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

**Abstract**

**1. Introduction**

Mental health disorders are a leading cause of global disease burden, prompting the need for effective and personalized treatment strategies. With a multitude of treatment methods ranging from pharmacotherapy to community-based and spiritual interventions, determining the most effective option for a given individual remains a challenge. In recent years, machine learning (ML) has shown promise in enhancing clinical decision-making and predicting treatment outcomes (Sajjadian et al., 2021)

**The challenge of predicting which treatment will work for which patient in mental health care continues to be a significant obstacle in clinical practice. As Van Mens et al. (2023) highlight, optimizing healthcare systems requires preventing the allocation of too many resources to some patients and too little to others, essentially maximizing the opportunity for appropriate care at an individual level.**

Machine learning approaches are particularly effective in capturing complex relationships between demographic, clinical, and psychosocial variables that traditional statistical models might overlook. For instance, dynamic prediction models such as the Oracle algorithm have demonstrated superior performance over traditional expected treatment response models in identifying patients likely to benefit from therapy(Bone et al., 2021)​. Furthermore, studies leveraging ML have shown that personalized prediction using routine clinical data can aid in early identification of treatment resistance, guiding more targeted interventions (Christian A. Webb, 2020)

Recent research by Van Mens et al. (2023) demonstrated that machine learning models provide a robust and generalizable approach in automated risk signaling technology to identify cases at risk of poor treatment outcomes. Their multisite study showed strong external validation, indicating that models developed in one clinical setting can perform similarly when applied to another site.

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 individualized mental health care planning and policy.

## ****2. Related Work****

Multiple studies have explored the application of machine learning in mental health treatment. (Bone et al., 2021) developed a dynamic model using session-by-session symptom data and showed improved predictive accuracy over traditional expected response models, reaching an AUC of 0.81 by session seven​.

Van Mens et al. (2023) conducted a multisite study across three different mental health care 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 validations, demonstrating good generalizability across different clinical settings.

In a comprehensive review(Sajjadian et al., 2021) analyzed 59 studies 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, highlighting the need for careful validation​.

(Taubitz et al., 2022) demonstrated 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 explored by Van Mens et al. provide a framework for allocating limited healthcare resources, which has proven cost-effective for depression and anxiety. Their research emphasizes the importance of routine outcome monitoring (ROM) to observe patterns of early treatment response and identify which patients will likely not benefit from their current treatment.

(Christian A. Webb, 2020) emphasized 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​.

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

## ****3 Methodology****

This study utilizes data from the Wellcome Global Monitor (WGM) 2020 Mental Health Module, focusing on individuals who self-reported having experienced anxiety or depression (MH7A = 1). The target outcome is the self-reported perceived helpfulness of eight distinct treatment approaches, captured in variables MH9A through MH9H.

### **3.1 Data Selection and Features**

Only respondents who answered “Yes” to MH7A ("Have been anxious/depressed") were retained. Each respondent rated up to eight treatment methods on a 3-point scale: 1 = Very helpful, 2 = Somewhat helpful, 3 = Not helpful (variables MH9A to MH9H). These include:

* MH9A: Talking to a mental health professional
* MH9B: Engaging in religious/spiritual activities
* MH9C: Talking to friends or family
* MH9D: Taking prescribed medication
* MH9E: Improving healthy lifestyle behaviours
* MH9F: Changing work situation
* MH9G: Changing personal relationships
* MH9H: Spending time in nature/outdoors

Predictor features were selected from a combination of:

* **Demographics**: Age (Age, age\_mh), gender (Gender), education (Education), income level (Household\_Income, wbi), employment status (EMP\_2010), and subjective income status (Subjective\_Income).
* **Mental health experience**: Age first experienced anxiety/depression (age\_mh), if experienced more than once (MH7C), and whether others in the respondent’s life had experienced anxiety/depression (MH6).
* **Beliefs and behaviors**: Comfort with discussing mental health (MH5), attitudes toward science and treatment (e.g., MH3B, MH4B), and use of each coping method (MH8A–MH8H).

### **3.2 Preprocessing**

* **Missing Data**: Predictors with missing values were imputed using the mode for categorical and ordinal variables. Targets with code 99 (DK/Refused) were excluded.
* **Encoding**: Ordinal variables (e.g., MH3B, MH4B) were ordinal encoded. Nominal variables (e.g., Gender, Employment Status) were one-hot encoded.
* **Scaling**: Continuous variables like age were normalized using MinMax scaling.
* **Resampling**: To address class imbalance (e.g., fewer “Not helpful” responses), we applied bootstrap oversampling to minority classes during model training.

### **3.3 Modeling and Evaluation**

Three machine learning models were implemented using 5-fold cross-validation:

* Random Forest Classifier
* Multi-layer Perceptron (Neural Network)
* XGBoost Classifier

Each model was trained independently for each target variable (MH9A–MH9H). Performance was evaluated using:

* Accuracy
* Weighted Precision, Recall, and F1 Score
* Cross-Validation Accuracy

## ****4. Results and Discussion****

### **4.1 Random Forest**

This model consistently outperformed others across most MH9 targets, particularly MH9B and MH9H (both exceeding 90% accuracy). It demonstrated robustness in handling non-linear relationships and feature interactions.

### **4.2 Neural Network**

Neural Network performance was comparatively weaker, particularly on MH9C and MH9G, likely due to sensitivity to imbalanced data or limited feature expressiveness.

### **4.3 XGBoost**

XGBoost performed well and was close to Random Forest in most metrics, with particularly competitive scores in MH9B and MH9H.

| **Target** | **Best Accuracy** | **Best Model** |
| --- | --- | --- |
| MH9A | 0.8878 | Random Forest |
| MH9B | 0.9054 | Random Forest |
| MH9C | 0.8781 | Random Forest |
| MH9D | 0.8499 | Random Forest |
| MH9E | 0.9023 | Random Forest |
| MH9F | 0.8733 | Random Forest |
| MH9G | 0.8704 | Random Forest |
| MH9H | 0.9076 | Random Forest |

The findings are in line with prior research where ensemble models like Random Forest and XGBoost showed high generalizability and resilience to overfitting (Bone et al., 2021; Taubitz et al., 2022).

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

References

Bone, C., Simmonds-Buckley, M., Thwaites, R., Sandford, D., Merzhvynska, M., Rubel, J., Deisenhofer, A. K., Lutz, W., & Delgadillo, J. (2021). Dynamic prediction of psychological treatment outcomes: development and validation of a prediction model using routinely collected symptom data. *The Lancet Digital Health*, *3*(4), e231–e240. https://doi.org/10.1016/S2589-7500(21)00018-2

Christian A. Webb, Z. D. C. C. B. M. F. (2020). Supplemental Material for Personalized Prognostic Prediction of Treatment Outcome for Depressed Patients in a Naturalistic Psychiatric Hospital Setting: A Comparison of Machine Learning Approaches. *Journal of Consulting and Clinical Psychology*. https://doi.org/10.1037/ccp0000451.supp

Sajjadian, M., Lam, R. W., Milev, R., Rotzinger, S., Frey, B. N., Soares, C. N., Parikh, S. V., Foster, J. A., Turecki, G., Müller, D. J., Strother, S. C., Farzan, F., Kennedy, S. H., & Uher, R. (2021). Machine learning in the prediction of depression treatment outcomes: A systematic review and meta-analysis. In *Psychological Medicine* (Vol. 51, Issue 16, pp. 2742–2751). Cambridge University Press. https://doi.org/10.1017/S0033291721003871

Taubitz, F. S., Büdenbender, B., & Alpers, G. W. (2022). What the future holds: Machine learning to predict success in psychotherapy. *Behaviour Research and Therapy*, *156*. https://doi.org/10.1016/j.brat.2022.104116