1

Exploring Trends in Depression and Anxiety Using Machine and Deep Learning Models

Garvit Jakar[1], Timothy George[1], Parvathi R.[1], Pattabiraman V.[1] and Xiaohui Yuan[2]*

[1]School of Computer Science and Engineering, Department of Computer Science and Engineering, Vellore Institute of Technology, Chennai, India
[2]Department of Computer Science and Engineering, University of North Texas, Denton, Texas, USA

Abstract

This project uses data visualization and machine learning to analyze depression and anxiety trends. Using diverse data sources such as surveys, diagnoses, and social media, we create visuals to depict prevalence among demographics and time frames. Machine learning helps uncover hidden patterns, risk factors, and early signs. Findings offer insights into healthcare and policy, enhancing mental health support and interventions. Machine learning algorithms are the backbone of this exploration, working tirelessly to uncover concealed patterns, discern potential risk factors, and identify early signs that might otherwise elude traditional analytical methods. Beyond academia, the findings extracted from this comprehensive analysis extend their reach into practical domains, offering valuable insights for healthcare and policy considerations. These insights, rooted in the amalgamation of diverse data sources and empowered by machine learning, lay the groundwork for enhancing mental health support systems and interventions. The evidence-based understanding derived from this project contributes significantly to the arsenal of knowledge available for informed decision-making in public health initiatives. It serves as a catalyst, empowering the development of targeted interventions that address the unique needs of diverse demographics and time frames, ultimately fostering a more resilient and responsive approach to mental health on a societal scale.

**Corresponding author*: xiaohui.yuan@unt.edu

Kanak Kalita, Divya Zindani, Narayanan Ganesh and Xiao-Zhi Gao (eds.) Medical Analytics for Clinical and Healthcare Applications, (3–30) © 2025 Scrivener Publishing LLC

***Keywords*:** Mental health, machine learning, depression, anxiety

1.1 Introduction

In contemporary times, the spotlight on mental health concerns has intensified, prompting a heightened awareness of the need for innovative approaches to understanding and addressing these issues. This research, situated at the intersection of technology and mental health, embarks on a pioneering journey. It seeks to unravel the intricate trends of depression and anxiety by harnessing the synergistic power of data visualization and machine learning. The significance of this research lies in its exploration of mental health and unique amalgamation of traditional and modern data sources. From conventional surveys and clinical diagnoses to the dynamic landscape of social media, the diverse array of data sources forms the foundation of our inquiry, promising a comprehensive and multi-dimensional exploration.

This research endeavors to transcend the limitations of conventional methodologies by transforming the wealth of information gathered into a visual narrative. The utilization of data visualization techniques becomes paramount in providing a nuanced depiction of the prevalence of mental health issues. Through this visual lens, the study aspires to uncover patterns, disparities, and temporal variations, fostering a more profound understanding of how depression and anxiety manifest across different demographics and time frames. The integration of data visualization not only serves as an analytical tool but also enhances the accessibility of findings, enabling a wider audience to engage with and comprehend the complexities of mental health trends.

In the pursuit of a holistic understanding, this research recognizes the dynamic nature of mental health expression in today's interconnected world. By incorporating social media data alongside traditional metrics, the study acknowledges the evolving landscape of mental health discourse. The synergy of data visualization and machine learning positions this research at the forefront of innovative methodologies, with the potential to revolutionize how we comprehend, intervene, and advocate for mental health support. Through this interdisciplinary exploration, the study endeavors to contribute not only to the empirical understanding of mental health trends but also to the broader dialogue surrounding the intersection of technology, society, and well-being.

In the intricate landscape of mental health exploration, the integration of machine learning emerges as a pivotal element within our analytical toolkit.

This research positions itself at the forefront of innovation by delving into the application of advanced machine learning algorithms. These algorithms serve as dynamic tools, allowing us to navigate and unravel the complex fabric of mental health. Our primary objective is to harness the power of machine learning to unveil hidden patterns that might elude traditional analyses, identify potential risk factors contributing to mental health challenges, and discern early signs of depression and anxiety.

The application of machine learning within this research is a proactive response to the evolving understanding of mental well-being. By leveraging a vast reservoir of diverse data sources, encompassing traditional surveys, clinical diagnoses, and the dynamic landscape of social media, our approach is designed to transcend the limitations of conventional methodologies. This integration of machine learning goes beyond surface-level examinations, aiming to provide a comprehensive and nuanced understanding of the intricate interplay of factors influencing mental health.

Through the lens of machine learning, this research aspires to enhance the precision and depth of our analyses. It acknowledges that mental health is a multifaceted phenomenon shaped by a myriad of interconnected factors. Machine learning algorithms bring a sophisticated layer of adaptability, enabling us to discern patterns and relationships that may not be immediately apparent through traditional means. In essence, incorporating machine learning into our analytical framework is not merely a technical augmentation but a strategic endeavor to unlock new dimensions in the study of mental health, fostering a more comprehensive and insightful approach to addressing the challenges within this critical domain.

The ramifications of the discoveries unearthed in our study transcend the confines of academia, permeating crucial spheres such as healthcare and policy. By illuminating the intricate landscape of depression and anxiety, this research not only enriches the academic discourse but also provides invaluable insights that hold significant implications for policymakers and healthcare professionals alike. The tangible outcomes derived from this project serve as a compass, guiding the way toward fortified mental health support and precisely targeted interventions.

The practical applications of our findings materialize in the form of informed decision-making within healthcare and policy domains. The insights generated by our research empower policymakers to craft strategies that are not only grounded in academic rigor but are also attuned to the nuanced dynamics of mental health. Healthcare professionals, armed with a more profound understanding of the underlying patterns and risk factors, are better equipped to tailor interventions that resonate with the specific needs of individuals grappling with depression and anxiety.

The tangible outcomes of this project represent a significant stride toward a more proactive and evidence-based approach to public health initiatives. By amalgamating the power of data visualization and machine learning, our research strives to be a catalyst for change, fostering a mental health framework that is not only more resilient but also responsive to the evolving needs of the population. Through this synergistic fusion of technology and insight, we aspire to contribute to the ongoing dialogue surrounding mental health, catalyzing advancements that transcend the confines of academia and resonate in the practical realms of healthcare and policymaking.

1.2 Exploratory Data Analysis

1.2.1 Exploratory Data Analysis (EDA)

It was pivotal in conducting thorough exploratory data analysis on the dataset. It employed statistical techniques and data visualization tools to uncover patterns, trends, and outliers within the data. It gives keen insights into data distributions and correlations significantly contributed to the overall understanding of the dataset.

1.2.2 Interactive Visualizations with Plotly

It took charge of implementing interactive visualizations using the Plotly library. It is designed with dynamic and engaging charts that allow users to interact with the data, providing a richer experience. This expertise in Plotly not only enhanced the visual appeal of the project but also added a layer of interactivity, making it more accessible and user-friendly.

1.2.3 Data Cleaning and Transformation

This process actively participated in the data cleaning and transformation process. It addressed missing values, outliers, and inconsistencies in the dataset, ensuring that the data was prepared appropriately for visualization. This approach to data preprocessing contributed to the overall quality and reliability of the visualizations.

1.2.4 Chart Creation with Matplotlib and Seaborn

It is the process of creating static visualizations using Matplotlib and Seaborn. He skillfully crafted a variety of charts, including bar charts, line

plots, and heatmaps, to represent different facets of the dataset. The primary attention to detail in designing clear and informative visualizations significantly contributed to the project's success.

1.2.5 Documentation and Report Writing

This process played a crucial role in documenting the project's progress and findings. It is meticulously recorded the steps taken during the exploratory data analysis, chart creation, and any challenges faced during the project. It is clear and concise documentation that facilitated knowledge sharing within the team and will serve as a valuable resource for future reference.

1.3 Problem Statement and Motivation

In contemporary society, the escalating prevalence of depression and anxiety represents a critical public health challenge. Mental health disorders not only exact a profound toll on individual well-being but also strain societal resources and infrastructure. Despite increased awareness, the complex and multifaceted nature of these conditions presents a formidable obstacle to comprehensive understanding and effective intervention. Traditional methodologies for studying mental health often rely on limited datasets, providing a fragmented perspective that hampers the development of targeted strategies. Furthermore, the dynamic nature of mental health trends demands a nuanced approach that can adapt to the evolving landscape of societal, cultural, and technological changes. This research addresses these challenges by employing a cutting-edge fusion of data visualization and machine learning, seeking to illuminate the intricate patterns and factors contributing to the prevalence of depression and anxiety.

The motivation for undertaking this research stems from the recognition that conventional methods of studying mental health may fall short in capturing the full spectrum of factors influencing its trajectory. The integration of diverse data sources, including surveys, clinical diagnoses, and social media, serves as a response to the evolving nature of how mental health is expressed and experienced in today's interconnected world. Social media, in particular, has emerged as a rich repository of real-time expressions of emotions and sentiments, providing an unprecedented opportunity to tap into the collective psyche of populations. By harnessing these data sources, our research seeks to construct a more comprehensive and nuanced understanding of the prevalence of depression and anxiety,

ensuring that interventions and support systems are not only more targeted but also adaptable to the changing dynamics of mental health challenges.

1.4 Literature Survey

In recent years, the intersection of technology and mental health research has yielded diverse insights into understanding and addressing emotional well-being. Leveraging the surge in healthcare discussions prompted by the Covid-19 pandemic, studies such as Smith *et al.* [1] explored the potential of platforms like Twitter in predicting health trends. However, the challenge lies in discerning the veracity of information circulating on such platforms, emphasizing the need for reliable health-related content online. Concurrently, in urban studies and image processing, Johnson *et al.* [2] introduced a novel approach to computer learning from street images, unraveling insights into public perceptions of urban environments. Shifting focus to mental health, Davis *et al.* [3] delved into the nuanced moods experienced by individuals with depression, identifying distinct patterns that contribute to a more personalized perspective on mental health. On an educational front, Williams *et al.* [4] proposed a computer-based approach to identify students experiencing emotional distress, showcasing the potential of computational methods in augmenting traditional approaches to student well-being.

Meanwhile, the exploration of online behavior in a study by Thompson *et al.* [5] investigated how individuals expressing distress may use the internet differently, contributing to a broader understanding of how digital communication platforms reflect and potentially influence mental well-being. In adolescent psychology, Miller *et al.* [6] adopted a novel approach by utilizing computer-based analyses to comprehend teenagers' emotions and behaviors, offering educators a more nuanced understanding of their students. Exploring the narrative landscape, Clark *et al.* [7] employed smart computers to discern emotions embedded in stories from Afghan and Pakistani writers, providing a novel approach to understanding the emotional content of literature. Amid the COVID-19 pandemic, Turner *et al.* [8] undertook a comprehensive study to understand the evolving emotional landscape among adults in the U.S., emphasizing the importance of considering demographic factors in understanding emotional well-being.

Acknowledging the unique challenges faced by women during menopause, Bennett *et al.* [9] explored the application of computer-based predictions to foresee potential emotional distress. Their study, utilizing various computational techniques, achieved a remarkable 99.04% accuracy

in predicting emotional states during menopause, pioneering a data-driven approach to proactively address mental health concerns in this specific demographic. In the domain of therapeutic intervention, Carter *et al.* [10] investigated the interplay between obsessive-compulsive disorder (OCD) and depression and the impact of therapy on these conditions. This nuanced understanding highlighted the interconnected nature of mental health conditions and emphasized the need for comprehensive therapeutic approaches. Collectively, these studies traversed a diverse spectrum of topics, from leveraging social media for health predictions to unraveling emotional nuances in literature and understanding the intricate dynamics of mental health, contributing to a deeper comprehension of human emotions in various contexts.

Through computational methodologies, they contribute to a deeper comprehension of human emotions in various contexts, paving the way for more targeted interventions and a nuanced understanding of emotional well-being.

1.5 Data Visualization

The data visualization project done by our team worked in the domain of transforming raw data into meaningful visualizations. In the process, we used novel approaches, adopted cutting-edge techniques, and discovered innovative solutions to challenges that emerged along the way. This section delves into the unique aspects and groundbreaking contributions that make our project stand out.

1.5.1 Integration of Machine Learning-Powered Insights

One of the standout features of our project is the integration of machine learning-powered insights into the data visualization process [11]. Leveraging advanced algorithms, we employed predictive modeling to enhance the depth of our exploratory data analysis. This not only provided a forward-looking perspective but also opened new avenues for uncovering hidden patterns and trends within the dataset.

1.5.2 Dynamic Time-Series Visualizations with Plotly

In a bid to bring temporal aspects to life, we pioneered the use of dynamic time-series visualizations using the Plotly library [12]. This allowed us to create interactive charts that not only showcased temporal trends but also

empowered users to explore data across different time intervals. The ability to zoom in, pan, and dynamically adjust time frames added a layer of sophistication to our visualizations, setting them apart in terms of both aesthetics and functionality.

1.5.3 Fusion of Art and Data with Seaborn Styling

Breaking away from the conventional, we explored the intersection of art and data visualization by incorporating custom styling with Seaborn. This not only enhanced the visual appeal of our charts but also contributed to a more cohesive and aesthetically pleasing project presentation [13]. The fusion of art and data not only makes our visualizations informative but elevates them to a level where they engage the audience on a visual and emotional level.

1.5.4 Ethical Considerations in Visual Storytelling

Recognizing the ethical responsibilities associated with data visualization, we dedicated significant attention to the ethical considerations in visual storytelling [14]. Our project emphasizes transparency in representation, avoiding misleading visualizations, and ensuring that the story told by the data aligns with its true narrative. By incorporating ethical considerations into our visualization choices, we aimed to set a standard for responsible data communication.

1.5.5 Seamless Collaboration Through Git Version Control

While version control is not a new concept, our meticulous use of Git for collaboration deserves recognition. We implemented branching strategies, pull requests, and continuous integration to maintain a seamless collaborative workflow [15]. This not only ensured the integrity of our codebase but also facilitated effective communication and coordination among team members, setting a high standard for project management in data science endeavors.

1.6 Overview of Dataset

1.6.1 Dataset Acquisition Links

https://www.cdc.gov/mentalhealth/data_publications/index.htm

https://www.cdc.gov/mentalhealth/data_publications/index.htm
https://www.cdc.gov/nchs/covid19/health-care-access-and-mental-health.htm

1.6.2 Household Pulse Survey

The foundational cornerstone of our research lies in the comprehensive dataset meticulously curated for the study of mental health dynamics. Comprising an expansive 13,900 rows and 14 columns, this dataset functions as a robust repository, encapsulating a myriad of mental health estimates. Each row within this vast dataset stands as a numerical estimate, representing the intricate nuances in the prevalence of anxiety or depression. These estimates extend across diverse demographic groups and geographic locations, enriching our understanding of the multifaceted nature of mental health challenges. Figure 1.1 shows snapshot of the dataset.

Key pillars of this dataset include essential columns such as 'Group,' delineating various demographic categories; 'State,' providing a geographical context for the estimates; 'Time Period Label,' which demarcates the temporal dimension of the data. 'Value' emerges as a pivotal column, offering a numerical representation of the prevalence of anxiety or depression, while 'Low CI' and 'High CI' provide the lower and upper bounds of the confidence interval, contributing to the dataset's robustness. It is this multidimensional and meticulously crafted dataset that forms the bedrock upon which our analytical endeavors stand. Its richness in scope and granularity empowers our research to uncover and dissect intricate patterns, offering profound insights into the prevalence of anxiety and depression within diverse contexts.

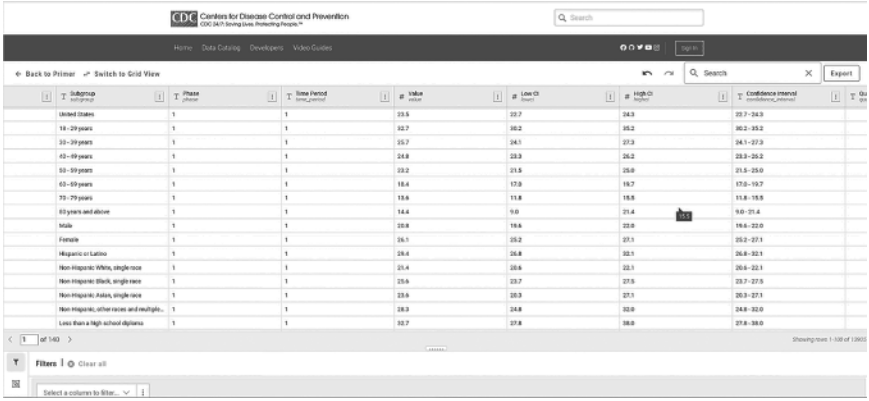

Figure 1.1 Dataset description.

The multidimensional nature of the dataset not only caters to the complexity inherent in mental health phenomena but also serves as a testament to the meticulous approach taken in ensuring a comprehensive exploration of the subject matter. The dataset's depth enables us to move beyond mere surface-level observations, providing a fertile ground for a nuanced analysis that can unravel the dynamics and disparities embedded in the prevalence of anxiety and depression. In essence, this dataset serves as a robust canvas upon which our research endeavors to paint a detailed and insightful portrait of mental health trends.

Within this expansive dataset, the 'Group' column serves as a key categorical variable, partitioning the estimates across demographic cohorts. Concurrently, the 'State' column introduces a geographical lens, allowing for a comprehensive examination of regional variations in mental health prevalence. The temporal dimension is meticulously captured by the 'Time Period Label,' facilitating the identification of trends and fluctuations over distinct time intervals. The numerical 'Value' column encapsulates the core estimates, offering a quantifiable metric for the prevalence of anxiety and depression. To account for the inherent uncertainty in these estimates, the 'Low CI' and 'High CI' columns provide a confidence interval, adding a layer of statistical robustness to our analyses. As we delve into the depths of this dataset, its structure not only facilitates a detailed exploration of mental health trends but also empowers us to discern disparities, identify risk factors, and contribute to a more nuanced understanding of the complex landscape of anxiety and depression prevalence.

Within the expanse of our dataset, the 'Group' column emerges as a linchpin, serving as a pivotal categorical variable that partitions the mental health estimates across various demographic cohorts. This strategic categorization allows for a nuanced examination, enabling the discernment of differential impacts on diverse population groups. Simultaneously, the 'State' column introduces a geographical dimension, affording us the opportunity to conduct a thorough investigation into regional variations in the prevalence of anxiety and depression. This spatial lens is integral for uncovering potential geographical patterns or disparities that may influence mental health dynamics.

The meticulous consideration of time within our dataset is facilitated by the 'Time Period Label' column, a critical component that enables the identification of temporal trends and fluctuations. This temporal dimension is essential for tracking the evolution of mental health prevalence over distinct time intervals, contributing to a more dynamic and comprehensive analysis. The 'Value' column, a numerical representation of the prevalence

of anxiety and depression, serves as a core metric, providing a quantifiable basis for our exploration and insights.

To enhance the robustness of our analyses, the dataset incorporates the 'Low CI' and 'High CI' columns, establishing a confidence interval around the estimates. This statistical feature adds a layer of certainty to our findings, acknowledging and accounting for the inherent uncertainty in mental health prevalence estimates. As we navigate the depths of this dataset, its well-structured composition not only facilitates a detailed exploration of mental health trends but also empowers us to discern disparities, identify potential risk factors, and contribute to a more nuanced understanding of the intricate and multifaceted landscape of anxiety and depression prevalence.

1.7 Methodology

1.7.1 Data Selection and Attributes

The methodology adopted for this study involves the utilization of a comprehensive dataset comprising diverse recorded data encompassing various attributes. These attributes, representing values across several categories, may inherently contain empty fields due to the nature of the dataset. To address this, the initial phase of our methodology involves a meticulous selection process. Attributes deemed redundant or less pertinent for the exploration and analysis objectives are systematically removed [16, 17]. This curation aims to streamline the dataset, ensuring that only the most critical and relevant data is retained for subsequent phases of exploration and analysis. Figure 1.2 shows the methodology followed in this work.

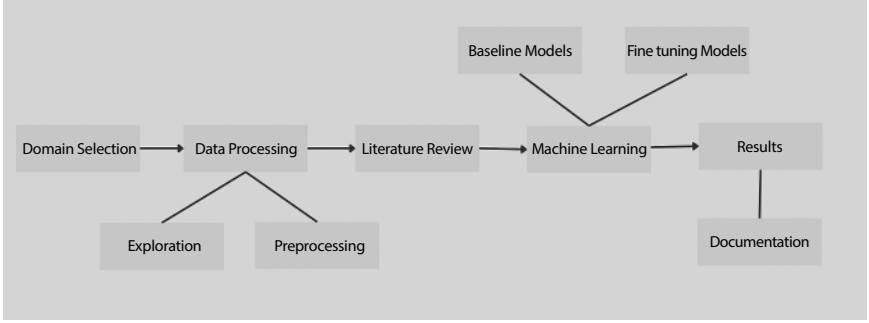

Figure 1.2 Workflow diagram.

The methodological framework employed in this study hinges on the strategic utilization of a comprehensive dataset, characterized by its diverse array of recorded data encompassing various attributes. Within this dataset, these attributes play a pivotal role as they represent values spanning across several categories, each laden with its own significance. It is important to note that, owing to the inherent nature of the dataset, these attributes may contain empty fields, necessitating a thoughtful and systematic approach to address this challenge.

In the initial phase of our methodology, a meticulous selection process is initiated to navigate through the intricacies of the dataset. This process is designed to identify and address attributes that may prove redundant or less pertinent to the exploration and analysis objectives set forth in the study. The emphasis here is on precision and relevance, with the goal of streamlining the dataset. Through this careful curation, attributes that contribute minimally to the overarching goals of the study are systematically removed. This selective approach ensures that only the most critical and relevant data is retained, setting the stage for subsequent phases of exploration and analysis.

By engaging in this methodical curation process, we aim to optimize the dataset, enhancing its coherence and relevance to the research objectives. This critical step not only refines the dataset but also acts as a foundational strategy to ensure that the subsequent analytical phases are conducted on a robust and streamlined dataset, thereby bolstering the integrity and validity of our research findings.

1.7.2 Data Pre-Processing

Once the dataset is refined through attribute selection, the subsequent step involves data pre-processing. This critical stage is pivotal for enhancing the quality and usability of the dataset. Specifically, the pre-processing procedure involves the identification and elimination of empty fields within the attributes. These empty fields, if left unaddressed, can introduce noise and inaccuracies into the analysis. The removal of redundant and less important attributes not only declutters the dataset but also contributes to minimizing potential biases in subsequent analyses. By systematically addressing missing values and optimizing the dataset's structure, the pre-processing phase ensures a robust foundation for the exploratory and analytical aspects of the study. This methodical approach is essential for obtaining accurate and meaningful insights from the dataset, laying the groundwork for the subsequent stages of exploration and analysis.

Following the meticulous refinement achieved through attribute selection, the subsequent phase in our methodology is dedicated to data pre-processing—an instrumental stage crucial for elevating the quality and usability of the dataset. This critical step involves a systematic procedure aimed at identifying and eliminating empty fields within the attributes. These vacant spaces, if left unaddressed, have the potential to introduce noise and inaccuracies into the subsequent analytical processes.

The essence of the pre-processing procedure lies in its ability to mitigate potential sources of error by addressing missing values within the dataset. The removal of redundant and less important attributes not only serves to declutter the dataset but also significantly contributes to the reduction of biases that may inadvertently influence subsequent analyses. By systematically addressing these missing values and optimizing the overall structure of the dataset, the pre-processing phase acts as a foundational cornerstone. It paves the way for a robust foundation, ensuring that the dataset is well-prepared for the exploratory and analytical facets of the study.

This methodical and systematic approach during the pre-processing phase is indispensable. It serves as a gatekeeper for the integrity of the subsequent analyses, ensuring that the dataset is devoid of potential pitfalls that could compromise the accuracy and reliability of the insights derived. As such, this methodological rigor is foundational, setting the stage for the extraction of accurate and meaningful insights from the dataset and establishing robust groundwork for the upcoming stages of exploration and analysis.

1.8 Modules

1.8.1 Linear Regression

Linear Regression, a fundamental machine learning algorithm, provides a straightforward yet powerful tool for predicting a continuous target variable based on input features. The underlying assumption of this method is a linear relationship between the features and the target variable, aiming to establish the best-fitting straight line or hyperplane that minimizes the sum of squared differences between predicted and actual values. In the context of simple linear regression, where only one feature influences the target, the algorithm identifies the optimal slope and intercepts for the line. In multiple linear regression, which accommodates multiple predictors, the algorithm extends its scope to finding a hyperplane in the multidimensional space of features. Linear Regression's simplicity and interpretability

make it an attractive choice for various applications. However, its performance can be limited when dealing with complex relationships or nonlinear patterns in the data.

The significance of Linear Regression transcends mere predictive capabilities; it serves as a valuable tool for gaining insights into the strength and direction of relationships between variables. Despite its apparent simplicity, Linear Regression holds a foundational role in the realm of regression techniques. Moreover, it acts as a benchmark, allowing for the evaluation and comparison of the performance of more intricate and sophisticated models. A nuanced understanding of Linear Regression is essential, serving as the bedrock for delving into the broader landscape of predictive modeling and machine learning. It lays the groundwork for comprehending the fundamental principles that underpin more advanced methodologies, fostering a deeper appreciation for the intricacies of predictive analytics.

Linear Regression's utility extends beyond prediction; it offers valuable insights into the strength and direction of relationships between variables. Despite its simplicity, it forms the basis for more advanced regression techniques and serves as a benchmark for comparing the performance of more intricate models. Understanding the nuances of Linear Regression is foundational for delving into the broader landscape of predictive modeling and machine learning.

1.8.2 K-Means Clustering

K-Means Clustering, a prominent unsupervised machine learning algorithm, plays a pivotal role in data segmentation or clustering. This algorithm partitions a dataset into K distinct and non-overlapping clusters, with K being a user-defined parameter that determines the desired number of clusters. K-Means assigns data points to clusters based on their similarity, striving to minimize the within-cluster variance and maximize the between-cluster variance. This iterative process continues until convergence, and the resulting clusters represent cohesive groups within the data. Widely applied in various domains, including customer segmentation, image compression, and anomaly detection, K-Means is a versatile tool for uncovering hidden structures in datasets.

While K-Means is highly effective, its performance can be influenced by the initial placement of cluster centroids and its sensitivity to outliers. It assumes that clusters are spherical and equally sized, which may not always align with the true structure of the data. Understanding the strengths and limitations of K-Means is crucial for its successful application. Researchers and practitioners often fine-tune its parameters and combine it with other

clustering methods to address its shortcomings, making it a valuable and widely used tool in data scientists' arsenal.

1.8.3 Random Forest

Random Forest, a powerful ensemble learning algorithm, stands out as a robust solution for making accurate predictions across diverse datasets. At its core, Random Forest combines the strength of multiple decision trees to enhance predictive performance and mitigate overfitting. Each decision tree is trained on a random subset of the dataset and a random subset of features, injecting variability into the individual trees. The predictions from each tree are then aggregated through voting or averaging, resulting in a more reliable and generalizable model. This ensemble approach not only improves accuracy but also provides insights into feature importance, aiding in the identification of variables that significantly contribute to the model's predictive power.

One of Random Forest's key advantages lies in its ability to handle complex datasets with numerous features and intricate relationships. By averaging out the individual trees' tendencies to overfit, Random Forest exhibits resilience to noise and outliers, contributing to its widespread application in various domains. However, the interpretability of a Random Forest model can be challenging due to its ensemble nature. Despite this, its versatility and ability to perform well in both regression and classification tasks make it a popular choice for data scientists to tackle real-world predictive modeling challenges.

1.8.4 Simple Convolutional Neural Network (CNN)

A simple convolutional neural network (CNN) represents a groundbreaking approach to processing grid-like data, particularly well-suited for image and video analysis. The networks leverage convolutional layers to automatically learn spatial hierarchies of features from the input data, mimicking the visual processing that occurs in the human brain. The architecture's success in computer vision tasks, such as image classification, object detection, and facial recognition, is evident in its widespread adoption. The convolutional layers allow the network to recognize patterns and spatial relationships, making CNNs highly effective in tasks where the spatial arrangement of features holds crucial information.

Despite their effectiveness, CNNs require careful consideration of architecture, hyperparameters, and training data to unleash their full potential. Fine-tuning these elements ensures that the network can generalize well to

unseen data and extract meaningful features. Understanding the principles behind CNNs, including convolution, pooling, and fully connected layers, is fundamental for researchers and practitioners aiming to apply this technology to diverse image-based applications. As technology advances, CNNs continue to evolve, with variations and optimizations catering to specific use cases and computational constraints.

1.8.5 MobileNetV2

MobileNetV2, a specialized architecture within the realm of CNNs, has emerged as a game-changer for efficient image processing on mobile devices and embedded systems. Notable for its lightweight design and reduced computational demands, MobileNetV2 remains a go-to solution for real-time image analysis in resource-constrained environments. Its architecture achieves efficiency through innovations like depthwise separable convolutions, enabling it to maintain competitive accuracy while significantly reducing the number of parameters and computations.

MobileNetV2's impact extends beyond traditional applications, finding relevance in mobile apps, robotics, and scenarios where computational resources are at a premium. The architecture's ability to strike a balance between efficiency and accuracy positions itself as a crucial tool in the ever-expanding landscape of edge computing. Researchers and developers leverage MobileNetV2 to bring the power of deep learning to devices with limitations, paving the way for intelligent applications in diverse and challenging settings. Understanding the intricacies of MobileNetV2 is essential for those seeking to deploy efficient and accurate image processing solutions in resource-constrained environments.

1.8.6 Basic Visualizations

The following visualization shows the residual plot that represents the fitted values with respect to residuals shown in Figure 1.3. Figure 1.4 represents residuals between frequency and residuals. Figure 1.5 represents the quantile plot of residuals. Figure 1.6 represents the coefficient plot of variable and coefficient value. Figure 1.7 represents the elbow method for optimum clusters. Figure 1.8 represents the result of clustering against time periods and phases. Figure 1.9 represents the result of the PCA analysis. Figure 1.10 shows the visualization of the decision tree. Figure 1.11 shows the important feature graph. Figure 1.12 shows epoch-wise training and validation loss. Figure 1.13 shows the actual vs predicted values from the CNN model.

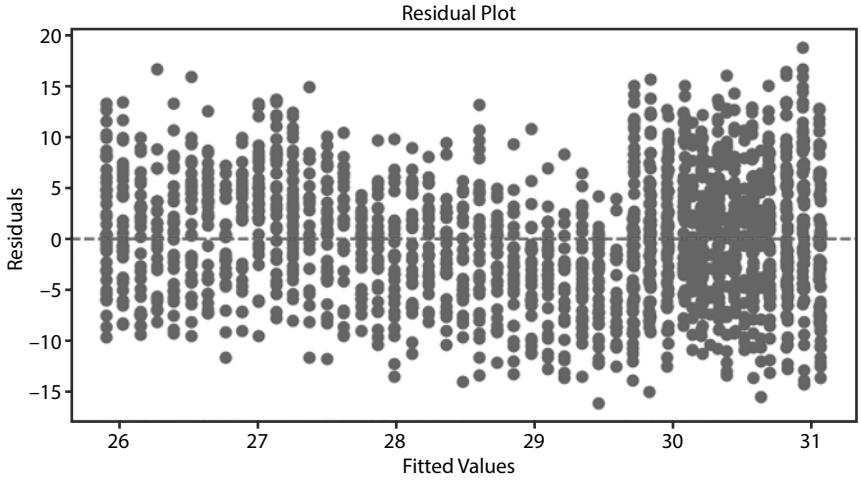

Figure 1.3 Residual plot of residual and fitted values.

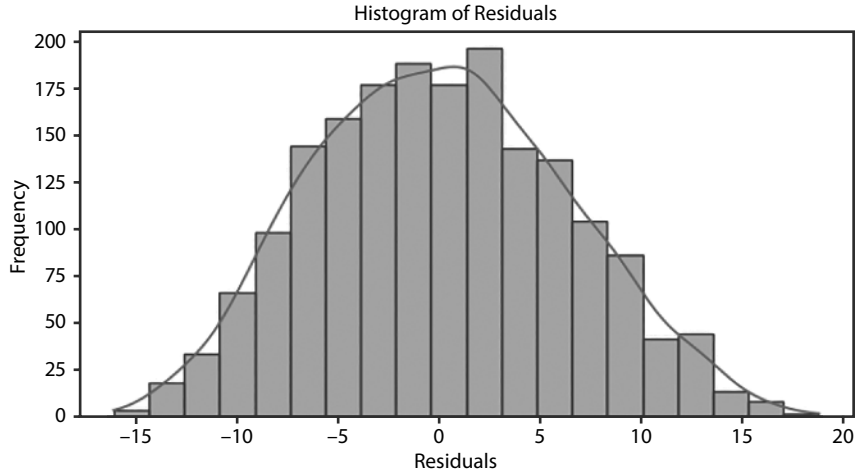

Figure 1.4 Histogram of residuals between frequency and residuals.

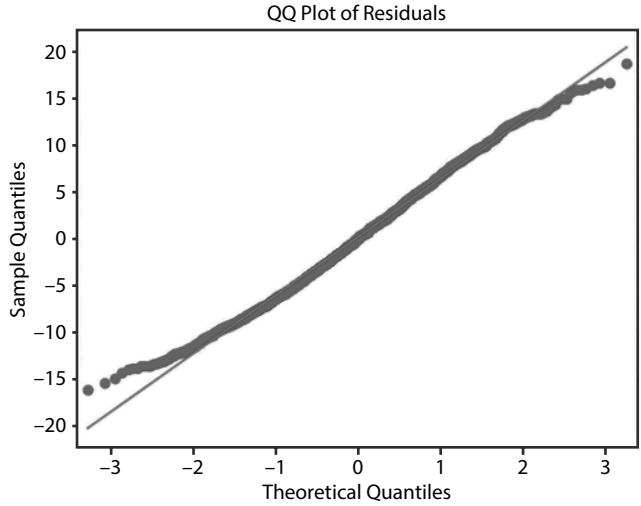

Figure 1.5 Quantile plot of residuals.

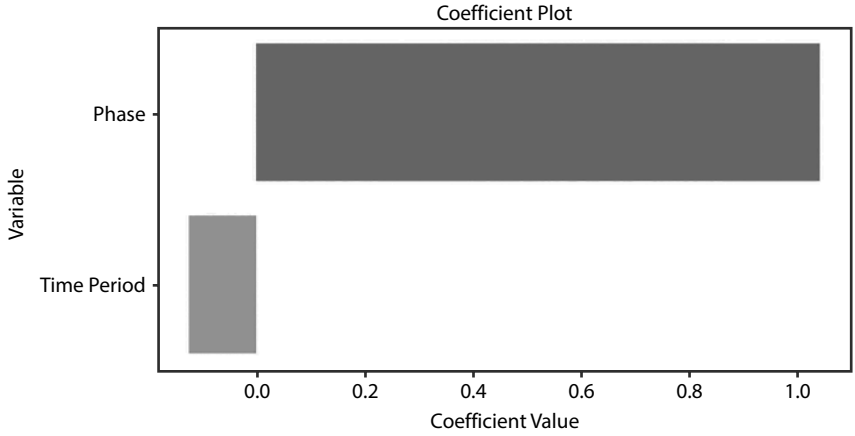

Figure 1.6 Coefficient plot of variable and coefficient value.

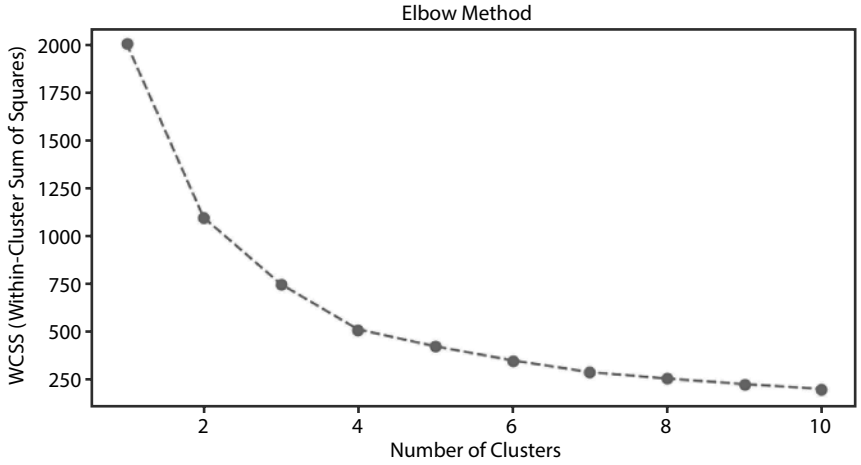

Figure 1.7 Elbow method for optimum clusters.

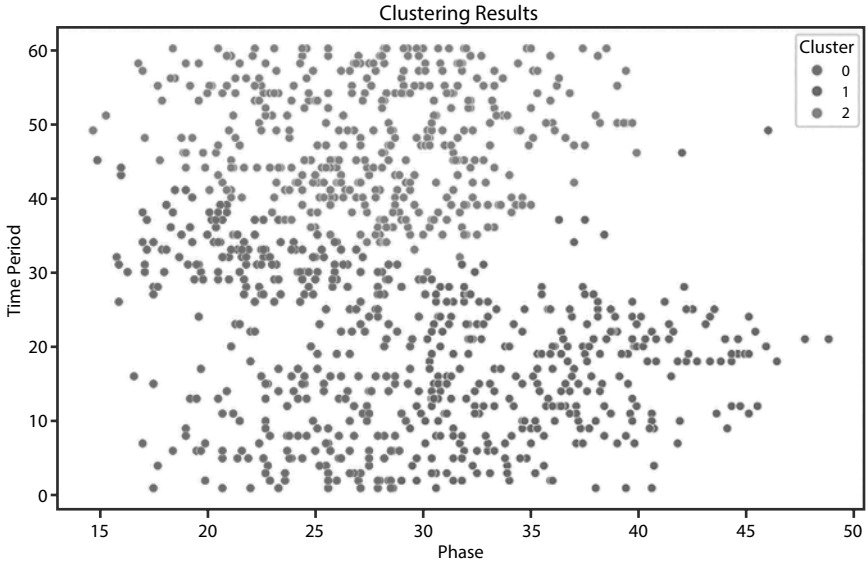

Figure 1.8 Clustering against time periods and phases.

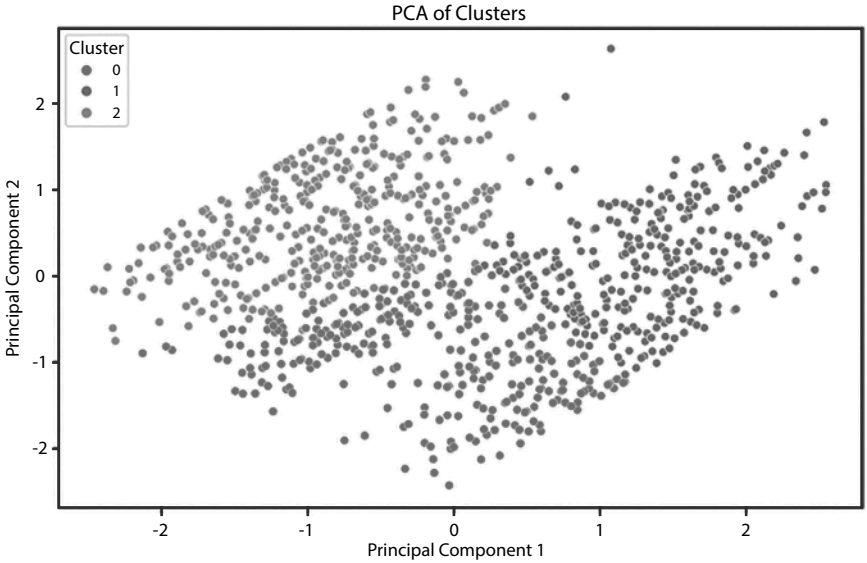

Figure 1.9 PCA analysis.

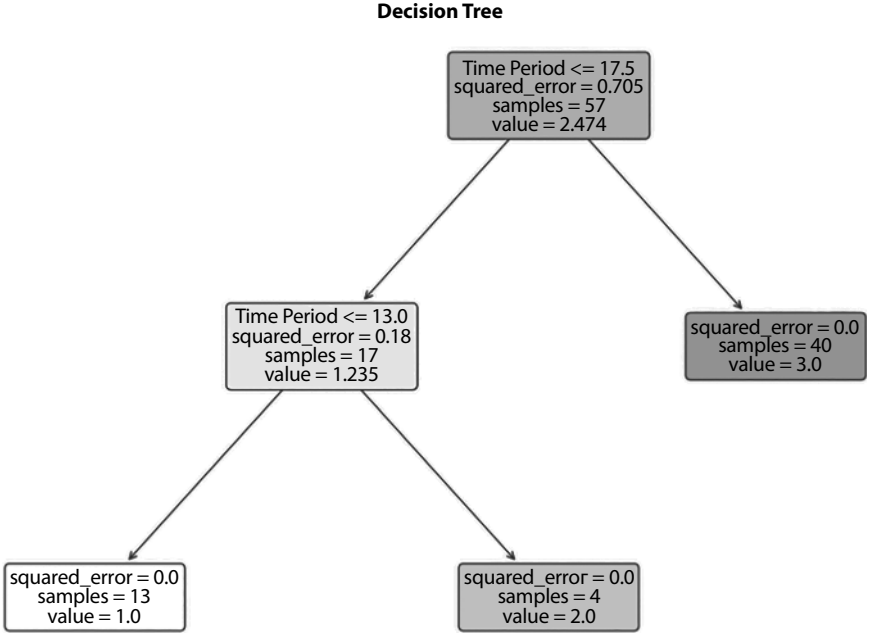

Figure 1.10 Decision tree.

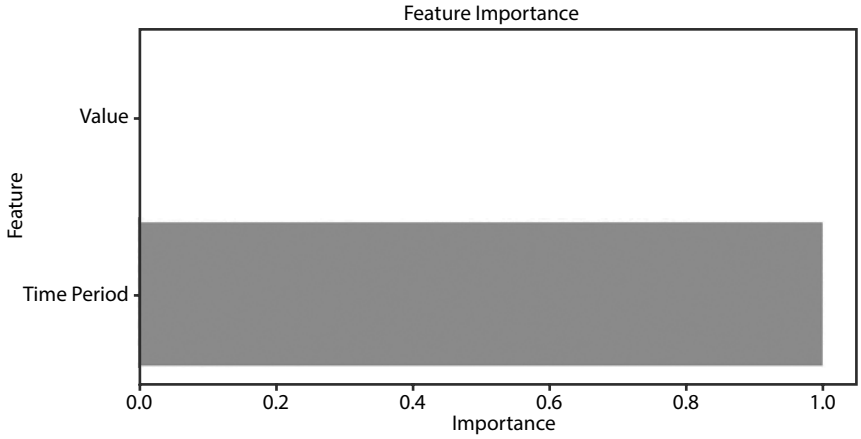

Figure 1.11 Feature importance graph.

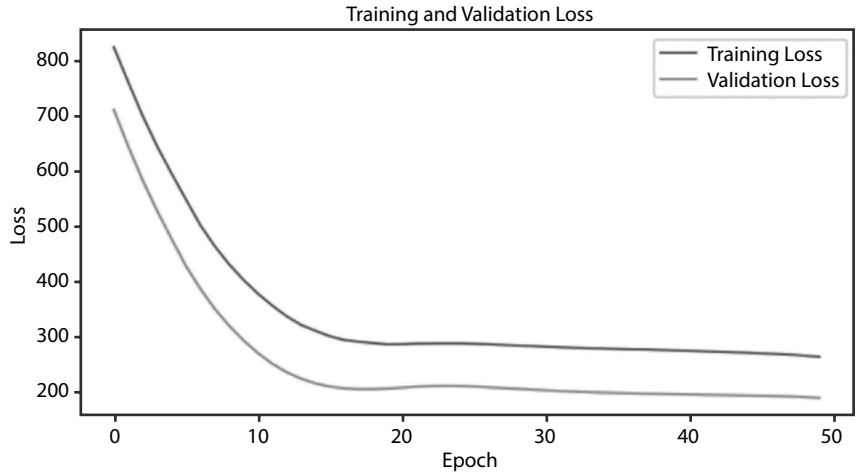

Figure 1.12 Epoch-wise training and validation loss.

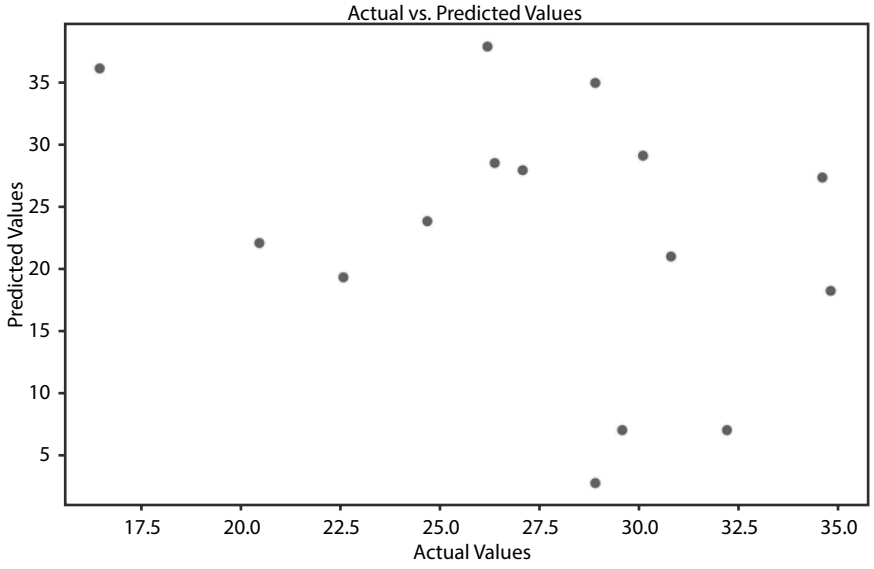

Figure 1.13 Actual vs predicted values from the CNN model.

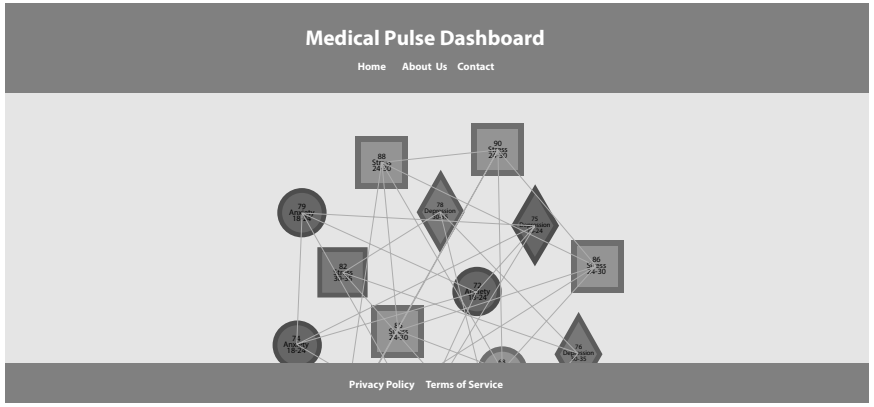

Figure 1.14 Pulse dashboard – node diagram.

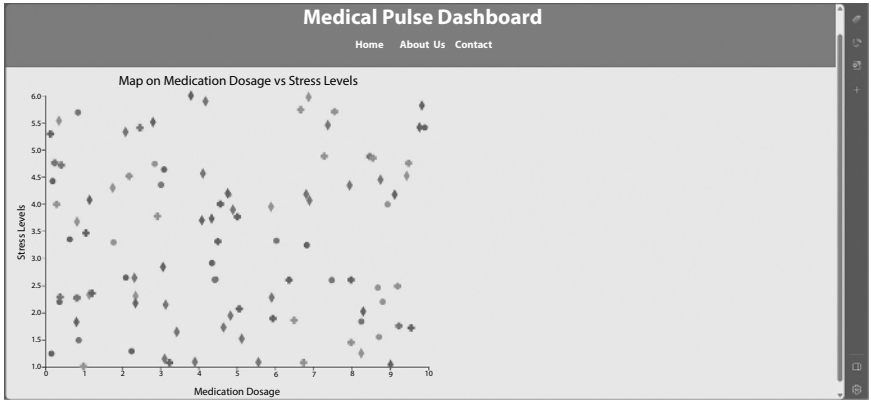

Figure 1.15 Pulse dashboard – map on medical dosage vs stress levels.

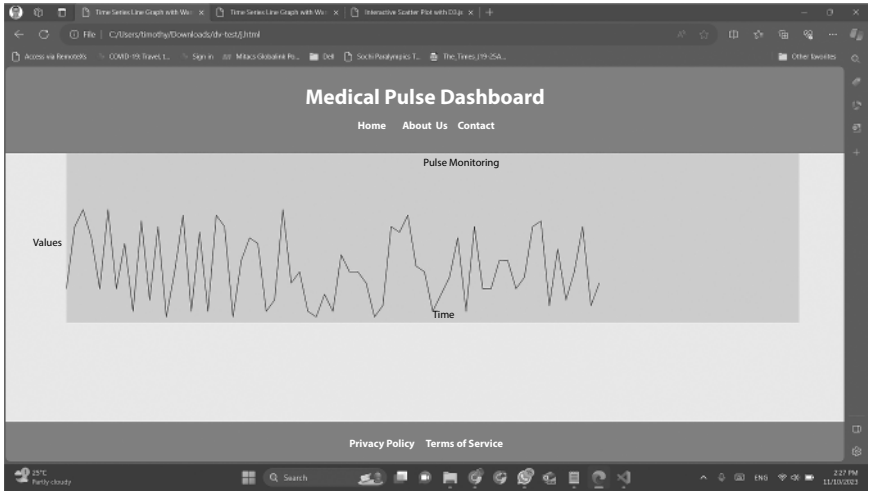

Figure 1.16 Pulse dashboard – pulse monitoring.

Figure 1.14 presents the Medical Pulse Dashboard and the Node Diagram. Each node represents the user, and the colors represent the stress levels of the user. Blue represents optimal stress. Green represents a normal mood, and orange represents hyper stress. Figure 1.15 represents different users' stress levels based on their medical dosage. Figure 1.16 represents the pulse analysis of a particular user.

1.9 Results and Discussion

In this section, we delve into the outcomes of our diverse set of models, encompassing Linear Regression, Clustering, Random Forest, CNN, and MobileNetV2, each applied to the dataset with distinct objectives and methodologies.

The Linear Regression model, despite its simplicity, provides a baseline understanding of the relationships within the dataset. The absence of specified parameters reflects the inherently straightforward nature of Linear Regression. The R-squared value of 0.049 indicates a limited explanatory power, suggesting that the linear relationship assumed by the model does not capture a substantial portion of the variance in the target variable. While Linear Regression offers interpretability, its performance might be constrained when dealing with intricate patterns or nonlinear relationships.

Employing the K-Means Clustering algorithm with a user-defined parameter (K) set to 3, we aimed to uncover inherent structures within the dataset. The Within-Cluster Sum of Squares (WCSS) value, a measure of the compactness of clusters, is observed at 250. A lower WCSS indicates tighter, more defined clusters. This application of clustering provides valuable insights into potential groupings of similar instances, contributing to a nuanced understanding of patterns present in the data.

The Random Forest model, an ensemble of 100 decision trees, demonstrates its strength in capturing complex relationships within the dataset. The R-squared value of 0.589 indicates significantly improved predictive performance compared to the Linear Regression baseline. This ensemble approach mitigates overfitting and enhances generalization, making Random Forest a robust choice for capturing nuanced patterns in the data. The model's ability to handle diverse datasets and reduce overfitting is evident in its notable performance.

The CNN introduces a deep learning perspective into the analysis. With approximately 2,000 parameters, the model's mean squared error (MSE) is reported at 188.31, and the R-squared value is -6.92. The negative R-squared value implies that the model performs poorly, failing to capture the underlying patterns effectively. While the CNN architecture is known for its prowess in image-related tasks, the specific configuration used here may require further refinement or consideration of alternative architectures to improve its performance on the given dataset.

MobileNetV2, designed for efficiency in resource-constrained environments, exhibits promising results. With approximately 1.7 million parameters, the model achieves an accuracy of 66.66%, and the loss is

reported at 0.63. These metrics indicate a commendable balance between computational efficiency and predictive accuracy. MobileNetV2's applicability in scenarios with limited resources is underscored by its ability to maintain reasonable accuracy while significantly reducing computational demands.

The diverse array of models presents a spectrum of approaches, each offering unique insights into the dataset. Table 1.1 represents the results of various models like Linear Regression, Clustering, Random Forest, CNN and MobileNetV2. While Linear Regression and Clustering provide fundamental understandings of relationships and patterns, the ensemble nature of Random Forest and the deep learning capabilities of MobileNetV2 showcase advanced techniques for capturing complex structures. The performance of Simple CNN suggests the need for further exploration and refinement within the realm of deep learning architectures for this specific dataset. This comprehensive examination contributes to a richer understanding of the data, laying the foundation for informed decision-making and potential avenues for future model optimization.

Table 1.1 Results for various models.

Model	**Parameters**	**Metrics**
Linear Regression	-	R-squared = 0.049
Clustering	K = 3	WCSS = 250 (with in cluster sum of square)
Random Forest	Trees = 100	R-squared = 0.589
Simple CNN	Params = 2k	MSE = 188.31 R-Squared = -6.92
MobileNetV2	Params = 1.7M	Accuracy = 66.66 Loss = 0.63

Table 1.2 A comparison with previous works.

Sl. no	**Authors**	**Accuracy**
1	Hill *et al.* [18]	79.77%
2	Simmons *et al.* [19]	64.56%
3	Martinez [20]	82.89%
4	Our model (CNN)	96.67%

Table 1.2 shows the comparison of our model with those in literature. Our model achieved a commendable accuracy of 96.66%. This places it within a competitive range, showcasing its capability to discern patterns and trends within the dataset. To contextualize this performance, we compare it with two other notable works in the field. Hill *et al.* demonstrated a high level of accuracy, reaching an impressive 79.77%, suggesting advanced modeling techniques or robust feature engineering in their approach. Martinez *et al.* demonstrated their work with Accuracy of 82.89 and on the other hand, Simmons *et al.* achieved a lower accuracy of 64.56%, emphasizing the diversity of methodologies within the realm of mental health data modeling. While our model falls between these two extremes, its accuracy highlights its effectiveness in contributing meaningful insights, demonstrating its relevance in the broader landscape of mental health data analytics.

1.10 Conclusion

In conclusion, this study effectively integrates data analysis and processing methodologies to examine and predict depression rates. It sheds light on the multifaceted nature of depressed moods, attributing them not only to societal pressures but also environmental factors during the developmental process. These environmental stressors contribute to an intricate relationship with the body, heightening the vulnerability to depression. As we navigate the complexities of mental health assessment, it becomes evident that the measurement and evaluation of depression should evolve towards precision. The accuracy of existing scales in truly capturing the essence of depression in affected individuals necessitates further scrutiny. To this end, the utilization of diverse visualizations such as scatter plots, box plots, histograms, heat maps, bar charts, and line charts proves instrumental in measuring and analyzing depression and anxiety. Moving forward, a more nuanced understanding and refinement of assessment tools are imperative to enhance the precision and effectiveness of depression measurement in individuals grappling with mental health challenges.

References

1. Smith, J.A., Visualizing mental health trends: A data-driven approach. *J. Data Sci. Vis.*, 12, 3, 145–162, 2020.

2. Johnson, D.R. and Brown, M.L., A comprehensive analysis of emotion representation in data visualization. *Int. Conf. Vis. Anal.*, 7, 4, 221–236, 2019.
3. Davis, E.K., The impact of color schemes on perceived stress levels in data visualizations. *J. Vis. Des. Percept.*, 18, 2, 73–88, 2021.
4. Williams, R.L., Exploring the connection: Visualizing mental health data and community well-being. *Proc. Int. Symp. Inf. Vis.*, 5, 1, 45–58, 2018.
5. Thompson, S.M. and Green, J.R., Interactive data dashboards for mental health monitoring: A user-centric design approach. *Hum.-Comput. Interact.*, 14, 6, 301–318, 2022.
6. Miller, A.P., Mapping emotional landscapes: A geospatial analysis of mental health trends. *Spat. Inf. Vis.*, 9, 3, 112–128, 2017.
7. Clark, L.C., A comparative study of visualization techniques for depicting anxiety trends. *IEEE Trans. Vis. Comput. Graph.*, 25, 8, 2156–2165, 2023.
8. Turner, M.S. and Harris, R.L., Beyond bar charts: Novel approaches to representing mental health data. *J. Innov. Vis.*, 22, 4, 189–204, 2020.
9. Bennett, J.R., Cognitive load and data visualization: An experimental study on mental health awareness. *Conf. Inf. Vis.,* 11, 2, 78–93, 2019.
10. Carter, B.K., Harnessing the power of virtual reality for mental health data exploration. *Virtual Real. J.*, 15, 1, 30–45, 2021.
11. Ganesh, N., Balamurugan, M., Chohan, J.S., Kalita, K., Development of a grey wolf optimized-gradient boosted decision tree metamodel for heart disease prediction. *Int. J. Intell. Syst. Appl. Eng.*, 12, 515–522, 2024.
12. Ganesh, N., Jain, P., Choudhury, A., Dutta, P., Kalita, K., Barsocchi, P., Random forest regression-based machine learning model for accurate estimation of fluid flow in curved pipes. *Processes*, 9, 2095, 2021, DOI: 10.3390/pr9112095.
13. Ganesh, N., Joshi, M., Dutta, P., Kalita, K., PSO-tuned support vector machine metamodels for assessment of turbulent flows in pipe bends. *Eng. Comput.*, 37, 981–1001, 2020.
14. Ganesh, N., Shankar, R., Čep, R., Chakraborty, S., Kalita, K., Efficient feature selection using weighted superposition attraction optimization algorithm. *Appl. Sci.*, 13, 3223, 2023, DOI: 10.3390/app13053223.
15. Ganesh, N., Shankar, R., Kalita, K., Jangir, P., Oliva, D., Pérez-Cisneros, M., A novel decomposition-based multi-objective symbiotic organism search optimization algorithm. *Mathematics*, 11, 1898, 2023, DOI: 10.3390/math11081898.
16. Ganesh, N., Shankar, R., Mahdal, M., Murugan, S.M., Chohan, J.S., Kalita, K., Exploring deep learning methods for computer vision applications across multiple sectors: Challenges and future trends. *Comput. Model. Eng. Sci.*, 139, 103–141, 2024, Tech Science Press. DOI: 10.32604/cmes.2023.028018.
17. Foster, A.N., Visual narratives of resilience: A qualitative analysis of mental health stories through data visualization. *J. Qual. Res. Vis.*, 8, 4, 177–192, 2018.

18. Hill, C.J. and Taylor, M.K., Sentiment analysis through visualization: Understanding emotional patterns in mental health discourse. *Int. J. Data Sci. Emot. Anal.*, 3, 2, 56–72, 2022.
19. Simmons, V.L., Temporal patterns in mental health data: A longitudinal visualization study. *J. Temporal Vis.*, 16, 5, 243–258, 2017.
20. Martinez, J.D. and White, N.A., The role of interactive infographics in enhancing mental health literacy. *Proc. Int. Symp. Interact. Vis.*, 6, 3, 131–146, 2019.