Analysing the Effectiveness of Anxiety and Depression Treatment in the UK: A Data-Driven Approach Using Feature Selection and Machine Learning

Anna Bienkowska[1], Ikram Ur Rehman[1] , Muazzam Ali Khan Khattak[2], Julie Wall[1]

[1]School of Computing and Engineering, University of West London, London, United Kingdom

[2]Department of Computer Science, Quaid-i-Azam University, Islamabad, Pakistan

21490999@student.uwl.ac.uk, ikram.rehman@uwl.ac.uk, muazzam.khattak@qau.edu.pk, julie.wall@uwl.ac.uk

Abstract

This research examines anxiety and depression treatment in the UK from 2012 to 2023. It aims to assist healthcare providers and researchers by offering evidence-based insights to optimise treatment strategies. More importantly, it seeks to improve therapy outcomes for patients referred to psychological services who are at risk of not achieving positive results from their treatment. With the increasing number of cases of anxiety and depression, advanced studies are needed to evaluate treatment effectiveness using demographic and clinical data supported by data analysis and machine learning techniques. Data was sourced from NHS Therapies for Anxiety and Depression, including 142 datasets, 11 annual reports, and seven interactive dashboards. Nine metrics were selected, analysed, and visualised. Then, using five metrics, predictive modelling was conducted with Linear Regression and Random Forest Regressor, both demonstrating strong predictive performance. The Random Forest model produced the best results with an 80/20 data split and 83 as the random state parameters. This model achieved a Mean Squared Error of 0.31, an R^2 value of 0.97, and a mean cross-validation score of 0.40. Linear Regression resulted in a prediction with a Mean Absolute Error of just 0.05 and a Root Mean Square Error of 0.045.

Introduction

Anxiety and depression are common mental health disorders that can cause specific symptoms characterised by significant disturbances in the thinking, behaviour, and emotional regulation of the affected individual. They can impact individuals differently, and each person's experience may differ (Baker et al., 2019). According to the World Health Organisation (WHO), there has been a substantial global increase in the prevalence of anxiety and depression (WHO, 2022). This increase imposes a significant cost on the global economy, estimated at US$1 trillion annually (Organization, 2022). In the period between 1990 and 2013, there was a nearly 50% rise in referrals of people struggling with anxiety and depression, soaring from 416 million to 615 million (WHO, 2016). Furthermore, the WHO reported that in 2019, approximately 970 million people worldwide, or one in eight individuals, experienced mental health disorders, with anxiety and depression emerging as the most common. Additionally, in 2020, due to the COVID-19 pandemic, the numbers continued to increase even more significantly, with anxiety cases rising by 26% and major depressive disorder by 28% (WHO, 2022).

In terms of the UK, research has indicated that one in four adults experiences at least one diagnosable mental health issue in any given year. Mental illness is the leading cause of illness and disability in the UK, resulting in an estimated loss of 91 million working days per year, almost 40% of people on disability benefits, and takes up around one-third of GPs' time (Oparina et al., 2024).

A significant aspect of anxiety and depression is that both conditions can be improved with mental health treatment. In 2008, a significant development of psychological therapies took place in England, including the launch of the NHS's Improving Access to Psychological Therapies (IAPT) program. This program enabled almost 1.2 million people to access services in 2021/22 (Oparina et al., 2024). Furthermore, the NHS Long Term Plan aims to expand NHS psychological therapy services and make them accessible to 1.9 million people annually by the end of 2023/24 (England, 2019).

With the continued advancements in information technology, the medical and health science fields have become overwhelmed with data over the past few decades. The healthcare industry has increasingly relied on Artificial Intelligence (AI) and machine learning to make sense of this vast amount of complex information and extract valuable insights. By analysing patient data, such as medical records and behavioural patterns, AI can assist mental health providers in making well-informed decisions (Su et al., 2020). The remainder of this paper is organised as follows: it commences by outlining the research potential and key research questions, followed by a review of the relevant literature. The methodology section then details the data analysis process, including visualisation and predictive modelling. This is followed by the presentation and analysis of results, a discussion addressing the research questions, and ultimately, the conclusion summarising the key findings and their implications.

Research Potential

Machine learning, a significant component of AI, has the potential to revolutionise mental healthcare. Mental illness is an area where perfect treatment in terms of therapy or prescribed medications has not been achieved (Kumar et al., 2021). For many years, psychologists and psychiatrists have strived to understand the factors involved in the response to medications or psychotherapies to personalise treatment choices. There is a growing interest in this area, with the aim of developing models that enhance treatment decisions using innovative statistical approaches from machine learning (Chekroud et al., 2021). This could potentially improve the effectiveness of mental health care by designing personalised treatment plans for patients, tailoring these plans to their individual needs, or offering alternative treatments (Vieira et al., 2022).

Research Questions

Four prominent research questions guided this study:
RQ1: Is there a publicly available dataset on anxiety and depression that can be utilised for statistical analysis, visualisation, and developing a predictive model to assess the efficacy of mental health services in England?
RQ2: What were the most common treatments for anxiety and depression in England from 2012 to 2023, and how have the key trends in the types of therapies offered evolved?
RQ3: How do various demographic factors, such as age, gender, and socioeconomic status, impact the treatment outcomes for patients with anxiety and depression, and what trends can be observed in treatment effectiveness over the last 11 years? What visual insights can be obtained from the data?
RQ4: How accurately can a predictive model forecast the effectiveness of the therapeutic interventions for anxiety and depression, and what are the key predictors influencing these outcomes?

Literature Review

This literature review investigates the potential of machine learning to predict the outcomes of psychotherapy treatments (Chen et al., 2023). The carefully selected research papers presented various investigations with different perspectives and methods for analysing and evaluating the effectiveness of psychological therapies, aiming to make these services more beneficial for more patients.

In 2024, Oparina *et al.* conducted an analysis of the effectiveness of anxiety and depression therapies available through the NHS in England. The study was based on a large dataset of 1,246,792 service users between April 2016 and December 2018. The methods used for this analysis were a Regression Framework, a set of statistical and machine learning techniques, a nonparametric method, and a nonparametric method strengthened by machine learning techniques such as Random Forest. The findings indicated that Cognitive Behavioural Therapy (CBT) was particularly effective. On average, reliable recovery increased by around 43% and reliability improvement by 38%. However, this research brought to attention the fact that this treatment was not beneficial to everyone.

The specific groups who were less likely to achieve reliable recovery included patients with more severe symptoms, as assessed by the Generalised Anxiety Disorder (GAD-7) questionnaire for anxiety and the Patient Health Questionnaire (PHQ-9) for depression severity. Additionally, factors such as having long-term health conditions, disabilities, being non-white British, non-religious, living in deprived areas, being unemployed, and being referred by general practitioners (GPs) rather than self-referral contributed to a decreased likelihood of reliable recovery (Oparina et al., 2024).

Similarly, Ewbank *et al.* (2020) investigated the relationship between patient variables and internet-enabled CBT treatment outcomes. This study differs from the previous one in a few aspects: it uses a different treatment delivery method, has fewer participants, and has a larger time frame. The number of individuals who started the treatment was 17,572, but later decreased to 14,899 due to missing data, especially from PHQ-9 and GAD-7 symptoms severity questionnaires. As an online therapy, most data included information from activities such as text messages. There were also approximately 90,934 patient session transcripts, modules, and workshops categorised into feature categories using a deep learning approach. The relationships between session features and treatment outcomes were analysed using multivariable Logistic Regression. These analyses showed that CBT therapy provided a reliable improvement rate of 63.4% and an engagement rate of 87.3%. The patient variables linked to improvement rate were the GAD-7 anxiety score level, patients who did not have prescribed medications for their condition, absence of long-term health conditions, older age, and a higher number of treatment sessions received by service users.

Despite many differences between these two research papers in many aspects, both concluded that the PHQ-9 depression severity score and having a long-term medical condition were connected to lower chances of their condition improving (Ewbank et al., 2020). Both papers also demonstrated that the GAD-7 and PHQ-9 scores are highly important and related to mental health treatment outcomes. In England, patients must complete these two questionnaires at the start of the service and just before each session. This practice helps service providers monitor and evaluate their condition and track their progress during treatment.

Green *et al.* (2015) aimed to predict positive and negative CBT treatment outcomes from independent pretreatment variables and identify key variables contributing to a positive outcome, including GAD-7 and PHQ-9 patient responses. The statistical method used for both tasks was Discriminant Analysis classification. This study initially included 7,388 patients who joined two health services in London from February 2009 to May 2012. This number later decreased to 4,393, including only patients who had fully completed the treatment. The predictive model determined the patients' positive or negative treatment outcome by considering five variables: the severity level of anxiety and depression at the start, ethnicity, gender, deprivation, and the number of sessions attended. The model accuracy for a positive outcome was 69.4%, and for a negative result, 79.3%. This research confirmed the findings of the first two studies mentioned, which reported that poorer outcomes were associated with elevated GAD-7 and PHQ-9 scores, socioeconomic deprivation, and treatment duration. It also highlighted the importance of analysing data collected at the beginning of treatment. Such analysis could provide essential information that helps to create a more individualised treatment plan for patients. Moreover, the high number of patients not completing their treatment highlights the importance of identifying their characteristics and understanding why they withdrew from the service (Green et al., 2015).

Bennemann *et al.* (2022) attempted to investigate this issue and aimed to identify the patients at risk of dropping out of their CBT psychotherapy treatment. The dataset was not extensive, consisting of 2,543 patients who joined a German clinic between 2007 and 2021. Treatment began with patients filling out PHQ-9 and GAD-7 questionnaires and attending therapy once a week. The investigation methodology involved comparing twenty-one machine learning algorithms using nested cross-validation to identify the best model and most crucial variables. The results showed that the most powerful model was an ensemble that used Random Forest and nearest-neighbour modelling. This model correctly identified 63.4% of cases at risk of discontinuing the treatment program before starting the treatment. It is essential to acknowledge that this research excluded important variables, such as PHQ-9 and GAD-7, due to missing data. Unfortunately, these scores were among the most crucial variables in predicting the treatment outcome in other research papers, and having this data would improve the model. The results included the main predictors of withdrawal, such as lower education and young age, which were previously associated with lower recovery rates, as shown in the other literature (Bennemann et al., 2022).

Anxiety and depression are primary mental health issues, and in recent decades, many treatments have proven to be effective. However, many patients experience relapses, which means a return to full symptoms following remission or recovery (Krijnen-de Bruin et al., 2022).

Lorimer *et al.* (2021) conducted a study focusing on developing a tool for identifying cases at risk of relapse. They also investigated the reason for deterioration after recovery from low-intensity CBT treatment completion. The dataset was small, with 317 patients in this study completing their treatment successfully and having monthly follow-ups for one year. During this period, 70% of them, totalling 223 individuals, experienced a relapse. The data was analysed using a machine learning approach, an ensemble of XGBoost algorithms. The results demonstrated good predictive accuracy using a cross-validation design. The Area Under the Curve (AUC) ranged from 71% to 75%, indicating a 71-75% probability that a positive outcome is correct (Positive Predictive Value, PPV). Additionally, there is a 56-74% chance that a negative result is correct (Negative Predictive Value, NPV). This study indicated that the crucial predictors that determine high-risk cases include treatment response, residual symptoms, young age, and unemployment, which were previously mentioned in the literature as predictors of poor mental outcomes and a high risk of dropping out from CBT psychotherapy treatment. This research indicates a positive response to CBT therapy. However, it also points out the necessity for additional interventions, such as ongoing therapy sessions or follow-up care, to maintain patient improvement. Given that this therapy was low-intensity, it may be beneficial to consider starting with a high-intensity treatment or switching to a different type of therapy during treatment. Furthermore, the study also emphasised the importance of implementing follow-up care after treatment completion (Lorimer et al., 2024).

Various therapies are available for anxiety and depression, categorised into low-intensity and high-intensity treatments. The types of therapies recommended to patients are based on the severity of their symptoms, which is determined by a questionnaire they must complete before joining the service. Lorenzo *et al.* tried to develop a predictive model to help mental health providers assign the type of therapy to patients based on their expected prognosis rather than solely relying on the questionnaires they complete. The dataset contained 622 patients with depression, and by using machine learning, developed a Prognostic Index (PI) to guide a selection of treatments of different intensities. Unemployment, severity of depressive symptoms, sleep problems and lower positive emotionality were associated with a lower likelihood of recovery across all available treatments. The PI integrated these variables, producing a classification accuracy of 73%.

In a study of patients with a high PI (75% of the sample), recovery rates were high and similar across treatments (79-86%) (Lorenzo-Luaces et al., 2017). However, in patients with the poorest prognosis, recovery rates were significantly better with CBT (60%) compared to Treatment as Usual (TAU) (39%) and Behavioural Therapy (BT) (44%). This model has the potential to help assign the most effective treatment, allow healthcare providers to tailor treatments more effectively and improve patient outcomes.

The literature indicates that more needs to be done to improve the effectiveness of psychological therapies. It highlights the importance of tailoring treatments to patients based on their symptoms, demographics, and clinical information. This paper explores the potential of machine learning in predicting treatment outcomes in medical and psychotherapy contexts, which could lead to significant advancements.

Methodology, Results & Analysis

This research is divided into data extraction, visualisation, and predictions. Methodology is carried out separately for each part, followed by Results and Analysis. This approach is taken because the completion of each part serves as the starting point for the next research phase.

Data

This section analyses the key measures such as activity, waiting times, patient demographics, and therapy outcomes within the IAPT program from 2012 to 2023. The data was extracted from multiple sources on the NHS website, including 142 datasets published by NHS England Digital, the 11 NHS Annual Reports, and the 6 NHS Dashboards (NHS England Digital, 2024).

Data preparation involved careful feature selection, ensuring all variables were considered, and minimising errors through rigorous double-checking. Two distinct datasets were designed: one for analysis and visualisation (nine metrics) and the other for predictive modelling (five metrics) (Jassim and Abdulwahid, 2021).

Metrics

Metrics include Activity Measures, Waiting time for the first treatment appointment, Gender, Age, Ethnicity, Employment status, Therapy Types (Low and High Intensity), Therapy Types (Various Treatment Types), and Outcome Measures.

Visualisation of all Metrics

Anaconda, a leading Python distribution platform, was used for data analysis alongside Jupyter Notebook.

The developed line charts, doughnut charts, and bar charts assisted in monitoring the metrics over eleven years and analysing the distribution of specific measurements over time.

Predictions

The dataset contained four critical demographic variables: age, gender, long-term condition, and ethnicity of the individuals referred to the psychological services, which were linked to outcome measures. The prediction was based solely on these dependent variables, which changed significantly over time and were highlighted during the analysis, visualisation, and other research presented in the literature review.

Due to data availability, the prediction dataset was created for four years, from 2017/2018 to 2020/2021, and contained these variables on five outcome measures for each year. Due to the need for dimension reductions, only "Reliable Improvement" was chosen for predictions, and all variables were grouped on one datasheet.

The prediction process was developed in Python and consisted of two methods: Linear Regression and Random Forest Regressor. Both algorithms were applied to the training dataset, and then the trained models were validated against a test dataset to assess their performance and accuracy (Singh et al., 2016).

The first method was Linear Regression, one of the simplest and most common statistical and machine-learning algorithms. Linear Regression is utilised to identify a linear relationship between one or more predictors (Maulud and Abdulazeez, 2020). For this method, the data was split in a 75/25 ratio. The second method used was the Random Forest Regressor. For this algorithm, the data was split using different ratios: 66/33, 75/25, and 80/20. To measure accuracy, metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Mean Square Error (MSE) and Coefficient of Determination (R^2) were used to assess the model's performance (Chicco et al., 2021). Furthermore, cross-validation was performed for the Random Forest Regressor.

The analysis comprised four steps: Step 1 applied linear regression, and Step 2 utilised the Random Forest Regressor. After removing the "Indeterminate" gender category from the dataset, linear regression and Random Forest Regressor were conducted again in Steps 3 and 4, respectively.

Step 1 - Linear Regression

Since the data for this analysis spans only four years, the Training and Test Years were prepared. The training data was based on the first three years: 2017, 2018, and 2019, and the test data was based on the year 2020. The "Reliable Improvement Rates" average was calculated, corresponding to the improvement rates for 2017/2018, 2018/2019 and 2019/2020; the training rate was calculated to correspond to

2020/2021, and then the data was reshaped for further processing. The next step involved creating a linear regression model, training it, and predicting 2020/2021 based on the rates from previous years. After training the model, it predicted the rate for 2020/2021 and compared it to the actual rate, providing a measure of how well the model has performed. The Error Metrics, both MAE and RMSE, were 3.41.

Step 2 - Random Forest Regressor

This process comprised five runs, during which the data split and random state were changed to identify the best result.

For the first run, 20% of the data was reserved for testing, and the split was done randomly but reproducibly by setting random_state=42. In the second run, the random state parameter was changed from 42 to 83. For the third run, the data split was changed from 80/20 to 66/33, while the random state remained at 83. Lastly, in the fourth run, the data split changed to 75/25, with the random state still at 83.

The fifth run involved using GridSearchCV with the Random Forest Regressor to determine the best combination of hyperparameters that minimises MSE on the validation set. The data split cv = 5 means that the data was split into five different folds: four training datasets and one remaining for testing against all four. The model with the best hyperparameters on the test set has the best MSE at 2.45. The best R^2 score for the best model was approximately 0.83, which indicates that the model explains 83% of the variance in the target variable.

The two tables below show the results from all five runs:

Random Forest Regressor	Data Split	Random State	Results
First Run	80/20	42	MSE = 43.68 R^2 = - 12.25
Second Run	80/20	83	MSE = 1.36 R^2 = 0.90
Third Run	66/33	83	MSE = 0.96 R^2 = 0.91
Fourth Run	75/25	83	MSE = 0.84 R^2 = 0.93

Table 1: Random Forest Regressor–Step 2

GridSearchCV	Data Split	Random State	Results
Fifth Run	CV – 5	83	Best MSE = 2.45 Best R^2 = 0.83

Table 2: GridSearch CV-Random Forest Regressor–Step 2

The Box Plot was also created, demonstrating that the predictions have remained consistently stable over the years, exhibiting only minor variations. However, some extremely low outliers are present. Furthermore, cross-validation was performed, and negative values were indicated for each Random Forest Regressor Run.

The outliers fall under the category labelled "Indeterminate" gender, as the value remained at zero for three years, with data available only for the fourth year (2020/2021). As a result, this category was eliminated from the dataset, and all procedures were conducted again on the revised dataset.

Step 3 - Linear Regression (revised dataset**)**

After repeating the exact predictions on a revised data set, the error metrics showed that MAE was 0.05 and RMSE was 0.045.

Step 4 - Random Forest Regressor (revised dataset**)**

The two tables below present the results from all five runs on a revised dataset:

Random Forest Regressor	Data Split	Random State	Results
First Run	80/20	42	MSE = 43.51 R^2 = -1.07
Second Run	80/20	83	MSE = 0.31 R^2 = 0.97
Third Run	66/33	83	MSE = 11.49 R^2 = 0.44
Fourth Run	75/25	83	MSE = 0.58 R^2 = 0.95

Table 3: Random Forest Regressor (revised dataset)–Step 4

GridSearchCV	Data Split	Random State	Results
Fifth Run	CV – 5	83	Best MSE =0.07 Best R^2 = 0.99

Table 4: GridSearch CV - Random Forest regressor tuned (revised dataset)–Step 4

Furthermore, cross-validation was performed for four runs, as presented below:

Random Forest Regressor	Data Split	Random State	Cross-validation (Mean CV Scores)
First Run	80/20	42	0.26
Second Run	80/20	83	0.40
Third Run	66/33	83	-0.53
Fourth Run	75/25	83	-36.31

Table 5: Random Forest Regressor

Discussion

This section presents the results for each research question.

Research Question 1

No suitable long-term data on anxiety and depression in England were found. To address this, new datasets were created by integrating annual NHS data from 2012 to 2023, enabling detailed exploration of historical trends. One dataset with nine metrics was used for analysis and visualisation, and another with five metrics was used for predictions.

Challenges during data compilation included inconsistencies in collection methods, discrepancies in categorisation, missing data, and unavailable information. Although these challenges were significant, they were addressed with careful attention to detail. They were overcome by meticulous data cleaning and validation, a deep understanding of the context in which the data was collected and reported, and an overall knowledge of England's psychological treatment delivery system.

Research Question 2

After analysis and visualisation of data, the most common treatments for anxiety and depression were identified and described. The data concerning high, low, and mixed intensity, visualised from 2015 to 2023, show a growing demand for mental health services and an increasing number of treatment sessions provided by the NHS. The most popular type is mixed therapy, which represents 39.5% of the total, followed by low-intensity therapy, which accounts for 35.7%, and high-intensity therapy, which makes up 24.8%. This distribution suggests a diverse approach to therapy types, with a balance between low and high intensity.

For the last nine years, guided self-help (books) has been the most used low-intensity therapy, representing 47.1% of all treatments provided. The "Other Low Intensity" therapy is the second most utilised, followed by Guided Self-Help (Computer) and Psychoeducational peer support, accounting for 23.7%,11.1%, and 10.4% of the therapy offered, respectively. These four types of therapies comprise 92.3% of the treatment delivered from the low-intensity category.

In 2020/2021, however, there were some changes, such as the number of psychological peer support therapies decreasing while Guided self-help (Computer) increased, indicating a shift in therapy delivery due to the impact of the COVID-19 pandemic. There was a rise in digital forms of therapy delivery and a transition to online platforms, with a decrease in in-person interaction.

Regarding High-Intensity Therapy, from 2015/2016 to 2022/2023, the NHS primarily offered CBT as the primary form of high-intensity therapy for psychological treatment, representing 54.5% of all therapy provided. In second place is "counselling for depression", followed by "other high intensity", accounting for 19% and 11% of the therapies offered, respectively. These three types of therapies comprise 84.5 % of treatments delivered from the high-intensity category. Furthermore, a new therapy was introduced in 20220/2021: Internet-enabled therapy due to the impact of the COVID-19 pandemic.

Research Question 3

After analysing and visualising the data, information about activity measures, waiting time, demographics, employment status and outcome measures was obtained and described.

The number of referrals to NHS psychotherapy services has grown yearly, particularly in 2019/2020 and 2021/2022, indicating rising demand. However, the rate of new referrals entering treatment was slower, suggesting potential bottlenecks in service delivery. Although more people were being referred, the number of people completing therapy remains low and has not improved proportionally. This slower growth in completed treatments indicated that the NHS may struggle to meet increasing demand.

From 2012 to 2023, the NHS psychological therapy services received 15,528,446 new referrals, with 10,628,643 (68%) entering treatment. However, only 5,796,057 (54.5%) of those who started treatment completed it. This suggests that, while many begin therapy, a significant number did not finish, possibly due to issues like long wait times or inadequate support. Addressing these challenges could improve treatment outcomes.

The analysis of the past 11 years revealed that 76% of patients received their initial treatment within the recommended 28 days. Around 16.1% experienced delays, waiting between 29 to 56 days, while 3.6% waited between 57 to 90 days, and 4.3% waited over 90 days. Despite these delays, the system accommodated most patients within the advised waiting period.

Over the past eleven years, the number of referrals has constantly risen for both men and women. The total distribution over the years has been 64.4% for women and 34.3% for men. Among female referrals from 2012 to 2021, the 18-35 age group represented the largest share at 49.7%, followed by 36-64 at 41.8%. The remaining categories are 65 and over (6.5%), 16-17 (1.9%), and under 16 (0.1%). For male referrals, the 36-64 age group totalled 46.3%, closely followed by 18-35 at 46.1%. The other age groups include 65 and over (5.9%), 16-17 (1.6%), and under 16 (0.1%). Until 2017/2018, the 36-64 age group had the highest male referrals before being surpassed by 18-35.

The analysis of ethnicities from 2012 to 2023 revealed that 72% of individuals identified as White, with this group showing the most significant rise over the years. The Asian and Asian British groups comprised 4.9%, while Black and Black British individuals comprised 3%. Those identifying

as Mixed accounted for 2.4%, and other ethnic groups represented 1.7%.
Employment status was monitored for referrals who completed treatment, with data collected at the beginning and end. The percentage of employed individuals was the highest and grew over the years. Initially, employment increased gradually, but by the end, it stabilised while still rising, suggesting that therapy helps maintain employment but does not significantly increase the number of employed individuals. At the start of treatment, unemployment and job-seeking showed more significant fluctuations, but by the end, the figures stabilised, indicating the therapy may reduce variability in unemployment, though not significantly. Over the past 11 years, 55.3% were employed at the start of treatment, decreasing slightly to 53.2% by the end. Meanwhile, the "Unemployed and seeking work" group went from 10.3% to 9.3%. Overall, both categories showed a reduction by the end of the treatment period.
The primary outcome measures are recovery, reliable improvement, and reliable recovery. The UK government aims for 50% of eligible referrals to move into recovery successfully, a target met since 2017/2018. The Reliable Improvement Rate has significantly increased, peaking around 2020/2021. In contrast, the Reliable Recovery Rate has declined since 2018/2019, indicating potential performance challenges.

Research Question 4

Predictions were based on age, gender, long-term condition, and ethnicity, linked to the "Reliable Improvement Rate" and observed from 2017 to 2021.
The Linear Regression method (Step 1) was applied using training data from 2017 to 2019, and testing was done with data from 2020. The model predicted a Reliable Improvement Rate of approximately 62.28 for 2020/2021, while the actual rate was 65.69. To assess the model's accuracy, metrics such as MAE and RMSE were calculated at 3.41. An additional comparison showed that while the model could predict general trends, it failed to capture significant data changes. Specifically, it underestimated the rise in the actual rate for 2020, indicating some predictive power but also limitations. The model's assumption of a constant rate of change may not be appropriate for data with sudden shifts, as observed in 2020.
To overcome these limitations, the Random Forest Regressor was used (Step 2), as it could be a better choice for handling this type of data, which has sudden changes or non-linear relationships. This process contained five different runs to improve the model. The initial run with the data split 80/20 and a random state parameter set to 42 resulted in an MSE of 43.68 and an R^2 score of -12.25, indicating that the model performed poorly. The second run had the same data split, but the random state was changed from 42 to 83, significantly improving the model. The results were MSE of 1.36 and R^2 of 0.90. The random state change resulted in a better data split for training and testing. These two runs achieved different results, suggesting that the model's performance highly depended on the specific train-test split.
The third run had a different data split, 66/33, but the same random state as the second run. The MSE dropped to 0.96, and R^2 increased slightly to 0.91, suggesting a better fit than the second run. Moreover, the fourth run, in which data was divided into a 75/25 split and again with the same random state parameter, achieved the lowest MSE of 0.84 and the highest R^2 of 0.93, indicating the best model performance among the four runs. This model performed best, fitting the data well and providing the most accurate predictions.
For the fifth run, we applied GridSearchCV to evaluate different hyperparameter combinations for the Random Forest Regressor. This process yielded a best MSE of 2.45 and an R^2 score of approximately 0.83, indicating that the model explained 83% of the variance in the target variable.
Given the low outliers noted in the box plot and the negative cross-validation values for each Random Forest Regressor run, the data was carefully revised by excluding the gender category marked "Indeterminate." All procedures were then redone using the updated dataset.
Data cleaning significantly improved model performance. The repeated application of linear regression (Step 3) yielded a predicted value of 65.94, closely matching the actual value of 65.89, with a low MAE of 0.05 and RMSE of 0.045, indicating high prediction. The Random Forest Regressor (Step 4) also showed marked improvement, particularly in the second run (80/20, 83 split), which reached an R^2 score of 0.97 and a MSE of 0.31, indicating a good fit. This also demonstrates that data quality is crucial for machine learning. The "Indeterminate" gender category, which only contained data for one year, introduced inconsistencies and overfitting. Removing it improved model reliability and helped standardise data patterns across all categories, resulting in more consistent and accurate predictions.

Conclusion

Over the last eleven years, treatment outcomes across Psychological Therapies services in England have improved. This progress includes exceeding the 50% target of the national average proportion of patients moving to recovery at the end of their treatment. This project demonstrated the effectiveness of various treatments and developed a prediction model based on patient demographics such as age, gender, ethnicity, and long-term mental health conditions. The findings indicate a high accuracy in predicting outcomes using these factors, suggesting that small changes in clinical practice could improve patient outcomes. The study shows that social-demographic variables can be effectively used in machine learning to forecast therapy outcomes, potentially benefiting non-responders and those at risk of dropout or relapse by offering alternative treatments.

Although this study benefited from the newly created dataset spanning eleven years to assess changes in clinical practice and patient outcomes, several limitations exist. Future research on more detailed data is recommended to include more social-demographic variables in designing predictive models. For example, specific types of interventions or sub-types of treatments, which are associated with outcome changes, should be included. Lastly, there is potential for more advanced methods, such as gradient boosting machines (GBM), support vector regression (SVR), and ensemble methods, to be applied in building predictive models, setting the foundation for more robust research in the future.

References

Baker, A., Simon, N., Keshaviah, A., Farabaugh, A., Deckersbach, T., Worthington, J.J., Hoge, E., Fava, M., Pollack, M.P., 2019. Anxiety Symptoms Questionnaire (ASQ): development and validation. General psychiatry 32.

Bennemann, B., Schwartz, B., Giesemann, J., Lutz, W., 2022. Predicting patients who will drop out of outpatient psychotherapy using machine learning algorithms. The British Journal of Psychiatry 220, 192–201.

Chekroud, A.M., Bondar, J., Delgadillo, J., Doherty, G., Wasil, A., Fokkema, M., Cohen, Z., Belgrave, D., DeRubeis, R., Iniesta, R., others, 2021. The promise of machine learning in predicting treatment outcomes in psychiatry. World Psychiatry 20, 154–170.

Chen, Y., Stewart, J.W., Ge, J., Cheng, B., Chekroud, A., Hellerstein, D.J., 2023. Personalized symptom clusters that predict depression treatment outcomes: A replication of machine learning methods. Journal of Affective Disorders Reports 11, 100470.

England, N.H.S., 2019. NHS mental health implementation plan 2019/20–2023/24. England: NHS England 57.

Ewbank, M.P., Cummins, R., Tablan, V., Bateup, S., Catarino, A., Martin, A.J., Blackwell, A.D., 2020. Quantifying the association between psychotherapy content and clinical outcomes using deep learning. JAMA psychiatry 77, 35–43.

Green, S.A., Honeybourne, E., Chalkley, S.R., Poots, A.J., Woodcock, T., Price, G., Bell, D., Green, J., 2015. A retrospective observational analysis to identify patient and treatment-related predictors of outcomes in a community mental health programme. BMJ open 5, e006103.

Krijnen-de Bruin, E., Scholten, W., Muntingh, A., Maarsingh, O., van Meijel, B., van Straten, A., Batelaan, N., 2022. Psychological interventions to prevent relapse in anxiety and depression: A systematic review and meta-analysis. PloS one 17, e0272200.

Kumar, P., Chauhan, R., Stephan, T., Shankar, A., Thakur, S., 2021. A machine learning implementation for mental health care. Application: smart watch for depression detection, in: 2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence). IEEE, pp. 568–574.

Lorenzo-Luaces, L., DeRubeis, R.J., Van Straten, A., Tiemens, B., 2017. A prognostic index (PI) as a moderator of outcomes in the treatment of depression: A proof of concept combining multiple variables to inform risk-stratified stepped care models. Journal of affective disorders 213, 78–85.

Lorimer, B., Kellett, S., Giesemann, J., Lutz, W., Delgadillo, J., 2024. An investigation of treatment return after psychological therapy for depression and anxiety. Behavioural and cognitive psychotherapy 52, 149–162.

Oparina, E., Krekel, C., Srisuma, S., 2024. Talking Therapy: Impacts of a Nationwide Mental Health Service in England.

Organization, W.H., 2022. WHO guidelines on mental health at work. World Health Organization.

Su, C., Xu, Z., Pathak, J., Wang, F., 2020. Deep learning in mental health outcome research: a scoping review. Translational Psychiatry 10, 116.

Vieira, S., Liang, X., Guiomar, R., Mechelli, A., 2022. Can we predict who will benefit from cognitive-behavioural therapy? A systematic review and meta-analysis of machine learning studies. Clinical Psychology Review 97, 102193.

WHO, 2022. Mental disorders [WWW Document]. Mental disorders. URL https://www.who.int/news-room/fact-sheets/detail/mental-disorders (accessed 4.15.24).

WHO, 2016. Investing in treatment for depression and anxiety leads to fourfold return [WWW Document]. URL https://www.who.int/news/item/13-04-2016-investing-in-treatment-for-depression-and-anxiety-leads-to-fourfold-return (accessed 4.17.24).