**Predicting the Helpfulness of Mental Health Treatment Methods Using ML**

# Abstract

Mental health problems represent a significant global health challenge, needing effective and personalized treatment methods. This study explores the application of machine learning (ML) models to predict the perceived helpfulness of eight different mental health treatment approaches, based on data from the Wellcome Global Monitor 2020 Mental Health Module. By applying structured survey data, this study aims to develop predictive models that provide personalized mental health care planning. Three ML algorithms, Random Forest, neural network (MLP), and XGBoost were trained and evaluated across eight independent prediction tasks, each representing a unique treatment method, medication, lifestyle change, spiritual practices, etc. The performance was evaluated using cross validation accuracy, precision, recall, and F1 score. Results demonstrated that ensemble models, particularly Random Forest, consistently outperformed other methods in predicting treatment helpfulness. These results emphasise the potential of ML to inform clinical decisions, enhance treatment provision, and promote precise mental health treatment approaches at the population level.

# 1**. Introduction**

Mental health disorders are the main causes of global disease problems, requiring the need for effective and personalized treatment approaches. With multiple treatment methods ranging from medical therapy to community based and spiritual activities, determining the most effective option for an individual remains a challenge. In recent years, ML has shown a potential in improving clinical decision making and predicting treatment outcomes[1]. The challenge of predicting which treatment will work better for which individual in mental health care continues to be a challenge in clinical practice. Machine learning approaches are effective in capturing complex relations between demographic, clinical, and psychosocial variables. For instance, dynamic prediction models such as the Oracle algorithm have shown higher performance over traditional expected treatment outcome models in identifying patients likely to benefit from therapy[3]​. Furthermore, studies leveraging ML have shown that personalized prediction using routine clinical data can help in early identification of treatment resistance, guiding more targeted interventions [4]

Recent study [2] demonstrated that ML models provide a generalizable model in automatic risk signaling technology to identify cases at risk of poor treatment outcomes. Another emerging focus is predicting treatment adherence, which is critical for long-term success. Hybrid models, mixing classifiers like SVM, MLP, and Random Forest, have also been successfully applied to pre clinical datasets to predict stages of anxiety and suggestpersonalized recommendations, showing the potential of machine learning for proactive mental health treatment plans in both clinical and conflict settings [6].   
This study aims to apply ML models to predict the perceived helpfulness of eight different mental health treatment methods, using data from the Wellcome Global Monitor 2020 Mental Health Module. By analyzing survey responses and demographic variables, the objective is to inform more personalized mental health treatment planning and policy.

# **2. Related Work**

Many studies have explored the application of machine learning in mental health treatment. [3] developed a dynamic model using session by session symptom data and demonstrated improved predictive accuracy over traditional expected response models, with an AUC of 0.81 by session seven​. [2] conducted a study across three different mental health organizations in the Netherlands using routinely collected clinical data. Their study employed a least absolute shrinkage and selection operator (LASSO) regression to predict treatment outcomes, achieving an area under the curve (AUC) between 0.77 and 0.80 in both internal and external evaluations, showing good generalizability across different clinical settings.

In a comprehensive review[1] analyzed 59 papers and found that machine learning could predict depression treatment outcomes with reasonable accuracy, mean balanced accuracy up to 0.75, although higher study quality was inversely related to reported accuracy, which highlights the need for careful validation​. [8] shown that the effectiveness of gradient boosting models in predicting clinical change in psychotherapy using routinely collected outpatient data. Their model achieved a balanced accuracy of 69%, identifying key predictors such as functioning level and symptom severity​.

The stepped care principles discovered by [2] provides a framework for allocating limited healthcare resources, which is cost effective for depression and anxiety. Their research highlights the importance of routine outcome monitoring (ROM) to observe patterns of early treatment response and identify which patient will likely not benefit from their current treatment.

[4] demonstrated the potential of personalized prediction tools in real world psychiatric settings. Using elastic net regularization, they achieved an R² of 0.38 and developed a prognosis calculator to predict treatment outcomes based on baseline features​. The [5] study focused on predicting treatment using ML algorithms such as logistic regression, k-nearest neighbors (KNN), decision trees, and random forest. These models were trained on outpatient datasets and highlighted the importance of early dropout prediction. [6] introduced a hybrid mental health prediction model using Support Vector Machine, Multilayer Perceptron, and Random Forest algorithms to classify pre clinical anxiety stages in a conflict-affected region. The model reached a remarkable 98.13% accuracy, underscoring the feasibility of early stage intervention planning even in resource constrained or high stress environments​.

[7] conducted a systematic analysis across 24 different machine learning studies. The models achieved classification accuracy of 74%. Their findings emphasized the greater success of ML in anxiety related disorders and highlighted the need for robust cross validation practices and interpretability in clinical applications**​**

These studies collectively highlight the applicability of ML in predicting mental health treatment outcomes and support its use in informing patient specific treatment planning.

# 3. Methodology

This study employed data from the Wellcome Global Monitor 2020 Mental Health Module to develop machine learning models that predict the helpfulness level of different mental health treatment methods among individuals who self-reported experiencing anxiety or depression. The aim was to identify patterns within demographic, behavioral, and experiential data that relate to perceptions of treatment effectiveness.

## 3.1 Data Source and Target Variables

The dataset was filtered to include only respondents who answered affirmatively to item MH7A, indicating they had experienced anxiety or depression. The target variables comprised self-reported helpfulness scores of eight distinct treatment approaches, encoded from variables MH9A to MH9H. Each was measured on an ordinal scale (1 = Very helpful, 2 = Somewhat helpful, 3 = Not helpful). The eight treatment types included professional advice, medication, social support, lifestyle changes, work and relationship changes, spiritual activities, and time spent in nature.

## 3.3 Data Preprocessing

To confirm analytic validity and model robustness, a complete preprocessing pipeline was conducted. Initially, entries with missing responses to key outcome variables coded as 99, denoting “Don’t know” or “Refused” were excluded from the training data. Predictor variables with missing data were imputed using the mode, which was appropriate given the mainly categorical nature of the variables.

Categorical features were encoded based on their scale: ordinal variables were transformed using ordinal encoding to preserve inherent order, while nominal features like global region, employment type were encoded using one-hot encoding.   
Given the high class imbalance in several MH9 variables particularly for the ‘Not helpful’ category, bootstrap oversampling was applied to augment underrepresented classes within each fold of the training data. This approach was chosen to reduce bias and improve minority class recall without introducing high synthetic variability.

Each treatment method was modeled as a separate prediction task, allowing the analysis to capture treatment specific patterns in helpfulness level.   
3.4 Modeling Approach and Evaluation  
Three machine learning models were selected for the prediction Random Forest, Neural Network, and XGBoost. These models were chosen for their strengths Random Forest and XGBoost are robust ensemble models well suited for structured survey data, while neural networks offer nonlinear modeling capacity suitable for capturing subtle patterns across high dimensional features.

All models were trained using k-fold cross-validation to ensure generalizability and to reduce overfitting. Performance was evaluated separately for each of the eight treatment targets using multiple evaluation metrics, including Cross validation accuracy, precision, recall, and F1-score.

# 4. Results and Discussion

The three machine learning models, Random Forest, neural network, and XGBoost were trained and evaluated independently across eight target variables (MH9A–MH9H), each representing a different treatment method. All models were evaluated using five-fold cross validation, and results were reported based on accuracy and weighted performance metrics.

Across all targets, Random Forest consistently demonstrated superior predictive performance**,** achieving the highest accuracy in most of the eight treatment categories. Specifically, it performed best in predicting helpfulness perceptions for MH9B (religious/spiritual activities) and MH9H (spending time in nature), with both exceeding 90% accuracy. This aligns with findings from prior literature indicating Random Forest’s robustness in handling non linear interactions and mixed data types [3] [8]

Neural networks: while it performs low relative to ensemble methods, it’s still shown a good predictive performance. The model exhibited low stability and lower accuracy, particularly on MH9C (talking to friends/family) and MH9G (changing personal relationships) relative to the other two models. This may be attributed to MLP's sensitivity to imbalanced class distributions and the limited expressiveness of certain features without deeper representation learning orlarger training samples.

XGBoost performed competitively and in some instances nearly matched Random Forest in predictive accuracy, particularly for MH9E (healthy lifestyle behaviors) and MH9H. Its ability to handle structured data and prevent overfitting through gradient boosting made it a strong secondary performer.

| **Target** | **Accuracy** | **Precision** | **Recall** | **F1 Score** |
| --- | --- | --- | --- | --- |
| MH9A | 0.88784 | 0.886333 | 0.886018 | 0.886333 |
| MH9B | 0.905367 | 0.904873 | 0.904704 | 0.904873 |
| MH9C | 0.878058 | 0.87468 | 0.874221 | 0.87468 |
| MH9D | 0.849872 | 0.850847 | 0.848848 | 0.850847 |
| MH9E | 0.90234 | 0.899743 | 0.899444 | 0.899743 |
| MH9F | 0.873345 | 0.873331 | 0.873121 | 0.873331 |
| MH9G | 0.87039 | 0.868149 | 0.867725 | 0.868149 |
| MH9H | 0.907557 | 0.90598 | 0.905658 | 0.90598 |

The predictive performance trends also reflect the nature of the treatments themselves. Interventions e.g., MH9A: professional help; MH9D: prescribed medication were easier to predict than those relying on interpersonal or environmental changes e.g., MH9C and MH9G. This aligns with studies suggesting that treatments involving professional facilitation are more consistently evaluated and perceived across respondents [7][2].

Additionally, the high performance on religious practices and spending time in nature may show stronger consensus in how spiritual and nature-based interventions are perceived, possibly reflecting culturally consistent associations with wellbeing. These findings open paths for culturally adapted recommendations and treatment prioritization in low resource or non-clinical settings.

## **5. Conclusion and Future Work**

This study highlights the application of machine learning in predicting the perceived helpfulness of mental health treatments, based on demographic and behavioural data from the Wellcome Global Monitor 2020. Random Forest models were the most effective, with accuracy over 88% for most treatments. From a neuroscience perspective, these findings demonstrate how machine learning can uncover complex, non linear relationships in mental health perceptions, potentially informing personalized treatment approaches and supporting neuropsychological research. This approach fills the gap between public perception and clinical decision making.

Future research should:

* Study neural mechanisms underlying treatment perceptions, combining ML with neuroimaging or EEG data.
* Investigate the influence of cognitive biases on treatment perception and explore ways to mitigate them using machine learning.  
  Machine learning offers valuable tools for advancing neuroscience research and improving personalized mental health care by linking perceptions to neural processes.

# References

[1] M. Sajjadian *et al.*, “Machine learning in the prediction of depression treatment outcomes: A systematic review and meta-analysis,” Dec. 12, 2021, *Cambridge University Press*. doi: 10.1017/S0033291721003871.

[2] K. Van Mens, J. Lokkerbol, B. Wijnen, R. Janssen, R. de Lange, and B. Tiemens, “Predicting Undesired Treatment Outcomes With Machine Learning in Mental Health Care: Multisite Study,” *JMIR Med Inform*, vol. 11, no. 1, Jan. 2023, doi: 10.2196/44322.

[3] C. Bone *et al.*, “Dynamic prediction of psychological treatment outcomes: development and validation of a prediction model using routinely collected symptom data,” *Lancet Digit Health*, vol. 3, no. 4, pp. e231–e240, Apr. 2021, doi: 10.1016/S2589-7500(21)00018-2.

[4] Z. D. C. C. B. M. F. Christian A. Webb, “Supplemental Material for Personalized Prognostic Prediction of Treatment Outcome for Depressed Patients in a Naturalistic Psychiatric Hospital Setting: A Comparison of Machine Learning Approaches,” *J Consult Clin Psychol*, 2020, doi: 10.1037/ccp0000451.supp.

[5] M. Bajaj, P. Rawat, Diksha, S. Vats, V. Sharma, and L. Gopal, “Prediction of Mental Health Treatment Adherence using Machine Learning Algorithms,” in *2023 International Conference on Computational Intelligence, Communication Technology and Networking, CICTN 2023*, Institute of Electrical and Electronics Engineers Inc., 2023, pp. 716–720. doi: 10.1109/CICTN57981.2023.10141520.

[6] E. S. Mohamed, T. A. Naqishbandi, S. A. C. Bukhari, I. Rauf, V. Sawrikar, and A. Hussain, “A hybrid mental health prediction model using Support Vector Machine, Multilayer Perceptron, and Random Forest algorithms,” *Healthcare Analytics*, vol. 3, Nov. 2023, doi: 10.1016/j.health.2023.100185.

[7] S. Vieira, X. Liang, R. Guiomar, and A. Mechelli, “Can we predict who will benefit from cognitive-behavioural therapy? A systematic review and meta-analysis of machine learning studies,” Nov. 01, 2022, *Elsevier Inc.* doi: 10.1016/j.cpr.2022.102193.

[8] F. S. Taubitz, B. Büdenbender, and G. W. Alpers, “What the future holds: Machine learning to predict success in psychotherapy,” *Behaviour Research and Therapy*, vol. 156, Sep. 2022, doi: 10.1016/j.brat.2022.104116.