Research suggests that depression contributes significantly to disability-adjusted life years (DALYs) and healthcare expenditures, underscoring the urgency of addressing this public health challenge comprehensively (Reddy, 2010). By elucidating the diverse manifestations of depressive disorders, healthcare systems can develop targeted interventions aimed at early detection, prevention and treatment, thereby mitigating the far-reaching consequences of untreated depression.

Furthermore, cultural and contextual factors play a pivotal role in shaping the experience and expression of depressive symptoms, emphasizing the need for culturally sensitive and contextually relevant approaches to assessment and care. Sociocultural influences, such as stigma surrounding mental health, access to resources and social support networks, profoundly influence individual's help-seeking behaviour and treatment outcomes. Therefore, adopting a culturally informed framework is essential for delivering equitable and effective mental healthcare services that resonate with diverse populations.

In addition to clinical interventions, efforts to address depression must encompass multifaceted strategies spanning public health initiatives, policy advocacy and community engagement. Promoting mental health literacy, fostering resilience and reducing structural barriers to care are integral components of a comprehensive approach to combating depression on both individual and societal levels. By fostering collaboration across sectors and mobilizing resources, stakeholders can work synergistically to create environments conducive to mental well-being, fostering resilience and facilitating recovery for individuals affected by depressive disorders. Ultimately, by embracing a holistic and inclusive approach, we can strive towards a future where depression is recognized, understood and effectively addressed, empowering individuals to lead fulfilling and meaningful lives.

1.4 Depression Screening Tools

At present, there are several screening tools available for the evaluation of depressive symptoms. Though it is important to note that not all are diagnostic instruments, they all facilitate the identification of patients with relevant symptoms. Triaging patients and/or monitoring their progress, they serve the fundamental basis of our argument for the designation of AI centred tools.

Currently used tools include:

- **PHQ-9 (Patient Health Questionnaire-9):** Assesses the presence and severity of depressive symptoms using nine questions. It is based on the criteria for major depressive disorder in the DSM-5 and is common within hospital settings (Siniscalchi et al., 2020).
- **PHQ-8 and PHQ-2:** Based on PHQ-9, the Patient Health Questionnaire also includes shorter versions such as PHQ-2 (two questions) and PHQ-8 (nine questions) (Kroenke et al., 2001; Kroenke et al., 2009).
- **BDI-II (Beck's Depression Inventory-II):** A self-report inventory that measures the severity of depression symptoms, including cognitive, affective and somatic symptoms (Steer et al., 1999).

- **CES-D (Center For Epidemiologic Studies Depression Scale):** A self-report scale designed to measure depressive symptoms in larger populations; more commonly used in research studies (Lewinsohn et al., 1997).
- **GDS (Geriatric Depression Scale):** Specifically designed for older adults, this self-report scale screens for depression in individuals over the age of 65 (Shin et al., 2019).
- **Hamilton Depression Rating Scale (HAM-D):** Unlike self-report measures, the HAM-D is a clinician-administered tool used to assess the severity of depressive symptoms. It consists of 17 items evaluating mood, guilt, suicidal ideation, insomnia andother depressive features. The HAM-D is frequently employed in clinical trials and psychiatric evaluations to quantify symptom severity and track treatment response (Hamilton, 1960).
- **Hospital Anxiety And Depression Scale (HADS):** Developed to assess anxiety and depression in non-psychiatric hospital settings, the HADS is a self-report questionnaire comprising 14 items (7 for anxiety and 7 for depression). It screens for symptoms of anxiety and depression separately, making it particularly useful in medical settings where comorbidities are common (Zigmond & Snaith, 1983).

These screening tools provide valuable insights into the presence and severity of depressive symptoms, facilitating early detection, intervention and monitoring across diverse populations. Integrating multiple assessment measures is often tailored to specific demographics to enhance the accuracy and comprehensiveness of screening. However, exploring this process further falls outside the scope of our study.

1.5 Concepts Behind AI Screening Tools

As various AI screening tools exist, the scope of our research must be defined through our aims to encourage the use of passive screening tools in popular culture. Consequently, rather than suggesting the development of a tool that is solely commercial, we both develop and demonstrate the use of an open-source tool that can be further improved, integrated into other programs and contribute towards a standardised screening method that parses textual input in public forums.

Passive textual parsing, a process involving the automated analysis of written content such as text messages and social media posts, holds significant promise for the screening of depression. Employing NLP techniques, this approach enables the extraction of detailed insights from language patterns and linguistic markers associated with depressive symptoms. Although there are several advantages, a key advantage lies in the potential for early detection and continuous monitoring, allowing for the identification of subtle shifts in an individual's mental well-being over time. This proactive approach facilitates early intervention and support, potentially preventing the escalation of depressive symptoms (Medich et al., 2023). This form of analysis also contributes to a more holistic understanding of individuals by examining expressions, thoughts and emotions across various contexts (Guo, 2022). Traditional depression screening tools often rely on specific questions with smaller data samples, whereas passive data analysis offers a broader examination of individual's experiences and communications. Additionally, this approach reduces the stigma associated with face-to-face assessments, as individuals may express themselves more freely through written communication, such as online posts or private messages. The privacy afforded by passive data parsing fosters an environment in which individuals feel more comfortable sharing their feelings and experiences, potentially leading to more accurate and insightful assessments.

When implementing AI screening tools, it is understood that consideration is also given to cultural nuances and linguistic variations that may affect the expression of depressive symptoms. Cultural differences in language use expressions of emotion and perceptions of mental health can influence the effectiveness and generalizability of screening algorithms across diverse populations. Therefore, incorporating cultural sensitivity into the development and validation of AI tools is crucial to ensure that they are applicable across different cultural contexts (Al-Mosaiwi & Johnstone, 2018). Analysing the context in which language patterns or expressions occur, provides a more nuanced understanding of the triggers and stressors contributing to depressive symptoms. This contextual awareness enhances the precision of assessments and supports a personalized approach to mental health care (Guo, 2022).

Unlike current screening tools, a major benefit of AI tools is that they can monitor individual’s longitudinal mental health trajectories by analysing changes in language patterns and emotional expressions both prospectively and retrospectively. By leveraging predictive analytics techniques, AI screening tools can consequently anticipate the likelihood of future depressive episodes based on historical data and early warning signs. This proactive approach empowers individuals and healthcare providers to take preventive measures and intervene before symptoms escalate, ultimately improving mental health outcomes (Salas-Zárate et al., 2022). This real-time feedback loop enables healthcare providers to tailor interventions to an individual’s evolving needs, optimizing treatment plans and maximizing therapeutic outcomes. For example, if an individual's language patterns indicate a deterioration in mood despite ongoing treatment, healthcare providers can promptly adjust medication dosages, therapy modalities or support services to address emerging symptoms and prevent further decline. Furthermore, the predictive analytics of AI screening tools can extend beyond individual-level interventions to population-level mental health management. By aggregating and analysing anonymised data from large cohorts, these tools can identify trends and patterns in depressive symptomatology across different demographics, geographic regions or socio-economic groups. Such insights enable policymakers and public health authorities to allocate resources effectively, implement targeted prevention strategies and design community-based interventions to address prevalent risk factors and mitigate the burden of depression on a broader scale.

Moreover, the integration of AI-based tools, such as chatbots or virtual assistants, represents a synergistic approach to enhancing depression screening and intervention efforts. These AI-driven platforms have the potential to revolutionize the delivery of mental health support by providing accessible, scalable and personalized interventions to individuals in need. One key advantage of integrating AI-based tools is the ability to offer timely and contextually relevant interventions based on real-time analysis of individual’s language patterns and emotional expressions. Chatbots and virtual assistants equipped with natural language processing capabilities can engage in interactive conversations with users, offering empathetic responses, coping strategies and links to additional resources based on the individual's expressed needs and concerns. Furthermore, these AI-driven platforms can serve as a valuable complement to traditional mental health services, particularly in contexts where access to in-person care is limited or stigmatized. Individuals may feel more comfortable disclosing sensitive information or seeking support through anonymous online platforms, thereby overcoming barriers to help-seeking behaviour and receiving more timely assistance.

However, as with any technology-mediated intervention, it is understood ethical considerations surrounding privacy, consent and responsible data use must be built into the creation of such technologies. Safeguarding individual’s sensitive information and ensuring data security and confidentiality, are paramount to building trust and maintaining the integrity of the screening process. Robust privacy policies, transparent data practices and informed consent procedures should be integral components of AI-based mental health platforms to uphold individual’s rights and autonomy. However, we propose that open-source software is the only way to substantiate legitimacy as proprietary technology cannot be validated. Moreover, ongoing monitoring and evaluation of AI-driven interventions are necessary to assess their effectiveness, safety and user satisfaction. Iterative refinement based on user feedback and empirical evidence can help optimize the performance and usability of these platforms, ensuring that they meet the diverse needs of individuals seeking mental health support.

As research and development in this field progresses, AI-based tools hold immense promise for augmenting depression screening practices and expanding access to timely, evidence-based interventions. By embracing innovative technologies while upholding ethical standards and respecting individual privacy, healthcare providers can harness the full potential of AI to improve mental health outcomes and promote well-being for all.

1.6 An Argument Over Ethics

Although the scientific evaluation of communications data often raises issues with ethical committees, it is difficult to accept that a tool designed for improving healthcare services would overstep ethical boundaries when those designed by marketing, banking and governmental institutions do not.

Popular socialised examples might include:

- **Search Engines:** Contextual analysis that is currently carried out by search engines, incorporate user's search history, location and other factors to both tailor and subsequently improve the results returned based on a user's anticipated questions.

- **IOT Devices:** Collect demographic, location, behavioural and endless other user settings/preference-based data (Boumil et al., 2012; Véliz, 2019). These details are on sold and cross referenced with other market data to companies that use this to increase targeted advertising.

    - **Law Enforcement:** Utilise data in the same manner as marketing departments, but for the subjective application of law against 'potential' criminality based on rules stipulated by the government of the day (McCue, 2010).

    - **Banking And Financial Auditing:** Cross reference data available for financial tests to facilitate more accurate loan schemes, financial auditing and more informed investment decisions (Hasan et al., 2020; Subrahmanyam, 2019).

Although the above methods are accepted as social norms, it becomes difficult under any circumstances to argue that a tool designed for improving healthcare services would in any way stipulate a breach in ethics, when the above examples do not. Additionally, to participate in user forums, various terms and agreements often require acceptance. In all circumstances where this is the case, users often waive ownership but this is of course dependent on the forum.

Furthermore, Ethics Committees:

- May consist of members with diverse backgrounds and perspectives, leading to subjective interpretations of ethical guidelines and variability in decision-making. This inconsistency could result in researchers facing different approval requirements or facing rejection for similar studies, depending on the composition and biases of the ethics committee.
- Can impose overly stringent regulations that stifle innovation and discourage exploration of novel research questions. The fear of facing rejection or lengthy approval processes may deter researchers from pursuing innovative studies that could yield valuable insights into communication dynamics and behaviours.

- May inadvertently discourage researchers from conducting studies involving sensitive topics or marginalized populations due to concerns about potential ethical implications or negative repercussions. This could limit our understanding of important social issues and hinder efforts to address inequalities and injustices through empirical research.

With these points considered, we summarily argue that the contribution of AI tools offset their detractors and given how socially accepted other industries are with their purveying of data, concern over studies such as this are a moot point.

1.7 Research Aims And Questions

We broke our study into five key aims:

I. **Model And Script Development**

Due to the absence of a classification tool, we are seeking to initially build one alongside pre-processing scripts to parse our data prior to training. Given the absence of available tools in the public space, we are further seeking to release this into the public sphere for other researchers on completion.

II. **Model Training And Comparison**

We are aiming to train our model using similar parameters to Yates et al. (2017) Reddit Self-reported Depression Diagnosis toolset for a degree of standardisation. This will ensure the use of a previously validated pre-labelled dataset relevant to our population.

As model optimisation can continue endlessly, and research into depressive models at the time of writing have not covered the following two parameters, we will focus on the following:

1. The number of posts per user.
2. Post length (words).

These parameters were selected for the following reasons:

a) **Dimensionality Reduction:** By incorporating post count and post length as features, it is possible to reduce the dimensionality of the input space. This reduction was chosen to improve the model's computational efficiency as large datasets and/or complex features can significantly increase training time.

b) **Predictive Power:** Post count and post length can serve as strong indicators of certain outcomes or behaviours. For instance, in social media sentiment analysis, longer posts might be associated with more nuanced opinions or detailed explanations, while higher post counts could signify increased engagement or interest in a particular topic. By leveraging these features, the binary classification model can capture valuable signals that contribute to better predictions.

c) **Interpretability:** Post count and post length are intuitive and easily interpretable features that provide insights into user behaviour. For stakeholders who may not have expertise in machine learning or natural language processing, these features offer a straightforward way to understand the model's predictions and rationale. Additionally, visualizations or analyses of these features can help identify patterns and trends in the data, enhancing the interpretability of the model's results.

III. Data Labelling

When parsing a new unseen Reddit dataset, we will seek to identify and label users dichotomously based on prior training, applying a depressed or non-depressed label for concurrent aims.

IV. Identify Keywords

We will seek to identify the strongest relationships (based on keywords) within our binary classifications rather than focusing on keyword frequency for the following reasons:

a) **Relevance Over Commonality:** TF-IDF prioritizes words that are both frequent within a document (term frequency) and rare across all documents in the corpus (inverse document frequency). This means that TF-IDF focuses on terms that are not only frequent but also distinctive or unique to individual documents. As a result, TF-IDF-based keyword selection emphasizes words that are more likely to capture the essence or topic of a document, rather than simply identifying commonly occurring terms. At the time of writing, this method has not been applied to the evaluation of keywords in depressive research.

b) **Handling Stopwords:** Frequency-based keyword selection methods may inadvertently prioritize common stopwords (e.g., 'and', 'the', 'of') that appear frequently but carry little semantic or contextual meaning. In contrast, TF-IDF downweighs such stopwords because they tend to have high document frequencies across the corpus, making TF-IDF-based keyword selection more robust to the presence of stopwords and other noise in the data.

c) **Addressing Document Length Discrepancies:** Frequency-based methods may bias keyword selection towards longer documents as they inherently contain more words and thus more occurrences of terms. TF-IDF mitigates this bias by normalizing term frequencies by the length of the document, ensuring that keywords are selected based on their proportional representation within each document rather than absolute frequency.

V. Evaluate User Labels For Changes/Differences:

We will evaluate user's initial post labels to obtain the primary classification of the original population. We will then evaluate user posts for label changes over time and whether these changes are significant between users. These will all be conducted for the following reasons:

a) **Baseline Assessment:** By evaluating user's initial labels, we establish a baseline understanding of the population's distribution of labels related to depression. This baseline provides crucial context for subsequent analyses and allows us to assess the prevalence of depression according to our classification criteria within the user population.

b) **Identifying Trends:** Examining changes in post labels over time enables us to identify trends or patterns in user behaviour and mental health status. Understanding how user's labels evolve or fluctuate over time provides insights into potential changes in their mental well-being such as improvements, deterioration or fluctuations in symptoms.

c) Comparative Analysis: Evaluating changes in post labels between users enables comparative analysis to identify differences in label dynamics across both user groups. This analysis can help uncover demographic, behavioural, or contextual factors that may influence changes in mental health status or response to interventions.

In doing so, we will simultaneously be seeking answers to the following research questions:

- How does the number of posts per user and/or post length affect our model's performance?

- How can TF-IDF be utilized to identify meaningful keywords for binary classification in depression research on Reddit?
- Do user’s post labels change over time, and are these changes significant between users?

1.8 Summary And Potential Contributions/Outcomes

This research proposal aims to expedite the screening process for depression while exploring novel approaches to data analysis in this domain. The systematic approach involves:

I. Model And Script Development:

- Developing a classification tool and pre-processing scripts for data parsing before training.
- Commitment to releasing the tool to for other researchers upon completion which at the time of writing, is unavailable.

II. Model Training And Comparison:

- Training the model using parameters similar to Yates et al. (2017) for standardisation.
- Focusing on two parameters - number of posts per user and post length - for dimensionality reduction, predictive power and interpretability.

III. Data Labelling:

- Dichotomous labelling of users as depressed/non-depressed based on prior training.

- Applying labels concurrently when parsing new, unseen Reddit datasets.

IV. **Identifying Keywords:**

- Emphasizing the identification of strong relationships based on keywords within binary classifications.
- Preferring TF-IDF-based keyword selection over frequency-based methods for relevance, handling of stop-words and addressing document length discrepancies.

V. **Evaluate User Labels For Changes/Differences:**

- Conducting a baseline assessment of user's initial post labels to understand the distribution of depression-related labels within the population.
- Analysing changes in post labels over time to identify trends, patterns and significant differences between users.
- This evaluation informs subsequent analyses, identifies improvements or deteriorations in mental well-being and uncovers factors influencing mental health status or response to interventions.

The results of this study will contribute to the field of both Computer Science and Psychology through the demonstration of:

- Reduced cost methods on a Governmental, Corporate and individual level for the treatment of depression.

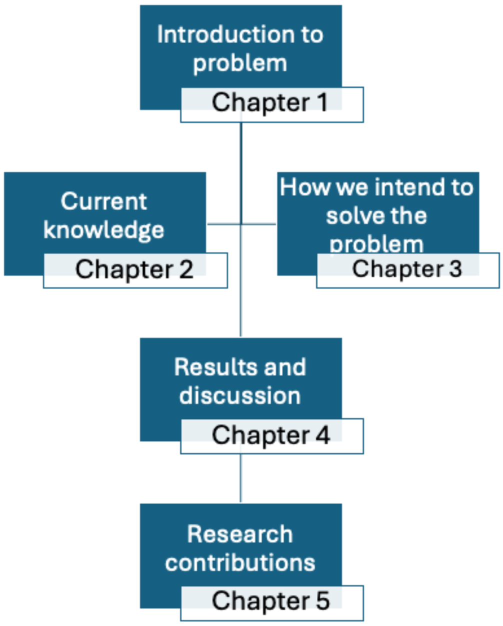
- More efficient 'time to treatment' for depression and a theoretical improvement in the associated health outcomes.
- Relevant model that can be redressed for other mental health issues.
- A novel approach to the problem of mental health issues.
- Unique approach to the value of words linked to depression.
- Knowledge contributions in the field of psychology and computer science through the quantitative analysis of social network data with a large sample size (n = >2,000,000)
- Establish further support for the use of self-reporting in healthcare research (Parimi et al., 2010; Sanchez-Villegas et al., 2008; Tlachac & Rundensteiner, 2020; Workman et al., 2019)
- Demonstrated use of automated analytical technology for the evaluation of depression.
- Demonstration of a more sensitive screening tool for the identification of clients experiencing depression.

Furthermore, improving current methods of psychological evaluation (as stipulated in the literature review summary) will have a profound impact on several communities if adopted (Corrigan et al., 2014; Parcesepe & Cabassa, 2013; Zhang et al., 2019). Not only will this research demonstrate a valid theoretical improvement in diagnostic efficiency, but the hypothesized reduction of time to treatment may improve quality of life by reducing the time to treatment and recovery time (Brotsky & Giles, 2007; Corrigan et al., 2014). Our suggested contribution also has a very low-cost barrier to implementation, for further research, can reduce the knowledge burden on healthcare providers and with further development improve the accuracy of current diagnostic tools (Andersson et al., 2021; Islam et al., 2018).

1.9 Thesis Structure

Figure 1

Thesis Structure And Design



This thesis has been broken into five chapters as outlined in Figure 1.

1.9.1 Chapter 1: Introduction To The Problem

This chapter serves as the introductory section of the thesis, setting the stage for the research that follows. It begins by discussing the concept of psychology as a social construct, highlighting the societal influences and perceptions that shape our understanding of psychological phenomena. The focus then shifts to psychological testing, with an emphasis on it’s significance and role in mental health assessment. The chapter further delves into the definition of depression, exploring it’s complexities and manifestations.

Moreover, it discusses depression screening tools, examining their development, reliability and validity in identifying depressive symptoms. The introduction also introduces the concept of artificial intelligence (AI) screening tools and their potential applications in mental health assessment. Additionally, the chapter presents an argument for the involvement of ethicists in the development and implementation of AI-based screening tools, addressing ethical considerations and implications.

Furthermore, the chapter outlines the specific aims and objectives of the research, providing readers with a clear understanding of the intended outcomes and contributions to the field. It concludes with a brief overview of relevant publications and the structure of the thesis, providing readers with a roadmap for navigating the subsequent chapters and sections.

1.9.2 Chapter 2: Current Knowledge

In this chapter, a comprehensive review of the existing literature related to the research topic is presented. The review begins with an examination of clinical interviews as a traditional method of assessing mental health, discussing their strengths, limitations and relevance in diagnosing depression. Following this, the validity of self-reporting measures in mental health assessment is explored, highlighting the reliability and challenges associated with relying on individual's subjective reports of their symptoms.

Moreover, the chapter delves into the emerging field of natural language processing (NLP) and it's applications in mental health screening. It discusses the potential of NLP techniques to analyse and interpret textual data, such as social media posts or online forum discussions, for identifying depressive symptoms. Additionally, the review identifies barriers to the adoption of algorithms in mental health assessment, including issues related to data privacy, algorithm bias and interpretability.

Furthermore, the chapter examines strategies for overcoming these barriers, such as improving transparency in algorithmic decision-making processes and addressing concerns related to data security and confidentiality. It also highlights the role of decision support systems in facilitating clinical decision-making and enhancing the accuracy of mental health assessments.

Lastly, the chapter compares the effectiveness of binary and multifactor NLP approaches for mental health screening, evaluating their respective strengths and limitations. By synthesizing and analysing the literature on these topics, this chapter provides a comprehensive understanding of the current state of research in the field of mental health assessment and establishes the groundwork for the subsequent chapters of the thesis.

1.9.3 Chapter 3: How We Intend To Solve The Problem

This chapter provides a detailed overview of the research methodology employed in the thesis. It begins by outlining the design and reasoning behind the research, elucidating the rationale for the chosen approach and it's alignment with the research objectives. Following this, the chapter discusses the datasets and data collection methods used in the study, providing insight into the sources of data and the procedures for gathering relevant information.

Moreover, the chapter presents the research tools utilized in both Study 1 and Study 2, highlighting their functionality and relevance to the research objectives. It then delves into the specifics of Study 1, focusing on model development and training. This includes pre-processing steps, model training procedures and methods for comparing model performance, providing readers with a comprehensive understanding of the technical aspects of the research.

Furthermore, the chapter details the research and analysis procedures undertaken in Study 2, emphasizing the steps involved in labelling posts, conducting text analysis and identifying keywords using scripts. It also discusses the use of mixed-effects logistic regression analysis to explore the relationship between variables and examine the factors influencing the occurrence of depressive symptoms in online forums.

Additionally, the chapter addresses issues of reliability, discussing the measures taken to ensure the validity and consistency of the research findings. It also acknowledges limitations and assumptions inherent in the research methodology, providing transparency and insight into potential areas of concern or bias.

Lastly, the chapter discusses ethical considerations related to the research, outlining the principles and guidelines followed to ensure the protection of participant’s rights and confidentiality. By providing a comprehensive overview of the research methodology, this chapter enhances the transparency and credibility of the study, laying the groundwork for the interpretation and discussion of the findings in subsequent chapters.

1.9.4 Chapter 4: Results And Discussion

Chapter 4 presents the culmination of the research endeavors, beginning with the unveiling of results from both Study 1 and Study 2. Study 1 outcomes focus on the performance and efficacy of the sentiment analysis model developed and trained in the preceding chapters. This segment scrutinizes the model's accuracy, performance metrics and noteworthy trends observed during evaluation. Following the presentation of Study 1findings, the subsequent paragraphs delve into a thorough examination and critique of these outcomes, grounding the discussion in the broader context of the research objectives and existing literature.

Moving forward, the discourse seamlessly transitions to the findings derived from Study 2, which include the outcomes of text analysis and mixed-effects logistic regression analysis. This phase unveils the identification of keywords linked to depressive symptoms in online forums and explores the intricate factors influencing their occurrence. The ensuing discussion section scrutinizes these revelations, dissecting their significance within the realm of mental health assessment and intervention. It examines the practical implications of the identified keywords, contemplates potential avenues for further exploration and reflects on the overall contributions of the study to the domain.

In summary, Chapter 4 encapsulates the research journey, showcasing the empirical findings and engaging in critical discourse to unravel their implications. From the performance evaluation of sentiment analysis models to the revelation of keywords indicative of depressive symptoms, this chapter provides a comprehensive synthesis of the research outcomes. Moreover, it underscores the significance of the study in advancing understanding in the field of mental health assessment and intervention, paving the way for future research endeavors and practical applications.

Chapter 5: Research Contributions

Chapter 5 serves as a synthesis of the research findings and offers concluding remarks, delineating the key insights gleaned throughout the study.

In the initial section, Chapter 1 and 2 are revisited, encapsulating the foundational concepts introduced in the introduction and literature review. This retrospective analysis contextualizes the subsequent discussions, reinforcing the trajectory of the research journey.

Following this retrospective, the focus shifts to a comprehensive overview of the results obtained in Chapter 4. This segment provides a succinct summary of the empirical findings derived from Study 1 and Study 2, encapsulating the performance of sentiment analysis models and the identification of keywords associated with depressive symptoms in online forums.

Subsequently, the chapter delineates potential avenues for future research, offering insights into promising directions for further exploration. These include investigations into information access and classification reviews, which hold promise for advancing the understanding and application of mental health assessment tools in digital environments.

Finally, the chapter concludes by synthesizing the key takeaways from the study and emphasizing their implications for both research and practice. It underscores the significance of leveraging technology and interdisciplinary approaches in addressing mental health challenges in the digital age, ultimately paving the way for future advancements in the field.

CHAPTER 2: LITERATURE REVIEW

As recent as 2018, 20% of Australia's population were affected by a mental health condition, with depression or depressive-like symptoms accounting for 10.4% (ABS, 2018). Descriptively, 62.1% of this group between 15-64 were employed, whereas those without depression saw employment rise to 79.5%. Internationally, Weissman et al. (1996) identified that despite the observation of different ratios between countries, the mean age of depression onset fell between 24.8-34.8 years, with rates of major depressive disorder (MDD) higher in women than men and higher in high versus low-income countries (n=30,000). This was further corroborated using a diagnostic interview developed by the World Health Organisation, with Bromet et al. (2011) later contributing that mean lifetime and 12-month prevalence was 14.6% and 5.5% in higher income countries respectively, versus 11.1% and 5.9% in lower income countries (Andrade et al., 2003).

Although quality of life is generally believed superior in higher income countries, this is also dependent on how it is measured/reported, with Kahneman and Deaton (2010) finding correlations between income, education and 'life-evaluation', but not acute 'emotional experiences' (n = 450,000). Although perhaps ambiguous, when reconceived as 'longitudinal' and 'acute experiences', higher income countries appear to offer better longitudinal experiences at the expense of the acute. Furthermore, this study helped to establish support for the theory that annual income does not contribute to overall happiness beyond ~$75,000USD (in 2010) but is subject to cost of living changes (Kahneman & Deaton, 2010; Morris et al., 2021; Rizal et al., 2022; Zhang & Xiang, 2019).

Irrespective of these issues, most healthcare models require those suffering depression seek their own counselling and treatment. Although this is consistent, it presents as a central issue due to the following:

- Depression effects over 10% of people worldwide, is chronic and is one of the leading catalysts for disability and morbidity (Gao et al., 2018; McLaughlin, 2011; Tao et al., 2021).
- Depression develops subtly until its manifested as MDD, substantially increasing suicide risk whilst also reducing help-seeking behaviour due to associated agoraphobia (McGarry et al., 2009).
- Stigmatisation causes individuals to avoid diagnosis/treatment, further compounding the risk of MDD (Parcesepe & Cabassa, 2013; Zhang et al., 2019).

Additionally, these points were further exacerbated by the emergence of COVID-19, with many countries underequipped for a surge in MDD and anxiety (Santomauro et al., 2021).

Depression can be further broken down into:

- Individual level detractors:

- Reduced work performance, employment, earnings, marital quality, social interaction and increased mortality (Santini et al., 2020; Weissman et al., 1996).

- Societal level detractors
  - Across the US in 2018, depression cost approximately US$36.2 billion in medical care alone and significantly increased primary and secondary care requirements (Proudman et al., 2021).
  - Research suggests that the economic burden associated with suicide amounts to US$13.4 billion, constituting 4% of the overall economic impact on adults with Major Depressive Disorder (MDD) (Greenberg et al., 2021).
  - Depressed employees exhibit a 2% productivity loss over their counterparts with an annual financial cost difference of approximately US$13billion (Greenberg et al., 2003; Stewart et al., 2003)

With this knowledge, it is debatable whether current healthcare interventions/methods can/are adapting fast enough to accommodate this population and given the associated costs, whether an automated system makes more economic sense on a governmental level than an increase in practicing professionals.

2.1 Clinical Interviews

Clinical interviews are commonplace within the field of psychology, for the diagnosis of mental health disorders. Processes up to the determination of psychologists, will stipulate whether these interviews take place on an unstructured, semi-structured or fully structured platform; with the intention of yielding more accurate information. Irrespectively, the complexity of social interactions is difficult to control for and although testing might occasionally prove reliable and accurate, factors that can influence diagnosis still include:

- Rapport
- Honesty
- Diagnostic tools/tests
- Personal beliefs and values
- Drug use
- Assumptions

Irrespective of Diagnostic and Statistical Manuals assumed validity; it is a social/legal construct. As defensible as it's evidence suggested in the past, several conflicts left DSM-4 facing difficulties in repeatability, resulting in it's full review (Hyman, 2010). These circumstances included:

- Significant differences in rater-agreement, ranging from .3 to .77 on certain illnesses, but conversely by less than .03 on others (Solanto & Alvir, 2009).
- Questionable assumptions related to drug induced states and mood disorders whereby:
    - Induced mania and hypomania was classified as a 'mood disorder' under drug induced 'specifiers'; whereas depression was not, despite both drug and alcohol use demonstrably causal (Boden & Fergusson, 2011; Joyce, 2008, p. 857; Peck et al., 2005).
    - Mood disorders were assumed to be episodic, despite the acknowledgement of Major Depression as having increased severity and a chronic diagnosis, escaping this through independent classification (Joyce, 2008, pp. 855-856).

DSM-5 however superseded this edition with less conflict. Since it's release, it has improved it's consistency and uniformity in psychiatric conditions across clinical settings, facilitating more effective communication between healthcare professionals, researchers and clinicians through more common language. It has not experienced the same inquiries to date.

Clinicians (specifically psychologists and psychiatrists), now heavily rely on DSM-5 to guide treatment planning and interventions. The more detailed diagnostic criteria assist healthcare professionals in formulating effective and evidence-based treatment strategies tailored to specific diagnoses. Moreover, it plays a crucial role in research and clinical trials by providing more standardized criteria for participant recruitment and outcome assessment. It's use promotes collaboration among multidisciplinary healthcare teams, facilitating communication among professionals with diverse backgrounds and expertise in mental health.

In addition to it's clinical applications, it also serves as a valuable educational tool for mental health professionals, with the manual's acceptance and use contributing to the continuity across training globally. Furthermore, DSM-5 supports continuity of care by providing a standardized framework for documenting and communicating diagnostic information. This ensures that individuals receive consistent and comprehensive care as they move between different healthcare providers or settings. Lastly, DSM-5's influence extends to insurance and reimbursement processes, where adherence to it's criteria facilitates accurate diagnoses for coverage determination, enhancing accessibility to mental health treatment.

Consequently, despite it's initial caveats, the diagnostic statistical manual upon which clinical interviews rely, has progressively built upon previous iterations as a commonly accepted standard in mental health practice. Although it has contained conflicting classification methods in the past due to the complexities of it's subject, we posit that computerised solutions could only offer further augmentation.

2.2 Validity Of Self-Reporting

Epidemiological studies evaluating the efficacy of self-reporting in medically compromised populations are mostly based on symptom scales, introducing arbitrary cut-offs determined by characteristics subject to significant variation (age, pathology, depression type) (Sanchez-Villegas et al., 2008). Although the variation between testing methods might affect continuity, self-reporting has proven useful among population studies due to it's ease of use and application across fields (Sanchez-Villegas et al., 2008; Varraso et al., 2007).

To date, studies have focused on the efficacy of self-reported depression given the necessity cited by many for a preventative healthcare model (Sanchez-Villegas et al., 2008; Tlachac & Rundensteiner, 2020). So far, the majority of studies have seen support for self-diagnosis; with Sanchez-Villegas et al. (2008) identifying 74.2% of participants self-diagnosed correctly against formal evaluations. Tlachac and Rundensteiner (2020) similarly reported a true positive rate of ~74% against the patient health questionnaire (PHQ-9), with medical investigations into physical ailments demonstrating even greater accuracy in the reporting of hip aetiology (~95%) and sinusitis (98%) (Parimi et al., 2010; Workman et al., 2019).

Although it might be argued that self-diagnosis could prove terminal in physical ailments specifically in the case of false negatives, the same cannot be said of mental illnesses. Nor is it suggested that self-diagnosis be depended upon for official diagnosis.

Stigmatisation has been a persistent reason for the avoidance of mental-health diagnosis cross-culturally (Parcesepe & Cabassa, 2013; Zhang et al., 2019). Although the exact reasons are beyond the scope of this review, persons with mental illnesses often shift towards online communities, where their experiences can be shared with greater anonymity (Corrigan et al., 2014). According to Brotsky and Giles (2007). This is not just mediated by the ability of the individual to manage disclosures with agency, but the possibility of meeting a like-minded group making absent the perceived threat of stigmatisation that formal consultations bring.

With the growing presence of online communities and ambiguity of mental health diagnosis, it could also be argued that there is a risk many self-selecting groups will reinforce each other’s views, leading to an increase in false positives (Joyce, 2008, p. 857). Irrespectively, due to possible or perceived stigmatising barriers that prevent many seeking help, we suggest that if mental health diagnostics were made more accessible online, it might attract new clients that would otherwise avoid care entirely (Lanseng & Andreassen, 2007; Zhang et al., 2019)

Although the argument for self-reporting as a valid interface might still seem paradoxical, the concept that individuals are seeking care online, is indicative of an unfulfilled need. Given the compounding factors involved in depression and the recurrent theme of societal withdrawal, we feel the validity of using a self-diagnostic scale for the analysis of depression is significant enough to justify support.

2.3 Natural Language Processing

Natural language processing (NLP) combines the fields of computer science, artificial intelligence and linguistics in an effort to read and interpret human language (Luo & Chong, 2020). Simplistically, it applies algorithms to discern meaning from the structure of sentences and is best utilised when applied to problems involving explicit rules. Recent developments however have extended it’s capacity to the analysis of images (statistical NLP) and audio, enabling it’s utility to expand into medical imaging, report analysis, triage and drug evaluation among others (Cho et al., 2022; Leightley et al., 2020; Steinkamp & Cook, 2021; Yin & Wong, 2021).

Other features of NLP include:

- **Sentiment Analysis:** Can determine the emotional valence of text, aiding applications in social media monitoring, customer feedback analysis and market research.

- **Language Modelling:** Can predict word sequences likelihood through speech and written text.
- **Text Classification:** Can assign predefined categories to documents, supporting spam detection, topic categorization and sentiment analysis.
- **Machine Translation:** Automatically translates text across languages, promoting global accessibility and cross-language communication.
- **Information Retrieval**: Involves extracting relevant information from large datasets, benefiting search engines, document summarization and question-answering systems.
- **Speech Recognition:** Transforms spoken language into written text, contributing to voice assistants, transcription services and accessibility tools.
- **Question Answering Systems**: Generate contextually relevant responses to user queries, applicable in virtual assistants and information retrieval.
- **Conversational Agents (Chatbots, Virtual Assistants Etc):** Engage in natural language conversations, further extending NLP's impact across diverse applications.
- **Large Language Models (LLMs):** Similar to conversational agents, these models take in larger amounts of textual data and can understand and generate human-like text with more intricate patterns and contextual nuances. At the time of writing, these tools

are in their infancy however, LLM's such as GPT-4, are showing the most promise relative to other methods, with more accurate sentiment analysis, text generation and conversational AI results than other methods (Zahid et al., 2024).

Despite these solutions appearing as stand-alone programs however, they are more often used synergistically, empowering various programs with the capacity to navigate the complexities of human language in various contexts more effectively.

Although Western medicine is somewhat resistant to technological creep largely derived from the Hippocratic oath, it is argued much of the resistance is trivial given modern healthcare's extension beyond direct caregivers into administrative and payor organisations (Harris, 1997). Given the availability of information as a result of the internet, privacy arguments face greater difficulties given machine learning has demonstrated the ability to identify disorders, both less invasively and with similar accuracy, to skilled professionals (Andersson et al., 2021; Islam et al., 2018). Furthermore, it's contention for the evaluation of communications data is second to none due to it's omnipresent nature, retrospective capabilities and elimination of social barriers (rapport, omission etc).

In contrast to individualised psychological screening, natural language processing also has the capacity to be performed on enormous, dynamic, input-based datasets occurring in real-time, at a very low cost.

Atop these reduced costs, it would also relieve the manpower burden among corporate entities and can assist more specifically within the critical time frame major depressive disorder takes to develop – leading to earlier interventions.

2.4 Barriers To The Adoption Of Algorithms

Algorithms, whether individuals are conscious or unaware of their modalities, are present within the majority of software. Although this is known within the computer industry, users without this knowledge have distinct privacy concerns, often centring around metadata and how it is stored/used. Uniform Resource Locators inherently deconstruct our privacy expectations, providing valuable geolocation, time and descriptive information to third parties; yet are utilised regularly by anyone using the internet (Ferreira & Aguiar, 2019). This is not to say that more specific information is not protected by robust security protocols such as encryption, which facilitated online banking, but that the concept of internet privacy seems to be often applied erroneously (Bhatt et al., 2021, pp. 132-144).

It is well publicised that Alphabet uses Google accounts, which total nearly 1.8 billion, to form digital identities for personalised advertising, collecting data through web-browsers and mobile devices when users are signed in (Google, 2022). Although the same process occurs for Apple, who alongside Alphabet dominates this market space, much of their information falls under patents, legitimising their security method's obfuscation under data protection arguments. Despite this and relevant to this study, users not only persist in their use of social media irrespectively, but sign terms of agreement willingly (Facebook, 2022; Reddit, 2020; Twitter, 2022).

Beyond security and privacy issues, professional barriers will understandably arise within fields that see technological creep as a threat. Given the capacity of machine learning to evaluate different variables with higher accuracy than professionals, it is foreseeable these skills might also be feared redundant with time (Cheng et al., 2021; Hassan et al., 2021).

2.5 Overcoming Barriers To Algorithms

Understandably, NLP poses challenges and concerns related to privacy and efforts are often made to address them in several ways:

- Anonymisation/Pseudonymisation: Across most industries, personal information in text data can/is anonymized or pseudonymized, mapping identifiable details to more generic terms. This ensures that individuals cannot be readily identified through the language data being processed.
- Metadata: NLP applications aim to collect and process only the minimum necessary information to perform a specific task. By minimizing data collection, the risk of exposing sensitive or personally identifiable information is reduced.

- Consent and Transparency: Users are often informed about how their language data will be used in the terms and conditions on most websites and are prompted to provide explicit consent, giving users more transparency/control.
- Security and storage: Most countries have stringent laws on data storage, such as the GDPR (General Data Protection Regulation) or HIPAA (Health Insurance Portability and Accountability Act), with encryption, access and data transfer protocols mitigating against potential breaches.
- Regulatory Compliance: Adherence to privacy regulations and standards, ensures that NLP applications comply with legal frameworks that protect user privacy.

Our research only serves to demonstrate what can be done with NLP, with the aim of socialising the integration of more effective algorithms into current software, ideally to increase the screening of clients that are currently overlooked. Objectively, we are seeking to enable psychologists to render services earlier and highlight other areas of interest for relevant research. If adopted, this may help to reduce the onset of MDD and it’s consequences.

Itemised in our research, we aimed to provide evidence to:

- Improve the rate of screening in online forums.

- Decrease the time required for the evaluation of depression in clinical interviews.
- Demonstrate a capacity to reduce time to treatment.
- Contribute to diagnostic accuracy.
- Identify new key areas for research.

We are currently aware that social media effects mood negatively, primarily through upward social comparison and fear of missing out (Fabris et al., 2020; Pang, 2021). Given social media's cross-cultural pervasiveness and relevance to this study, we are also seeking to contribute to a framework by which platforms seek to reduce these factors.

Although drug interventions also exist, behavioural interventions have superior long-term outcomes with minimal side effects (David et al., 2008). Therefore, we also posit that our endeavours, in addition to those within this field of research, places patients at less risk.

2.6 Decision Support Systems In Focus

Various real-life examples of companies embracing text analysis for mood disorder screening exist, with Johns Hopkins University citing an increase in the use of machine learning algorithms to analyze social media posts for signs of depression and anxiety (Kim et al., 2021). Over the course of their study, Kim et al. (2021) retrieved publications on the intersection of social media and machine learning (ML) in mental health from the Scopus and Web of Science databases. Their analysis focused on understanding the distribution of publications to gauge productivity across sources, countries, institutions, authors and research subjects through the generation of a keyword co-occurrence network to visually represent the evolving trends in this field. As a result of this analysis, 565 papers spanning from 2015 to 2020 were identified suggesting consistent growth over this five year period; with the most prolific sources being Scopus and Web of Science records, Lecture Notes in Computer Science and the Journal of Medical Internet Research (Kim et al., 2021).

Various attempts at AI support systems for Mental Health are currently in development, with Tutun et al. (2023) highlighting several investigations that focus on the developed Decision Support Systems (DSSs) tailored for the diagnosis and treatment of mental disorders. The Computerized Texas Medication Algorithm Project serves as an example, supporting decisions related to major depressive disorders and integrating seamlessly into clinical settings (Trivedi et al., 2004). Razzouk et al. (2006) introduced the SADDESQ, a clinical DSS for diagnosing schizophrenia spectrum disorders based on variables like psychosis symptoms and seizure frequency, whilst the Sequenced Treatment Alternatives to Relieve Depression is another DSS aiding doctors in determining optimal medication dosage and timing (Sinyor et al., 2010).

Beyond these studies, endeavours to design and integrate AI-based DSSs for more accurate mental disorder diagnoses have also emerged, with Stewart et al. (2020) introducing a DSS based around tree-based learning to identify children at high risk of suicide. Chen et al. (2020) utilizing deep learning to screen and score dementia patients. Zhang et al. (2020)

employing biological markers and genetic data for an early diagnosis of mental disorders through a deep learning framework. Mueller et al. (2011) detected ADHD using an SVM model based on responses to the BSI tool; and Nie et al. (2012) diagnosed mental disorders utilizing an SVM model constructed via the SCL-90-R tool.

In the realm of healthcare companies, IBM Watson Health has embarked on pioneering endeavours in the creation of Natural Language Processing (NLP) powered tools dedicated to mental health screening. The multifaceted approach involves harnessing NLP techniques to scrutinize vast datasets of patient records, in a time efficient manner in which professionals cannot compete, discerning intricate linguistic cues intricately linked to mood disorders (van Hartskamp et al., 2019). By tapping into the wealth of textual information within these records, IBM Watson endeavours to augment the capabilities of healthcare professionals by providing a nuanced understanding of patient's mental well-being. This innovative application not only facilitates early detection of potential mental health issues but also empowers healthcare practitioners to intervene promptly, thereby enhancing the overall efficacy of mental health care delivery. By integrating advanced language processing and healthcare analytics, IBM Watson Health contributes significantly to the ongoing paradigm shift in mental health diagnostics and treatment.

Lesser-known companies such as Woebot, have also harnessed the power of NLP for therapeutic interactions through a chatbot interface. Operating as a virtual conversational agent, Woebot engages users in meaningful dialogue, monitoring their language patterns. Using sentiment analysis algorithms, the platform discerns emotional cues and provides personalized support and resources for effective emotional well-being management; highlighting the tangible impact of NLP in democratising mental health support, offering a scalable and accessible solution to a broader demographic. Woebot's approach not only exemplifies the seamless integration of technology into mental health care but also underscores the potential of NLP to enhance user experiences and outcomes in the realm of emotional well-being. Furthermore, it has the capacity to reduce vicarious trauma in clinicians (Kounenou et al., 2023)

Despite the ongoing efforts in the development of AI-based DSS for mental disorder detection, significant gaps persist both in the literature and on a clinical level. Many existing diagnostic tools are still paper-based, necessitating individuals visit clinics in person with the analysis process often requiring extensive manual effort from mental health professionals. Additionally, a limited number of studies explore the intricate relationships among variables employed to assess mental health. Most of the studies so far have been dedicated to specific disorders, lacking a comprehensive approach that considers multiple disorders simultaneously. Moreover, ethical design elements are frequently overlooked in the creation of these AI-based DSS.

2.7 Current Sentiment Analysis Tools For Screening Depression

Research into the field of sentiment analysis and mood disorder screening has been more generalised up until now, with the refinement of current machine learning models taking priority over specific disorders (Gkotsis et al., 2017; Yates et al., 2017). With the advent of LLM’s, more advanced techniques have been introduced such as transformer-based models like BERT and GPT. As these have been able to identify more nuanced differences across a wider range of data sources, it appears unlikely resources will be reallocated towards specific disorders in the near future (Rathje et al., 2024).

Furthermore, given the commercial interest, it is also unlikely that the knowledge and tools created by this competition will be made available in open source products (Salas-Zárate et al., 2022; Yates et al., 2017).

2.8 Binary Versus Multifactor NLP For Mental Health Screening

Binary and multifactor natural language programs designed for mental health screening exist and cater to distinct purposes, each carrying their own unique set of advantages.

Binary Natural Language Program

Binary language programs are significantly easier to write as well as implement. They categorize responses into two distinct categories, such as positive or negative and can be quickly interpreted. One of the key advantages is their cost-effectiveness, with the streamlined nature reducing algorithmic complexity, subsequently minimizing development time and resource requirements. Binary systems excel in swift identification, enabling the rapid categorization of individuals who may require immediate attention. This attribute is particularly vital in mental health scenarios where timely intervention is imperative.

The caveat of binary systems however comes with the increased generalisation of diagnosis (Straw & Callison-Burch, 2020). Upon scrutinizing mental health terminology within GloVe and Word2Vec embeddings, Straw and Callison-Burch (2020) identified distinct biases across various dimensions including religion, race, gender, nationality, sexuality and age. Across 52 papers, none comprehensively addressed the identified bias areas during model development. Furthermore, only a solitary article appeared in more than one research database, underscoring the entrenched isolation of research within disciplinary silos, posing a hindrance to fostering cross-disciplinary collaboration and effective communication (Straw & Callison-Burch, 2020).

Multifactor NLP Analysis

Multifactor analysis stands out for it’s capacity to undertake a more holistic assessment. By considering various dimensions and factors, it provides a more nuanced evaluation of an individual's mental health, potentially leading to a more accurate understanding, facilitating more tailored recommendations and interventions based on specific needs. Although there are several benefits to multifactor methodology, primarily it reduces the likelihood of false positives and negatives. This stands in contrast to binary systems, which may oversimplify the screening process, potentially leading to misinterpretation of responses, leading to concurrent misdiagnosis and potentially negative patient outcomes. Multifactor programs are also well-suited for prolonged monitoring and tracking of mental health trends, possessing the ability to adapt to changes in an individual's condition over time to offer a more dynamic/evolving assessment.

Fit For Purpose

The differences in implementation however may demand more extensive resources, professional oversight and ongoing development and computational power. Organizations must carefully weigh these resource demands against the benefits obtained. Additionally, binary programs can offer higher user engagement for corporate implementation due to their simplicity.

In conclusion, the selection between binary and multifactor natural language programs for mental health screening hinges on the specific goals, available resources and preferences of the implementing organization or individual. Both approaches bring their own merits, with the optimal solution striking a balance between simplicity and a comprehensive understanding of mental health.

For our study, a binary classification was deemed more appropriate as our intention was not to diagnose, but screen for depression. With the end goal of increasing the rate of clinical consultations, this classification model made more sense, with false positives posing no risk given the identification would be substantiated by the professional's post-analysis.

CHAPTER 3: RESEARCH METHODOLOGY

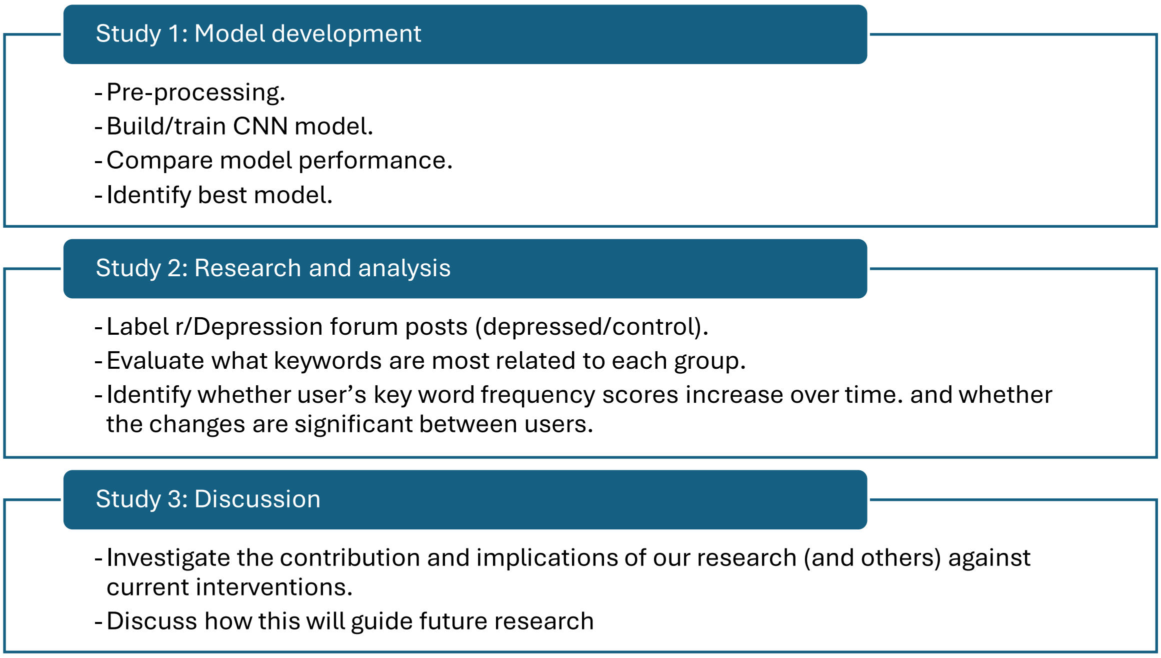
In the following section, we delve into our methodology, providing a comprehensive overview of how we will conduct our study. This serves as a detailed roadmap, offering insight into the various steps we will take to address our research questions. Foremost, we outline the research designs employed in our study, elucidating the overarching framework, followed by the specifics of our data collection methods and how we gathered the necessary information to conduct our analysis sectionally. Furthermore, we provide insight into the characteristics of our study participants, elucidating their inclusion criteria and recruitment procedures, finally demonstrating the analytical techniques, ranging from statistical methods to qualitative data analysis procedures. By detailing our analytical framework on this basis, we believed our readers would gain better insight into how we derived our conclusions and interpreted our findings.

In essence, this methodology section serves as a foundational component of our research, laying the groundwork for the subsequent results and interpretation. Through a meticulous level of documentation and transparency, we felt we could further ensure the rigor and integrity of our study, fostering confidence in the validity and reliability of our findings.

3.1 Design And Reasoning

Figure 2

Flowchart Of Our Overall Research Model



The overall structure of the study will be broken down into three smaller parts with the following reasoning (see Figure 2).

Study 1: Model Development

Table 1

Score Achieved By Yates Et Al. (2017). From: Yates, A., Cohan, A., & Goharian, N. (2017). Depression And Self-Harm Risk Assessment In Online Forums, Pg. 7.

Method	Precision	Recall	F1
BoW - MNB	0.44	0.31	0.36
BoW - SVM	0.72	0.29	0.42
Feature-rich - MNB	0.69	0.32	0.44
Feature-rich - SVM	0.71	0.31	0.44
FastText	0.37	**0.70**	0.49
User model - CNN-E	0.59	0.45	0.51
User model - CNN-R	**0.75**	0.57	**0.65**

In the initial phase of Study 1, our focus is on developing pre-processing scripts followed by the building/training of a Convolutional Neural Network (CNN) model. Although other methods exist, CNN models significantly outperformed their alternatives in the RSDD (Table 1). As our research necessitates the development of a model to be used in Study 2, and we will be building ours around similar parameters on the same dataset, a CNN is more suitable given the time constraints.

Once the CNN model is trained, we will proceed to compare it’s performance against alternative models. This comparison allows us to evaluate the effectiveness of different models in achieving our research objectives, ultimately comparing the influence of our chosen metrics of interest (post-length and number) for accuracy and reliability.

Subsequently, we will identify the most promising model based on the highest F1 score, to be used on the classification task in Study 2.

Study 2: Research And Analysis

In Study 2, our attention shifts towards the classification of forum posts from the r/Depression community into dichotomous categories: depressed and control. This classification enables us to differentiate between posts made by individuals exhibiting depressive symptoms and those made by individuals who do not. Following classification, we undertake an evaluation of the keywords most closely associated with each group. This analysis provides valuable insights into the distinguishing characteristics of depressed and non-depressed individuals within the online community based on textual input.

Additionally, we investigate whether user's post labels change over time and whether these changes are statistically significant between users. By tracking these changes, we aim to discern patterns indicative of shifts in user's mental states and behaviours. This analysis contributes to our understanding of how individuals engage with online platforms and the potential implications for their mental health.

Study 3: Discussion

Lastly, in the systematic review component of our research, we delve into investigating the contributions and implications of our findings against current interventions for depression. This will allow for the situation of our research within the broader context of existing literature and identify gaps or areas for further exploration. Furthermore, we discuss how our research insights can guide future investigations and inform the development of more effective interventions for depression management.

3.2 Datasets And Data Collection

Both studies utilise different datasets from Reddit, through an agreement between Yates et al. (2017), Reddit and the University of Southern Queensland. We felt it was best to break down their use into study-based categories for interpretation.

3.2.1 Study 1

The dataset for this study was selected due to it’s use within the Reddit Self-reported Depression Diagnosis (RSDD) developed by Yates et al. (2017), for the classification of depression based on self-reported data from Reddit forum posts. This allowed us to use a pre-validated and pre-labelled dataset for training, facilitating research completion within the projected time constraints.

The pool of potential control users within this dataset were identified by selecting only those users who had never posted in a subreddit related to mental health and never used a term related to depression or mental health. For each diagnosed user and potential control user, the probability of posting in each subreddit was calculated. Diagnosed users were matched with control users who had the smallest Hellinger distance between their subreddit post probability distributions, excluding control users with 10% more or fewer posts than the diagnosed user. This approach ensured that diagnosed users were matched with control users who were interested in similar subreddits and had similar activity levels, preventing bias based on the subreddits users were involved in or based on how active the users are on Reddit. The RSDD dataset differs from prior work creating self-reported diagnoses datasets in several ways: it is an order of magnitude larger, posts were annotated to confirm that they contained claims of a diagnosis and a realistic number of control users were matched with each diagnosed user (Yates et al., 2017).

The RSDD was also chosen due to its increasingly standardised means in screening for self-reported depression in Reddit datasets and it's inclusion of pre-classified Reddit posts from ~9,000 users (diagnosed as depressed) alongside ~107,000 matched controls. Posts related to mental health, or featuring depression-related keywords, were excluded from the diagnosed user's dataset, while a selection process was applied to control user's data to ensure the absence of such posts.

Explicit Details

The cohort of users with self-reported depression within Yates et al. (2017):

- Possess a post featuring a specific diagnosis pattern (e.g., "I received a diagnosis of") in conjunction with a mention of depression and do not transgress any exclusion criteria.

Exclusion standards were applied at both the user and post levels. On the user level, those with fewer than 100 posts in non-mental health subreddits prior to their self-reported diagnosis post are excluded. At the post level, two exclusion stipulations are implemented:

- References to depression needed to fall within 80 characters of the section of the post that corresponds to a diagnosis pattern; posts failing to meet this criterion were excluded.
- Any mention of depression was recognized as the occurrence of particular terms like "depression", "depresion", "depressive disorder", "major depressive", or "mild depressive."
- Posts matching a negative diagnosis pattern (e.g., "mother was diagnosed with") were dismissed.

Users meeting both the diagnosis pattern and exclusion criteria are considered for inclusion. Annotators from the crowd assessed these posts to ascertain if the user genuinely asserted a diagnosis of depression.

A mental health (MH) post is any post created in a mental health-related subreddit or matching a mental health pattern. All MH posts from diagnosed users were excluded and users without MH posts became potential control users.

Control users were selected by aligning candidates with diagnosed users. Each diagnosed user is paired with the 12 control users exhibiting the smallest Hellinger distance between their subreddit post probability distributions, excluding those with 10% more or fewer posts than the diagnosed user.

3.2.2 Study 2

Reddit provided and approved the use of it's user data for our research and approval had also been cleared by University of Queensland's Ethics Committee, prior to it's use. The protection of user privacy and anonymity was prioritised in the handling of this dataset, with usernames replaced with unique identifiers, ensuring the preservation of anonymity for temporal evaluations. Moreover, all personal identifying information within the dataset were either removed or obfuscated before it's receipt for this study.

Approximately 2,000,000 datasets from the user forum 'r/depression' were included in the analysis. The substantial volume of data allowed for in-depth exploration and insights into patterns related to depression and associated linguistic expressions within the online community. These measures collectively uphold the ethical considerations of privacy while facilitating valuable research into mental health aspects in online forums.

The dataset included key fields including 'uid' (unique identifier), 'id,' 'anon' (anonymized user data), 'created_utc' (timestamp), 'retrieved_on' and 'posts.' The inclusion of 'posts' allowed for the primary analysis of linguistic expressions associated with depression within the 'r/depression' forum; with all inclusions handled with consideration for privacy and ethical standards, aligning with the approved research protocols. 'Anon', ‘created_utc’ and ‘posts’ were the only variables utilised for the purposes of this study.

It is through this dataset that we intend to test our most successful model post-training and discuss it’s results along with further evaluation.

3.3 Research Tools

Across all studies, a common thread emerged in the tools utilized, encompassing a range of methodologies and approaches. These tools, essential for the research endeavours, played a pivotal role in data collection, analysis and interpretation. Their consistent application ensured methodological coherence and facilitated comparative analyses between studies. From pre-processing scripts to advanced machine learning algorithms, the tools employed spanned the spectrum of research techniques, reflecting the interdisciplinary nature of the investigations. As integral components of the research process, these tools contributed to the rigor and validity of the findings, underscoring their significance in advancing knowledge within the respective fields of study:

- 2020 Macbook (M1 chipset)
- Anaconda Data Science Platform
- Xrealstats (Mac)
- Spyder (v5.4.3)

- Microsoft Excel

3.4 Study 1: Model Development

As our study did not include a current CNN model for the labelling of depressive symptoms, our initial task was to design a novel algorithm that could fulfill our research purposes. The dataset selected to train our model was provided by Yates et al. (2017) RSDD, in agreement with the University of Southern Queensland.

Due to the increased difficulty for interpretation when evaluating JSON files, as well as the study's demand for lower training times, we opted to pre-process our dataset prior to our model's training.

3.4.1 Pre-Processing: Script 1

We initially converted the dataset into a CSV format for increased legibility. For each user, it extracted the first 40 posts along with their 'ID' and 'label', storing this information in a structured list with reduced contents, to reduce processing time during model training.

Details

1. **Prompt for JSON File Path:** The script began by asking the user to input the path to the JSON file containing user data.

2. **Initialise Data Storage:** An empty list called `users_data` was initialised to store the extracted user data.

3. **Read JSON Data:** The script opened/read the JSON file specified by the user, iterating through each line of the file, loading the JSON data and appending it to the 'users_data' list.

4. **Handle File Not Found:** If the specified JSON file was not found, the script would print an error message, exiting the program.

5. **Prepare CSV Data Structure:** A list called 'CSV_data' was initialised with headers for 'Posts', 'ID' and 'Label' to structure the CSV file.

6. **Extract User Data:** For each user in the 'users_data**'** list, the script extracted the first 40 posts, concatenating them into a single string and appending this string along with the user's ID and label to the 'csv_data' list.

7. **Write CSV File:** The script opens a new CSV file in write mode and writes the data from the 'csv_data' list into the file.

3.4.2 Pre-Processing: Script 2

This script read data from script 1, removing unlabelled rows. When completed, it printed out the count of samples before and after deletion, along with the number of samples deleted; finally indicating the number of samples not equal to 'control' or 'depression' that were deleted from the CSV file.

Details

1. **Prompt For The Input And Output File Path:** The script asks for input and output paths for the CSV files.

2. **Open Input And Output CSV Files:** The script opened the input CSV file for reading and the output CSV file for writing.

3. **Read The Header Row Of The Input CSV File:** The header row of the input CSV file is read and stored.

4. **Initialise Counters:** Variables to count the number of samples before deletion and the number of samples deleted are initialised.

5. **Iterate Through Each Row Of The Input CSV File:** Iterates through each row of the input CSV file, checking the label in the third column.

6. **Write Header Row To The Output CSV File:** The header row of the input CSV file is written to the output CSV file.

7. **Check If The Label Is 'Control' Or 'Depression':** If the label in the current row is 'control' or 'depression', the row is written to the output CSV file and the count of samples before deletion is incremented.

8. **Check If The Label Is Not 'Control' Or 'Depression':** If the label in the current row is not 'control' or 'depression' and the number of samples deleted is less than the specified limit, the count of samples deleted is incremented and the row is not written to the output CSV file.

9. **Return Counts Of Samples Before Deletion And Samples Deleted:** After iterating through all rows, the counts of samples before deletion and samples deleted are returned from the function.

10. **Count samples in the input CSV file:** The script counts the total number of samples in the input CSV file.

11. **Prompt The User For The Number Of Samples To Delete:** Prompts for the number of samples to delete that are not 'control' or 'depression'.
12. **Delete Samples:** The function to delete unlabelled samples is called with the input CSV file path, output CSV file path and number of samples to delete as arguments.
13. **Print Summary Information:** Prints the total number of samples in the input CSV file, the number of samples deleted and the number of remaining samples after deletion. It finishes with a printout of the number of samples deleted from the CSV file.

3.4.3 Model Training

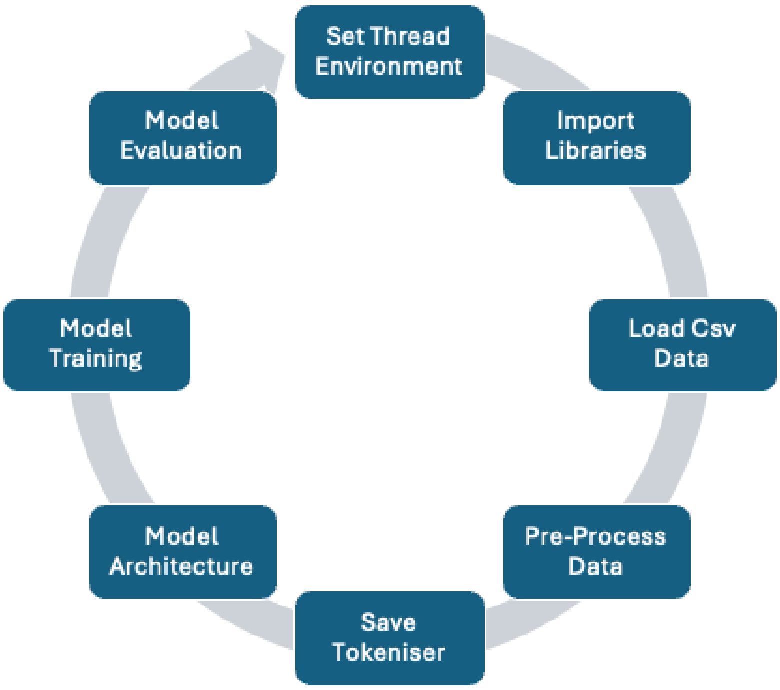
The primary script used for this study was created in Python using a thread configuration for Apple's M1 chipset to improve parallel processing speed, with interoperation and intraoperation threads set to 8 and 16 respectively. Data sampling was set to 50, 100 and 150 words per post (wpp), with Random Forest Technique used across 1000, 1250, 1500, 1750, 2000, 2250 and 2500 posts per user respectively. The script applied the change in variables incrementally.

Model architecture remained consistent across all tests, with an embedding layer of 100 dimensions, convolutional layer of 256 and a kernel size of 7. The Keras classifier applied the following fixed parameters (Epochs = 3, Batch_size = 32, Stratified K-Fold cross validation = 5).

3.4.4 Model Training: Script 1

Figure 3

CNN Script Structure



As described above, our Python script implemented a CNN model for text classification and performed the following steps in chronological order (Figure 3):

1. **Set Thread Environment:** The script began by setting the environment variables for TensorFlow to optimise performance on an M1 Mac, specifying the number of interop and intraop threads.

2. **Imported Libraries:** Necessary libraries, including TensorFlow, NumPy, CSV handling tools andKeras-related modules, were imported.

3. **Loaded CSV Data:** The script defined a function 'load_CSV_data' to load data from CSV files. It processed the data, considering user

posts and labels and randomly selected Y posts for users with more than X posts.

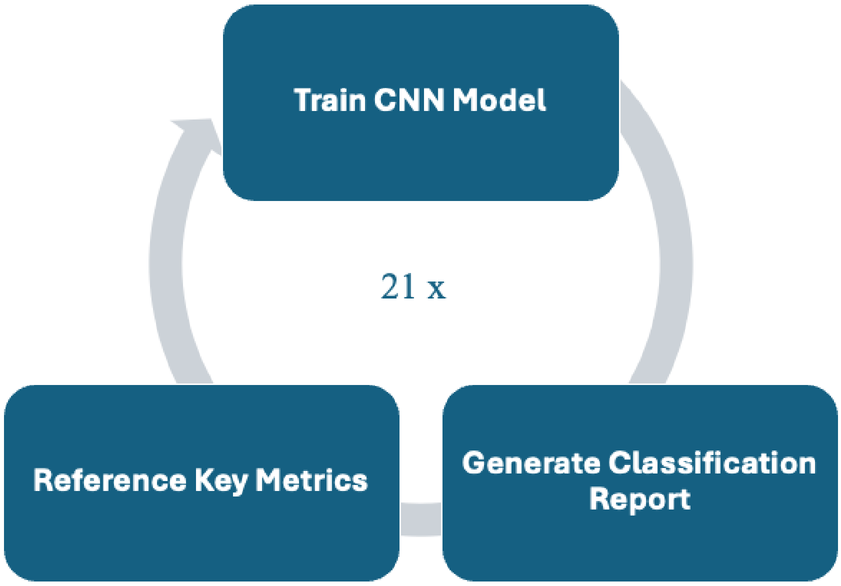
4. **Pre-Processed Data:** Another function ‘preprocess_data’ pre-processed the loaded data, mapping labels to binary values (0 for 'control' and 1 for 'depression'). Text data was tokenised and padded to create sequences suitable for model input.
5. **Saved Tokeniser:** The trained tokeniser was saved to a file using the Pickle library.
6. **Model Architecture:** The script defined a CNN model with an embedding layer, convolutional layer, global max pooling and dense layers. The model was compiled with binary crossentropy loss and accuracy metrics.
7. **Model Training:** The Keras model was created using the specified architecture and the training data was fit to the model using Stratified K-Fold cross-validation. The trained model was saved to a file.
8. **Model Evaluation:** The model was evaluated on the test data and accuracy was printed. Additionally, a classification report was generated, providing detailed metrics for each class (control and depression).

In summary, our script loaded, pre-processed, trained and evaluated a CNN model for text classification, identifying depression-related content in user posts.

3.4.5 Compare Model Performance

Figure 4

Modelling Workflow



After each training cycle, the classification report generated will communicate metrics relevant to each class (control and depression). As we are anticipating 21 iterations at the time of writing, we will reference the precision, recall and F1 scores for each model, with the best model determined based on the best F1 score (Figure 4).

We felt that selecting a model with a higher F1 score was more advantageous, given our goal of achieving a balance between precision and recall is paramount. The F1 score, being the harmonic mean of precision and recall, offers a holistic assessment by considering both false positives and false negatives. This balance is particularly crucial in fields like medical diagnosis, where missing positive cases (false negatives) and incorrectly diagnosing negative cases (false positives) can have equally severe consequences. Therefore, the F1 score provides a comprehensive evaluation of model performance, ensuring that neither type of error is disproportionately favoured.

Moreover, the F1 score is especially valuable in handling imbalanced datasets, where one class significantly outweighs the other. In such cases, precision or recall alone may not adequately capture the model's performance, as they may be skewed by the dominant class. By incorporating both precision and recall, the F1 score offers a more robust measure that is less affected by class imbalance. This stability is essential for ensuring reliable model evaluation, particularly in real-world applications where class distributions may vary widely.

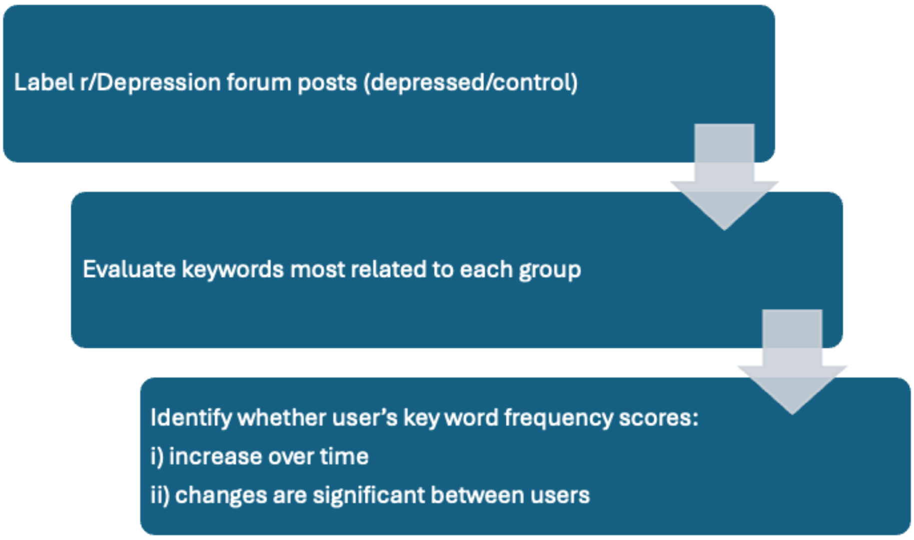
Furthermore, the F1 score simplifies model evaluation and comparison by condensing precision and recall into a single metric. This makes it easier to assess overall model performance, especially when dealing with multiple models or classification thresholds. Additionally, the F1 score is less sensitive to changes in classification thresholds compared to precision or recall alone, enhancing it’s utility in scenarios where the optimal threshold is uncertain. Overall, selecting a model with a higher F1 score ensures a balanced trade-off between precision and recall, resulting in more reliable and interpretable model evaluations.

At the end of the first portion of the study, model scores would be compared using a scatter plot, with the best scoring candidate selected for study 2.

3.5 Study 2: Research And Analysis

Figure 5

Study 2 Flowchart



The second portion of our study is observational; evaluating ~2,000,000 user posts from the r/Depression dataset provided by Reddit. As in study 1, Reddit's data has also been made available to USQ for research purposes and was cleared by ethics prior to our use (see Data and Ethics). Reddit's provided posts span a wide range of age brackets, professions and backgrounds and were ideal for our circumstance given the study aims.

The best performing CNN model trained during Study 1 will be used for Study 2 in an effort to capture the most complex patterns whilst performing more efficiently across larger datasets (Bhandari et al., 2022; da Silva et al., 2022).

3.5.1 Label Posts: Script 1

The script written for this portion of the study applied our CNN sentiment analysis model for text classification on the described dataset and performed the following steps in chronological order:

1. **Import Libraries:** The necessary Python libraries were imported, including Pandas for data manipulation, TensorFlow for machine learning and the required modules for text preprocessing (Tokenizer, pad_sequences) and model loading.
2. **Load Pre-Trained Model And Tokeniser:** The pre-trained CNN model was loaded using TensorFlow's 'load_model' function. Additionally, a tokeniser, which was used during the training of the model, was loaded from a pickled file.
3. **Define Preprocessing Function:** A function was defined to pre-process a CSV file containing text data to make predictions using the loaded model taking the input CSV file path, output CSV file path and the tokeniser as parameters.
4. **Preprocess CSV Data:** The CSV file was read into a Pandas DataFrame and rows with missing or non-textual data in the 'posts' column filtered out. The textual data was then pre-processed, tokenised and padded to match the sequence length used in the successful candidate.
5. **Make Predictions:** The pre-processed text data was fed into the loaded CNN model to make predictions. A threshold of 0.5 was

applied to convert the model's continuous output into binary predictions.

6. **Save Labelled CSV File:** The predicted labels were added as a new column to the DataFrame and the labelled data was saved to a new CSV file.
7. **Specify Input and Output Paths:** The input and output CSV file paths were specified and a function called to perform the necessary labelling.

The resulting output illustrated below is an example of what will be expected post-processing (Table 2-3).

Table 2

Model Labelling Example.

Post Label	Value
Depressed	1
Control	0

Table 3

Example CNN Output.

User Post	Prediction
"Feeling utterly lost and alone right now. It's like no matter how hard I try, I just can't shake off this overwhelming sadness. Everything feels pointless..."	1
"Just finished a great workout and feeling energized!"	0
"It's exhausting pretending to be okay all the time when deep down, I feel like I'm falling apart."	1
"Had a productive day at work and feeling accomplished."	0
"Sometimes I wonder if anyone would even notice if I disappeared. Would anyone care?"	1
"I love fishing every now and then."	0

3.5.2 Text Analysis

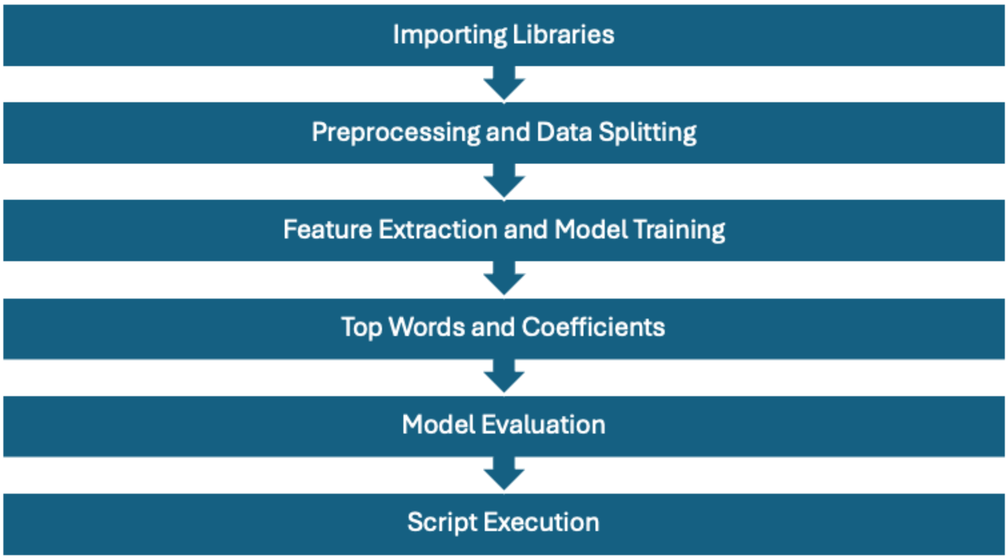
Our second study has several aims (Figure 5). To identify the keywords most strongly related to each label, the proportion of user’s initial post labels (depressed/control) and whether the likelihood of ascertaining a different post label changed over time.

To facilitate this, we created two separate scripts.

3.5.3 Identify Keywords On A Linear Scale: Script 1

Figure 6

Simplified Flowchart Of Keyword Identification



We are initially seeking to perform a sentiment analysis on user posts using a TF-IDF vectorizer and Logistic Regression model. Overall, we are aiming to analyse user posts related to depression, identify key terms associated with depressed and control users and evaluate the performance of the sentiment analysis model using the following code structure below (Figure 6):

1. **Importing Libraries:** We imported Pandas for data manipulation, TfidfVectorizer for feature extraction, 'train_test_split' for data splitting, 'LogisticRegression' for building the classification model and'stopwords' and 'nltk' for natural language processing.

2. **Preprocessing and Data Splitting:** We created a function that loads the data from an Excel file, dropped rows with missing values in the 'posts' or 'Predicted_Label' columns, converted 'Predicted_Label' to numeric type (separating data into depressed and control groups) and splitting the data into training and testing sets.
3. **Top Words and Coefficients Extraction:** We designed a function to retrieve the top N related words and their coefficients for a specified class label (0 or 1) based on the TF-IDF vectoriser and Logistic Regression model.
4. **TF-IDF vectorisation and model training:** We Designed a TF-IDF function that transformed the training and testing data, initialising a Logistic Regression model and training the model on the TF-IDF-transformed training data, whilst making predictions on the test set.
5. **Top Words and Coefficients Display:** The script displayed the top 10 words and their coefficients for depressed and control users, providing insights into the most influential terms for each class.
6. **Model Evaluation:** The accuracy of the model on the test set was calculated and a classification report was generated, including precision, recall and F1-score for each class.
7. **Main Script Execution:** The main script specified the path to the Excel file, calling the pre-processing function and then executing the sentiment analysis using TF-IDF and Logistic Regression.

3.5.4 Script 2 (Mixed-Effects Logistic Regression Analysis):

Our second script aims to perform a mixed-effects logistic regression analysis on our dataset using the predicted labels, elapsed time and anonymised user ID's using the code structure below:

1. **Importing Libraries:** The script imports necessary libraries, including 'Pandas' for data manipulation, 'Statsmodels' for statistical modeling and 'Numpy' for numerical operations.
2. **Data Loading And Preprocessing:** The script read/converted data from an Excel file into a Pandas DataFrame. It converted the UTC column to datetime format, handling time zone information, whilst calculating elapsed time from the minimum timestamp. Time periods were redefined without losing time zone information. The anonymous user ID column was converted to string type to ensure consistent data types and rows with missing or NaN values in any column were dropped.
3. **Creating A New DataFrame:** A new DataFrame was created with 'Predicted_Label' (converted to float), 'elapsed_time' and 'anon'.
4. **Mixed-Effects Logistic Regression:** The script performed a mixed-effects logistic regression using the MixedLM class from statsmodels using the Predicted_Label as the dependent variable, elapsed time as the fixed effect and random effects introduced for each pair within 'anon'. The model was fit using the fit method.
5. **Displaying Regression Results:** The summary of the mixed-effects logistic regression results is printed, including coefficients, standard

errors, z-values and p-values.

It is expected both scripts will be able to facilitate the sought outcomes stipulated within Study 2.

3.6 Reliability

Reliability of the study's methods were ensured under their relevant subheadings.

- Study 1 model:
  - Cross-validation: The training dataset was already split by the original researchers and labelled into several subsets, with the model trained on each with the largest dataset used for testing to ensure against overfitting.
  - Data-validation: The original pre-labelled data used in the RSDD was used to train our model, with the only anticipated difference between the RSDD model and ours being a reduction in training data used for computational efficiency.
- Study 2:
  - TF-IDF: Although implemented to identify keyword strength with binary classifications, if our model is functioning correctly, it is anticipated a clear contextual disparity will appear between classifications.

Third-party reliability tests:

- A panel of experts reviewed portions of work as they were produced on a 'chapter by chapter' basis.

3.7 Limitations And Assumptions

Given the complexity of psychological assessments, the following limitations and assumptions have been considered:

- Self-confession for this experiment was assumed true. Although this might appear arbitrary, there is substantial evidence that already suggests self-reporting has a high degree of accuracy. This was controlled for by a stipulation in the training data used for our model in section 3.2: Datasets and data collection.
- Our model will be trained on the RSDD which captures a subpopulation of depressed people within Reddit exclusively and may not be a representative sample of the population as a whole or transferrable to other platforms.
- This study will be limited to the evaluation of the Reddit forum 'r/Depression'. Reddit is a typical text based social platform allowing the novel algorithm to be used without issue. Using this dataset:
    - controls for selection bias due to random sampling;
    - is not necessarily comparable to the conclusions drawn from other platforms;

- has >2,000,000 data entries, providing a good basis for testing.

3.8 Ethics

- Both Reddit and Yates et al. (2017) have provided/approved the use of our data for research. This has been cleared by USQ's Ethics Committee under the research ID:

  ETH2023-0483: *Sentiment Analysis for Early Detection of Depression on Reddit, a Social Media Platform*

- Reddit user data is publicly available, with rights to any content created or submitted by the user waived as "irrevocable, transferable and sub-licensable" by Reddit or any third party (See: '4. Your Content: https://www.redditinc.com/policies/user-agreement-october-15-2020'). Consequently, use of this data for research has also been agreed to by users under the 'terms and conditions' of Reddit when making an account.
- Anonymity:
  - Usernames have been replaced with unique ID's prior to their receipt to enable temporal evaluations with anonymity.
  - The rest of the data is not sensitive, with all personally identifying information removed/obfuscated prior to its receipt.
- The dataset has been stored encrypted for improved security.

- The data will be used in the hope the results will be published in peer-reviewed journals.

CHAPTER 4: RESULTS AND DISCUSSION

This section serves as the cornerstone of our research, where we present and interpret the findings derived from our analyses. In this section, we delve into the outcomes of our study, providing insights, interpretations and implications that elucidate the significance of our research endeavours. Through a comprehensive examination of the results, we aim to address the research questions and contribute to the broader body of knowledge in our field.

Within this section, we systematically organize and present the results obtained from our data analyses in chronological order, ensuring clarity, coherence and relevance to our research objectives. We will explore the patterns, trends and relationships uncovered through our analyses to offer a more detailed account of our empirical findings and contextualise the results within the existing literature, drawing connections to previous studies and theoretical frameworks to enrich our understanding and interpretation.

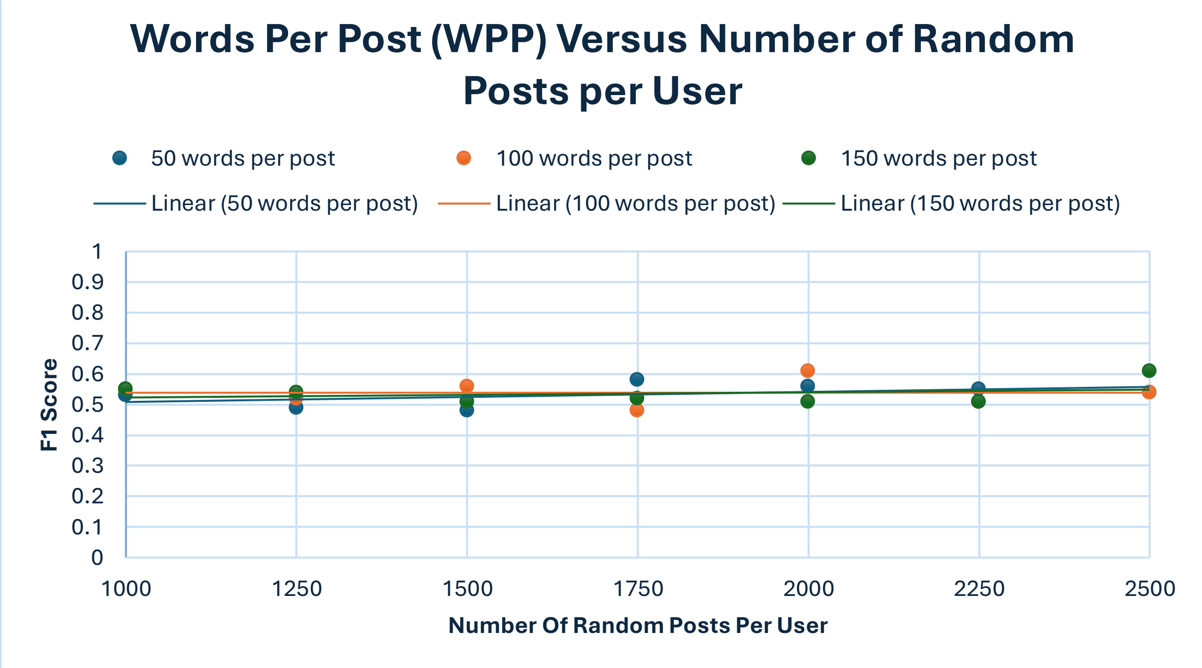
Moreover, the discussion component of this section following each study will provide a platform for critical reflection, synthesis and interpretation of the results in light of the research questions and objectives. Through this, we will examine the implications, limitations and future directions in the hope of fostering scholarly discourse. Through a nuanced exploration of the results and discussions, we endeavour to contribute novel insights, advance theoretical understanding and inform practical applications in our field of inquiry.

4.1 Study 1: Results

After training each CNN model and comparing F1 scores, our Two-Factor ANOVA (Without Replication), examining the impact of the number of user posts versus post length on F1 suggested neither had a significant influence; reporting F = .75, P = >0.05 and F = .03, P = >.05 respectively (Figure 7, Table 4). Overall, the ANOVA suggested that our model's performances were not significantly influenced by variations in word count (per post) or post length selection in the dataset.

Figure 7

Two-Factor ANOVA Revealed No-Significant Effect Of Word Count Or Random Post Number On Model Performance (F1).



Post-hoc evaluation revealed the best models produced F1 scores of 0.61, differentiated only by a Recall score of 0.52 > 0.51 (Table 4-5).

Table 4

Our Two Factor Anova Indicated No Significant Effect Of Random User Post Selection And/Or Post Length.

SUMMARY				
Post Length	*Count*	*Sum*	*Average*	*Variance*
1000	3	1.63	0.543333333	0.000133333
1250	3	1.55	0.516666667	0.000633333
1500	3	1.55	0.516666667	0.001633333
1750	3	1.58	0.526666667	0.002533333
2000	3	1.68	0.56	0.0025
2250	3	1.57	0.523333333	0.000533333
2500	3	1.69	0.563333333	0.001633333
50 words per post	7	3.73	0.532857143	0.00132381
100 words per post	7	3.77	0.538571429	0.001714286
150 words per post	7	3.75	0.535714286	0.001328571

ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Rows	0.007114286	6	0.001185714	0.745508982	0.624267701	2.996120378
Columns	0.000114286	2	5.71429E-05	0.035928144	0.964812978	3.885293835
Error	0.019085714	12	0.001590476			
Total	0.026314286	20				

Table 5

Final Primary Candidate Scores For CNN-R Model.

Model Candidates			
	Precision	Recall	F1
Weighted average	0.85	0.51	0.61
	0.85	0.52	0.61

Table 6

Model Results Table (F1 Scores).

Random Posts (N)	50 words per post	100 words per post	150 words per post
1000	0.53	0.55	0.55
1250	0.49	0.52	0.54
1500	0.48	0.56	0.51
1750	0.58	0.48	0.52
2000	0.56	0.61	0.51
2250	0.55	0.51	0.51
2500	0.54	0.54	0.61

As the purpose of this study was to identify an appropriate model for use in Study 2, the highest scoring candidate was selected (Precision = 0.85, Recall = 0.52, F1 = 0.61).

This candidate was trained using similar parameters to that developed by Yates et al. (2017); but we were unable to achieve the target score (Table 1).

4.2 Study 1: Discussion

Our first study provided crucial insights into the effectiveness of sentiment analysis models for the identification of depressive symptoms among users. Despite 21 training iterations, with changes to the word count/random post number (per post), neither of the two factors significantly influenced model performance (Table 4, Table 6). Although Bailly et al. (2022) also found their model was less influenced by the size of their dataset, this could have been the result of too small a step size, with Prusa et al. (2015) identifying that classifiers for Twitter sentiment improved during steps up to 81,000 instances (our equivalent of posts), making a more conclusive deduction difficult without more breadth.

Current research however indicates that, with concern to the medical domain, classifier performance is more dependent on how well the training data represents the original distribution than size (Althnian et al., 2021). In addition, discussions persist in regards to the definition of a small dataset, with the majority only reaching consensus on the fact that too little data results in worse classification performance, making ideal dataset sizes difficult to define at present (Dris et al., 2019). To date, Shawe-Taylor et al. (1993) introduced the Probably Approximately Correct (PAC) metric to determine the minimum sample size for desired accuracy, with others approaching the issue by utilizing algorithmic information theory to define small datasets (Dris et al., 2019). As this was beyond the scope of our study, there is evidently room for future research into what is an ideal data set size for training models. What we can conclude from this at the time of this study is that from the ranges selected, there were minimal changes. Due to a lack of more computational power at the time of this study, as well as time constraints, we were not able to expand further as assumed and so were limited in this area.

Despite the small differences identified however, the results of our final model still elucidated the existence of a factor(s) that influenced F1 score, emphasizing the importance of considering multiple metrics when training models to discern relevant factors (Table 5, Table 6). In our case, although the difference between F1 scores was minor for our competing models, the separation from the score reported by Yates et al. (2017) indicated a variation strong enough for future consideration. That said, we deviated by training 40 posts per user which could explain this difference.

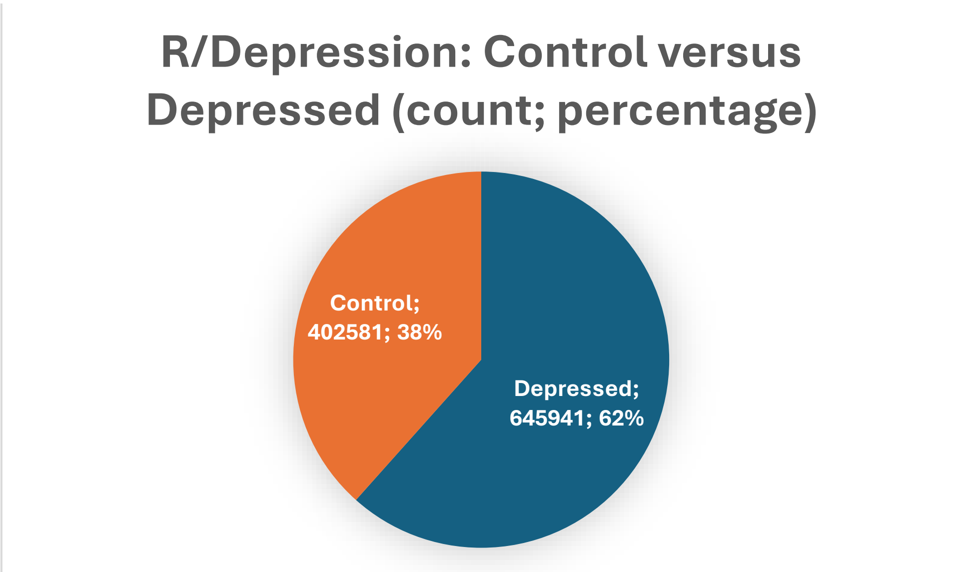
Given that the goal of Study 1 was to identify an appropriate model for use in Study 2, the selection of the highest F1 candidate reflected our decision to prioritize a balance between precision and recall. In many contexts, particularly those characterized by class imbalances, which were the case for us, the F1 score holds greater significance given the increased likelihood to identify positive cases while minimizing both false positives and false negatives whilst controlling against extreme differences in accuracy/precision (Hicks et al., 2022). We felt that, given it's potential use in future research, the simplicity and interpretability of the F1 score also made it a more convenient metric for decision-making and communication, allowing others to discern our model's performance against their requirements more effectively (Hicks et al., 2022).

4.3 Study 2: Results

Following the execution of our second script, our model processed a total of 1,048,522 posts categorising 38% (402,581) and 62% (645,941) posts with control and depressed labels respectively. (Figure 8).

Figure 8

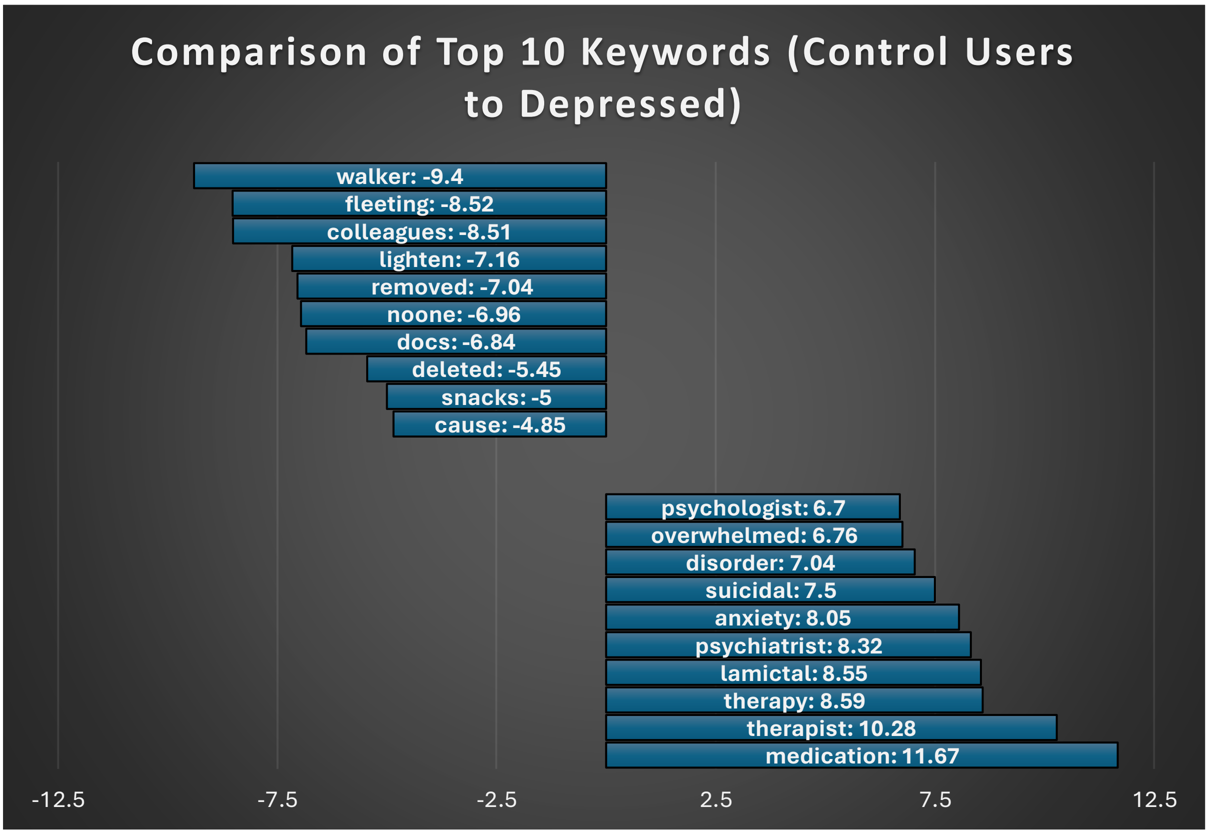
Apportioned Control Versus Depressed Posts (Count; Percentage).



Our TF-IDF script determined the top 10 words associated with depression/control by examining the magnitude of coefficients, producing a list of keywords in a dichotomous, linear relationship, spanning from control, through to depressed keywords (Precision: 0.82, Recall: 0.79, F1: 0.81). The diversity of keywords within the control group were visibly unrelated to the theme consistent within the depressed cohort, suggesting valid precision and recall reporting (Figure 9).

Figure 9

TF-IDF Vectorization Returned The Top 10 Keywords In A Linear Relationship. More Negative Scores Indicated A Stronger Relationship With The Control Cohort, Whereas More Positive Scores Indicated A Stronger Relationship With The Depressed Cohort.



The mixed-effects logistic regression model offered compelling insights into the temporal dynamics and individual variations within our dataset. When elapsed time was 0, the log-odds of a user receiving a predicted label indicative of depression were approximately 0.67. This suggests that, on average, initial posts across all users conformed to our model's criteria for a depressive label (P < 0.001) (Table 8). Interestingly, the analysis of group variance underscored differences in baseline log-odds (0.03) emphasizing the diverse characteristics among individual users (Table 8).

Furthermore, as time progressed, there was a consistent decrease in the log-odds of predicted depressed labels compared to control users. The coefficient associated with elapsed time (-0.00) signifies the impact of time on the likelihood of depressive syntax, with statistical significance (P < 0.001) (see Table 8). This uniform trend implies a potential shift in the frequency of predicted depressed labels over time, suggesting evolving usage patterns among users but with minimal practical impact given the magnitude of the coefficient.

Table 8

Mixed Linear Regression Results.

Model:	MixedLM				Dependent Variable:	Predicted_Label
No. Observations:	1048480				Method:	REML
No. Groups:	163429				Scale:	0.1741
Min. group size:	1				Log-Likelihood:	-611488.7504
Max. group size:	133688				Converged:	Yes
Mean group size:	6.4					
	Coef.	Std.Err.	z	P>|z|	[0.025	0.975]
Intercept:	0.673	0.001	590.9	0	0.671	0.675
elapsed_time:	0	0	-6.8	0	0	0
Group Var	0.034	0.001				

4.4 Study 2: Discussion

The results of our second study illuminates significant distinctions between control and depressed labels within the Reddit community, with our TF-IDF script successfully identifying the top 10 words associated with each label alongside a high precision, recall and F1 score. Notably, the diversity of keywords within the control group starkly contrasted those in the depressed cohort, supporting the efficacy of our approach (Yang & Leskovec, 2011).

Although other strategies could have been used, we felt TF-IDF vectorization helped to capture the importance of each term relative to their category (control/depressed) but more importantly served as a means of indirect validation. As the most important key-words unique to each category were weighted against common terms (increasing discrimination), we could conclude that if our model was ineffective, there should be no observable/thematic differences between groups. This was demonstrably not the case, with all key words in the depressed groups distinctive/synonymous with depression and all control words appearing random (Figure 9).

When reviewing the labels manually, many of the posts identified by our model could be recategorized due to the absence of an obvious contextual relationship and/or keywords associated with their determined class. Although this made our model's determination curious the overall semantic hierarchy determined by our tool suggested ability to identify more nuanced patterns than the human eye. While several arguments could be formed against our approach, we believe it is more important to highlight the assumptions upon which these arguments are likely based. For instance, they would most likely assume that practitioners and current screening tools are more accurate. These are of course defensible, primarily due to the realities of human error but also due to the goals of our model and research, which is to produce a screening tool, not a diagnostic tool. The primary benefit of artificial intelligence is it’s ability to replicate the same learning patterns exhibited by humans, primarily for the identification of patterns to be used in the decision making process (Ahmad et al., 2021). Given our model was trained using a CNN, which Ahmad et al. (2021) posited was the most used ML technique in biomedical research at the time, it lends further support to the potential validity of our model and it’s concurrent results. It is also important to accept that we use big data tools to identify patterns that we as a species cannot, due to our inability to extract meaning from so many datapoints (Ahmad et al., 2021).

Focusing on our results, research has to date, extensively explored the linguistic and behavioural cues associated with depression on social media platforms. Gkotsis et al. (2017) conducted an integrative review focusing on the detection of depression and mental illness on social media. This comprehensive review examined how individuals express depressive symptoms through online interactions across various platforms. By analysing linguistic markers and behavioural patterns using different methods, the study identified similar words and expressions commonly used by individuals experiencing depression across multiple platforms including Twitter, Facebook, Reddit and other online forums. Examples of keywords found include phrases such as "I feel hopeless", "I'm worthless", "I'm so sad", "I'm alone" and "I can't stop crying." These expressions reflect emotional distress, self-criticism, social withdrawal and negative thought patterns commonly associated with depression. De Choudhury et al. (2021) and Coppersmith et al. (2015) have also investigated the predictive power of language-based features extracted from social media posts for identifying depression but with a similar focus on keywords. These studies analysed large datasets of social media posts to identify linguistic characteristics and keyword usage associated with depression. Through similar NLP techniques, researchers were able to detect linguistic markers indicative of depression such as ‘depressed’, ‘sad’, ‘lonely’, ‘worthless’ and ‘suicidal’; with Smith et al. (2020) also finding keywords such as ‘sad’, ‘lonely’ and ‘hopeless’ in social media posts of individuals diagnosed with depression, distinguishing them from non-depressed counterparts.

The findings from Smith et al. (2020) and the referenced studies by De Choudhury et al. (2021) and Coppersmith et al. (2015) indicate a strong correlation between certain keywords and depression when analyzing social media posts. Specifically, words like ‘sad’, ‘lonely’, ‘hopeless’, ‘worthless’ and ‘suicidal’ are clearly indicative of depressive states. When considering our keywords, the discovery of ‘anxiety’, ‘suicidal’, ‘disorder’ and ‘overwhelmed’ aligns with this thematic pattern, suggesting that these terms may carry significant weight across platforms in individuals struggling with mental health issues like depression. Although terms like ‘medication’, ‘therapist’, ‘therapy’, ‘psychiatrist’, ‘lamictal’ and ‘psychologist’ also appeared but were not identified in other studies, they might indicate the discussion of actions taken or sought by individuals experiencing depressive symptoms, reinforcing their relevance in understanding and addressing mental health concerns. As these studies did not apply the same techniques for weighting, we argue our findings highlight the potential for using a TF-IDF tool for identifying and supporting individuals at risk of depression in future research, which we feel offers a different perspective into mental health expression within the digital space.

Our mixed-effects logistic regression model additionally unveiled temporal dynamics alongside individual variations within the dataset that support current research, with most initial posts meeting our model’s standard for a depressive label. Although this could be explained in a number of ways, the majority of research argues that users engage online forums as support systems when suffering from depression, with Griffiths, Calear and Banfield (2009) and Moore et al. (2017) highlighting how online forums offer anonymity and accessibility, making them attractive options for individuals hesitant to seek face-to-face support due to stigma or logistical barriers. As the sense of anonymity provided by online platforms allows users to share their experiences and feelings more candidly, we argue this fully supports our inclination towards passive text analysis for depression and the use of online forums by the community in general. Moreover, the accessibility enables users to seek support more conveniently, regardless of time or location, so can be particularly beneficial for those experiencing social isolation or difficulty accessing traditional mental health services due to geography.

As further examination revealed decreases in the log-odds of depressed labels over time, although the practical impact of our trend was minimal given the small significance level, we argue that forum engagement helps to successfully reduce the occurrence of depressive syntax, which could be interpreted as a reduction in symptoms. This is supported by several studies that have suggested that participating in online forums can lead to improvements in depressive symptoms among users, with both Griffiths, Calear, Banfield, et al. (2009) systematic review and Houston et al. (2002) finding that participants reported significant reductions in depressive symptoms after engaging with online forums. Interesting, this trend also holds consistent the provision of services offered by clinicians such as cognitive behavioural therapy and interpersonal therapy but at a reduced cost to the client (Bessière et al., 2010; Gautam et al., 2020; Markowitz & Weissman, 2004; Pantic, 2014).

It is important to argue that we hold this position as it has also been demonstrated outside the boundary of internet forums, with the engagement of friends, family, or peers allowing individuals to feel understood, supported and cared for, triggering the release of neurotransmitters like oxytocin and dopamine, which can alleviate stress and depressive symptoms (Berkman et al., 2000). Moreover, open communication in informal settings provides opportunities for individuals to gain fresh perspectives, insights and coping strategies, enhancing their ability to manage their symptoms effectively. This exchange of emotional support fosters a sense of connectedness and belonging, which is essential for combating feelings of loneliness and isolation commonly experienced by those with depression (Joiner & Katz, 1999). Therefore, we suggest that the relatively minor impact of our trend could be attributed to the limited sample size. With a longer observation period of posts, we anticipate that this trend may gain further support.

Finally, although we don't feel the observation is necessary for significant expansion, the variance between users highlighted diverse baseline log-odds, emphasizing the inherent heterogeneity of the cohort due to individual differences. Within a depressed internet forum, the observed diversity in baseline log-odds among users likely reflects the inherent heterogeneity of the population. Depression manifests differently in individuals due to various factors such as genetic predisposition, life experiences, coping mechanisms and access to resources. As a result, users entering the forum may have varying degrees of severity, symptom presentation and personal backgrounds. Some individuals may have more pronounced symptoms and lower baseline levels of well-being, leading to higher log-odds of depressive expressions in their posts. Conversely, others may experience milder symptoms or possess stronger resilience factors, resulting in lower log-odds.

This variance consequently underscores the complex nature of depression and the unique experiences of each forum participant. Understanding this heterogeneity is crucial for providing tailored support and interventions within the forum setting and substantiates the ongoing need for a diverse range of approaches and resources to effectively address the multifaceted needs of individuals struggling with depression. By recognizing and accommodating these individual differences, forums can foster a supportive environment where users feel understood, validated and empowered to navigate their journey towards recovery. Given the complexity of human interaction however, we argue that professionals will always be relied on to carry out clinical follow-ups.

In summary, our findings offer valuable insights into the use of a TF-IDF as a linguistic tool, which provide an alternative view of depressed keywords, whilst substantiating use in future research methods. Furthermore, the complex interplay we found between elapsed time, individual user characteristics and the likelihood of predicted depressed labels supported current research into the use of online forums and community groups as an outlet of improving depressed symptoms and other mental health disorders in general.

CHAPTER 5: CONCLUSIONS

As we approach the culmination of this thesis, it is essential to reflect on the journey undertaken and the insights gleaned throughout our exploration. Our investigation has traversed the intricate landscape of mental health assessment, delving into the nuances of depressive symptoms identification in online forums. From examining sentiment analysis models to analysing keyword patterns associated with depression, each step of our research has contributed to a deeper understanding of the complex interplay between technology, psychology and mental health.

In this concluding section, we synthesize the key findings from our research endeavours and contextualize them within the broader framework of mental health assessment and intervention. By revisiting the objectives outlined at the outset of this thesis, we aim to elucidate the implications of our findings and chart a path for future inquiry and practice in mental health screening.

5.1 Chapter 1 And 2

In this thesis, we embarked on a comprehensive exploration of mental health assessment and intervention, beginning with an in-depth investigation into the foundational concepts and methodologies. We started by examining psychology as a social construct, acknowledging the profound impact of societal norms, cultural values and historical context on our understanding of mental health. From there, we delved into the intricacies of psychological testing, highlighting it’s role in diagnosing and assessing mental health disorders, particularly depression. Through a nuanced discussion, we defined depression and explored the various dimensions of this complex condition, laying the groundwork for our subsequent exploration of depression screening tools.

Building upon our understanding of depression, we further delved into the landscape of depression screening tools. We surveyed existing methodologies and technologies, exploring their strengths, limitations and potential for improvement. In doing so, we introduced the novel concept of AI screening tools, which harness the power of artificial intelligence and machine learning to enhance the accuracy and efficiency of mental health assessment. Our exploration of the concepts behind AI screening tools laid the foundation for a deeper examination of their ethical implications, setting the stage for interdisciplinary dialogue between psychology, technology and ethics.

We then conducted a comprehensive review of the existing literature surrounding mental health assessment methodologies. We started by examining traditional approaches such as clinical interviews and self-reporting measures, assessing their validity and reliability in diagnosing depression. From there, we explored the emerging field of natural language processing (NLP) and its potential applications in mental health screening. Our review also addressed barriers to the adoption of algorithms in clinical practice, highlighting challenges related to accuracy, privacy and bias. Additionally, we discussed strategies for overcoming these barriers and maximizing the potential of AI in mental health assessment, focusing on decision support systems and the debate between binary and multifactor NLP approaches.

Through our chronological exploration of these chapters, we have laid the groundwork for a comprehensive understanding of mental health assessment and intervention. By bridging the gap between psychology, technology and ethics, we aim to contribute to the development of innovative solutions that enhance mental health care delivery and promote holistic well-being for individuals and communities.

5.2 Results

The findings from our studies provide significant insights into the effectiveness of sentiment analysis models for identifying depressive symptoms among users in online forums. Our first study focused on evaluating sentiment analysis models and their performance in identifying depressive symptoms. Despite varying the word count/random post number per post during training, we found minimal influence on our model's performance. This contrasts with findings by Bailly et al. (2022) but aligns with the assertion that classifier performance in the medical domain is more dependent on how well the training data represents the original distribution (Althnian et al., 2021). We also noted challenges in defining ideal dataset sizes for training models, with consensus only reached on the detrimental effects of too little data (Dris et al., 2019).

Our second study delved into the identification of keywords associated with depression in online forums using TF-IDF analysis. We successfully identified top words associated with both control and depressed labels, highlighting significant distinctions between the two groups. The diversity of keywords in the control group starkly contrasted those in the depressed cohort, supporting the efficacy of our approach. Additionally, the TF-IDF vectorization helped capture the importance of each term relative to their category and served as a means of indirect validation.

While our model demonstrated the ability to identify nuanced patterns indicative of depressive symptoms, there were instances where manual review revealed posts that could or could not be recategorized due to contextual ambiguity. Nonetheless, our tool showcased the potential to identify patterns beyond human perception, emphasizing the role of artificial intelligence in aiding decision-making processes. We also argue that online forums play a vital role in providing support for individuals experiencing depression, aligning with research that suggests participating in such forums can lead to improvements in depressive symptoms.

Furthermore, our findings suggest a strong correlation between certain keywords and depression, consistent with existing literature on linguistic cues associated with depressive states on social media platforms. The thematic pattern identified in our study reinforces the relevance of terms such as ‘anxiety’, ‘suicidal’, ‘disorder’ and ‘overwhelmed’ in understanding mental health issues like depression. Additionally, our exploration of temporal dynamics and individual variations within the dataset supports the use of online forums as outlets for improving depressive symptoms, echoing the findings of previous studies.

In conclusion, our research contributes to the understanding of mental health assessment in online forums and highlights the potential of sentiment analysis models and TF-IDF analysis in identifying depressive symptoms. In the future, it is imperative to conduct additional research aimed at determining the ideal dataset sizes for training models and to persist in examining the significance of online forums in aiding individuals with depression. Ultimately, our discoveries emphasize the criticality of harnessing technology and employing interdisciplinary methodologies to tackle mental health obstacles in the era of digital advancements.

5.3 Future Research Directions

Building upon the insights gained from our studies, several avenues for future research emerge, offering opportunities to deepen our understanding of mental health assessment in online forums and advance the development of innovative interventions. One promising direction is to further investigate the optimal dataset sizes for training sentiment analysis models. While our study shed light on the minimal influence of varying word count/random post number per post, there remains a need to explore how different dataset sizes affect model performance across diverse populations and linguistic contexts. Future research could employ larger and more diverse datasets to ascertain the robustness of sentiment analysis models and their generalizability across various demographic and cultural groups.

Additionally, there is a pressing need to enhance the interpretability and transparency of sentiment analysis models in mental health assessment. While our TF-IDF analysis successfully identified keywords associated with depression, manual review revealed instances of contextual ambiguity. Future research could focus on refining natural language processing techniques to improve the accuracy and reliability of model predictions, while also providing explanations for the decisions made. Incorporating explainable artificial intelligence methodologies into sentiment analysis models could enhance their interpretability, enabling clinicians and researchers to better understand the factors influencing model predictions and identify potential biases.

Furthermore, exploring the temporal dynamics of depressive symptoms expression in online forums presents a fertile area for future investigation. Our study revealed decreases in the log-odds of depressed labels over time, suggesting potential changes in symptom severity or coping mechanisms among forum participants. Future research could employ longitudinal approaches to examine how depressive symptoms evolve over time within online communities, identifying patterns of resilience and vulnerability and informing targeted interventions. Additionally, investigating the role of moderators and peer support networks in mitigating depressive symptoms within online forums could offer valuable insights into effective strategies for promoting mental health and well-being in digital spaces.

Finally, there is a need to continue exploring the potential of online forums as platforms for supporting individuals with depression and other mental health concerns. While our study and existing literature highlight the positive impact of forum engagement on depressive symptoms, further research is warranted to understand the mechanisms underlying this effect and optimize the design and implementation of online support interventions. Collaborative efforts between researchers, clinicians and forum administrators are essential to develop evidence-based guidelines and best practices for fostering supportive and inclusive online communities for mental health support.

As model training is also heavily dependent on infrastructure, with larger models requiring more computational power, we also identified the potential need for more research into factors influencing training to ensure the effective utilisation of resources (Prusa et al., 2015). Our model, although imperfect, was created using very limited resources, was based on a subjective classification process with very noisy labels but was still able to effectively parse a Reddit forum as a screening tool without any negative effect to users. As the process of evaluating noisy labels using machine learning improves, it is expected that tool accuracy may increase further. Additionally however, with the advent of LLM's, we anticipate they will take over this process; with commercial models able to self-learn at a rate individual research teams are unlikely to compete.

Furthermore, the comparison with a model developed by Yates et al. (2017) highlights the challenges and complexities inherent in replicating or achieving similar performance levels across different studies or datasets. Despite efforts to emulate the parameters and methodologies of a previously successful model, the inability to achieve identical scores underscores the nuances and intricacies involved in model development and training. This observation underscores the need for continuous refinement and adaptation of models to suit specific contexts and datasets, acknowledging the inherent variability and complexity of real-world data.

To date, we anticipate two primary challenges in:

- Information access
- Classification reviews

5.3.1 Information Access

Any model provided for a service is ultimately subject to not only the way it was built but primarily on what it was built. As deep neural networks continue to improve, which they are at a rapid rate, we anticipate future models will continue to improve in accuracy (Song et al., 2023). This accuracy however will still be determined by the data available for training, as well as the data upon which it is being tested. As was mentioned during our systematic review, a model trained on another service (Facebook, X etc) or even another forum, will not necessarily perform well or at all, elsewhere.

Although our reasoning is multifaceted, we posit this is primarily the result of the following:

1. **Domain-Specific Vocabulary**: Different areas or domains have their own unique terminologies, jargon and linguistic nuances. A model trained on one domain might not understand or generalize well to the terminologies of another domain.
2. **Data Distribution**: The distribution of data in one domain might be significantly different from another domain. Models trained on one dataset learn patterns specific to that distribution and may fail to generalize to data with a different distribution. As we performed our analysis on a Reddit forum specifically related to depression, further research could be conducted on other forums, as well as other domains.
3. **Semantic Differences**: Even if the words are the same across domains, their meanings or contexts can vary. For instance, the word "apple" could refer to a fruit in one domain and a technology company in another. Although this might not appear to be related

to our domain, models do require exposure to different contexts to understand these nuances.

4. **Task Differences**: NLP tasks can vary widely across different domains. A model trained for sentiment analysis on social media data might not perform well for medical text classification tasks. Each task may require different features, representations or modelling techniques.
5. **Annotation Quality And Quantity**: The quality and quantity of annotations in different datasets can vary. Models heavily rely on labelled data during training and if the quality or quantity of annotations differs significantly between datasets, it can impact the model's performance.
6. **Biases and Cultural Differences**: Models trained on one dataset may inherit biases present in that dataset. These biases may not generalise well to a different domain, where different biases or cultural nuances exist.
7. **Cross-Linguistic Variances:** Human languages exhibit diverse structures, vocabularies and cultural nuances. For instance, expressions or idioms in English may not have direct equivalents in other languages like Spanish or Mandarin. Therefore, when deploying NLP models across different linguistic contexts, such as translating between languages or analysing multilingual data, it is crucial to account for these variations. Models trained solely on one

language may struggle to accurately interpret or generate content in another, highlighting the necessity for multilingual training and robust linguistic understanding in NLP systems.

To mitigate these issues, domain adaptation techniques such as fine-tuning on the target domain or using techniques like transfer learning can be employed. These methods involve updating the parameters of a pre-trained model on a new dataset to adapt it to the target domain (Ebbehoj et al., 2022; Waisberg et al., 2023). However, even with these techniques, performance may vary depending on the similarity between the source and target domains.

5.3.2 Classification Reviews

DSM-5 undergoes continuous scrutiny and assessment by mental health professionals, researchers and pertinent organizations.

The release of the DSM-5 in 2013 marked a significant advancement over it's predecessor as it integrated changes in diagnostic criteria, based on advancements in scientific research and evolving perspectives on mental health disorders. Much like it's predecessors, the DSM-5 undergoes periodic review and refinement in response to emerging evidence and the evolution of psychiatric practices.

While there are not any communicated plans for a comprehensive overhaul of the DSM-5, future updates to specific diagnostic criteria or the inclusion of new disorders may transpire as research advances and our comprehension of mental health disorders deepens. These modifications typically entail a meticulous review process, involving input from domain experts and consideration of the latest scientific findings. Consequently, any tools built on prior definitions might necessitate additional reviews and re-training.

5.4 Conclusion

Our research highlights significant advancements in leveraging sentiment analysis models for identifying depressive symptoms in online forums. The first study demonstrated that while varying word count and post numbers during training had minimal impact on model performance, the effectiveness of sentiment analysis is more closely tied to the representativeness of the training data. This aligns with Althnian et al. (2021) and contrasts Bailly et al. (2022), suggesting that sufficient, representative data is crucial for classifier performance. We also identified challenges in determining optimal dataset sizes, yielding support for prior findings by Dris et al. (2019) about the drawbacks of insufficient data.

In our second study, the use of TF-IDF analysis effectively differentiated keywords associated with depressive and non-depressive posts, reinforcing the model's capacity to capture meaningful linguistic patterns. Although some contextual ambiguities required manual review, the tool demonstrated potential in uncovering nuanced depressive indicators that may surpass human detection capabilities. Our findings emphasize the value of online forums as supportive platforms for individuals with depression and affirm the significance of keywords like 'anxiety' and 'suicidal' in identifying mental health issues. Moving forward, further research is needed to refine dataset sizes and continue exploring the role of online forums in mental health support, underscoring the importance of integrating technology and interdisciplinary approaches to address mental health challenges in the digital age.

References

ABS. (2018). Mental health. https://www.abs.gov.au/statistics/health/mental-health/mental-health/latest-release

Ahmad, Z., Rahim, S., Zubair, M., & Abdul-Ghafar, J. (2021). Artificial intelligence (AI) in medicine, current applications and future role with special emphasis on its potential and promise in pathology: present and future impact, obstacles including costs and acceptance among pathologists, practical and philosophical considerations. A comprehensive review. *Diagn Pathol*, *16*(1), 24. https://doi.org/10.1186/s13000-021-01085-4

Al-Mosaiwi, M., & Johnstone, T. (2018). In an Absolute State: Elevated Use of Absolutist Words Is a Marker Specific to Anxiety, Depression, and Suicidal Ideation. *Clin Psychol Sci*, *6*(4), 529-542. https://doi.org/10.1177/2167702617747074

Alarcón, R. D. (2009). Culture, cultural factors and psychiatric diagnosis: review and projections. *World Psychiatry*, *8*(3), 131-139. https://doi.org/10.1002/j.2051-5545.2009.tb00233.x

Althnian, A., AlSaeed, D., Al-Baity, H., Samha, A., Dris, A. B., Alzakari, N., Abou Elwafa, A., & Kurdi, H. (2021). Impact of Dataset Size on Classification Performance: An Empirical Evaluation in the Medical Domain. *Applied Sciences*, *11*(2), 796. https://www.mdpi.com/2076-3417/11/2/796

Andersson, S., Bathula, D. R., Iliadis, S. I., Walter, M., & Skalkidou, A. (2021). Predicting women with depressive symptoms postpartum with machine learning methods. *Sci Rep*, *11*(1), 7877. https://doi.org/10.1038/s41598-021-86368-y

Andrade, L., Caraveo-Anduaga, J. J., Berglund, P., Bijl, R. V., De Graaf, R., Vollebergh, W., Dragomirecka, E., Kohn, R., Keller, M., Kessler, R. C., Kawakami, N., Kilic, C., Offord, D., Ustun, T. B., & Wittchen, H. U. (2003). The epidemiology of major depressive episodes: results from the International Consortium of Psychiatric Epidemiology (ICPE) Surveys. *Int J Methods Psychiatr Res*, *12*(1), 3-21. https://doi.org/10.1002/mpr.138

Bailly, A., Blanc, C., Francis, É., Guillotin, T., Jamal, F., Wakim, B., & Roy, P. (2022). Effects of dataset size and interactions on the prediction performance of logistic regression and deep learning models. *Computer Methods and Programs in Biomedicine*, *213*, 106504. https://doi.org/https://doi.org/10.1016/j.cmpb.2021.106504

Baker, L. A. (2007). THE BIOLOGY OF RELATIONSHIPS: WHAT BEHAVIORAL GENETICS TELLS US ABOUT INTERACTIONS AMONG FAMILY MEMBERS. *De Paul Law Rev*, *56*(3), 837-846.

Bargh, J. A., & Morsella, E. (2008). The Unconscious Mind. *Perspect Psychol Sci*, *3*(1), 73-79. https://doi.org/10.1111/j.1745-6916.2008.00064.x

Berkman, L. F., Glass, T., Brissette, I., & Seeman, T. E. (2000). From social integration to health: Durkheim in the new millennium. *Soc

Sci Med, *51*(6), 843-857. https://doi.org/10.1016/s0277-9536(00)00065-4

Bessière, K., Pressman, S., Kiesler, S., & Kraut, R. (2010). Effects of internet use on health and depression: a longitudinal study. *J Med Internet Res*, *12*(1), e6. https://doi.org/10.2196/jmir.1149

Bhandari, H. N., Rimal, B., Pokhrel, N. R., Rimal, R., Dahal, K. R., & Khatri, R. K. C. (2022). Predicting stock market index using LSTM. *Machine Learning with Applications*, *9*. https://doi.org/10.1016/j.mlwa.2022.100320

Bhatt, C., Kumar, N., Bashir, K. A., & Alazab, M. (2021). Implementing Data Analytics and Architectures for Next Generation Wireless Communications. 132-144. https://doi.org/10.4018/978-1-7998-6988-7

Blumenthal-Barby, J. S., & Krieger, H. (2015). Cognitive biases and heuristics in medical decision making: a critical review using a systematic search strategy. *Med Decis Making*, *35*(4), 539-557. https://doi.org/10.1177/0272989x14547740

Boden, J. M., & Fergusson, D. M. (2011). Alcohol and depression. *Addiction (Abingdon, England)*, *106*(5), 906-914. https://doi.org/10.1111/j.1360-0443.2010.03351.x

Boumil, M. M., Dunn, K., Ryan, N., & Clearwater, K. (2012). Prescription data mining, medical privacy and the First Amendment: the U.S. Supreme Court in Sorrell v. IMS health Inc. *Ann Health Law*, *21*(2), 447-491, 444 p preceding i.

Bromet, E., Andrade, L. H., Hwang, I., Sampson, N. A., Alonso, J., de Girolamo, G., de Graaf, R., Demyttenaere, K., Hu, C., Iwata, N., Karam, A. N., Kaur, J., Kostyuchenko, S., Lepine, J. P., Levinson, D., Matschinger, H., Mora, M. E., Browne, M. O., Posada-Villa, J., . . . Kessler, R. C. (2011). Cross-national epidemiology of DSM-IV major depressive episode. *BMC Med*, *9*, 90. https://doi.org/10.1186/1741-7015-9-90

Brotsky, S. R., & Giles, D. (2007). Inside the "pro-ana" community: a covert online participant observation. *Eat Disord*, *15*(2), 93-109. https://doi.org/10.1080/10640260701190600

Chand, S. P., & Arif, H. (2023). Depression. In *StatPearls*. StatPearls Publishing

Copyright © 2023, StatPearls Publishing LLC.

Chen, S., Stromer, D., Alabdalrahim, H. A., Schwab, S., Weih, M., & Maier, A. (2020). Automatic dementia screening and scoring by applying deep learning on clock-drawing tests. *Scientific Reports*, *10*(1), 20854. https://doi.org/10.1038/s41598-020-74710-9

Cheng, P. M., Montagnon, E., Yamashita, R., Pan, I., Cadrin-Chenevert, A., Perdigon Romero, F., Chartrand, G., Kadoury, S., & Tang, A. (2021). Deep Learning: An Update for Radiologists. *Radiographics*, *41*(5), 1427-1445. https://doi.org/10.1148/rg.2021200210

Cho, A., Min, I. K., Hong, S., Chung, H. S., Lee, H. S., & Kim, J. H. (2022). Effect of Applying a Real-Time Medical Record Input Assistance System With Voice Artificial Intelligence on Triage Task

Performance in the Emergency Department: Prospective Interventional Study. *JMIR Med Inform*, *10*(8), e39892. https://doi.org/10.2196/39892

Coppersmith, G., Dredze, M., Harman, C., & Hollingshead, K. (2015). *From ADHD to SAD: Analyzing the Language of Mental Health on Twitter through Self-Reported Diagnoses*. https://doi.org/10.3115/v1/W15-1201

Corrigan, P. W., Druss, B. G., & Perlick, D. A. (2014). The Impact of Mental Illness Stigma on Seeking and Participating in Mental Health Care. *Psychol Sci Public Interest*, *15*(2), 37-70. https://doi.org/10.1177/1529100614531398

da Silva, D. G., Geller, M. T. B., Moura, M. S. d. S., & Meneses, A. A. d. M. (2022). Performance evaluation of LSTM neural networks for consumption prediction. *e-Prime - Advances in Electrical Engineering, Electronics and Energy*, *2*. https://doi.org/10.1016/j.prime.2022.100030

David, D., Szentagotai, A., Lupu, V., & Cosman, D. (2008). Rational emotive behavior therapy, cognitive therapy, and medication in the treatment of major depressive disorder: a randomized clinical trial, posttreatment outcomes, and six-month follow-up. *J Clin Psychol*, *64*(6), 728-746. https://doi.org/10.1002/jclp.20487

De Choudhury, M., Gamon, M., Counts, S., & Horvitz, E. (2021). Predicting Depression via Social Media. *Proceedings of the*

International AAAI Conference on Web and Social Media, *7*(1), 128-137. https://doi.org/10.1609/icwsm.v7i1.14432

Dris, A. B., Alzakari, N., & Kurdi, H. (2019, 18-20 June 2019). A Systematic Approach to Identify an Appropriate Classifier for Limited-Sized Data Sets. 2019 International Symposium on Networks, Computers and Communications (ISNCC),

Ebbehoj, A., Thunbo, M., Andersen, O. E., Glindtvad, M. V., & Hulman, A. (2022). Transfer learning for non-image data in clinical research: A scoping review. *PLOS Digit Health*, *1*(2), e0000014. https://doi.org/10.1371/journal.pdig.0000014

Fabris, M. A., Marengo, D., Longobardi, C., & Settanni, M. (2020). Investigating the links between fear of missing out, social media addiction, and emotional symptoms in adolescence: The role of stress associated with neglect and negative reactions on social media. *Addict Behav*, *106*, 106364. https://doi.org/10.1016/j.addbeh.2020.106364

Facebook. (2022). https://www.facebook.com/terms.php

Ferreira, R., & Aguiar, R. L. (2019). Repositioning privacy concerns: Web servers controlling URL metadata. *Journal of Information Security and Applications*, *46*, 121-137. https://doi.org/10.1016/j.jisa.2019.03.010

Fox, J. R., Park, B., & Lang, A. (2016). When Available Resources Become Negative Resources. *Communication Research*, *34*(3), 277-296. https://doi.org/10.1177/0093650207300429

Gabarron, E., Dorronzoro, E., Rivera-Romero, O., & Wynn, R. (2019). Diabetes on Twitter: A Sentiment Analysis. *J Diabetes Sci Technol*, *13*(3), 439-444. https://doi.org/10.1177/1932296818811679

Gao, S., Calhoun, V. D., & Sui, J. (2018). Machine learning in major depression: From classification to treatment outcome prediction. *CNS Neurosci Ther*, *24*(11), 1037-1052. https://doi.org/10.1111/cns.13048

Gautam, M., Tripathi, A., Deshmukh, D., & Gaur, M. (2020). Cognitive Behavioral Therapy for Depression. *Indian J Psychiatry*, *62*(Suppl 2), S223-s229. https://doi.org/10.4103/psychiatry.IndianJPsychiatry_772_19

Gkotsis, G., Oellrich, A., Velupillai, S., Liakata, M., Hubbard, T. J. P., Dobson, R. J. B., & Dutta, R. (2017). Characterisation of mental health conditions in social media using Informed Deep Learning. *Scientific Reports*, *7*(1), 45141. https://doi.org/10.1038/srep45141

Google. (2022). Overview of Google identity management. https://cloud.google.com/architecture/identity/overview-google-authentication

Greenberg, P. E., Fournier, A. A., Sisitsky, T., Simes, M., Berman, R., Koenigsberg, S. H., & Kessler, R. C. (2021). The Economic Burden of Adults with Major Depressive Disorder in the United States (2010 and 2018). *Pharmacoeconomics*, *39*(6), 653-665. https://doi.org/10.1007/s40273-021-01019-4

Greenberg, P. E., Kessler, R. C., Birnbaum, H. G., Leong, S. A., Lowe, S. W., Berglund, P. A., & Corey-Lisle, P. K. (2003). The economic burden of depression in the United States: how did it change between 1990 and 2000? *J Clin Psychiatry*, *64*(12), 1465-1475. https://doi.org/10.4088/jcp.v64n1211

Griffiths, K. M., Calear, A. L., & Banfield, M. (2009). Systematic review on Internet Support Groups (ISGs) and depression (1): Do ISGs reduce depressive symptoms? *J Med Internet Res*, *11*(3), e40. https://doi.org/10.2196/jmir.1270

Griffiths, K. M., Calear, A. L., Banfield, M., & Tam, A. (2009). Systematic review on Internet Support Groups (ISGs) and depression (2): What is known about depression ISGs? *J Med Internet Res*, *11*(3), e41. https://doi.org/10.2196/jmir.1303

Guo, J. (2022). Deep learning approach to text analysis for human emotion detection from big data. *31*(1), 113-126. https://doi.org/doi:10.1515/jisys-2022-0001 (Journal of Intelligent Systems)

Hamilton, M. (1960). A rating scale for depression. *J Neurol Neurosurg Psychiatry*, *23*(1), 56-62. https://doi.org/10.1136/jnnp.23.1.56

Harris, R. E. (1997). The need to know versus the right to know: privacy of patient medical data in an information-based society. *Suffolk Univ Law Rev*, *30*(4), 1183-1218. https://www.ncbi.nlm.nih.gov/pubmed/16086487

Hasan, M. M., Popp, J., & Oláh, J. (2020). Current landscape and influence of big data on finance. *Journal of Big Data*, *7*(1), 21. https://doi.org/10.1186/s40537-020-00291-z

Hassan, S.-U., Shabbir, M., Iqbal, S., Said, A., Kamiran, F., Nawaz, R., & Saif, U. (2021). Leveraging Deep Learning and SNA approaches for Smart City Policing in the Developing World. *International Journal of Information Management*, *56*. https://doi.org/10.1016/j.ijinfomgt.2019.102045

Hicks, S. A., Strümke, I., Thambawita, V., Hammou, M., Riegler, M. A., Halvorsen, P., & Parasa, S. (2022). On evaluation metrics for medical applications of artificial intelligence. *Sci Rep*, *12*(1), 5979. https://doi.org/10.1038/s41598-022-09954-8

Houston, T. K., Cooper, L. A., & Ford, D. E. (2002). Internet support groups for depression: a 1-year prospective cohort study. *Am J Psychiatry*, *159*(12), 2062-2068. https://doi.org/10.1176/appi.ajp.159.12.2062

Hyman, S. E. (2010). The diagnosis of mental disorders: the problem of reification. *Annual review of clinical psychology*, *6*(1), 155-179. https://doi.org/10.1146/annurev.clinpsy.3.022806.091532

Islam, M. R., Kabir, M. A., Ahmed, A., Kamal, A. R. M., Wang, H., & Ulhaq, A. (2018). Depression detection from social network data using machine learning techniques. *Health Inf Sci Syst*, *6*(1), 8. https://doi.org/10.1007/s13755-018-0046-0

Joiner, T., & Katz, J. (1999). Contagion of Depressive Symptoms and Mood: Meta-analytic Review and Explanations From Cognitive, Behavioral, and Interpersonal Viewpoints. *Clinical Psychology: Science and Practice*, *6*, 149-164. https://doi.org/10.1093/clipsy.6.2.149

Joyce, P. R. (2008). Classification of mood disorders in DSM-V and DSM-VI. *Australian and New Zealand journal of psychiatry*, *42*(10), 851-862. https://doi.org/10.1080/00048670802363667

Kahneman, D., & Deaton, A. (2010). High income improves evaluation of life but not emotional well-being. *Proc Natl Acad Sci U S A*, *107*(38), 16489-16493. https://doi.org/10.1073/pnas.1011492107

Kang, G. J., Ewing-Nelson, S. R., Mackey, L., Schlitt, J. T., Marathe, A., Abbas, K. M., & Swarup, S. (2017). Semantic network analysis of vaccine sentiment in online social media. *Vaccine*, *35*(29), 3621-3638. https://doi.org/10.1016/j.vaccine.2017.05.052

Kelly, B. C., Harris, E., & Vuolo, M. (2017). Psychosocial Influences of the Escalation of Deviance: The Case of Prescription Drug Sniffing. *Deviant Behav*, *38*(8), 941-956. https://doi.org/10.1080/01639625.2016.1229934

Kim, J., Lee, D., & Park, E. (2021). Machine Learning for Mental Health in Social Media: Bibliometric Study. *J Med Internet Res*, *23*(3), e24870. https://doi.org/10.2196/24870

Kounenou, K., Kalamatianos, A., Nikoltsiou, P., & Kourmousi, N. (2023). The Interplay among Empathy, Vicarious Trauma, and Burnout in

Greek Mental Health Practitioners. *Int J Environ Res Public Health*, *20*(4). https://doi.org/10.3390/ijerph20043503

Kralj Novak, P., Smailovic, J., Sluban, B., & Mozetic, I. (2015). Sentiment of Emojis. *PLoS One*, *10*(12), e0144296. https://doi.org/10.1371/journal.pone.0144296

Kroenke, K., Spitzer, R. L., & Williams, J. B. (2001). The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med*, *16*(9), 606-613. https://doi.org/10.1046/j.1525-1497.2001.016009606.x

Kroenke, K., Strine, T. W., Spitzer, R. L., Williams, J. B., Berry, J. T., & Mokdad, A. H. (2009). The PHQ-8 as a measure of current depression in the general population. *J Affect Disord*, *114*(1-3), 163-173. https://doi.org/10.1016/j.jad.2008.06.026

Lanseng, E. J., & Andreassen, T. W. (2007). Electronic Healthcare: A Study of People's Readiness and Attitude toward Performing Self-Diagnosis. *International Journal of Service Industry Management*, *18*, 394-417. https://doi.org/10.1108/09564230710778155

Leightley, D., Pernet, D., Velupillai, S., Stewart, R. J., Mark, K. M., Opie, E., Murphy, D., Fear, N. T., & Stevelink, S. A. M. (2020). The Development of the Military Service Identification Tool: Identifying Military Veterans in a Clinical Research Database Using Natural Language Processing and Machine Learning. *JMIR Med Inform*, *8*(5), e15852. https://doi.org/10.2196/15852

Leppink, J. (2017). Cognitive load theory: Practical implications and an important challenge. *J Taibah Univ Med Sci*, *12*(5), 385-391. https://doi.org/10.1016/j.jtumed.2017.05.003

Lewinsohn, P. M., Seeley, J. R., Roberts, R. E., & Allen, N. B. (1997). Center for Epidemiologic Studies Depression Scale (CES-D) as a screening instrument for depression among community-residing older adults. *Psychol Aging*, *12*(2), 277-287. https://doi.org/10.1037//0882-7974.12.2.277

Luo, J. W., & Chong, J. J. R. (2020). Review of Natural Language Processing in Radiology. *Neuroimaging Clin N Am*, *30*(4), 447-458. https://doi.org/10.1016/j.nic.2020.08.001

Markowitz, J. C., & Weissman, M. M. (2004). Interpersonal psychotherapy: principles and applications. *World Psychiatry*, *3*(3), 136-139.

McCue, C. (2010, October). Connecting the Dots : Data Mining and Predictive Analytics in Law Enforcement and Intelligence Analysis.

McGarry, J., Kim, H., Sheng, X., Egger, M., & Baksh, L. (2009). Postpartum depression and help-seeking behavior. *J Midwifery Womens Health*, *54*(1), 50-56. https://doi.org/10.1016/j.jmwh.2008.07.003

McLaughlin, K. A. (2011). The public health impact of major depression: a call for interdisciplinary prevention efforts. *Prev Sci*, *12*(4), 361-371. https://doi.org/10.1007/s11121-011-0231-8

Medich, M., Cannedy, S. L., Hoffmann, L. C., Chinchilla, M. Y., Pila, J. M., Chassman, S. A., Calderon, R. A., & Young, A. S. (2023). Clinician and Patient Perspectives on the Use of Passive Mobile Monitoring and Self-Tracking for Patients With Serious Mental Illness: User-Centered Approach. *JMIR Hum Factors*, *10*, e46909. https://doi.org/10.2196/46909

Moore, D., Drey, N., & Ayers, S. (2017). Use of Online Forums for Perinatal Mental Illness, Stigma, and Disclosure: An Exploratory Model. *JMIR Ment Health*, *4*(1), e6. https://doi.org/10.2196/mental.5926

Morris, R. W., Kettlewell, N., & Glozier, N. (2021). The increasing cost of happiness. *SSM Popul Health*, *16*, 100949. https://doi.org/10.1016/j.ssmph.2021.100949

Mueller, A., Candrian, G., Grane, V. A., Kropotov, J. D., Ponomarev, V. A., & Baschera, G.-M. (2011). Discriminating between ADHD adults and controls using independent ERP components and a support vector machine: a validation study. *Nonlinear Biomedical Physics*, *5*(1), 5. https://doi.org/10.1186/1753-4631-5-5

Nalah, A., Bala, N., Ishaya, D., & Daniel. (2013). A Conceptual Overview of Deviance and Its Implication to Mental Health: a Bio psychosocial Perspective. *International Journal of Humanities and Social Science*, *2*, 01-09. https://doi.org/10.1037/rnd0000001

Nie, D., Ning, Y., & Zhu, T. (2012, 4-7 Dec. 2012). Predicting Mental Health Status in the Context of Web Browsing. 2012 IEEE/WIC/ACM

International Conferences on Web Intelligence and Intelligent Agent Technology,

Pang, H. (2021). Unraveling the influence of passive and active WeChat interactions on upward social comparison and negative psychological consequences among university students. *Telematics and Informatics*, *57*. https://doi.org/10.1016/j.tele.2020.101510

Pantic, I. (2014). Online social networking and mental health. *Cyberpsychol Behav Soc Netw*, *17*(10), 652-657. https://doi.org/10.1089/cyber.2014.0070

Parcesepe, A. M., & Cabassa, L. J. (2013). Public stigma of mental illness in the United States: a systematic literature review. *Adm Policy Ment Health*, *40*(5), 384-399. https://doi.org/10.1007/s10488-012-0430-z

Parimi, N., Lane, N. E., Bauer, D., Hochberg, M. C., Nevitt, M. C., & Study of Osteoporotic, F. (2010). Accuracy of self-reported diagnosis of hip replacement. *Arthritis Care Res (Hoboken)*, *62*(5), 719-724. https://doi.org/10.1002/acr.20111

Peck, J. A., Reback, C. J., Yang, X., Rotheram-Fuller, E., & Shoptaw, S. (2005). Sustained reductions in drug use and depression symptoms from treatment for drug abuse in methamphetamine-dependent gay and bisexual men. *Journal of Urban Health*, *82*(S1), i100-i108. https://doi.org/10.1093/jurban/jti029

Proudman, D., Greenberg, P., & Nellesen, D. (2021). The Growing Burden of Major Depressive Disorders (MDD): Implications for Researchers and Policy Makers. *Pharmacoeconomics*, *39*(6), 619-625. https://doi.org/10.1007/s40273-021-01040-7

Prusa, J., Khoshgoftaar, T. M., & Seliya, N. (2015, 9-11 Dec. 2015). The Effect of Dataset Size on Training Tweet Sentiment Classifiers. 2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA),

Rathje, S., Mirea, D.-M., Sucholutsky, I., Marjieh, R., Robertson, C. E., & Van Bavel, J. J. (2024). GPT is an effective tool for multilingual psychological text analysis. *Proceedings of the National Academy of Sciences*, *121*(34), e2308950121. https://doi.org/doi:10.1073/pnas.2308950121

Razzouk, D., Mari, J. J., Shirakawa, I., Wainer, J., & Sigulem, D. (2006). Decision support system for the diagnosis of schizophrenia disorders. *Brazilian Journal of Medical and Biological Research*, *39*.

Reddit. (2020). Reddit User Agreement. https://www.redditinc.com/policies/user-agreement-october-15-2020

Reddy, M. S. (2010). Depression: the disorder and the burden. *Indian J Psychol Med*, *32*(1), 1-2. https://doi.org/10.4103/0253-7176.70510

Rizal, H., Said, M. A., Abdul Majid, H., Su, T. T., Maw Pin, T., Ismail, R., & Shah Zaidi, M. A. (2022). Health-related quality of life of younger and older lower-income households in Malaysia. *PLoS One*, *17*(2), e0263751. https://doi.org/10.1371/journal.pone.0263751

Salas-Zárate, R., Alor-Hernández, G., Salas-Zárate, M. D. P., Paredes-Valverde, M. A., Bustos-López, M., & Sánchez-Cervantes, J. L. (2022). Detecting Depression Signs on Social Media: A Systematic Literature Review. *Healthcare (Basel)*, *10*(2). https://doi.org/10.3390/healthcare10020291

Sanchez-Villegas, A., Schlatter, J., Ortuno, F., Lahortiga, F., Pla, J., Benito, S., & Martinez-Gonzalez, M. A. (2008). Validity of a self-reported diagnosis of depression among participants in a cohort study using the Structured Clinical Interview for DSM-IV (SCID-I). *BMC psychiatry*, *8*(1), 43-43. https://doi.org/10.1186/1471-244X-8-43

Santini, Z. I., Jose, P. E., York Cornwell, E., Koyanagi, A., Nielsen, L., Hinrichsen, C., Meilstrup, C., Madsen, K. R., & Koushede, V. (2020). Social disconnectedness, perceived isolation, and symptoms of depression and anxiety among older Americans (NSHAP): a longitudinal mediation analysis. *The Lancet Public Health*, *5*(1), e62-e70. https://doi.org/10.1016/s2468-2667(19)30230-0

Santomauro, D. F., Mantilla Herrera, A. M., Shadid, J., Zheng, P., Ashbaugh, C., Pigott, D. M., Abbafati, C., Adolph, C., Amlag, J. O., Aravkin, A. Y., Bang-Jensen, B. L., Bertolacci, G. J., Bloom, S. S., Castellano, R., Castro, E., Chakrabarti, S., Chattopadhyay, J., Cogen, R. M., Collins, J. K., . . . Ferrari, A. J. (2021). Global prevalence and burden of depressive and anxiety disorders in 204 countries and territories in 2020 due to the COVID-19 pandemic.

The Lancet, *398*(10312), 1700-1712. https://doi.org/10.1016/S0140-6736(21)02143-7

Shawe-Taylor, J., Anthony, M., & Biggs, N. L. (1993). Bounding sample size with the Vapnik-Chervonenkis dimension. *Discrete Applied Mathematics*, *42*(1), 65-73. https://doi.org/https://doi.org/10.1016/0166-218X(93)90179-R

Shin, C., Park, M. H., Lee, S. H., Ko, Y. H., Kim, Y. K., Han, K. M., Jeong, H. G., & Han, C. (2019). Usefulness of the 15-item geriatric depression scale (GDS-15) for classifying minor and major depressive disorders among community-dwelling elders. *J Affect Disord*, *259*, 370-375. https://doi.org/10.1016/j.jad.2019.08.053

Siniscalchi, K. A., Broome, M. E., Fish, J., Ventimiglia, J., Thompson, J., Roy, P., Pipes, R., & Trivedi, M. (2020). Depression Screening and Measurement-Based Care in Primary Care. *J Prim Care Community Health*, *11*, 2150132720931261. https://doi.org/10.1177/2150132720931261

Sinyor, M., Schaffer, A., & Levitt, A. (2010). The Sequenced Treatment Alternatives to Relieve Depression (STAR*D) Trial: A Review. *The Canadian Journal of Psychiatry*, *55*(3), 126-135. https://doi.org/10.1177/070674371005500303

Smith, L., Jacob, L., Yakkundi, A., McDermott, D., Armstrong, N. C., Barnett, Y., López-Sánchez, G. F., Martin, S., Butler, L., & Tully, M. A. (2020). Correlates of symptoms of anxiety and depression and mental wellbeing associated with COVID-19: a cross-sectional study of UK-based respondents. *Psychiatry Res*, *291*, 113138. https://doi.org/10.1016/j.psychres.2020.113138

Solanto, M. V., & Alvir, J. (2009). Reliability of DSM-IV Symptom Ratings of ADHD: Implications for DSM-V. *Journal of attention disorders*, *13*(2), 107-116. https://doi.org/10.1177/1087054708322994

Song, H., Kim, M., Park, D., Shin, Y., & Lee, J. G. (2023). Learning From Noisy Labels With Deep Neural Networks: A Survey. *IEEE Trans Neural Netw Learn Syst*, *34*(11), 8135-8153. https://doi.org/10.1109/tnnls.2022.3152527

Stead, G. B. (2004). Culture and career psychology: A social constructionist perspective. *Journal of Vocational Behavior*, *64*(3), 389-406. https://doi.org/https://doi.org/10.1016/j.jvb.2003.12.006

Steer, R. A., Clark, D. A., Beck, A. T., & Ranieri, W. F. (1999). Common and specific dimensions of self-reported anxiety and depression: the BDI-II versus the BDI-IA. *Behav Res Ther*, *37*(2), 183-190. https://doi.org/10.1016/s0005-7967(98)00087-4

Steinkamp, J., & Cook, T. S. (2021). Basic Artificial Intelligence Techniques: Natural Language Processing of Radiology Reports. *Radiol Clin North Am*, *59*(6), 919-931. https://doi.org/10.1016/j.rcl.2021.06.003

Stevens, J. R. (2017). Replicability and Reproducibility in Comparative Psychology. *Front Psychol*, *8*, 862. https://doi.org/10.3389/fpsyg.2017.00862

Stewart, S. L., Celebre, A., Hirdes, J. P., & Poss, J. W. (2020). Risk of Suicide and Self-harm in Kids: The Development of an Algorithm to Identify High-Risk Individuals Within the Children's Mental Health System. *Child Psychiatry & Human Development*, *51*(6), 913-924. https://doi.org/10.1007/s10578-020-00968-9

Stewart, W. F., Ricci, J. A., Chee, E., Hahn, S. R., & Morganstein, D. (2003). Cost of lost productive work time among US workers with depression. *JAMA*, *289*(23), 3135-3144. https://doi.org/10.1001/jama.289.23.3135

Straw, I., & Callison-Burch, C. (2020). Artificial Intelligence in mental health and the biases of language based models. *PLoS One*, *15*(12), e0240376. https://doi.org/10.1371/journal.pone.0240376

Subrahmanyam, A. (2019). Big data in finance: Evidence and challenges. *Borsa Istanbul Review*, *19*(4), 283-287. https://doi.org/https://doi.org/10.1016/j.bir.2019.07.007

Tao, X., Chi, O., Delaney, P. J., Li, L., & Huang, J. (2021). Detecting depression using an ensemble classifier based on Quality of Life scales. *Brain Inform*, *8*(1), 2. https://doi.org/10.1186/s40708-021-00125-5

Thomas, K., Nilsson, E., Festin, K., Henriksson, P., Lowen, M., Lof, M., & Kristenson, M. (2020). Associations of Psychosocial Factors with Multiple Health Behaviors: A Population-Based Study of Middle-Aged Men and Women. *Int J Environ Res Public Health*, *17*(4). https://doi.org/10.3390/ijerph17041239

Tlachac, M. L., & Rundensteiner, E. (2020). Screening For Depression With Retrospectively Harvested Private Versus Public Text. *IEEE J Biomed Health Inform*, *24*(11), 3326-3332. https://doi.org/10.1109/JBHI.2020.2983035

Tolentino, J. C., & Schmidt, S. L. (2018). DSM-5 Criteria and Depression Severity: Implications for Clinical Practice. *Front Psychiatry*, *9*, 450. https://doi.org/10.3389/fpsyt.2018.00450

Trivedi, M. H., Kern, J. K., Grannemann, B. D., Altshuler, K. Z., & Sunderajan, P. (2004). A Computerized Clinical Decision Support System as a Means of Implementing Depression Guidelines. *Psychiatric Services*, *55*(8), 879-885. https://doi.org/10.1176/appi.ps.55.8.879

Tutun, S., Johnson, M. E., Ahmed, A., Albizri, A., Irgil, S., Yesilkaya, I., Ucar, E. N., Sengun, T., & Harfouche, A. (2023). An AI-based Decision Support System for Predicting Mental Health Disorders. *Information Systems Frontiers*, *25*(3), 1261-1276. https://doi.org/10.1007/s10796-022-10282-5

Twitter. (2022). Twitter Terms of Service. https://twitter.com/en/tos

van Hartskamp, M., Consoli, S., Verhaegh, W., Petkovic, M., & van de Stolpe, A. (2019). Artificial Intelligence in Clinical Health Care Applications: Viewpoint. *Interact J Med Res*, *8*(2), e12100. https://doi.org/10.2196/12100

Varraso, R., Jiang, R., Barr, R. G., Willett, W. C., & Camargo, C. A., Jr. (2007). Prospective study of cured meats consumption and risk of chronic obstructive pulmonary disease in men. *Am J Epidemiol*, *166*(12), 1438-1445. https://doi.org/10.1093/aje/kwm235

Véliz, C. (2019). Wellcome Trust–Funded Monographs and Book Chapters

Medical privacy and big data: A further reason in favour of public universal healthcare coverage. In T. C. de Campos, J. Herring, & A. M. Phillips (Eds.), *Philosophical Foundations of Medical Law*. Oxford University Press

© Carissa Véliz 2019.

Waisberg, E., Ong, J., Kamran, S. A., Paladugu, P., Zaman, N., Lee, A. G., & Tavakkoli, A. (2023). Transfer learning as an AI-based solution to address limited datasets in space medicine. *Life Sci Space Res (Amst)*, *36*, 36-38. https://doi.org/10.1016/j.lssr.2022.12.002

Weissman, M. M., Bland, R. C., Canino, G. J., Faravelli, C., Greenwald, S., Hwu, H. G., Joyce, P. R., Karam, E. G., Lee, C. K., Lellouch, J., Lepine, J. P., Newman, S. C., Rubio-Stipec, M., Wells, J. E., Wickramaratne, P. J., Wittchen, H., & Yeh, E. K. (1996). Cross-national epidemiology of major depression and bipolar disorder. *JAMA*, *276*(4), 293-299. https://www.ncbi.nlm.nih.gov/pubmed/8656541

Workman, A. D., Parasher, A. K., Blasetti, M. T., Palmer, J. N., Adappa, N. D., & Glicksman, J. T. (2019). Accuracy of Self-reported Diagnosis of Chronic Rhinosinusitis. *Otolaryngol Head Neck Surg*, *160*(3), 556-558. https://doi.org/10.1177/0194599818811328

Yager, J., Kay, J., & Kelsay, K. (2021). Clinicians' Cognitive and Affective Biases and the Practice of Psychotherapy. *American Journal of Psychotherapy*, *74*(3), 119-126. https://doi.org/10.1176/appi.psychotherapy.20200025

Yang, J., & Leskovec, J. (2011). *Patterns of temporal variation in online media* Proceedings of the fourth ACM international conference on Web search and data mining, Hong Kong, China. https://doi.org/10.1145/1935826.1935863

Yates, A., Cohan, A., & Goharian, N. (2017). Depression and Self-Harm Risk Assessment in Online Forums.

Yin, Z., & Wong, S. T. C. (2021). Artificial intelligence unifies knowledge and actions in drug repositioning. *Emerg Top Life Sci*, *5*(6), 803-813. https://doi.org/10.1042/ETLS20210223

Zahid, I. A., Joudar, S. S., Albahri, A. S., Albahri, O. S., Alamoodi, A. H., Santamaría, J., & Alzubaidi, L. (2024). Unmasking large language models by means of OpenAI GPT-4 and Google AI: A deep instruction-based analysis. *Intelligent Systems with Applications*, *23*, 200431. https://doi.org/https://doi.org/10.1016/j.iswa.2024.200431

Zhang, S., & Xiang, W. (2019). Income gradient in health-related quality of life - the role of social networking time. *Int J Equity Health*, *18*(1), 44. https://doi.org/10.1186/s12939-019-0942-1

Zhang, Z., Lin, W., Liu, M., & Mahmoud, M. (2020, 16-20 Nov. 2020). Multimodal Deep Learning Framework for Mental Disorder

Recognition. 2020 15th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2020),

Zhang, Z., Sun, K., Jatchavala, C., Koh, J., Chia, Y., Bose, J., Li, Z., Tan, W., Wang, S., Chu, W., Wang, J., Tran, B., & Ho, R. (2019). Overview of Stigma against Psychiatric Illnesses and Advancements of Anti-Stigma Activities in Six Asian Societies. *Int J Environ Res Public Health*, *17*(1). https://doi.org/10.3390/ijerph17010280

Zigmond, A. S., & Snaith, R. P. (1983). The hospital anxiety and depression scale. *Acta Psychiatr Scand*, *67*(6), 361-370. https://doi.org/10.1111/j.1600-0447.1983.tb09716.x

APPENDIX

Part of this research was published by the following conference:

- 11th International Conference on Behavioural and Social Computing (2024)