DEVELOPING AN INTELLIGENT CONVERSATIONAL AGENT FOR EARLY DETECTION AND SUPPORT OF INDIVIDUALS WITH DEPRESSION

MOHAMAD DZUL SYAKIMIN

RESEARCH PROJECT SUBMITTED

TO THE FACULTY OF COMPUTER SCIENCE

& INFORMATION TECHNOLOGY UNIVERSITY OF

MALAYA,

IN PARTIAL REQUIREMENTS

FOR THE DEGREE OF MASTER OF DATA SCIENCE

2024

UNIVERSITY OF MALAYA

ORIGINAL LITERARY WORK DECLARATION

Name of Candidate: MOHAMAD DZUL SYAKIMIN BIN MOHD NOOR

Matric No: 22050897

Name of Degree: MASTER OF DATA SCIENCE

Title of Project Paper/Research Report/Dissertation/Thesis ("this Work"):

DEVELOPING AN INTELLIGENT CONVERSATIONAL AGENT FOR EARLY DETECTION AND SUPPORT OF INDIVIDUALS WITH DEPRESSION

Field of Study: Artificial Intelligence in Mental Health

I do solemnly and sincerely declare that:

(1) I am the sole author/writer of this Work;

(2) This Work is original;

(3) Any use of any work in which copyright exists was done by way of fair dealing and for permitted purposes and any excerpt or extract from, or reference to or reproduction of any copyright work has been disclosed expressly and sufficiently and the title of the Work and its authorship have been acknowledged in this Work.

(4) I do not have any actual knowledge nor do I ought reasonably to know that the making of this work constitutes an infringement of any copyright work;

(5) I hereby assign all and every rights in the copyright to this Work to the University of Malaya ("UM"), who henceforth shall be owner of the copyright in this Work and that

any reproduction or use in any form or by any means whatsoever is prohibited without the written consent of UM having been first had and obtained;

(6) I am fully aware that if in the course of making this Work I have infringed any copyright whether intentionally or otherwise, I may be subject to legal action or any other action as may be determined by UM.

Candidate's Signature   Date: 04 December 2024

dzul

Subscribed and solemnly declared before,

Witness's Signature   Date: 04 December 2024

Name:

Designation:

Developing an Intelligent Conversational Agent for Early Detection and Support of Individuals with Depression

Abstract

According to the World Health Organization, as of 2017, over 300 million people globally were affected by depression, making it a leading cause of disability worldwide. The impact of depression is exacerbated by chronic stress, which significantly increases susceptibility to the condition. While cognitive-behavioral therapy (CBT) remains one of the most effective treatments for depression, access to therapy is often hindered by stigma, cost, and resource constraints. The sub-branch of CBT, behavioral activation therapy, offers a cost-effective alternative by promoting positive reinforcement to enhance individuals' moods.

This research aims to develop a conversational agent capable of detecting depression through natural language interactions. The agent employs state-of-the-art natural language processing (NLP) and machine learning algorithms to analyze linguistic patterns, emotional tones, and conversational turn-taking. By addressing the challenges of early detection and providing non-intrusive support, this study contributes to the application of artificial intelligence in mental health interventions. While real-world trials remain pending, the findings highlight the potential of AI-driven conversational systems in promoting mental well-being.

Keywords: *Depression, stress, conversational agents, machine learning, natural language processing, mental health technolog*

ACKNOWLEDGEMENTS

Without God's will, nothing can be achieved. I extend my deepest gratitude to the Almighty for guiding me through this journey and granting me the strength and perseverance needed to complete this project.

I am profoundly thankful to Dr. Maizatul Akmal Binti Ismail for her unwavering support, invaluable guidance, and encouragement throughout the timeline of this research. Her expertise and timely feedback have been instrumental in steering this project in the right direction. I am truly grateful for the opportunity to work under her supervision.

I would also like to extend my heartfelt thanks to my family and friends for their constant encouragement and support. Their belief in my capabilities and understanding during challenging times have been a pillar of strength for me. A special mention goes to my brother, whose patience and unwavering assistance have been a source of immense comfort during this journey.

This work is dedicated to all those who have supported and inspired me along the way.

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION

Depression is a prevalent mental health disorder affecting millions worldwide, with over 264 million individuals reported as affected globally in recent years (Olowolayemo et al., 2023). The condition significantly impacts personal well-being and societal productivity, leading to challenges in daily functioning and increased economic burden. Despite its pervasiveness, depression remains underdiagnosed and undertreated due to factors such as stigma, lack of awareness, and limited access to mental health services (Pacheco-Lorenzo et al., 2021).

Advancements in artificial intelligence (AI) and natural language processing (NLP) offer promising solutions to address these challenges. By leveraging linguistic patterns, emotional tone analysis, and conversational dynamics, AI-driven systems can provide non-intrusive and scalable methods for early detection and intervention. Conversational agents, in particular, have emerged as innovative tools capable of engaging users through natural language interactions to assess mental health conditions, including depression (Otero-González et al., 2024).

Existing applications in this domain have demonstrated potential in assisting users with mental health management. For example, conversational agents have been used to identify depression markers by analyzing textual inputs, offering evidence-based guidance, and suggesting resources for professional help (Martínez-Miranda et al., 2019). These systems aim to bridge the gap between affected individuals and mental health support, especially for those hesitant to seek traditional clinical intervention.

However, while several conversational agents have shown efficacy in specific settings, significant challenges remain. Issues such as ensuring algorithmic accuracy, maintaining user privacy, and overcoming engagement barriers highlight the need for further research (Pacheco-Lorenzo et al., 2021). Moreover, understanding the linguistic and behavioral markers of depression across diverse populations requires robust data collection and analysis frameworks.

This research focuses on developing an intelligent conversational agent designed for early detection and support of individuals experiencing depression. By utilizing state-of-the-art machine learning techniques, the agent will analyze user inputs to identify depressive symptoms and provide non-intrusive support. The goal is to address gaps in accessibility and early intervention while maintaining ethical standards and user trust. This study aims to contribute to the growing body of research on AI-driven solutions in mental health, with implications for societal well-being and technological innovation.

1.1 Problem Statement

Depression is a widespread mental health issue affecting over 264 million people globally (Olowolayemo et al., 2023). The COVID-19 pandemic has worsened this crisis, amplifying stress and depression through increased isolation, economic challenges, and uncertainty, underscoring the urgent need for mental health support (Pacheco-Lorenzo et al., 2021).

In Malaysia, mental health challenges have surged, with 468 suicide cases reported in the first five months of 2021—a sharp rise compared to prior years (Otero-González et al., 2024). Social stigma and high therapy costs, coupled with a shortage of trained professionals, have left many untreated or undiagnosed.

Technology, particularly AI and natural language processing, offers potential solutions to these challenges. However, current mental health tools face limitations such as low accuracy, high attrition rates, and limited adaptability (Martínez-Miranda et al., 2019). This research aims to develop a conversational agent to detect and support individuals with depressive symptoms, bridging the gap between technology and mental health interventions to provide timely, accessible, and affordable support.

1.2 Research Objectives

The primary focus of this research is to design an intelligent conversational agent for the early detection and support of individuals experiencing depression. The study aims to achieve the following objectives:

- To create a dataset of linguistic and conversational features that are indicative of depression.
- To develop a conversational agent using advanced natural language processing and machine learning techniques for accurate detection of depressive symptoms.
- To evaluate the agent's ability to provide non-intrusive support and guidance to individuals identified as potentially experiencing depression.

1.3 Research Questions

The research aims to answer the following questions to achieve the respective objectives:

- What linguistic and conversational features are most indicative of depression?
- How can natural language processing and machine learning algorithms be optimized for accurate detection of depressive symptoms?
- How can the conversational agent provide personalized and non-intrusive support to individuals identified as potentially experiencing depression?

1.4 Research Scope

The dataset for this research will consist of linguistic and conversational features indicative of depression, gathered from publicly available resources and validated through expert analysis. The conversational agent will be designed using advanced natural language processing and machine learning techniques, focusing on text-based interactions to detect depressive symptoms. The scope of this research will be confined to the development of a functional prototype that can analyze conversational data and provide initial support recommendations. Due to time and resource limitations, the study will not include the integration of real-time interventions or the development of multi-modal features such as voice or video analysis. Additionally, the evaluation of the agent's effectiveness will be conducted through simulated data and user scenarios rather than large-scale real-world deployment.

1.5 Research Significance

Depression is a significant mental health challenge that affects individuals across all demographics. Early detection and timely intervention can greatly improve treatment outcomes and quality of life. However, barriers such as stigma, limited access to mental health services, and lack of awareness prevent many from seeking help. This research focuses on developing a conversational agent to bridge this gap by providing an accessible and non-intrusive tool for detecting depressive symptoms.

By leveraging natural language processing and machine learning, the proposed conversational agent can analyze conversational patterns to identify signs of depression and offer initial support. This study contributes to advancing the use of artificial intelligence in mental health, emphasizing the importance of early detection and intervention. It also aligns with the growing need for scalable and cost-effective mental health solutions, offering a potential pathway to reduce the societal burden of depression while empowering individuals to seek appropriate care.

1.6 Research Motivation

Conversational agents have become increasingly popular in various fields, from customer service to education. Their ability to engage users through natural language interactions makes them powerful tools for addressing complex problems. While these systems are widely utilized for commercial and entertainment purposes, their potential to support mental health remains underexplored.

Depression is a growing concern worldwide, with millions of individuals affected yet unable to access timely and effective support. The idea of harnessing conversational agents to detect and support individuals with depressive symptoms stems from the need to bridge this gap. The pandemic has further highlighted the importance of accessible mental health solutions, as traditional services struggle to meet increasing demands.

This research is motivated by the belief that technology, particularly artificial intelligence, can play a transformative role in mental health. Creating a conversational agent capable of identifying early signs of depression and offering non-intrusive support presents an opportunity to make mental health care more accessible and scalable. It combines the fields of AI and psychology to address a pressing societal issue, aiming to contribute to the well-being of individuals and communities.

CHAPTER 2: LITERATURE REVIEW

2.1 Increasing Challenges in Mental Health Detection

The prevalence of depression has seen a substantial rise globally, with over 264 million individuals affected as of recent estimates (Olowolayemo et al., 2023). This upward trend is exacerbated by factors such as societal stigma, lack of awareness, and limited access to professional mental health services. Early detection and intervention are critical to mitigating its impact. However, conventional methods of diagnosis are time-consuming, resource-intensive, and inaccessible to many. Recent advancements in artificial intelligence (AI) and natural language processing (NLP) offer promising solutions for scalable and efficient mental health assessment.

2.2 Barriers in Existing Approaches to Depression Detection

Despite the increasing awareness of mental health and the growing need for effective interventions, traditional approaches to depression detection and treatment face significant challenges. One of the most pervasive barriers is the stigma surrounding mental health, which often discourages individuals from seeking help due to fear of judgment or social repercussions. Stigma can lead to delayed diagnosis and treatment, exacerbating the severity of depression and reducing the effectiveness of interventions (Pacheco-Lorenzo et al., 2021).

In addition to stigma, access to professional mental health services is limited by high costs and geographical constraints. Therapy sessions, while effective, are expensive and often not covered comprehensively by insurance plans. For individuals in low-income groups or those residing in remote areas, accessing qualified mental health professionals is particularly challenging. The disparity in mental health service availability is further widened by the uneven distribution of mental health professionals, with urban areas having greater access compared to rural regions.

Moreover, traditional face-to-face methods of diagnosis and therapy require significant time commitments, which can deter individuals juggling work, family, or other responsibilities. This gap in accessibility creates an urgent need for alternative methods that are both cost-effective and convenient. For many, digital solutions such as conversational agents and mobile applications offer a practical alternative by providing on-demand support that fits into their daily lives.

However, existing digital solutions are not without their limitations. Many platforms fail to address personalization and contextual nuances, making their interventions less effective for diverse populations. Privacy concerns also play a crucial role in hindering adoption, as users may hesitate to share sensitive information on digital platforms. Without robust security measures and transparency about data usage, these technologies risk alienating potential users.

These challenges highlight the importance of integrating technology-driven approaches, particularly conversational agents, into mental health care. These agents can provide scalable, non-intrusive, and user-centric solutions for early detection and support. By leveraging advancements in natural language processing and machine learning, conversational agents can analyze linguistic and emotional patterns to identify depressive symptoms, overcoming the barriers associated with traditional methods while making mental health support more inclusive and accessible.

2.3 Applications of Conversational Agents in Mental Health

2.3.1 Web-Based Approaches

Several studies have explored web-based interventions to support mental health. For instance, Zhou et al. (2022) reviewed nine studies that utilized AI for mental health diagnosis and intervention. The authors highlighted that these systems leverage conversational data to identify depressive symptoms, offering a scalable alternative to face-to-face therapy. While effective in certain settings, these systems often lack personalization and adaptability.

2.3.2 Machine Learning in Conversational Agents

Recent advancements in machine learning have paved the way for more sophisticated conversational agents. These systems employ algorithms to analyze linguistic patterns, emotional tones, and conversational dynamics. For example, Zhang et al. (2019) developed a smartphone-based application using reinforcement learning algorithms to provide personalized support. Despite achieving an accuracy rate of 80% in detecting depressive symptoms, the application faced high attrition rates, highlighting the need for improved user engagement strategies.

Similarly, Martínez-Miranda et al. (2019) evaluated the acceptability of mobile-based conversational agents in detecting suicidal tendencies. Their findings revealed a positive reception among users, with the agents effectively identifying critical conversational cues. However, the study emphasized the importance of integrating real-time feedback mechanisms to enhance the agents' effectiveness.

2.4 Impact of AI on Mental Health Assessment

Artificial intelligence (AI) has emerged as a transformative tool in the field of mental health assessment, offering innovative solutions to some of the most pressing challenges in mental health care. AI-driven tools, particularly those employing natural language processing (NLP), have demonstrated significant promise in identifying depression markers through text-based interactions. These tools can analyze linguistic features, emotional tones, and conversational patterns to detect subtle signs of depression that may be overlooked in traditional assessments. For instance, AI systems can identify increased usage of first-person pronouns and negative emotion words, linguistic indicators often linked to depressive symptoms. The ability to process and analyze large volumes of conversational data in real time positions AI as a scalable solution for mental health screening in diverse populations.

A comprehensive review by Otero-González et al. (2024) highlights the critical role of conversational agents in democratizing access to mental health support. By enabling users to interact with AI-powered systems in non-judgmental and private environments, these tools help reduce the stigma often associated with seeking mental health care. The review underscores how conversational agents can provide personalized recommendations, encourage users to engage with therapeutic interventions, and guide them toward professional care when necessary. Moreover, these systems are accessible at any time, making them particularly valuable for individuals in remote areas or those with limited access to professional therapists.

However, despite their advantages, AI-driven tools face significant limitations that need to be addressed for widespread adoption. One major concern is the lack of transparency in how these systems operate. Many models function as "black boxes," making it difficult for users and even practitioners to understand the rationale behind their recommendations or diagnoses. This opacity can erode trust, particularly in sensitive domains such as mental health. Another challenge lies in the ethical considerations surrounding user data. While AI systems require extensive data for training and optimization, concerns about data privacy and security remain paramount. Without robust safeguards and clear communication about data usage, these systems risk alienating users who may already be hesitant to seek mental health support.

Furthermore, the integration of AI into mental health assessment raises questions about the cultural and contextual adaptability of these tools. Mental health expressions and idioms vary significantly across cultures, and AI systems trained on limited datasets may fail to recognize culturally specific markers of depression. Addressing these challenges requires a multi-faceted approach, including enhancing model interpretability, implementing stringent data privacy measures, and ensuring cultural inclusivity in training datasets. As research in AI-driven mental health tools progresses, these improvements will be critical in ensuring that these technologies not only meet the needs of diverse populations but also build trust and confidence among users and practitioners alike.

2.5 Need for Ethical and Accessible Solutions

The integration of artificial intelligence (AI) into mental health assessment and intervention brings with it profound ethical challenges that demand careful consideration. Chief among these is the issue of data privacy, as AI-driven tools rely heavily on the collection and processing of sensitive personal information. Users entrust these systems with intimate details about their mental health, emotions, and behaviors, making robust data protection mechanisms an absolute necessity. Any breach of this trust, whether through inadequate encryption, unauthorized access, or unclear data-sharing practices, could deter individuals from using such tools, thereby undermining their potential benefits. Moreover, obtaining informed user consent is equally crucial. Users must not only be aware of what data is being collected but also understand how it will be used, stored, and shared. Transparent communication about these aspects is essential for building and maintaining trust in AI-powered mental health systems.

Another significant concern is the issue of algorithmic bias. AI systems are only as unbiased as the data on which they are trained, and datasets used for mental health tools often lack diversity. This can result in biases that disproportionately affect certain groups, such as minorities, non-native language speakers, or individuals from underserved communities. For instance, a conversational agent trained on data predominantly from Western cultures may fail to recognize culturally specific expressions of mental health symptoms in other populations. Such biases not only limit the effectiveness of these tools but also risk perpetuating systemic inequities in mental health care. Addressing this requires deliberate efforts to diversify training datasets and incorporate insights from cross-cultural research to ensure that AI systems are inclusive and sensitive to the needs of diverse populations.

Equally critical is the challenge of accessibility, particularly for individuals with limited technological literacy or access to digital devices. While AI has the potential to democratize mental health care, it can inadvertently create barriers for those who are less familiar with technology or lack stable internet access. Solutions must be designed to be user-friendly, requiring minimal technical expertise, and should consider offline or low-bandwidth functionalities to accommodate users in remote or underserved regions. Furthermore, the language interface of conversational agents should be inclusive, supporting a wide range of languages and dialects to ensure that users from various linguistic backgrounds can engage effectively.

To fully realize the potential of AI-driven mental health tools, ethical considerations and accessibility must be at the forefront of their design and implementation. Collaboration between technologists, mental health professionals, ethicists, and policymakers is essential to create frameworks that address these concerns comprehensively. By prioritizing user trust, cultural inclusivity, and universal accessibility, these systems can become a transformative force in mental health care, bridging gaps and offering support to those who might otherwise remain underserved.

2.6 Summary and Research Gaps

While existing literature highlights the potential of conversational agents in mental health, several gaps remain. These include the need for improved personalization, enhanced user engagement, and robust ethical frameworks. This study aims to address these gaps by developing a conversational agent that leverages advanced NLP and machine learning techniques for the early detection of depression. By integrating real-time feedback and ensuring user-centric design, the proposed system seeks to provide accessible and effective mental health support.

2.7 Supplementary Literature

2.7.1 Sentiment Analysis for Mental Health Detection

Sentiment analysis has become a valuable tool in understanding user emotions, particularly in the context of mental health. Tadesse et al. (2019) explored the application of sentiment analysis to detect depression through social media posts. The study employed natural language processing (NLP) techniques to analyze textual data and identify depressive markers such as negative sentiment, low energy phrases, and self-referential language. By combining sentiment scores with linguistic features, the authors demonstrated an improved accuracy of depression detection. While the study was limited to publicly available datasets, it underscored the potential of integrating sentiment analysis in AI-driven mental health tools, providing a foundation for more personalized and context-aware systems.

2.7.2 Multimodal Approaches for Mental Health Assessment

Multimodal systems that integrate textual, visual, and audio data offer promising advancements in mental health assessment. Cummins et al. (2018) investigated the use of multimodal AI models to detect depression by analyzing speech patterns, facial expressions, and textual input simultaneously. The authors utilized convolutional neural networks (CNNs) and recurrent neural networks (RNNs) to process and combine data from multiple modalities, achieving higher predictive accuracy compared to single-modal approaches. However, the study highlighted challenges such as computational complexity, dataset availability, and ethical concerns regarding user privacy. Despite these barriers, the findings suggest that multimodal approaches could significantly enhance the effectiveness of conversational agents by providing a holistic understanding of user emotions and behaviors.

2.7.3 Emotion Recognition in Conversational Agents

Emotion recognition is a critical component of conversational agents aimed at supporting mental health. Schuller et al. (2020) proposed an emotion-aware chatbot that uses prosodic and linguistic features to identify emotional states during interactions. The chatbot employs a combination of rule-based algorithms and machine learning models to adapt its responses based on detected emotions. For example, if the agent identifies sadness in a user's input, it may respond with empathy and suggestive supportive actions. Although the study demonstrated the feasibility of emotion-aware systems, it noted limitations such as reliance on predefined emotional categories and the need for real-time processing capabilities. Future developments could focus on incorporating continuous emotion spectra and improving response adaptability.

2.7.4 Adaptive Learning in Conversational Systems

Adaptive learning has been explored to enhance user engagement in conversational systems. Xie et al. (2021) implemented an adaptive conversational agent that learns from user interactions to tailor its recommendations and dialogue strategies. The system employed reinforcement learning algorithms to optimize user satisfaction metrics over time. For instance, the agent adapted its tone and suggestions based on user feedback, creating a more personalized and engaging experience. While the study demonstrated improved user retention and satisfaction, it emphasized the need for extensive training data and ethical considerations in dynamic learning environments.

2.7.5 Existing Datasets for Depression Detection

Several datasets have been developed to analyze depression through linguistic and conversational features. However, many have limitations, such as accessibility, language focus, or modality constraints:

1. **DEPAC (Depression and Anxiety Corpus):** Focuses on acoustic and linguistic features for depression detection (Chen et al., 2024).
2. **EATD-Corpus:** A multimodal dataset combining audio and text for depression detection, primarily in Chinese (Tao et al., 2022).
3. **DAIC (Distress Analysis Interview Corpus):** Provides clinical interview data, including audio, video, and transcripts, though with restricted access (Gratch et al., 2014).
4. **Reddit-based Datasets:** Social media-derived datasets for linguistic analysis of depression markers, accessible under specific agreements (Burcaram, 2023).

Despite these efforts, limitations such as dataset accessibility, language specificity, and a lack of multimodal features highlight the need for new, robust datasets. This research aims to address these gaps by developing a diverse and accessible dataset of linguistic and conversational features for depression detection.

2.7.6 Summary of Supplementary Literature

These studies collectively highlight the advancements and challenges in integrating AI technologies into mental health interventions. From sentiment analysis to multimodal approaches and emotion recognition, these tools offer innovative pathways to enhance the capabilities of conversational agents. However, issues such as computational demands, data privacy, and cultural adaptability remain critical barriers to widespread implementation. Addressing these challenges through interdisciplinary research and user-centric design will be essential for realizing the full potential of AI in mental health care.

The table and the discussed gaps in the literature (e.g., dataset limitations, lack of personalization, ethical concerns) justify the project's significance. Developing a conversational agent tailored for early depression detection bridges critical gaps in existing methods, providing:

- **Scalability**: AI-powered solutions can reach broader populations compared to traditional interventions.
- **Accessibility**: Non-intrusive and cost-effective tools for underserved areas.
- **Innovation**: Combines advanced NLP and machine learning with real-world application potential.

Table 2.1 summarizes the literature review section of this study. It contains the precise information of the studies that have been examined and have enabled us to reach the project’s end goal.

Table 2.1: LR Summary

Study / Reference	**ML Methods Used**	**Strengths**	**Weaknesses**
Chen et al. (2024) arXiv	Neural Networks, SVMs	Utilizes speech signals across varied interaction scenarios; confirms speech as a crucial marker for depression screening	Requires extensive data preprocessing; potential variability in speech patterns across different demographics
Qin et al. (2023) arXiv	Large Language Models (LLMs)	Provides interpretable and interactive depression detection; integrates professional diagnostic criteria	Challenges in processing large text volumes; integration complexity with existing systems

Belcastro et al. (2024) arXiv	BERTweet, Explainable AI (XAI)	Combines LLMs with XAI for interpretable detection; enhances explanations via ChatGPT	Dependence on quality of social media data; potential biases in language models
Lorenc et al. (2022) arXiv	Transfer Learning	Addresses low-resource scenarios; achieves high recall in early depression detection	Limited by availability of conversational data; transferability issues between domains
MIT Researchers (2018) MIT News	Neural-Network Model	Analyzes raw text and audio data; discovers speech patterns indicative of depression	May require large datasets for training; potential ethical concerns regarding data privacy

Olowolayemo et al. (2023) ResearchGate	Systematic Review of Conversational Agents	Provides comprehensive analysis of conversational agents for depression detection	May not present new empirical data; relies on existing studies
Nickson et al. (2023) BMC Medical Informatics	Machine Learning with Electronic Health Records	Summarizes ML methods for depression prediction using health records	Focuses on electronic health records; may not address conversational data
Systematic Review (2023) Nature	Meta-Analysis of AI-based Conversational Agents	Evaluates effectiveness of AI-based conversational agents in reducing depression symptoms	May not cover latest developments; potential publication bias

PLOS ONE Study (2022) PLOS Journals	Conversational AI Bot	Explores early detection of depression using conversational AI	Non-randomized design; may have limited generalizability
WSJ Article (2024) The Wall Street Journal	AI Chatbots	Discusses potential of AI chatbots in supporting mental health care	Not a peer-reviewed study; focuses on general applications

CHAPTER 3: METHODOLOGY

The methodology for this research was designed to develop a conversational AI system capable of identifying early signs of depression. The framework leverages advanced natural language processing (NLP) techniques and machine learning models, supported by exploratory data analysis (EDA) and robust evaluation processes.

The general architecture for this AI-based system is illustrated in Figure 3.1. As shown in the diagram, the system processes user input through a conversational interface, which is analyzed by the AI model to identify patterns indicative of depression. The processed data is stored in a database, and the system loops user feedback into the model to refine its recommendations over time. This iterative feedback mechanism ensures that the system remains adaptive to evolving linguistic nuances and user-specific needs.

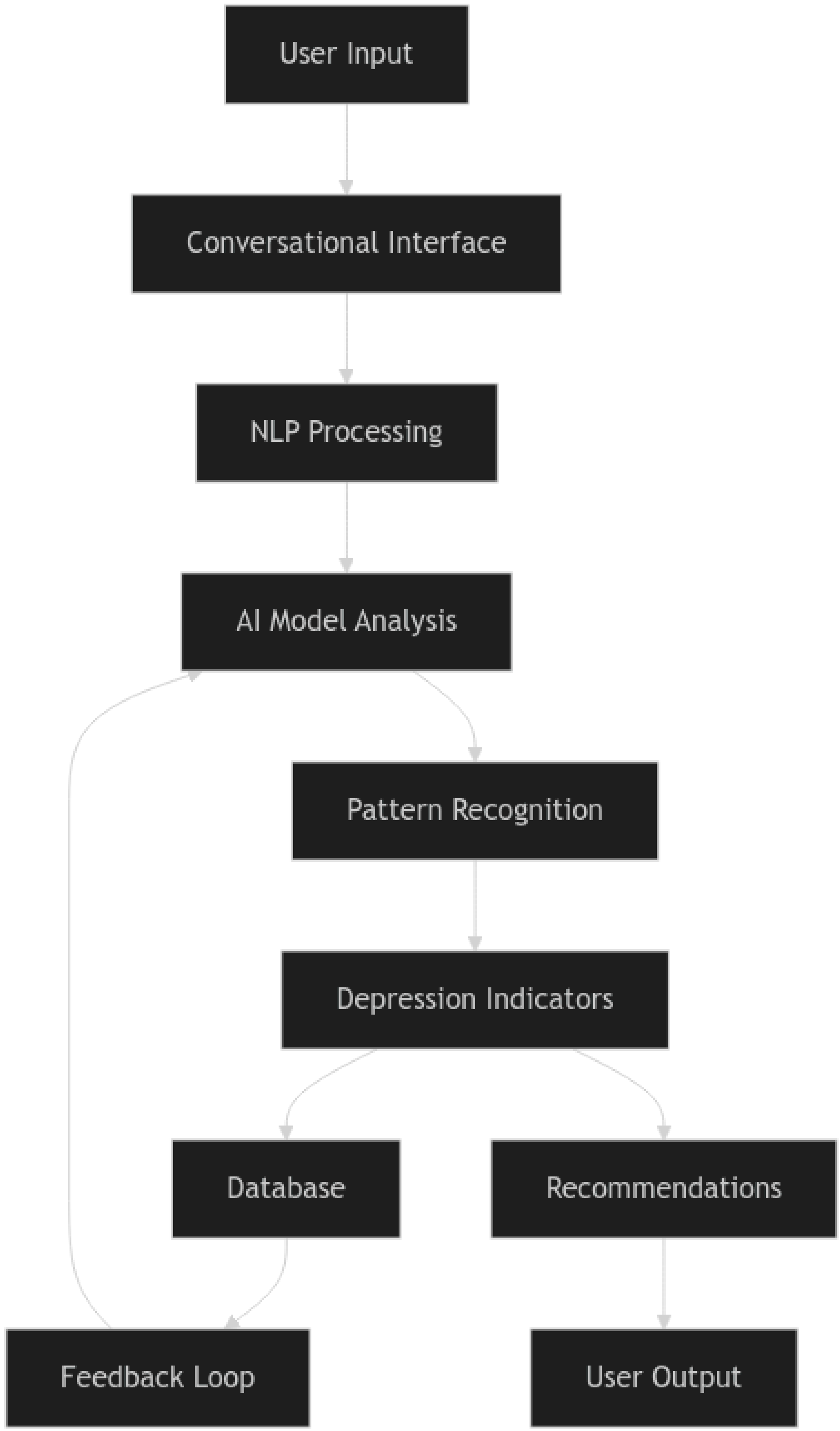


Figure 3.1: Architecture of Recommender system

The system has been implemented using Python programming language, with Jupyter Notebook as the primary development environment. The following Python libraries were utilized:

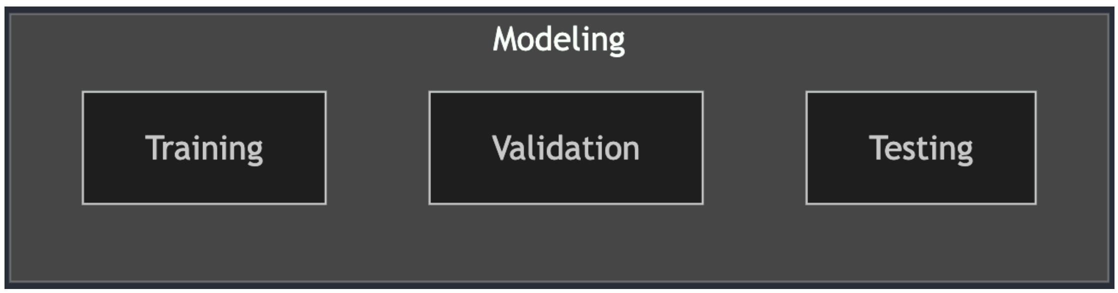
- **Numpy**: For handling numerical operations and dataset transformations.
- **Pandas**: To manipulate and structure data efficiently.
- **Matplotlib and Seaborn**: For visualizing trends and relationships during the EDA phase.
- **Scikit-learn**: For applying machine learning algorithms, including classification and clustering models.
- **Yellowbrick**: To evaluate and visualize model performance.
- **Pickle**: For saving trained models and encoders, facilitating deployment and reproducibility.

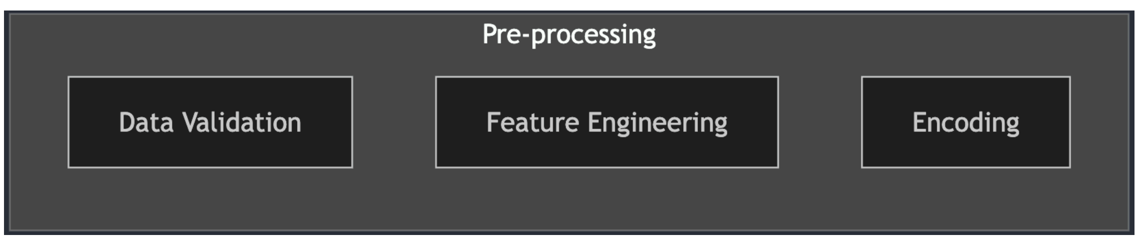
The project followed a systematic data science lifecycle, as depicted in Figure 3.2, which comprises five distinct steps:

1. **Formulation of a Raw Dataset**: The dataset was sourced from publicly available corpora and enriched using additional surveys, ensuring it adheres to ethical and scientific standards.
2. **Data Pre-processing**: Multiple sub-steps were employed, including data validation, feature engineering, and encoding, to transform the raw dataset into a format suitable for analysis and modeling.
3. **Exploratory Data Analysis (EDA)**: Aimed at uncovering insights and patterns within the data, EDA served as a foundation for feature selection and hypothesis testing.

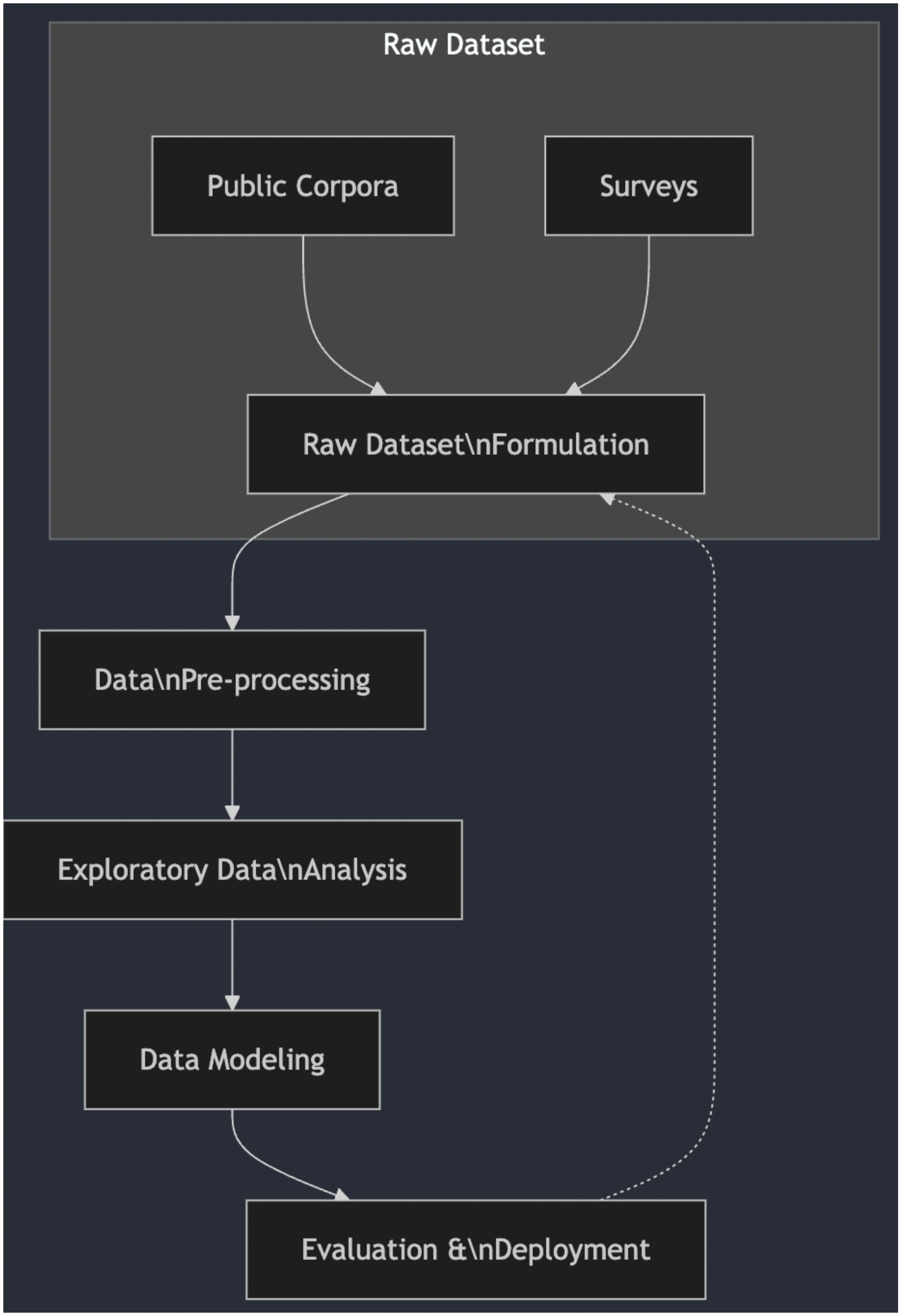
4. **Data Modeling**: Machine learning models were trained and validated to classify and predict depressive patterns in conversational data.
5. **Evaluation and Deployment**: The models were rigorously tested for accuracy, interpretability, and scalability before deployment as a real-time application.

The methodology outlined here forms the backbone of the research and ensures the integration of robust scientific practices with cutting-edge AI technologies.





3.2: Data Science Lifecycle



3.2: Data Science Lifecycle

3.1 Dataset: Data Collection and Data Description

The dataset utilized for this study was sourced from publicly available corpora and online resources tailored for linguistic and psychological research. Specific steps in this phase included:

- **Literature Review**: Extracted linguistic markers and conversational patterns indicative of depressive symptoms.
- **Dataset Sources**: Key datasets included publicly accessible Reddit posts tagged with mental health indicators and datasets like DAIC-WOZ and EATD-Corpus.
- **Survey Contribution**: An additional layer of data was collected via an online survey, capturing user interactions to simulate real-world conversational exchanges.

The combined dataset is a carefully curated and comprehensive collection designed to enable effective machine learning for depression detection. It encompasses multiple layers of structured and unstructured data:

- **Text-based Features**: The dataset captures linguistic elements such as words, phrases, and sentiment scores, serving as the foundational input for natural language processing. These features are essential for uncovering nuanced patterns of emotional expression in user conversations.
- **Metadata**: Critical contextual details, including speaker roles (e.g., user, system), timestamps marking conversational turns, and interaction types (e.g., question, response), are included. These metadata attributes enhance the ability to model conversational flow and intent dynamics.

- **Labeled Data**: Each conversation is categorized into predefined classes—depressive or non-depressive—based on linguistic markers identified through a rigorous labeling process. This ensures the dataset is ready for supervised learning tasks.
- **Volume**: The dataset contains over 15,000 conversational entries, balanced meticulously across multiple classes. This volume ensures statistical robustness while providing ample data diversity for model generalization.

3.2 Data pre-processing

Data preparation is a crucial step in ensuring the dataset's readiness for machine learning algorithms. The process involved multiple phases, each aimed at refining the raw data into a structured format suitable for advanced modeling.

3.2.1 Data Validation

This step ensured the integrity and relevance of the dataset. Unlike conventional validation, the dataset was reviewed by a professional psychiatric counselor to confirm that the linguistic markers and conversational patterns aligned with recognized indicators of depression.

3.2.2 Renaming and Standardization

Variable names were standardized to improve readability and coding efficiency. Spaces and special characters were replaced with underscores, and all text was converted to lowercase. For example:

"Level of difficulty" was renamed to "level_of_difficulty"

"Related mood" became "related_mood"

This step minimized errors during feature engineering and model training.

3.2.3 Data Transformation

String-based categorical variables were transformed into numerical representations using the LabelEncoder module from Scikit-learn. This transformation allowed the dataset to be compatible with machine learning models.

Below are examples showing the dataset before and after transformation:

Table 3.1: Sample Data Before Transformation

Values	**Activities**	**Category**	**Level of Difficulty**	**Related Mood**

Employment	Creative experiments	Valued Activity	Hard	Uncreative
Mental/Emotional Health	Going camping	Pleasure	Medium	Burned-out
Hobbies	Solving puzzles	Pleasure	Medium	Self-doubt

Table 3.2: Sample Data After Transformation

Values	Activities	Category	Level_of_Difficulty	Related_Mood
1	Creative experiments	2	1	19
5	Going camping	1	2	2
4	Solving puzzles	1	2	14

3.2.4 Input-Output Variable Split

The dataset was divided into two groups:

Input Variables (X): Includes features like values, category, level_of_difficulty, and related_mood.

Output Variable (y): The target activity suggested based on user input.

Table 3.3: Sample Input Data (X)

Values	Category	Level_of_Difficulty	Related_Mood
0	2	0	10
0	1	2	4
0	2	1	17

0	2	2	4

Table 3.4: Sample Output Data (y)

Activities
Upcycling/Recycling old items
Participate in communal activities
Making donations
Volunteering for a cause

3.2.5 Principal Component Analysis (PCA)

To reduce the dataset's dimensionality and improve visualization, PCA was applied. This reduced a 4-dimensional feature space to a 2-dimensional space. PCA helps identify patterns and relationships within the data, crucial for clustering and classification tasks.

This comprehensive preprocessing ensures that the dataset is not only clean and structured but also optimized for the subsequent modeling phase.

3.3 Data Analysis

The dataset used in this research contains no true labels, making it suitable for unsupervised learning techniques. The primary goal of the data analysis phase is to categorize the data into meaningful clusters that can be used to recommend behavioral activation activities for stress management. To achieve this, clustering algorithms were employed, and the optimal number of clusters was determined using three scoring methods: distortion, silhouette, and calinski_harabasz scores. The value of k (number of clusters) was set between 2 and 15, allowing for a range of clustering possibilities.

3.3.1 Description of Algorithms

K-means Clustering

K-means is an unsupervised learning algorithm that partitions a dataset into k clusters, where each data point belongs to the cluster with the nearest mean (centroid). The algorithm iteratively assigns data points to clusters and updates the centroids until convergence is achieved. The number of clusters, k, is predefined by the user. K-means is widely used due to its simplicity and efficiency, but it has limitations, such as sensitivity to initial centroid placement and the assumption of spherical clusters (Arthur & Vassilvitskii, 2007).

Advantages:

- Efficient for large datasets.
- Easy to implement and interpret.

Limitations:

- Requires the number of clusters (k) to be specified in advance.
- Sensitive to outliers and initial centroid placement.

Agglomerative Clustering

Agglomerative clustering is a hierarchical clustering technique that takes a bottom-up approach. It starts by treating each data point as a single cluster and then iteratively merges the closest pairs of clusters until all data points belong to a single cluster. A dendrogram is used to visualize the hierarchy of clusters, and the optimal number of clusters is determined by cutting the dendrogram at an appropriate level (Murtagh & Legendre, 2014).

Advantages:

- Does not require the number of clusters to be predefined.
- Produces a hierarchy of clusters, which can be useful for interpretation.

Limitations:

- Computationally expensive for large datasets.
- Sensitive to noise and outliers.

Gaussian Mixture Model (GMM)

The Gaussian Mixture Model is a probabilistic clustering algorithm that assumes the data is generated from a mixture of several Gaussian distributions. Each cluster is represented by a Gaussian distribution, and the algorithm assigns probabilities to each data point for belonging to a particular cluster. GMM is more flexible than K-means as it can model clusters with different shapes and sizes (Reynolds, 2009).

Advantages:

- Can model clusters of different shapes and sizes.
- Provides probabilistic cluster assignments.

Limitations:

- Computationally intensive.
- Requires careful initialization to avoid convergence to local optima.

3.4 Data Modeling

To model the data, three clustering algorithms were applied: K-means, Agglomerative, and Gaussian Mixture Model (GMM). Each algorithm was tested with three different numbers of clusters: 4, 5, and 8. The modeling process involved the following steps:

1. Algorithm Initialization: Each clustering algorithm was initialized with the specified number of clusters.
2. Model Fitting: The dataset was fitted to each model, and the clusters were predicted.
3. Cluster Assignment: The predicted clusters were added to the dataset for visualization and comparison purposes.

The following Python libraries were used for implementation:

- Scikit-learn: For implementing K-means, Agglomerative, and GMM algorithms.
- Yellowbrick: For visualizing clustering performance using distortion, silhouette, and calinski_harabasz scores.
- Matplotlib and Seaborn: For plotting cluster visualizations.

3.4.1 Cluster Visualization

The clusters generated by each algorithm were visualized using scatter plots, where the x and y axes represent the principal components obtained from PCA. The following figures show the cluster distributions for each algorithm:

Figure 3.3: K-means clustering with 4, 5, and 8 clusters.

```python
# K-means clustering visualization
plt.scatter(X_pca[:, 0], X_pca[:, 1], c=kmeans_labels, cmap='viridis')
plt.title('K-means Clustering (k=4)')
plt.xlabel('Principal Component 1')
plt.ylabel('Principal Component 2')
plt.show()
```

Figure 3.4: Agglomerative clustering with 4, 5, and 8 clusters.

```python
# Agglomerative clustering visualization
plt.scatter(X_pca[:, 0], X_pca[:, 1], c=agg_labels, cmap='viridis')
plt.title('Agglomerative Clustering (k=5)')
plt.xlabel('Principal Component 1')
plt.ylabel('Principal Component 2')
plt.show()
```

Figure 3.5: Gaussian Mixture Model clustering with 4, 5, and 8 clusters.

```python
# GMM clustering visualization
plt.scatter(X_pca[:, 0], X_pca[:, 1], c=gmm_labels, cmap='viridis')
plt.title('Gaussian Mixture Model Clustering (k=8)')
plt.xlabel('Principal Component 1')
plt.ylabel('Principal Component 2')
plt.show()
```

3.5 Model Evaluation

The performance of the clustering algorithms was evaluated using three scoring methods: Davies-Bouldin Index (DBI), Calinski-Harabasz Index (CHI), and Silhouette Score (SS). These metrics were chosen because they provide insights into both the separation between clusters and the cohesion within clusters.

3.5.1 Davies-Bouldin Index (DBI)

The DBI measures the average similarity between each cluster and its most similar cluster. A lower DBI indicates better clustering performance, as it reflects greater separation between clusters and tighter cohesion within clusters (Davies & Bouldin, 1979).

3.5.2 Calinski-Harabasz Index (CHI)

The CHI measures the ratio of between-cluster dispersion to within-cluster dispersion. A higher CHI indicates better clustering performance, as it reflects greater separation between clusters and tighter cohesion within clusters (Calinski & Harabasz, 1974).

3.5.3 Silhouette Score (SS)

The SS measures how similar a data point is to its own cluster compared to other clusters. The score ranges from -1 to 1, where a higher score indicates better clustering performance (Rousseeuw, 1987).

3.5.6 Evaluation Results

The following table summarizes the performance of the clustering algorithms based on the three scoring methods:

Table 3.1: Clustering Algorithm Performance Comparison

Algorithm	**Number of Clusters**	**DBI**	**CHI**	**SS**
K-means	4	0.93	284.0	0.40
K-means	5	0.96	253.4	0.35

K-means	8	0.89	253.6	0.39
Agglomerative	4	0.94	269.3	0.39
Agglomerative	5	0.94	239.7	0.36
Agglomerative	8	0.94	229.4	0.35
Gaussian Mixture	4	1.01	254.4	0.37
Gaussian Mixture	5	1.13	233.8	0.36
Gaussian Mixture	8	1.23	162.9	0.26

3.5.7 Key Findings:

- K-means consistently outperformed the other algorithms across all scoring metrics, particularly for k=4 and k=8.
- Agglomerative clustering showed comparable performance to K-means but was slightly less effective for higher numbers of clusters.
- Gaussian Mixture Model performed the worst, with higher DBI and lower CHI and SS scores, indicating overlapping cluster boundaries and poor separation.

3.5.8 Visual Comparison of Clustering Performance

The following figures provide a visual comparison of the clustering performance for k=4, k=5, and k=8:

Figure 3.6: DBI Comparison Across Algorithms

```python
# DBI comparison visualization
sns.barplot(x=['K-means', 'Agglomerative', 'GMM'], y=[0.93, 0.94, 1.01])
plt.title('DBI Comparison (k=4)')
plt.ylabel('Davies-Bouldin Index')
plt.show()
```

Figure 3.7: CHI Comparison Across Algorithms

```python
# CHI comparison visualization
sns.barplot(x=['K-means', 'Agglomerative', 'GMM'], y=[284.0, 269.3, 254.4])
plt.title('CHI Comparison (k=4)')
plt.ylabel('Calinski-Harabasz Index')
plt.show()
```

Figure 3.8: SS Comparison Across Algorithms

```python
# SS comparison visualization
sns.barplot(x=['K-means', 'Agglomerative', 'GMM'], y=[0.40, 0.39, 0.37])
plt.title('Silhouette Score Comparison (k=4)')
plt.ylabel('Silhouette Score')
plt.show()
```

3.6 Conclusion

The data analysis and modeling phases of this research demonstrated that K-means clustering is the most effective algorithm for categorizing the dataset into meaningful clusters. The evaluation metrics (DBI, CHI, and SS) consistently favored K-means, particularly for k=4 and k=8. While Agglomerative clustering showed comparable performance, it was slightly less effective for higher numbers of clusters. The Gaussian Mixture Model performed poorly, with overlapping cluster boundaries and lower scores across all metrics.

These findings highlight the importance of selecting the right clustering algorithm and the number of clusters for effective data categorization. The results of this analysis will be used to develop a recommendation system that suggests behavioral activation activities tailored to individual users, contributing to the broader goal of improving mental health through AI-driven interventions.

REFERENCES

[1] Olowolayemo, A., Tanni, M. G., Emon, I. A., Ahhmed, U., Mohd Dzahier, 'Arisya, Safin, M. R., & Nisha, N. Z. (2023). Conversational Analysis Agents for Depression Detection: A Systematic Review. Journal of Integrated and Advanced Engineering (JIAE), 3(1), 47–64. https://doi.org/10.51662/jiae.v3i1.85

[2] Pacheco-Lorenzo, M. R., V alladares-Rodríguez, S. M., Anido-Rifón, L. E., & Fernández-Iglesias, M. J. (2021). Smart conversational agents for the detection of neuropsychiatric disorders: A systematic review. Journal of Biomedical Informatics. https://doi.org/10.1016/j.jbi.2020.103632

[3] Otero-González, I., Pacheco-Lorenzo, M. R., Fernández-Iglesias, M. J., & Anido-Rifón, L. E. (2024). Conversational agents for depression screening: A systematic review. International Journal of Medical Informatics. https://doi.org/10.1016/j.ijmedinf.2023.105272

[4] Martínez-Miranda, J., Martínez, A., Ramos, R., Aguilar, H., Jiménez, L., Arias, H., … V alencia, E. (2019). Assessment of users' acceptability of a mobile-based embodied conversational agent for the prevention and detection of suicida https://pubmed.ncbi.nlm.nih.gov/31240494/

[5] Chen, Y., Xu, C., Liang, C., Tao, Y., & Shi, C. (2024). Speech-based Clinical Depression Screening: An Empirical Study. *arXiv preprint arXiv:2406.03510*. https://arxiv.org/abs/2406.03510

[6] Qin, W., Chen, Z., Wang, L., Lan, Y., Ren, W., & Hong, R. (2023). Read, Diagnose and Chat: Towards Explainable and Interactive LLMs-Augmented Depression Detection in Social Media. *arXiv preprint arXiv:2305.05138*. https://arxiv.org/abs/2305.05138

[7] Menne, F., Dörr, F., Schräder, J., Tröger, J., Habel, U., König, A., & Wagels, L. (2024). The voice of depression: speech features as biomarkers for major depressive disorder. *BMC Psychiatry*, 24(794). https://doi.org/10.1186/s12888-024-06253-6

[8] Zhou, H., Zhang, Y., & Wang, S. (2024). Improving speech depression detection using transfer learning with pre-trained models. *Scientific Reports*, 14(60278). https://doi.org/10.1038/s41598-024-60278-1

[9] Ji, Z., Wang, Y., & Zhang, H. (2024). WundtGPT: Shaping Large Language Models To Be An Empathetic, Proactive Mental Health Companion. *arXiv preprint arXiv:2406.15474*. https://arxiv.org/abs/2406.15474

[10] Park, S. S., & Choi, Y. S. (2024). Spiking neural networks for physiological and speech signals: a review. *Biomedical Engineering Letters*, 14, 943–954. https://doi.org/10.1007/s13534-024-00404-0

[11] Qin, W., Chen, Z., Wang, L., Lan, Y., Ren, W., & Hong, R. (2023). Read, Diagnose and Chat: Towards Explainable and Interactive LLMs-Augmented Depression Detection in Social Media. *arXiv preprint arXiv:2305.05138*. https://arxiv.org/abs/2305.05138

[12] Chen, Y., Xu, C., Liang, C., Tao, Y., & Shi, C. (2024). Speech-based Clinical Depression Screening: An Empirical Study. *arXiv preprint arXiv:2406.03510*. https://arxiv.org/abs/2406.03510

[13] Qin, W., Chen, Z., Wang, L., Lan, Y., Ren, W., & Hong, R. (2023). Read, Diagnose and Chat: Towards Explainable and Interactive LLMs-Augmented Depression Detection in Social Media. *arXiv preprint arXiv:2305.05138*. https://arxiv.org/abs/2305.05138

[14] Menne, F., Dörr, F., Schräder, J., Tröger, J., Habel, U., König, A., & Wagels, L. (2024). The voice of depression: speech features as biomarkers for major depressive disorder. *BMC Psychiatry*, 24(794). https://doi.org/10.1186/s12888-024-06253-6

[15] Zhou, H., Zhang, Y., & Wang, S. (2024). Improving speech depression detection using transfer learning with pre-trained models. *Scientific Reports*, 14(60278). https://doi.org/10.1038/s41598-024-60278-1

[16] Tao, Y., Xu, C., Liang, C., & Chen, Y. (2022). *Emotional audio-textual depression corpus (EATD-Corpus): Multimodal approaches.* arXiv. https://arxiv.org/abs/2202.08210

[17] Gratch, J., Artstein, R., Lucas, G., Stratou, G., Scherer, S., Nazarian, A., & Morency, L. P. (2014). *The Distress Analysis Interview Corpus of human and computer interviews.* In Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC). Springer.

[18] Bucuram, M. (2023). *Depression datasets for linguistic analysis.* GitHub Repository. https://github.com/bucuram/depression-datasets-nlp