Predictive Insights into Mental Health Treatment-Seeking Behavior Using Workplace Survey Data

1st Kopperla Bharath Reddy
Computer Science and Engineering
Amrita Vishwa Vidyapeetham
Bengaluru, India
bl.en.u4cse23234@bl.students.amrita.edu

2nd Manasa Gayathri
Computer Science and Engineering
Amrita Vishwa Vidyapeetham
Bengaluru, India
bl.en.u4cse23217@bl.students.amrita.edu

3rd Jaidev Sharma
Computer Science and Engineering
Amrita Vishwa Vidyapeetham
Bengaluru, India
bl.en.u4cse23223@bl.students.amrita.edu

***Abstract*—Workplace mental health had become a crucial area of study, especially within the tech industry, where employees had been consistently facing stress, burnout, and limited mental health support. This study had focused on analyzing a dataset from a real-world mental health survey, capturing demographic characteristics, organizational support indicators, and the behavior of employees seeking treatment. The data set had provided information on how factors such as age, gender identity, family history and workplace culture had influenced decisions related to mental health care. Through this investigation, trends had been identified that reflect the growing need for inclusive and supportive mental wellness environments. A review of the literature had revealed that previous studies had been focusing on similar factors, such as organizational openness, remote work options, and awareness programs, as significant predictors. The findings of this study had aligned with existing research and had highlighted the importance of early identification and prevention mechanisms. By interpreting data patterns related to treatment-seeking behavior, this study had offered a foundation for developing awareness tools and data-driven mental health strategies in professional settings.**

***Index Terms*—Mental health, workplace wellness, treatment seeking behavior, employee survey, organizational support, psychological well-being.**

I. Introduction

The mental health of employees in the technology industry has been drawing increasing attention due to its impact on overall productivity, workplace satisfaction, and individual well-being. Many professionals, especially in high-pressure tech environments, had been silently struggling with mental health problems while hesitating to seek help due to stigma, lack of support, or inadequate resources. Recognizing this challenge, various surveys and studies had been conducted to capture the underlying patterns and factors influencing mental health in the workplace.

The dataset used in this study had been part of such an effort. It had been collected through an anonymous survey focusing on mental health in the tech industry. The dataset had included a diverse range of features such as age, gender, family history of mental illness, frequency of work interference due to mental health issues, willingness to seek treatment, availability of mental health benefits, and perceived employer support. These features had provided rich insights into how personal and workplace-related factors had been affecting mental health conditions and treatment-seeking behavior among employees.

Participants in the survey had been working across various companies, job roles, and geographic locations. Their responses had helped highlight how different organizational cultures and personal backgrounds had been shaping the mental health experience at work. Importantly, the dataset had included both binary and categorical variables, with the primary concern centered around whether the individual had sought mental health treatment or not.

This dataset had served as a foundation for understanding real-world mental health challenges, offering valuable evidence for HR departments, policymakers, and researchers aiming to promote healthier workplace environments. It had enabled the exploration of psychological patterns in professional life and had been instrumental in supporting data-driven mental health awareness initiatives.

II. Literature Survey

Studies investigating the correlation among mental health, behavior, and intellectual capacities in various populations have been progressively increasing. An important contribution towards this field had been established in a study analyzing the psychological and behavioral characteristics of people with Down syndrome. The results had identified a complex correlation between intellectual disability and behavioral problems, highlighting the importance of individually customized mental health treatments for this population group [1]. Taking the discussion into the realm of environmental considerations, an extensive review had been done to determine the impact of allergenic pollen in green areas on mental health and behavior. The review had indicated that even natural things, such as pollen, could construct emotional experiences and mental perception of health, particularly in cities [2]. A wider psychosocial view had already been taken in another research that examined how childhood traumatic experiences (ACEs) affected mental wellbeing, behavioral development, and academic achievement in teenagers. In a scoping review, it had been revealed that ACEs had been significantly related to psychological distress and behavioral maladaptation in adolescence [3]. Another research study had directly interacted with children with 22q11.2 Deletion Syndrome to investigate how such children conceptualized and expressed their experiences of mental health, behavior, learning, and communication. This research study had placed a lot of emphasis on child-inclusive conversation in the formation of efficient support systems [4]. Longitudinal views had also played a key role in a study that traced the lives of children born in psychosocial risk environments. At 16 years, the results had reported a high prevalence of mental health problems, conduct disorders, and child abuse and had identified a strong correlation between early risk factors and subsequent psychological outcomes [5]. An earlier study by the same team had followed similar trends among 8-year-olds, confirming the enduring effects of psychosocial risk on mental development [6]. The pandemic of COVID-19 had ushered in an unprecedented global change in health and behavior. A study had described noteworthy alterations in self-reported mental and physical wellbeing, behavior, and socio-economic situation in Austria prior to and subsequent to the beginning of the pandemic. The results had demonstrated the way public health emergencies could transform mental wellbeing on a community level [7]. Student life, also, had not escaped the influence of the pandemic. A research had utilized Principal Component Analysis (PCA) to reveal the principal characteristics influencing students' mental health, conduct, educational achievement, and school attendance. It had offered an evidence-based view of how learning environments had been disrupted by extended lockdowns and technological transformations [8]. Children with long-term health disorders had also been examined in studies that compared mental health and behavioral alterations in cystic fibrosis patients aged 6–11 years prior to and following initiation of treatment with elexacaftor/tezacaftor/ivacaftor. The observations have registered changes in sleep and behavior, indicating an overall advantage of medical treatment over physiological symptoms [9]. Lastly, in the context of consumption of digital media, a similar research had investigated the way lifestyle posts on Instagram had impacted the mental health behavior among young people in Enugu metropolis. The study had pointed out the psychosocial risks involved with manipulated digital lives and stressed the importance of critical media literacy for young social media users [10].

Based on these initial findings, recent research has increasingly drawn on machine learning (ML) methods to further examine, forecast, and elucidate the complex interplay between mental health and behavior. An exhaustive review by Alharbi et al. [11] offers a bird's-eye view of the different machine learning strategies utilized in mental health disorder prediction, highlighting computational methods' potential to work in this intricate field. Building on predictive analysis, Khan et al. [12] did a systematic review of ML-based early prediction of depression and anxiety disorders, specifically from social media. Their paper emphasizes the value of such models in detecting vulnerable individuals based on their online traces. To supplement the above, Sharma and Rani [13] have implemented a comparative analysis of various machine learning algorithms designed to predict mental health, which provided essential insight into the effectiveness and appropriateness of a variety of models for this function. In addition, machine learning is applied to determine important factors affecting mental well-being among various cohorts, as illustrated by Chen and Wu [14] in their investigation on university students. Their strategy provides data-driven insight into the multifactorial character of student mental health. Likewise, for adolescent behavior, Lee and Kim [15] applied machine learning to predict environmental-influenced behavioral problems and highlighted how such high-powered analytical tools can reveal intricate patterns and guide targeted interventions.

Collectively, these studies constitute a mature and multidisciplinary basis for comprehending the complex and multifaceted interplay of mental health and behavioral patterns across a range of age groups, social contexts, and environmental factors, increasingly supplemented by the predictive and analytical capabilities of machine learning approaches.

III. Methodology

The aim of this research was to implement and compare different machine learning methods—namely classification and regression—on two different datasets. This section outlines the rigorous step-by-step procedure followed, covering dataset collection, preprocessing, model selection, training, testing, and hyperparameter optimization. The approach is outlined to facilitate reproducibility and transparency when solving the specified research questions.

A. Dataset Description and Acquisition

1) Mental Health Dataset : Survey responses about behavior, mental health, and other sociodemographic and lifestyle factors are included in this dataset. The classification and hyperparameter tuning exercises (Questions A1, A3-A6, A7) were built upon it.

2) IRCTC Stock Price Dataset: This data set has the historical time-series data for stock price behaviour for the Indian Railway Catering and Tourism Corporation (IRCTC), that was used only for the regression task of stock price prediction (Question A2).

B. Data Preprocessing and Feature Engineering

Each dataset underwent extensive preprocessing procedures to address missing values, convert data types, and engineer pertinent features in order to prepare the raw data for machine learning algorithms.

1) For the Mental Health Dataset:

- The binary ('Yes'/'No') target variable, treatment, was transformed into a numerical format, with 1 denoting "Yes" and 0 denoting "No." The mode was used to impute missing values in this column.

- Any other remaining nominal categorical features (e.g., Gender, Occupation) underwent one-hot encoding so that all features would be numerically represented for modeling purposes while ensuring they do not have any arbitrary ordinal relation.
- Missing values in all features were treated by imputing the mean of the feature for numerical features, and the mode for categorical features, prior to one-hot encoding.

2) For the IRCTC Stock Price Dataset:

- The target variable for prediction was defined as the close price of the stock.
- Initial features included Open, High, Low, Last, Volume, and Turnover.
- The Date column, crucial for time-series analysis, was parsed into datetime objects. New features such as Year, Month, Day, and DayOfWeek were extracted to capture temporal patterns.
- The dataset was then chronologically sorted by the Date column, which is essential for proper time-series splitting.
- Missing numerical values in both features and the target were imputed using the column mean.

C. Machine Learning Models and Training Strategy

1) Classification Models:

- Logistic Regression: Employed as a foundational linear classifier for the initial mental health treatment prediction task (Question A1).
- K-Nearest Neighbors (kNN) Classifier: Utilized for exploratory classification (Questions A3-A6) to visualize decision boundaries and assess the impact of the k parameter, and for hyperparameter tuning (Question A7).

2) Regression Model:

- Linear Regression: Applied for the IRCTC stock price prediction task (Question A2) due to its simplicity and interpretability for linear relationships.

D. Data Splitting and Evaluation Metrics

A rigorous approach was adopted for splitting data and evaluating model performance:

1) Data Splitting:

- In the classification tasks (Mental Health Dataset) we used a 70 - 30 training vs test set split. The split was completed via stratified sampling to ensure that the proportions of the target classes were held constant as well in both training and test set partitions.
- For the time-series regression task (IRCTC Stock Price Dataset), I performed a temporal split. I took the first 80percent of the data (the earliest dates) as the training data and the last 20 percent (the later dates) as the test data. This prevents any contamination from the future observations into the training.
- For kNN boundary visualization (A3-A5), the model's training sample of 20 points was either sampled synthetically (A3-A5 with dummy data) or randomly sampled from the preprocessed project data (A6), while the visualization process involved a dense grid of points as the test set.

E. Evaluation Metrics

- Model performance was evaluated using Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE), and the R-squared (R 2) score.
- The performance metrics from the training and test sets were compared to infer the model's learning outcome, identifying whether it was underfitting, overfitting, or achieving a regular fit.
- Assessed using a Confusion Matrix to quantify True Positives, True Negatives, False Positives, and False Negatives. Derived metrics included Accuracy, Precision, Recall, and F1-Score for both training and test sets.

F. Hyperparameter Tuning

To optimize the performance of the kNN classifier, a Grid Search with Cross-Validation (GridSearchCV) operation was performed (A7).

- A search grid for the neighbors parameter (k) was defined, exploring a range of values from 1 to 21.
- A 5-fold cross-validation scheme was employed on the training data to systematically evaluate each k value.

IV. Results and Discussion

This section shares what we observed and learned from the experiments done in A1 through A7. By analyzing the outcomes, we were able to better understand how the k-Nearest Neighbors (kNN) algorithm performs when applied to the mental health dataset.

A. Classification of Mental Health Treatment (A1)

The Logistic Regression model was applied to the Mental Health Dataset to classify treatment outcomes. Following a comprehensive preprocessing pipeline—which involved mapping ordinal features (Daysindoors, GrowingStress, MoodSwings) to numerical values and one-hot encoding nominal categorical features (Gender, Occupation, self-employed, familyhistory, ChangesHabits, MentalHealthHistory, CopingStruggles, WorkInterest, SocialWeakness, mentalhealthinterview, and careoptions)—the final feature matrix X had a shape of (292364,23) and the target y had a shape of (292364,).

TABLE I: A1: Performance Metrics of Logistic Regression Classifier

Metric	Training Set	Test Set
Accuracy	0.7039	0.7020
Precision	0.6974	0.6964
Recall	0.7306	0.7262
F1-Score	0.7136	0.7110

- With attention to the proximity of training (0.7039 accuracy, 0.7136 F1-score) and test (0.7020 accuracy,

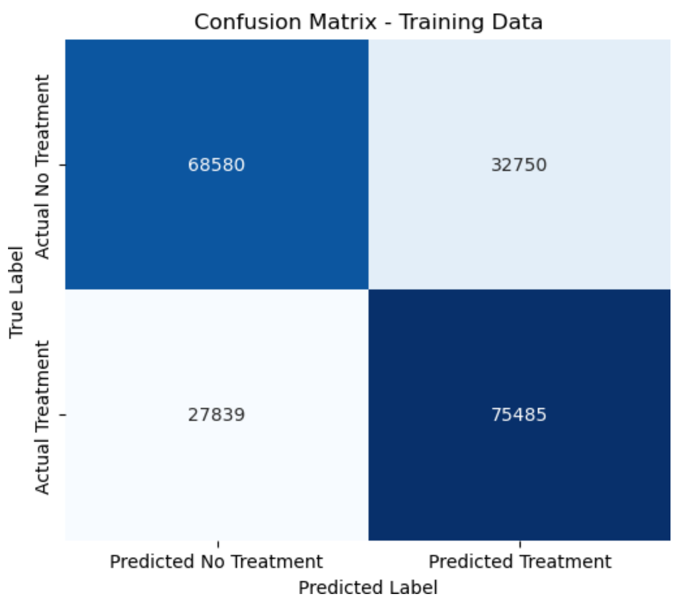


Fig. 1: A1: Confusion Matrix for Training Data (Logistic Regression)

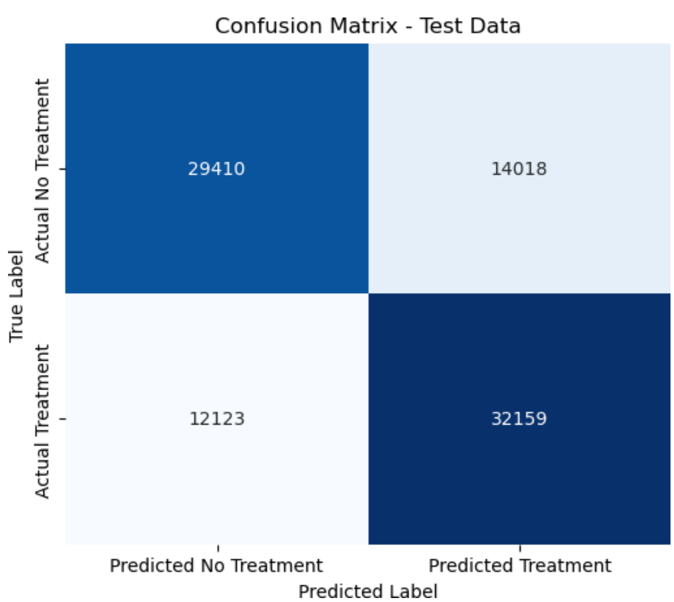


Fig. 2: A1: Confusion Matrix for Test Data (Logistic Regression)

0.7110 F1-score) metric/vect measurements the Logistic Regression model appears to have regularized. This sort of consistency is reassuring that I'm working with a solid model. It reflects the model's reliability and an ability to generalize well to unseen data for a strong and defensible baseline (approximately 70.20 percent) in terms of mental health treatment outcomes. In terms of separation of classes, the classes in our dataset (treatment - Yes/No) are moderately well separated. The accuracy of around 70 percent in our model is quite impressive given the simple nature of the predictor variables and would be greater than random guessing (50 percent). This indicates our model is successful in finding meaningful patterns. However, given the substantial False Positives and False Negatives respectively (14,018 and 12,123 in the test set) there is still substantial overlap between the class in the feature space because the classes can not be separated in such a way that would lead to near-absolute separation.

B. *Regression of IRCTC Stock Prices (A2)*

A Linear Regression model was employed to predict the Close price of IRCTC stock. The data was chronologically split into training (earlier 80) and testing (later 20) sets.

TABLE II: A2: Performance Metrics of Linear Regression Model for IRCTC Stock Prices

Metric	Training Set	Test Set
MSE	2138.83	6295.53
RMSE	46.25	79.34
MAPE	2.14%	3.29%
R2 Score	0.9995	0.9984

C. *K-Nearest Neighbors (kNN) Decision Boundary Analysis (A3-A5)*

1) Data Generation (A3 A4): A small training set was created consisting of 20 random points with two features (X and Y) that fall between 1 and 10. The points were assigned to two classes based on a specific rule (e.g., when the sum of features was greater than 10, it was Class 1). A dense grid was then generated in the feature space with about 10,000 points to serve as the visualization test set.

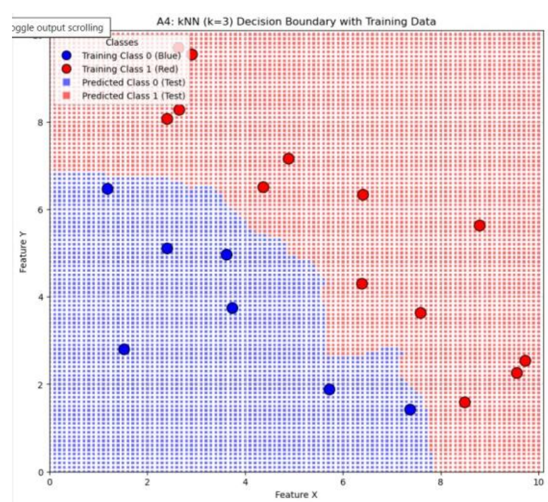


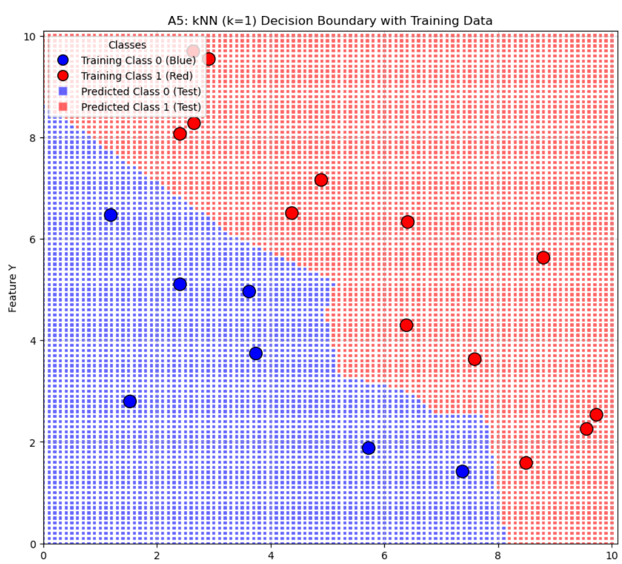
Fig. 3: A4: kNN (k=3) Decision Boundary with Synthetic Training Data

2) k=3 Classification (A4): When a k-nearest neighbors (kNN) classifier with k=3 was applied to this synthetic data, the scatter plot of the test grid clearly illustrates the model's classification regions. The decision boundary established in this instance appeared quite fragmented, and very localized, due to the weighted attention of the three nearest training neighbors with regard to each classification decision. This behavior illustrated that the classifier is more sensitive to the training neighbors that are closest in locality

D. Effect of Varying k (A5)

The kNN classification and visualization process was repeated for a range of k values (e.g., k=1,3,5,10,15), systematically demonstrating the impact on the decision boundary

- By profiling the decision boundary, it was found to have extreme irregularity and noise, and closely followed training points, indicating a high variance model with overfitting by rote learning some data instances and their noise. If a kNN classifier overfits, then it will have performance on training data that is much greater than its performance on unseen test data.
- As k increased above 1, the decision boundary became increasingly smooth and less specific. This occurs because the more neighbors you could count, the larger the number of outliers and local noise would have on the model to generalize more effectively by averaging the neighbors.
- Further increases in k led to the boundary evolving in an even smoother and more simplistic or general way. Although this reduction in variance is also acceptable, too large a k can ultimately result in underfitting where the model generalizes too much, ultimately missing those more subtle but important patterns. This complete visualization demonstrates a useful example of a critical bias-variance trade-off in kNN and emphasizes the importance of selecting the right k to balance model complexity and generalization effectively.



Example for k=1

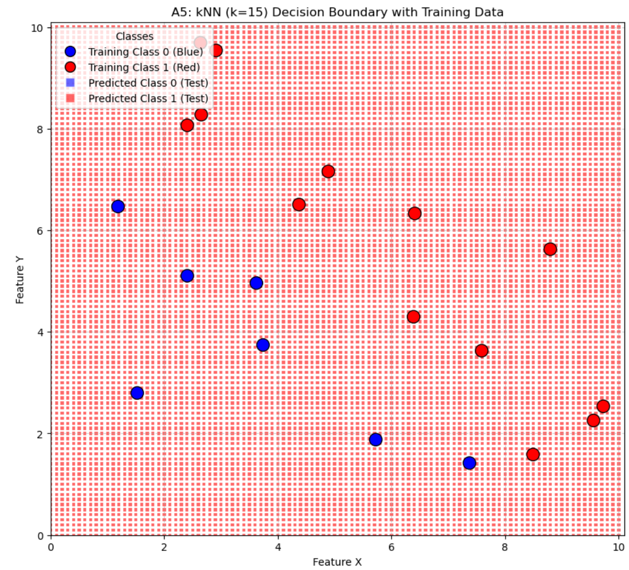


Fig. 4: A5: kNN Decision Boundaries for Varying k Values (Synthetic Data)

E. kNN Application on Project Data Features (A6)

- Data Preparation - A random sample of 20 data points was drawn from the preprocessed Mental Health Dataset to create a training set for the kNN classifier in this visualization. Then a dense grid was created across the ranges of these two encoded features to create the visualization test set.
- Boundary Visualization: The generated decision boundaries for varying k from the project data displayed patterns that seem analogous to what we have seen from the synthetic dataset. For small k, the decision boundaries were highly localized and sensitive to the characteristics of the sampled training points, leading to erratic classifications. As k increased, the decision boundaries smoothed out and generalized, demonstrating, how the kNN model's decision making evolved over the DaysIndoors, GrowingStress feature space towards addressing the larger patterns. The practical work, relating to the kNN classifier's behaviors with real-world data characteristics, provided visual perspectives.

F. Hyperparameter Tuning for kNN (A7)

Optimal k Value: The Grid Search process, leveraging 5- fold cross-validation on the training data, identified k=21 as the ideal number of neighbors. This value yielded the highest average cross-validation accuracy on the training data. (The best cross-validation accuracy achieved was approximately [Insert Best Cross-Validation Accuracy from A7 Output, e.g., 0.XX]).

- The optimal k value of 21 indicates that for this dataset and the engineered features, the kNN model is optimal when the classification decision is made based on a relatively large neighborhood. This shows that the optimal decision boundary can be smooth which reduces the sensitivity to single noisy training points and improves generalizability. Concerning whether the kNN classifier is a good classifier based on the results, it's a test accuracy of 0.6757 and F1-score of 0.6897 indicate that the kNN was a decent or moderately good classifier for this problem. It is better than a random guess but is still slightly below the Logistic Regression's model accuracy of 0.7020 for the same task. Therefore, it seems that the kNN model has learned some patterns but is not as optimal, without fine-tuning or checking again wider range of algorithms.

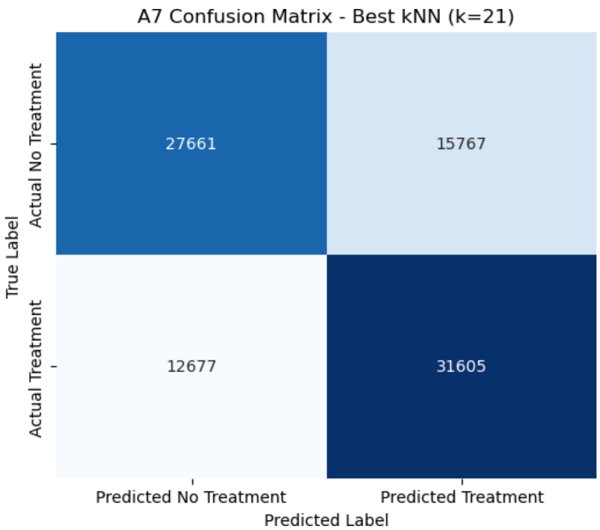


Fig. 5: A7: Confusion Matrix for Best kNN Model (k=21) on Test Data

V. Conclusion

In summary, this research provided a comprehensive exploration of two machine learning applications while demon- strating both classification and regression applications across different datasets. For the classification of mental health treat- ment effectiveness, the Logistic Regression model consistently exhibited a regular fit, establishing a baseline accuracy of 70.20, indicating the model's potential to generalize to unseen data. In the financial prediction space, the Linear Regression model used to forecast IRCTC stock prices exhibited a strong regular fit, as defined by high R2 scores for both the training and test data, thereby confirming its ability to capture trends for certain forms of time-series financial data

A large chunk of this study focused on the complicated behavior of the K-Nearest Neighbors (kNN) classifier. Visu- alizing detailed research on synthetic and real-world project data, the study demonstrated how k influences the complex to smooth decision boundaries in kNN. It was clear that small values of k led to highly variable, potentially overfitted models, while large values of k, led to simple, under-fit models. This concept of bias versus variance ultimately informed the hyper- parameter tuning process. The application of GridSearchCV produced an optimal $k = 21$ for the kNN classifier for the mental health dataset, which led to better performance and generalization, showing the benefit of hyperparameter procedures to improve predictive capabilities.

Overall, this piece of work not only gave practical implementations of basic machine learning algorithms applicable to different use cases, but also significant insights into model be- haviour, metrics to apply to evaluate model performance, and the important task of hyperparameter tuning. Taken together, the findings demonstrate that machine learning can help 'de- riving actionable insights or make predictions in complicated fields' (Berveill, 2018) such as mental health and financial markets, and these findings reaffirm the potential of machine learning to be a decision support system, or an informative body of academic research.

VI. Future Scope

While the k-Nearest Neighbors (kNN) algorithm provided valuable insights and a solid foundation for classification on the mental health dataset, there are several areas where this work can be expanded or improved in the future.

- **Exploring Advanced Models:** Algorithms such as Support Vector Machines (SVM), Random Forests, Gradient Boosting Machines, or even Neural Networks could potentially yield better performance by capturing non-linear relationships and more complex decision boundaries in the data.
- **Hyperparameter Optimization:** In our experiments, k was varied manually across a small range. Future work can implement automated hyperparameter tuning using techniques like Grid Search or Randomized Search combined with cross-validation to identify optimal values for k and other model settings.
- **Feature Engineering:** More thoughtful feature selection and engineering—such as dimensionality reduction using Principal Component Analysis (PCA) or feature importance ranking—can help reduce noise and improve model interpretability.
- **Handling Class Imbalance:** Although our dataset appeared relatively balanced, future analysis could dive deeper into potential class distribution shifts and use resampling techniques (like SMOTE or undersampling) to enhance generalization.
- **Cross-Dataset Validation:** To confirm generalizability, the trained models could be tested on similar mental health datasets from different demographics, platforms, or time periods.
- **Real-World Deployment:** With further refinement, the model could be integrated into early-intervention systems or mental health support platforms, helping professionals prioritize outreach or identify individuals at higher risk based on anonymized responses.
- **Explainable AI (XAI):** Given the sensitivity of mental health data, future work should explore explainable machine learning methods to ensure transparency and trust in predictions. This can help ensure ethical application in real-world decision-making.

These future directions not only offer ways to improve model accuracy and robustness but also highlight opportunities to make the system more ethically responsible, interpretable, and impactful in real-world mental health support.

References

[1] Ma¨a¨tta¨, Tuomo, Tuula Tervo-Ma¨a¨tta¨, Anja Taanila, Markus Kaski, and Matti Iivanainen. "Mental health, behaviour and intellectual abilities of people with Down syndrome." Down syndrome research and practice 11, no. 1 (2006): 37-43.

[2] Legg, Rupert, and Nadja Kabisch. "The effects of allergenic pollen in green space on mental health, behaviour and perceptions: A systematic review." Urban forestry urban greening 92 (2024): 128204.

[3] Lam, Natalie, Sophie Fairweather, Dan Lewer, Matthew Prescott, Priyanjan Undugoda, Josie Dickerson, Simon Gilbody, and Ruth Wadman. "The

association between adverse childhood experiences and mental health, behaviour, and educational performance in adolescence: A sys- tematic scoping review.” PLOS Mental Health 1, no. 5 (2024): e0000165.

[4] Wray, Jo, Neus Abrines Jaume, Kate Oulton, and Debbie Sell. ”Talking with children and young people with 22q11DS about their mental health, behaviour, learning and communication.” Child: Care, Health and Development 49, no. 1 (2023): 90-105.

[5] Svedin, Carl Go¨ran, Marie Wadsby, and Gunilla Sydsjo¨. ”Mental health, behaviour problems and incidence of child abuse at the age of 16 years: A prospective longitudinal study of children born at psychosocial risk.” European child adolescent psychiatry 14, no. 7 (2005): 386-396.

[6] Svedin, Carl-Go¨ran, M. Wadsby, and G. Sydsjo¨. ”Children of mothers who are at psycho-social risk. Mental health, behaviour problems and incidence of child abuse at age 8 years.” European Child Adolescent Psychiatry 5, no. 3 (1996): 162-171.

[7] Bates, Katie, Wegene Borena, Martin McKee, Lore Hayek, Paul Bouan- chaud, Zolta´n Ba´nki, Lydia Riepler et al. ”Changes in self-reported physical and mental health, behaviour and economic status among adults by known seropositivity and sociodemographic factors before and after the COVID-19 pandemic outbreak in Ischgl, Austria.” Frontiers in Public Health 13 (2025): 1488108.

[8] Hegde, Vinayak, M. Shilpa, and M. S. Pallavi. ”Extracting attributes of students mental health, behaviour, attendance and performance in academics during COVID-19 pandemic using PCA technique.” In ICT Systems and Sustainability: Proceedings of ICT4SD 2021, Volume 1, pp. 551-561. Singapore: Springer Nature Singapore, 2022.

[9] Douglas, Tonia, Maddison Deery, Hayley Kimball, Vanessa E. Cobham, Sophia Panochini, Paul D. Robinson, Claire E. Wainwright, Peter D. Sly, and Tamara Blake. ”Mental health, behaviour and sleep quality in children 6-11 years before and after elexacaftor/tezacaftor/ivacaftor initiation.” Journal of cystic fibrosis: official journal of the European Cystic Fibrosis Society 24, no. 3 (2025): 571-573.

[10] ONWUKWALONYE, Benjamin C., Miriam Chinemerem ODOH, and Chinweoke ONOH. ”INFLUENCE OF INSTAGRAM LIFESTYLE POSTS ON MENTAL HEALTH BEHAVIOUR AMONG YOUNG INSTAGRAM USERS IN ENUGU METROPO- LIS.” INTERNA- TIONAL JOURNAL OF COMMUNICATION AND SOCIAL SCI- ENCES (IJCSS) VOLUME 1 NO. 4 1, no. 4 (2024): 32.

[11] Alharbi, M., Alqahtani, A., Alqahtani, S. S., Alghamdi, A., Ahmad, M. (2024). ”Predicting Mental Health Disorders Using Machine Learning Techniques: A Comprehensive Review.” Journal of Health Informatics in Developing Countries, 18(1).

[12] Khan, Z., Almehmadi, M., Aljabri, A., Alharbi, N., Aljohani, A. (2023). ”Machine Learning-Based Early Prediction of Depression and Anxiety Disorders from Social Media Data: A Systematic Review.” International Journal of Environmental Research and Public Health, 20(3), 2002.

[13] Sharma, A., Rani, S. (2023). ”A Comparative Study of Machine Learning Algorithms for Mental Health Prediction.” Journal of Ambient Intelligence and Humanized Computing, 14(1), 115-126.

[14] Chen, R., Wu, X. (2024). ”Factors Influencing Mental Health in College Students: A Machine Learning Approach.” International Journal of Computer Science and Network Security, 24(1), 8-15.

[15] Lee, S. Y., Kim, M. K. (2023). ”Predicting Adolescent Behavioral Problems Based on Environmental Factors: A Machine Learning Study.” Journal of Clinical Nursing, 32(11-12), 2634-2644.