DATA-599 CAPSTONE PROJECT

Greenspace Team 3 - Therapeutic Alliance

PROJECT REPORT

Kohsin Chen
Bingshen Yang
Zerui Zhang
Zheng Zhang

June 25, 2024

Disclosure: Much of the information is considered confidential by Greenspace, so even though analyses were conducted, the results cannot be included in this report.

1. Executive Summary

Our capstone project, conducted in partnership with Greenspace Mental Health, a leader in measurement-based care, addresses the challenge of predicting and enhancing the therapeutic alliance between patients and therapists. Using advanced data science techniques, we aimed to refine the way patients and therapists are matched, ensuring a more effective therapeutic process. Our collaboration has leveraged their extensive experience and our analytical expertise from the Master of Data Science program at UBCO. Each team member brought a unique perspective to the project, informed by a rigorous academic and practical background in data science.

We began by exploring various statistical methods including supervised and unsupervised learning models. However, these traditional techniques initially struggled with the complex nature of our data. To improve our model's performance, we cleaned and organized the dataset comprising patients' historical mental test scores and treatment improvement scores, providing a robust foundation for predicting therapeutic outcomes. Building on this, we transitioned to more sophisticated models—deep neural networks (DNNs) and convolutional neural networks (CNNs), along with hybrid recommendation systems. These advanced approaches significantly improved the accuracy with which therapists can be recommended based on patient data and suggest optimal therapist-patient matches. This work demonstrates the potential of machine learning in mental health care, providing a pathway to personalized and effective treatment strategies.

In closing, our project has provided us with valuable insights into the potential of machine learning to support therapeutic relationships. Collaborating with Greenspace Mental Health, we aimed to explore the practical applications of data science in improving therapist-patient matches. While we recognize that our work is an initial step in a much broader field of research, we are hopeful that it will contribute to ongoing discussions and developments within the mental health community.

2. Introduction

In our capstone project at Greenspace Health, we aimed to address the critical elements of therapeutic alliances within mental health treatments. Building on the foundational literature which recognizes the Therapeutic Alliance (TA) as a significant predictor of treatment outcomes (Ardito & Rabellino, 2011), our project sought to enhance this alliance through predictive modeling and advanced data analytics.

Utilizing the comprehensive Measurement-Based Care (MBC) platform developed by Greenspace Health, we conducted extensive data analysis to identify the key predictors that influence the strength of therapeutic relationships. Our approach involved rigorous Exploratory Data Analysis (EDA) to understand the nuances within our dataset and employed a range of machine learning techniques including regression models, Random Forest, Neural Networks, and recommendation system to investigate and predict therapeutic alliance. Throughout the project, we developed two main components:

1. **Predictive Model:** Focused on identifying variables that could predict the TA score between patients and therapists.
2. **Therapist Recommendation System:** Aimed at matching patients with therapists based on similarity measurement and therapist effectiveness.

The core problem our project addressed was the variability in Therapeutic Alliance, which significantly impacts the effectiveness of psychological treatments. We examined the complexities involved in establishing and maintaining a robust TA, considering both intrinsic patient factors and therapist attributes. Our project was guided by the following research questions:

- What are the key predictors of a high Therapeutic Alliance score?
- How can we effectively match patients with therapists to optimize therapeutic outcomes?

Through our research, we have provided insights into the factors that enhance therapeutic relationships and developed models that can potentially improve the matching process of therapists and patients, thus fostering better treatment outcomes.

3. Background

Therapeutic alliance, defined as the collaborative and affective bond between a therapist and a patient, is a critical determinant in the success of psychological treatments. Extensive research has demonstrated that a strong therapeutic alliance significantly correlates with positive clinical outcomes across various psychotherapies. Ardito and Rabellino's study underscores the therapeutic alliance as a consistent predictor of successful treatment outcomes, emphasizing the importance of the relationship quality in influencing patient recovery trajectories [1]. Additionally, Martin et al. have explored various assessment methods for evaluating the quality of therapeutic alliances, identifying key factors such as empathy, agreement on therapy goals, and mutual collaboration as essential components of a robust therapeutic relationship [2].

Building on this foundation, recent advancements have incorporated machine learning and statistical algorithms to enhance treatment personalization. Schwartz et al. have developed models that integrate these technologies to recommend cognitive behavioral or psychodynamic therapies based on patient characteristics and predicted responses to treatment [3]. This approach provides insight that this method can also be applied to selecting suitable therapists for patients, thus optimizing treatment outcomes by tailoring therapy choices to individual profiles.

Furthermore, a randomized clinical trial by Constantino et al. tested the efficacy of matching patients to therapists based on the therapists' empirically derived strengths in treating specific mental health concerns [4]. The study revealed that such personalized matching significantly improves psychotherapy outcomes compared to traditional assignment methods. This evidence highlights the potential of data-driven approaches in refining therapist-patient matching processes to foster stronger therapeutic alliances and improve mental health care quality.

Additionally, many platforms are now utilizing advanced technologies for therapist matching, employing data-driven methods to enhance treatment outcomes [5]. These platforms typically consider various factors, including the specific needs of the patient, the professional background of the therapist, their preferred treatment methods, and their success rates in addressing particular mental health issues.

Greenspace Health has developed a comprehensive Measurement Based Care (MBC) platform tailored for mental health professionals, offering critical insights into patient outcomes and enhancing care decisions [6]. The platform’s assessments component allows patients to complete mental health condition-related questionnaires throughout their treatment. Therefore, we aim to leverage the advanced capabilities of the Greenspace Health MBC platform to build predictive models that assess and enhance the therapeutic alliance. We also hope to utilize the existing information on this platform for therapist matching.

4. Data

4.1 Dataset Description

The dataset comprises 15 tables delivered on Snowflake designed to capture the aspects of therapeutic interactions, including generalized information about psychological assessments, patient responses and evaluation made by therapists. Each table contains 260 thousand up to 20 millions rows. The datasets we used consisted of both structured and unstructured data with a primary focus on textual information. We utilized SQL queries to download the data to CSV files for further analysis.

4.1.1 Structured Data

- Patients' type (adult or child), session location (in-person or remote), suicidal tendencies, therapeutic alliance scores with their therapists, associated clinics and assigned therapists
- Patients' assessments code, submission date and score
- Therapists' associated clinics and average therapeutic alliance scores

4.1.2 Unstructured Data

- Information collected from patients regarding the actual responses provided to assessment questions
    - Patients' age, gender, occupation and other background informations
- Information based on the custom labels assigned to patients by their therapists
    - Patients' symptoms and treatment

4.2 Data Wrangling and Cleaning

We used *pandas* to structure the data into *DataFrames* for data manipulation and transformations.

4.2.1 Extract Data From Response Table

In our project, one of the key sources of data came from open-ended survey questionnaires collected by various clinics for each patient. Due to the sensitive nature of this information, which includes private and confidential responses, we are unable to share specific details of these responses here. However, it's important to note that the content of these surveys was initially believed to be highly informative for our modeling efforts, given their specific and detailed nature.

The surveys varied significantly across different clinics, each focusing on different aspects of treatment, which presented a challenge in terms of data consistency. To address this, after a thorough review and selection process, we decided to focus on extracting universally reported features that were consistently provided without missing values. Such features included age, gender, occupation, and whether the individual had a history of psychological disorders.

To address the challenge of categorizing the diverse and open-ended responses in the survey questionnaires, our team developed specific methods to systematically extract and classify key information from the data. Here's how we approached this task for different features:

- **Age**: We calculated the age of the patients by subtracting their date of birth from the current date, thus providing us with a precise numerical value that could be directly used in our models.
- **Occupation**: Keyword extraction techniques were being used. By identifying common occupational keywords in the responses, we could categorize individuals into broad occupational groups. This allowed us to consider occupation as a potential factor in the therapeutic relationship.
- **Mother Tongue and Ethnicity**: The classification of responses related to mother tongue and ethnicity involved creating a predefined mapping of keywords to categories. The responses were classified into groups such as "English", "French", or "Other" for mother tongue, and categories for ethnicity including "White", "Mixed Heritage", "Indigenous" etc. Each response was matched against this mapping to ensure consistent categorization.
- **Psychological History**: Similarly, for extracting information about mental health history, we developed a method to parse and categorize mentions of indications of suicidal tendencies or self-harm intentions.

These methods not only standardized the data for further analysis but also helped to maintain the confidentiality and privacy of the respondents by abstracting personal details into categorical data. This approach was crucial in enabling our predictive models to utilize this rich, yet sensitive, dataset effectively without compromising on ethical standards.

Challenge:

- **Age Calculation from Date of Birth**: Some calculated ages appeared implausibly high, spanning several centuries, while others were anomalously low, showing single-digit ages for respondents who were evidently employed. We even encountered negative values, indicating errors in data entry. These discrepancies necessitated rigorous data cleaning and validation steps to ensure the reliability of age-related analysis.
- **Diverse and Disordered Occupation Responses**: The occupation data presented significant diversity and complexity, with responses ranging broadly in specificity and relevance. The wide variation required us to develop a comprehensive categorization system, which involved manually creating classification groups and employing keyword extraction techniques.
- **Classification of Negative Psychological Records:** Handling the diverse psychological records was challenging due to the variety of conditions and non-standard terminologies used as well. We focused on grouping the most frequently mentioned issues, which allowed us to categorize only the most common phenomena. This approach highlighted the need for more sophisticated tools to manage the complexity and variability of mental health data effectively.

4.2.2 Extract Data From Tag Table

In efforts to enhance the predictive capabilities of our model for therapeutic alliance scores, a key focus was on extracting valuable demographic and clinical data from the extensive tags table. This data is crucial as it enriches our understanding of patient and therapist characteristics, enabling the model to identify significant predictors of therapeutic outcomes more effectively.

The tags table, which was initially designed to be highly customizable to accommodate the varying needs of different clinics and therapists, presented a complex dataset with a large volume of entries. The flexibility of the table, while beneficial for clinical use, introduced substantial challenges in data extraction for predictive modeling purposes. Many tags, due to their bespoke nature, were not directly useful for our analysis. This necessitated a meticulous process of sifting through thousands of tags to determine which were relevant to our research objectives.

To tackle the issue of relevancy and ensure accuracy in our dataset, we employed a multi-step approach to categorize the tags into two main groups: diagnosis and treatment modalities. Given the diversity in wording and terminology used across different entries, a significant effort was required to standardize and group tags effectively. We utilized keyword matching strategies, supplemented by manual review, to sort tags that explicitly mentioned key terms associated with various mental health diagnoses and recognized therapeutic approaches, such as Depression, Anxiety, CBT (Cognitive Behavioral Therapy), and DBT (Dialectical Behavior Therapy).

However, the inherent customization of the tag system meant that many tags were unique to specific clinics or therapists, leading to inconsistencies in how similar treatments or diagnoses were recorded. This variability posed a significant obstacle in creating a universally applicable model that could reliably predict outcomes based on these tags. In response to these challenges, our team developed a standardized set of criteria for tag inclusion in the analysis, focusing on those most frequently used and recognized across multiple data sources. This approach helped in mitigating the impact of less common, clinic-specific tags on the predictive accuracy of our model.

4.3 Data Manipulation

4.3.1 Identify Intake Assessments

To identify the assessments conducted when patients first join the program, known as intake assessments, we consider the first recorded data for all patients as their intake assessment score, specifically the first record of each type of assessment. Assessments categorized under category 1 are excluded because they are used to evaluate therapeutic alliance.

4.3.2 Identify Diagnosis

To further refine our analysis and identify more predictors, we calculate the frequency of each assessment given to all patients and determine the most frequently assigned assessment for each patient. This most common assessment is then used to infer the patient's diagnosis. For instance, if a patient most frequently completes the PHQ-9 assessment, they are classified as having depression. This approach allows us to utilize existing assessment data to approximate patient diagnoses, aiding in the identification of relevant predictors.

However, some patients' diagnoses are too general, such as those derived from functional scores, making it difficult to determine their specific conditions accurately, resulting in many NaN values. This method of inference is highly subjective and arbitrary, leading to a high rate of inaccuracy. Diagnoses should ideally be provided by psychiatric professionals, not inferred in this manner. Therefore, we must acknowledge this limitation before incorporating these inferred diagnoses into our analysis.

4.3.3 Assessment Completed Ratio

The assessment completed ratio is calculated by examining the assessments given to patients and determining whether they have completed the assessments (i.e. the scores are not null). This ratio helps to estimate their level of cooperation and compliance.

4.3.4 Caseload For Therapists

To calculate the caseload for each therapist, we determine the total number of patients (Total Patients) and the number of active patients (Active Patients). Total Patients is the count of distinct patients each therapist has treated. Active Patients is the count of patients who have submitted at least one assessment response since March 1, 2024. These calculations are then used to generate a report that includes the therapist's ID, total number of patients, and number of active patients.

However, the data retrieved has some inconsistencies. Some clinics never archive patients, resulting in unusually high numbers of total and active patients. Additionally, some therapist IDs have only one total patient, indicating they might not be true therapists. To address this, we filter out therapists with more than 200 active patients or with only 1 total patient.

4.3.5 Therapist Performance

To evaluate therapist performance comprehensively, we focused on two key metrics: the average improvement in patient scores and the ratio of improved patients. We began by aggregating the normalized score improvements for each patient from their first to last assessments. Using these aggregated results, we computed the average score improvement for each therapist. Additionally, we classified patient outcomes into three categories—improved, unchanged, and worsened—and calculated the count for each category per therapist. These counts were then normalized to percentages to reflect the proportion of each outcome type relative to the total number of patients seen by each therapist.

4.3.6 Patient Treatment Duration

A function sorts each group of data by response submission time, then calculates the duration of treatment. This structured data is then grouped by therapist ID, patient ID, and assessment code to provide detailed insights into the treatment outcomes.

4.4 Challenges

In the original data, we discovered that many clinics do not routinely use therapeutic alliance assessments, resulting in many therapists lacking TA scores. Only about 25% of therapists have TA scores given by their patients, limiting the availability of this important measure across the dataset.

Upon further examination, it was found that a single patient might be linked to multiple therapists. This could be due to transitions between therapists, or because some provider IDs represent intake workers or administrators rather than therapists. Additionally, the system does not indicate the specific times of these transitions. As a result, a patient's record may correspond to various providers, making it challenging to identify the primary caregiver. This ambiguity also makes it difficult to determine under whose care the patient improved and which provider the patient's TA score is referring to.

5. Proposed Solution

To address two different issues, we explore from different perspectives. This project primarily focuses on identifying predictors of therapeutic alliance and recommending therapists. Thus, the methods and results are divided into two sections.

5.1 Predictive Model for Therapeutic Alliance Score

Using the existing dataset, we aim to identify various independent variables and treat the therapeutic alliance score as the dependent variable. We analyze this using both supervised and unsupervised models to find the predictors of the therapeutic alliance score.

5.2 Personalized Therapist Matching

For therapist matching, we propose two approaches:

Assumption-Based Approach: This approach assumes that therapists with higher TA scores and better performance are better matches for patients. It does not consider real-world behaviors, as therapist assignments in practice might be random or based on other invisible factors. The focus here is on leveraging performance data and TA scores to recommend the most suitable therapists for patients.

Neural Network Models: This approach captures real-world matching behavior through a neural network model. While real-world matches may not always be the most suitable for patients, this method aims to account for the actual behavior patterns observed in therapist assignments. By doing so, it provides a more accurate reflection of how therapists are matched with patients in practice, though it acknowledges that these matches might not always be optimal.

These two approaches provide a comprehensive strategy for recommending therapists, balancing between theoretical assumptions and real-world data to enhance patient outcomes.

6. Project Methodology

6.1 Predictive Model for Therapeutic Alliance Score

6.1.1 Data

Attribute	**Description**
THERAPEUTIC_ALLIANCE_SCORE	Numeric score quantifies therapeutic relationship between a therapist and their patient
PATIENT_TYPE	Binary variable indicates whether the patient is Adult or Child
INOFFICE	Boolean value indicates whether the patient's sessions are conducted in person
SUICIDALITY_FLAG	Boolean value indicates whether the patient has suicidal tendencies
AVERAGE_INITIAL_SCORE	Numeric value calculated average patient initial assigned assessment
COMPLETED_RATIO	Numeric value calculated assessment completed ratio

DIAGNOSIS	Categorical variable classified patients diagnosis into 12 groups
Age	Integer value calculated ages between 0-100
Gender	Categorical variable classified Female, Male or Other
Occupation	Categorical variable classified into 9 groups

Table 1: Data Variables

6.1.2 T-Test

We divide the therapeutic alliance (TA) scores into high and low groups based on the median. To identify the differences between the two groups, a t-test is performed to determine if there are significant differences between the high and low TA score groups.The library used here is *Scipy*.

6.1.3 Supervised Model

Supervised learning is a machine learning technique that trains algorithms using labeled datasets, aiming to predict the outputs for new data by learning the mapping between inputs and outputs. This approach relies on a predefined set of labels that guide the algorithm in recognizing and understanding patterns and structures within the data [9].

In our project, we used different methods on the preprocessed dataset as Section 6.1.1 demonstrated and compared their results. We experimented with several models including random forest, decision tree, linear regression, ridge regression, and lasso regression, to identify which method performed best under our specific conditions. The library used for all the supervised models is *scikit-learn*. The result will be demonstrated specifically under Section 7.1.4.

6.1.4 Unsupervised Model

In this section, we preprocess the data, perform dimensionality reduction using Principal Component Analysis (PCA), and apply K-means clustering to identify patterns within the data. The library used for unsupervised models is *scikit-learn.* The steps are as follows:

- Feature Selection: We select relevant features from the dataset, including THERAPEUTIC_ALLIANCE_SCORE, COMPLETED_RATIO, DIAGNOSIS, PATIENT_TYPE, AVERAGE_INITIAL_SCORE, SUICIDALITY_FLAG, INOFFICE, Age, Gender, and Occupation.

- Categorical Data Handling: Categorical variables such as DIAGNOSIS, PATIENT_TYPE, SUICIDALITY_FLAG, INOFFICE, Gender, and Occupation are converted into numerical format using one-hot encoding. This step creates binary columns for each category, allowing these variables to be used in numerical computations.
- Data Standardization: We standardize the data using StandardScaler to ensure that each feature has a mean of zero and a standard deviation of one. Standardization is crucial for ensuring that the PCA and clustering algorithms perform optimally, as they are sensitive to the scale of the input data.
- Principal Component Analysis (PCA): We apply PCA to reduce the dimensionality of the data. By selecting two principal components, we transform the high-dimensional dataset into a two-dimensional space. This step captures the maximum variance within the data, simplifying the visualization and analysis.
- K-means Clustering: Finally, we use the K-means algorithm to cluster the standardized data into two clusters (n_clusters=2). This algorithm assigns each data point to one of the clusters based on their features, helping to uncover underlying patterns.

The combined use of PCA and K-means clustering allows us to explore and visualize complex relationships in the dataset, facilitating the identification of distinct groups within the data.

6.2 Personalized Therapist Matching

6.2.1 Recommendation System Algorithm

A hybrid of collaborative filtering and content-based filtering method [7] [8] was utilized for the recommendation system for therapist matching. Collaborative filtering considers the similarity between users, while content-based filtering takes into account features.
For each new patient, different intake assessments are assigned, which implicitly represent the patient's demographics and symptoms. By identifying patients in the database with the same assessments and similar scores, we can find patients similar to the new patient. These similar patients' therapists have varying levels of performance and therapeutic alliance (TA) scores. We assume that therapists with better performance and higher TA scores will yield better treatment outcomes for patients. Therefore, we designed two approaches to recommend therapists.

6.2.1.1 Approach 1

Algorithm Description:

- Initial Setup: the algorithm begins by setting the clinic ID and filtering patients from the same clinic in the database.
- Patient Matching Steps:
  - Criteria: Patients are matched based on patient type (adult or child), suicidality flag (true or false), and common intake assessment types.
  - Similarity Calculation:
    - Filter Patients: Include only those with the same patient type and suicidality flag as the new patient.
    - Determine Completed Assessments: Count the completed assessments for the new patient.
    - Iterate Through Assessment Levels: Use a loop to check from the current number of assessments down to fewer assessments:
      - Current Level Matching: Calculate the number of common assessments with the new patient.
      - Calculate Distance: If the number of common assessments matches the current level, calculate the weighted Euclidean distance. Divide by the number of common assessments to prioritize those with more common assessments. Add a small term to avoid division by zero.
      - Update Top Patients: Add these patients to the top patients list, ensuring they are unique.
      - Reduce Assessment Level: If fewer than 5 patients are found, reduce the number of common assessments required and repeat.
- Therapist Recommendation:
  - Identifying Therapists: Once the top 5 similar patients are identified, find their corresponding therapists.
  - Ranking Score Calculation: Multiply the therapists' TA scores by their weighted improvement rates to compute a combined score.
    - Weighted Improvement Rate Calculation:
      - Multiply the ratio of improved patients of each subset by the number of patients in that subset to get the weighted success rate for each group.
      - Sum these weighted success rates across all subsets

- Divide the total weighted success rate by the total number of patients to normalize the result, ensuring it reflects the overall weighted success rate.

By following this method, the algorithm ensures that the most suitable therapists are recommended based on a combination of similarity in patient assessments and the performance of therapists as indicated by their improvement rates and TA scores.

6.2.1.2 Approach 2

This approach was developed using Python, utilizing libraries such as *Pandas* for data manipulation and *Scikit-learn* for machine learning tasks. The cosine_similarity function was specifically utilized to compute the similarity between patients, which is essential for the recommendation engine. The methodology encompassed several key actions:

- Data Preparation: We filtered patient data by clinic ID to ensure context-specific recommendations. We selected features indicative of therapeutic outcomes, including patient types and mental health assessment scores.
- Feature Encoding and Normalization: We employed vectorization to encode categorical data, transforming it into a format suitable for modeling. Continuous variables were standardized using the StandardScaler to normalize features across different scales.
- Similarity Calculation and Therapist Effectiveness: We computed a similarity matrix using cosine similarity to identify patients with similar profiles. Therapist effectiveness was quantified based on historical patient outcome improvements.
- Recommendation Engine: The system recommends therapists by combining similarity scores with effectiveness ratings, prioritizing therapists who not only match the patient's profile but also have a proven track record.

6.2.2 Neural Networks

Neural networks were chosen for this recommendation task due to their ability to handle high-dimensional data and learn complex, nonlinear relationships inherent in the features of patients and therapists. The multidimensional nature of the data, comprising various patient and therapist features, makes neural networks particularly suitable as they excel in capturing interactions between multiple variables in large datasets.

To ensure reproducibility, the analysis follows precise steps in Python using scikit-learn, keras and tensorflow library. Data preparation involves clear categorization, encoding of variables, and standardization. Model configuration is detailed, including neural network layers, activation functions, and optimizer choices. The training process specifies train-validation splits, batch sizes, and early stopping criteria. These thorough measures enable consistent replication of the training and evaluation of the model.

6.2.2.1 Dense Neural Networks (DNN)

DNNs are used to process encoded patient and therapist data, transforming categorical data into a format suitable for deep neural network analysis. The process begins with the systematic categorization and encoding of input data, ensuring that all categorical information, such as patient types or therapy outcomes, is converted into numerical codes for neural network processing.

Dense layers are then used to thoroughly analyze these encoded inputs. Each neuron in these layers processes the entire data vector, enabling the network to capture complex and abstract patterns across the entire dataset. This comprehensive data processing is ideal for datasets where relationships and dependencies between features are intricate, as is often the case in healthcare settings.

To avoid the risk of overfitting associated with high-dimensional data, dropout regularization is incorporated within the dense layers. This technique randomly deactivates a percentage of neurons during training iterations, promoting model robustness and preventing the network from becoming overly dependent on particular features or patterns.

At the output stage, a softmax layer is employed to convert the logits from the dense layers into probabilities for each class. This probabilistic output is then refined using a custom filtering mechanism that adjusts probabilities based on the clinic affiliations of patients and therapists. It specifically zeroes out probabilities when the clinic of the patient and therapist do not match and normalizes these probabilities to ensure that predictions favor scenarios where both parties are affiliated with the same clinic.

6.2.2.2 Convolutional Neural Network (CNN)

In terms of implementation, unlike DNNs which rely on multiple dense layers, CNNs apply convolutional layers to focus on localized feature extraction. These convolutional layers analyze input data to identify local patterns and relationships, effectively capturing spatial hierarchies within the data. This localized processing is particularly beneficial for detecting subtle changes and patterns that may be overlooked by the global processing approach of dense layers. In addition, pooling layers in CNNs reduces dimensionality and compresses information, preserving only the most significant features, which helps in reducing computational complexity and enhancing model efficiency. This focus on local and hierarchical feature extraction allows CNNs to be effective for datasets with large and structured information.

7. Results and Analysis

In this part, different models may include different datasets. When not specified, the dataset used will be the one described in section 6.1.1.

7.1 Predictive Model for Therapeutic Alliance Score

7.1.1 Exploratory Data Analysis

The descriptive statistics for the Therapeutic Alliance Score are as follows:

Statistics	Mean	SD	Min	Q1	Median	Q3	Max
Score	81.97	15.67	0	73.44	85	94.5	100

Table 2: TA Score Statistics

These statistics and the histogram together highlight that while the therapeutic alliance scores are generally high across the dataset, there is a significant number of patients with scores clustered towards the upper end of the scale. This skewness might suggest a generally positive therapeutic engagement within the population studied, with a few outliers on the lower end.

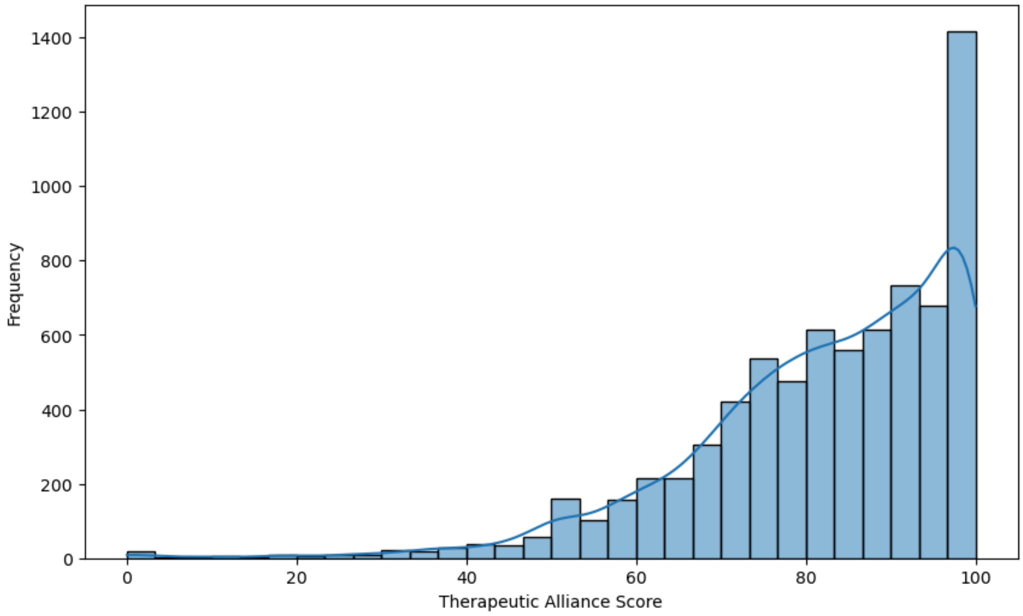


Figure 1: Distribution of TA Score

7.1.2 Correlation Exploration

7.1.2.1 Correlation between TA score and intake (initial) normalized score

The correlation coefficient between TA score and initial assessment normalized score is 0.138042. When exploring the relationship between average initial score and TA score, the correlation coefficient is 0.109899. Both scatter plots indicate that there is a weak positive correlation between therapeutic alliance scores and the types of scores analyzed (initial normalized assessment scores and average initial normalized scores). Despite this positive correlation, the relationship is not strong, implying that other factors might be more significant in determining therapeutic alliance scores.

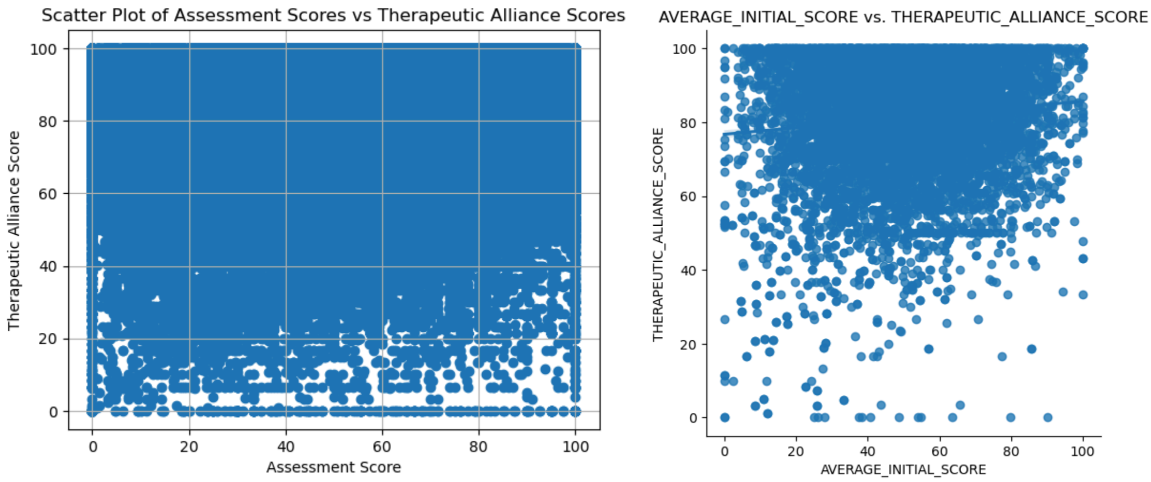


Figure 2: Scatter Plot of Initial Assessment Score/Average Initial Score V.S. TA Score

7.1.2.2 Correlation between TA score and complete ratio

The scatter plot and the weak positive correlation coefficient (0.171197) indicate that there is a minimal relationship between the completed ratio of assessments and TA score. While higher completed ratios may slightly correlate with higher TA score, the correlation is not strong enough to be conclusive.

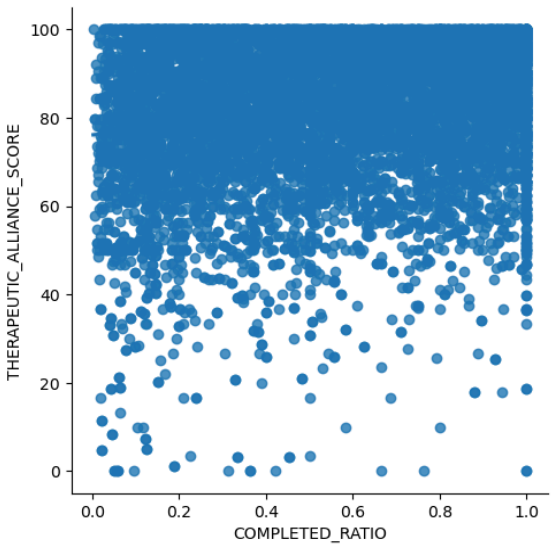


Figure 3: Scatter Plot of Complete Ratio V.S. TA Score

7.1.2.3 Correlation between TA score and therapist performance

The scatter plot reveals a weak correlation of 0.099 between therapists' average score improvements and their therapeutic alliance scores, suggesting that higher alliance scores do not strongly predict better patient improvements. Additionally, the correlation between the ratio of improved patients and therapeutic alliance scores is also low at 0.128.

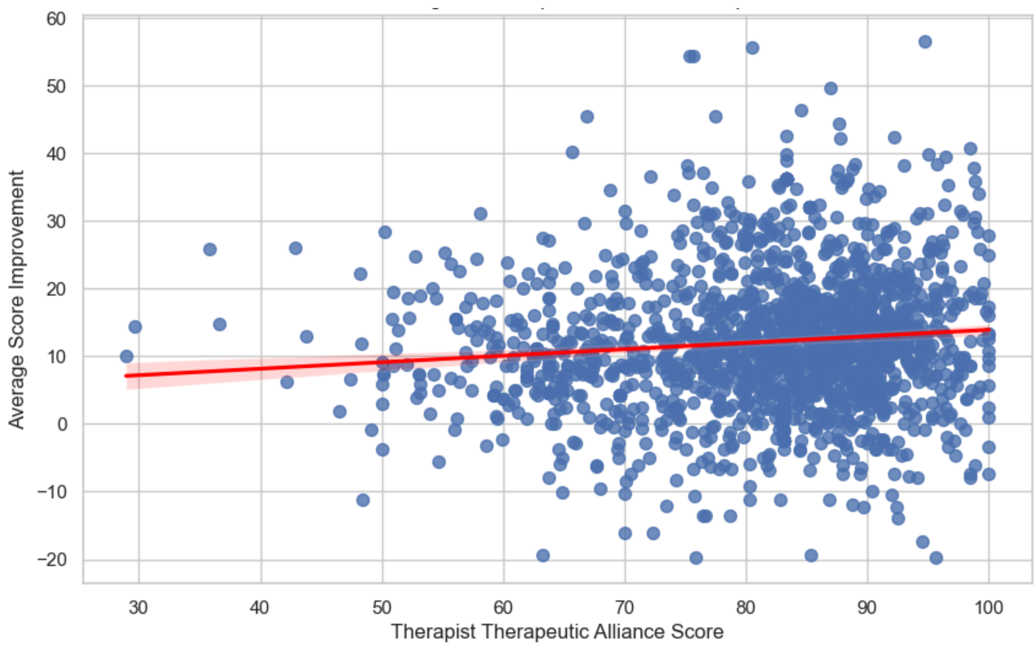


Figure 4: Scatter Plot of Average Score Improvement V.S. TA Score

7.1.3 Difference Between two groups

The p-values obtained from the t-test indicate the significance of the differences between the high and low TA score groups for each variable. The table showing the p-values for each variable can be found in Appendix A. These results suggest that factors like completed ratio, gender, specific diagnoses (e.g., PTSD, Anxiety_Depression), age, and certain occupations have significant differences between patients with high and low therapeutic alliance scores. This analysis helps in understanding which factors are more likely to influence the therapeutic alliance in patients.

7.1.4 Supervised Modeling Results

7.1.4.1 Random Forest

- NaN value filled with mean or most frequent value

  MSE: 191.353

  Adjusted R Squared: 0.325

The result shows the R squared of 0.32 which is relatively good. However, given the context that NaN values are filled with mean/most frequent values, this result was considered to be biased.

- NaN value row directly being dropped

MSE: 220.015
Adjusted R Squared: 0.207

The R-squared value is 0.207, which is lower than previous results. In this approach, we eliminated every row containing any NaN values, significantly reducing the dataset size. Despite this substantial data reduction, the remaining dataset is considered accurate and reliable.

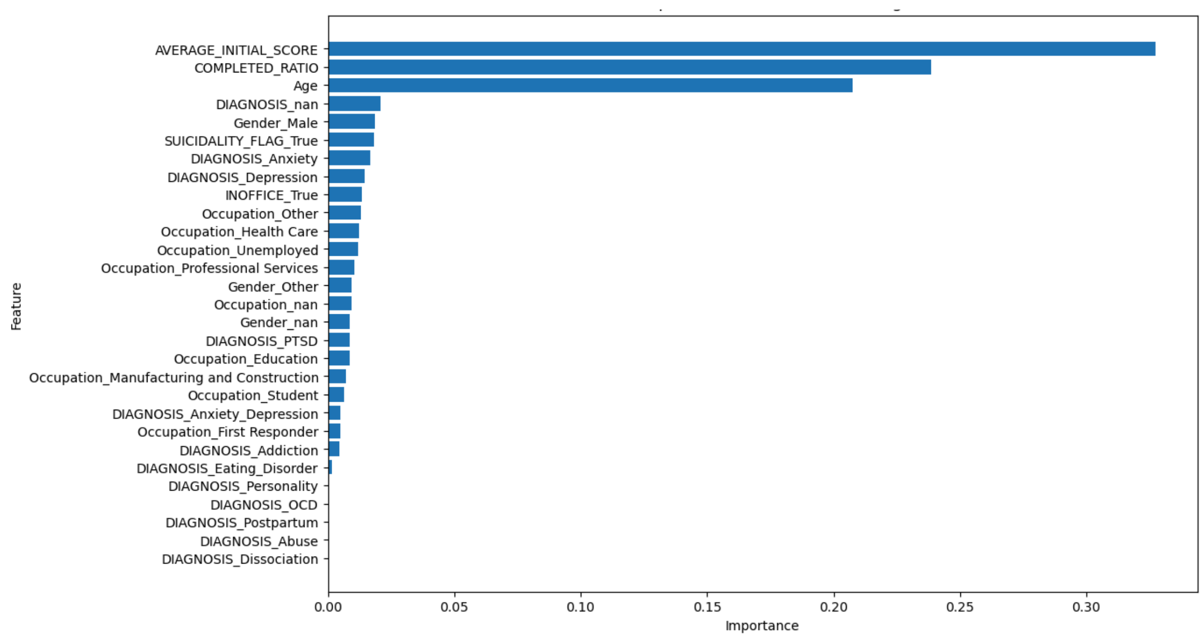


Figure 5: Feature Importances

Here is the feature importance chart produced after running the random forest model. As many features’ importance is much less compared with others, this inspired further exploration on feature selection.

7.1.4.2 Lasso + Random Forest

The combination with Lasso starts by employing Lasso’s linear model for rigorous feature selection, eliminating unimportant features, and then modeling with Random Forest. This approach can provide a more purified data input when the dataset contains a large number of redundant or irrelevant features, potentially enhancing model performance [10]. By implementing Lasso on our dataset, it selects 21 features out of 31 total features, and gives the following result:

MSE: 211.575
Adjusted R Squared: 0.178

The result here is worse than just using the random forest algorithm solely, with an adjusted R-squared to be 0.17. This could be primarily because of the excessive feature selection. Lasso utilizes L1 regularization to conduct feature selection. This process might inadvertently remove some features that are useful for predicting nonlinear relationships. If these features hold significance in the Random Forest model, removing them could lead to a loss of essential information, thereby reducing the effectiveness of the model.

7.1.4.3 Linear Regression

MSE: 250.204
Adjusted R Squared: 0.160

The linear regression analysis yielded an adjusted R-squared of 0.15. While this result is modest, it suggests that there may be a potential linear relationship within the feature being analyzed.

7.1.4.4 Ridge Regression

MSE: 285.119
Adjusted R Squared: 0.019

Compared with linear regression, ridge regression R squared has only 0.019, suggesting that the model explains a very small portion of the variance in the dependent variable, indicating a weak predictive ability. While ridge regression is particularly effective in situations where there is high multicollinearity among features if the features in the dataset do not exhibit high correlation, standard linear regression might be sufficient to handle the data, while the additional regularization imposed by ridge regression could become a burden. In such cases, linear regression, due to the absence of regularization, can freely adjust the coefficients to maximize the explanation of variance in the target variable. Conversely, the regularization in ridge regression could restrict this adjustment, potentially limiting the model’s ability to capture the essential variations in the data.

7.1.5 Unsupervised Modeling Results

7.1.5.1 Principal Component Analysis (PCA)

The Principal Component Analysis (PCA) performed on the dataset resulted in two principal components. The explained variance by each principal component is as follows: Principal Component 1: 7.08%/ Principal Component 2: 6.83%. Together, the first two principal components explain around 13.91% of the total variance in the dataset. This indicates that while these components capture some of the variability in the data, a significant portion of the variance is still unexplained by these two components.

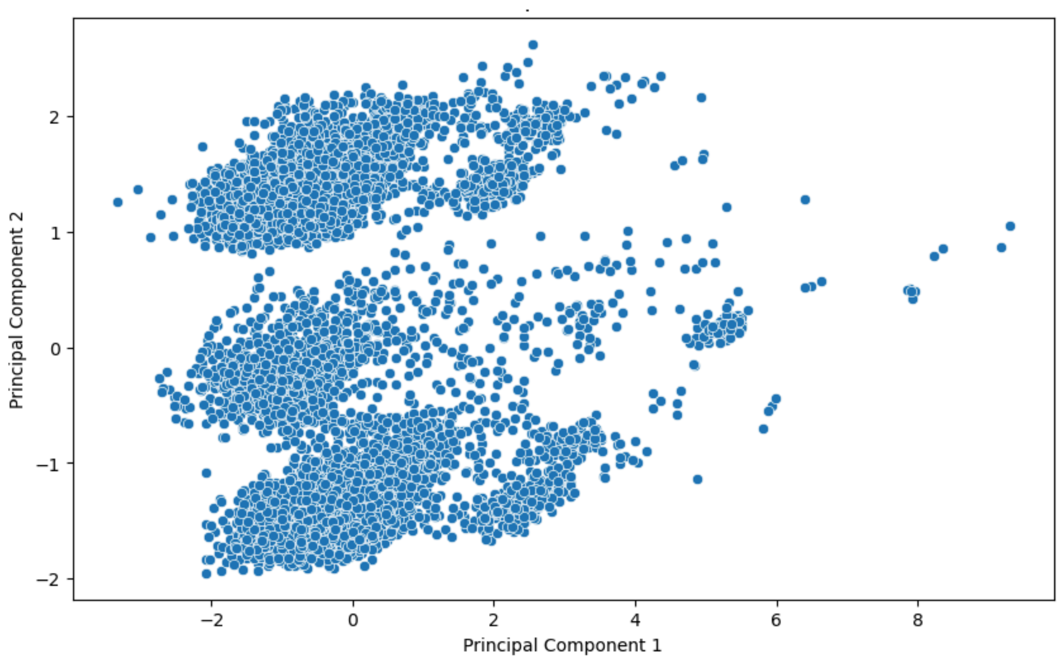


Figure 6: PCA of Therapeutic Alliance Data

7.1.5.2 K-means Clustering

Based on the cluster analysis results, the data has been grouped into two distinct clusters. Cluster 0 generally has a higher average therapeutic alliance score (82.41) and initial assessment score (54.49) compared to Cluster 1, which has an average therapeutic alliance score of 81.20 and an initial assessment score of 43.62. The average age in Cluster 0 is also slightly higher at 40.17 years compared to 38.49 years in Cluster 1. The diagnosis distribution reveals that Cluster 0 has a higher percentage of anxiety cases (43.38%) compared to Cluster 1 (38.99%), while Cluster 1 has a higher percentage of depression cases (43.89% vs. 39.44%) and PTSD cases (10.19% vs. 6.9then added to the 9%).

To help visualization, the resulting clusters are PCA-transformed data for visualization as shown below.

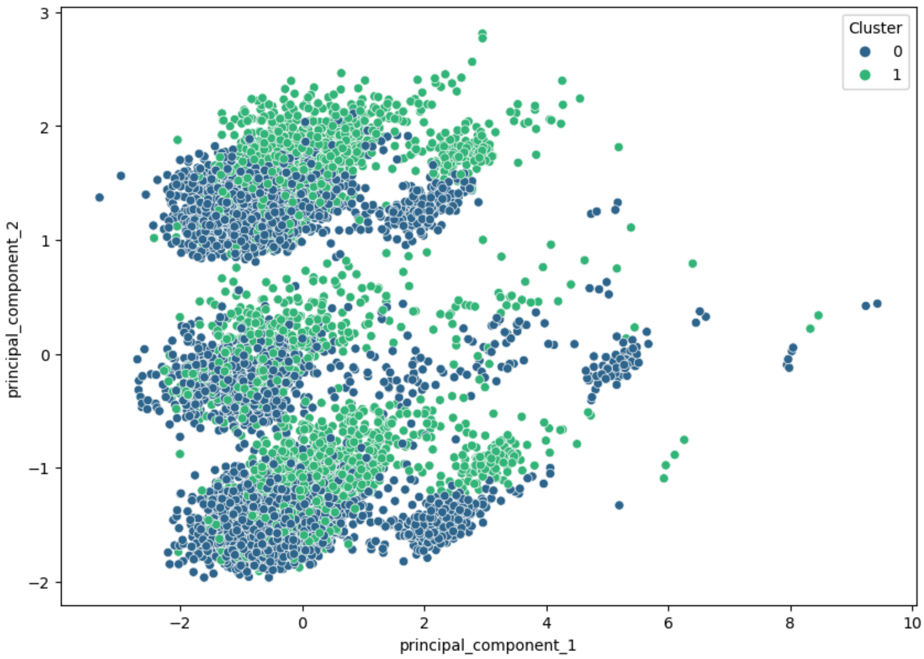


Figure 7: PCA of Therapeutic Alliance Data Custer

7.1.6 Discussion

7.1.6.1 Supervised Models

In our comparison of various supervised models, the performance of the random forest model was notably superior. We analyzed this outcome and identified the following key reasons for its success:

Robustness to Overfitting: Random forests are inherently resistant to overfitting, thanks to their mechanism of building multiple decision trees and averaging their results. This characteristic is particularly beneficial in handling our complex dataset, which has high dimensionalities and inherent variance in different assessments.

Handling of Non-linear Relationships: Unlike models such as linear or logistic regression, random forests can effectively capture nonlinear interactions between features. This capability is crucial given the intricate patterns and interactions present in our data.

Feature Importance Insights: Random forests provide valuable insights into feature importance, which has helped us identify the most significant predictors within our dataset.

This understanding has allowed us to fine-tune our model by focusing on the most impactful features.

Handling Mixed-Type Data: Random forests are adept at managing mixed-type data, including both numerical and categorical variables. This versatility ensures that the model can process and leverage all types of information available in our dataset.

These factors collectively contributed to the random forest model’s superior performance in our project, making it the preferred choice for handling our specific analytical needs.

7.1.6.2 Unsupervised Models

For unsupervised models, categorical variables are converted into multiple binary features using one-hot encoding. These features are often sparse (mostly zeros) and lack continuity between different categories, which conflicts with PCA's ability to capture continuous variations. One-hot encoding increases the dimensionality of the data, potentially making the originally high-dimensional space even sparser. This negatively impacts the effectiveness of PCA, as PCA is designed to handle continuous numerical data rather than sparse, high-dimensional data.These factors combined suggest that the underperformance of PCA in this context is likely due to the nature of categorical variables and the high dimensionality introduced by one-hot encoding.

K-means clustering is a widely used unsupervised learning algorithm that partitions data into distinct clusters based on similarity. It minimizes the within-cluster variance by iteratively assigning data points to the nearest cluster centroids and recalculating the centroids. This method is particularly effective for numerical data where the notion of "distance" between data points is well-defined. While K-means can be applied to this type of mixed data with appropriate preprocessing, it is not inherently the best fit due to distance metric limitation.K-means relies on the Euclidean distance, which may not be the most appropriate metric for high-dimensional, sparse data resulting from one-hot encoding of categorical variables.

Overall, while PCA and K-means clustering provided valuable insights, their effectiveness was limited by the nature of the data. Exploring alternative methods could enhance the analysis and lead to more meaningful and interpretable results.

7.1.6.3 Conclusion

The prediction performance for the Therapeutic Alliance (TA) score was not as strong as anticipated, largely due to the inherent complexity of TA and the limitations within our dataset. The nature of TA is multifaceted, encompassing various factors such as therapist characteristics, treatment methods, and the dynamics of the therapist-patient relationship. Each of these elements plays a crucial role in shaping the TA score, but capturing and quantifying these intricate interactions poses significant challenges.

Our dataset, while extensive, has several shortcomings that further complicate the predictive modeling. It contains numerous open-ended questions, which are difficult to standardize and analyze quantitatively. Additionally, the tags and questions were tailored by individual clinics, leading to inconsistencies across the dataset. This lack of uniformity complicates the creation of a cohesive model that accurately reflects the diverse data points.

Moreover, the dataset suffers from a significant amount of missing data, which can introduce biases and reduce the robustness of any predictive model. The absence of comprehensive therapist demographic information is another critical gap, as demographic factors can significantly influence the therapeutic process and outcomes.

Due to these complexities and data limitations, even though we experimented with various supervised and unsupervised models, the identified variables from the dataset were insufficient to yield high predictive power. Future efforts should focus on improving data collection practices, standardizing variables, and filling in critical data gaps to enhance the accuracy and reliability of TA score predictions.

7.2 Personalized Therapist Matching

7.2.1 Recommendation Systems

7.2.1.1 Approach 1

Based on the methodology described, after selecting the clinic ID and inputting the new patient's intake assessment and score, the system calculates the common assessment and Euclidean distance to identify the top 5 most similar patients within the clinic.

Next, using the therapists linked to these top 5 similar patients, the system calculates their scores based on the formula TA score multiplied by the weighted improvement rate. This calculation takes into account each therapist's therapeutic alliance score and their performance in improving patient outcomes.

The system then ranks the therapists according to these calculated scores and recommends the top 3 therapists. For each of these recommended therapists, additional details are provided, including:

- Therapist's Average TA Score: The average therapeutic alliance score for each recommended therapist is retrieved and displayed.
- Active Patients: The number of active patients currently under the care of each therapist is shown. This provides insight into the therapist's current workload.
- Total Patients: The total number of patients that each therapist has treated is also displayed, indicating the therapist's overall experience.
- Weighted Success Rate: This rate reflects the therapist's effectiveness in improving patient outcomes, adjusted for the number of patients they have treated and the types of assessments involved.
- Improvement Effects: For the new patient's specific assessment criteria, the system provides the improvement effects achieved by each therapist. This shows the percentage improvement for each assessment type based on the therapist's past performance.

By considering these factors, the system ensures that the recommended therapists are those who have demonstrated the best performance, thereby optimizing the likelihood of positive outcomes for new patients.

7.2.1.2 Approach 2

This approach utilizes cosine similarity for patient profile analysis and average score improvement to evaluate therapist performance. It is designed to be inclusive, capitalizing on available data while circumventing the limitations posed by the sparse availability of TA scores.

The choice to use average score improvements as a performance metric stems from the observation that only 25% of therapists have TA scores, with many large clinics not following TA assessment protocols. Employing average improvement scores allows for a broader inclusion of therapists and patients, thereby minimizing data exclusion and ensuring a comprehensive analysis. The underlying assumption is that these scores are indicative of therapist effectiveness, making them critical for accurate matching. This methodological shift is justified as it aligns with the data realities of the participating clinics and ensures that the model's recommendations are grounded in measurable outcomes—average score improvements.

The model outputs a list of five therapists for each patient, complete with confidence percentages that reflect the strength of each match based on similarity and effectiveness scores. Additionally, it provides information on each therapist's current caseload, allowing for an informed choice that considers availability as well as suitability.

The model's practical implications extend beyond individual matches, suggesting potential benefits in terms of improved patient outcomes, higher therapist engagement, and optimized clinic operations. This analytical approach ensures that the system is robust, equitable, and directly applicable to real-world clinical settings, where diverse data sets and operational practices often complicate straightforward analyses.

7.2.1.3 Test The Real World

The design of our algorithm makes it challenging to split data into traditional training and testing sets or perform cross-validation as there are no built-in functions available from standard Python libraries, nor are there any parameters within the model to fine-tune. In testing the effectiveness of our custom therapist-patient matching recommendation model, we conducted a manual validation using data from 20 new patients who were treated in June and had not been previously evaluated by the model. For each patient, we compared the therapist actually assigned to them against the list of therapists recommended by our system. The results were instructive: the first approach successfully matched 9 out of the 20 patients with their assigned therapists, while the second approach managed to correctly predict 5 out of the 20. These outcomes provide a direct measure of our model's predictive accuracy and offer valuable insights into its practical application and potential areas for refinement.

When digging deeper, we noticed an interesting phenomenon. In the same clinic, if the new patient intake assessment combination consists of popular assessments like GAD-7 or PHQ-9, the recommended therapists often differed from the therapists actually assigned to the patients. However, when the intake assessments involved less common combinations, such as TSQ/AUDIT/DAST-10, the recommended therapists were more likely to match the actual therapists assigned to the patients.This observation is reasonable because popular assessments like GAD-7 and PHQ-9 are widely used, resulting in a larger pool of therapists experienced with these assessments. Consequently, the recommendation system has a broader range of therapists to choose from, increasing the likelihood of recommending different therapists compared to the ones actually assigned. On the other hand, less common assessments like TSQ/AUDIT/DAST-10 are handled by a smaller, specialized group of therapists. The limited pool of therapists familiar with these assessments means the recommendation system is more likely to align with the actual assignments, as there are fewer options and a higher probability of overlap.

7.2.1.4 Comparison

	Approach 1	Approach 2
Measure similarity	Euclidean distance	Cosine similarity
How to Rank	TA score * weighted therapist performance	Average score improvement
Dataset	Only include therapists with TA score	Include all therapists
Therapist counts	Thousands	2-3 times than approach 1
Patient counts	10 Thousands	4 times than approach 1
Accuracy	9/20 = 0.45	5/20 = 0.25

Table 3: Recommendation Systems Approach Comparison

The two approaches to the recommendation algorithm follow a similar hybrid logic, combining elements of content-based and collaborative filtering to recommend therapists. Both strategies aim to identify the most similar patients and then rank their therapists based on calculated effectiveness for recommendations. The key difference between the two lies in how patient similarity and therapist effectiveness are computed. Approach 1 utilizes Euclidean distance for similarity measurement and a composite score that includes Therapeutic Alliance (TA) scores for ranking therapists, which restricts its application to a smaller dataset but achieves higher accuracy at 45%. Approach 2, employing cosine similarity and average score improvements for effectiveness ranking, accommodates a broader dataset including all therapists, thus expanding its scope to about 120K patients and 5-6K therapists, but with a reduced accuracy of 25%.

7.2.1.5 Limitation

The primary limitation of our recommendation model lies in its dependency on the availability and quality of data. For instance, Approach 1 requires Therapeutic Alliance (TA) scores, which limits its applicability to a subset of therapists and excludes a significant amount of patient data. Both approaches also assume that past therapist performance and patient improvement metrics are accurate predictors of future outcomes, which may not always hold true due to the dynamic nature of therapy and individual patient responses. Additionally, the models' reliance on specific computational methods (Euclidean distance and cosine similarity) to assess similarity might not capture all the nuances of patient-therapist compatibility. Moreover, the broader dataset used in Approach 2, while inclusive, may introduce variability that reduces accuracy. These factors underline the challenges in developing a universally effective recommendation system and highlight the need for continual model refinement and validation to improve reliability and applicability.

7.2.2 Neural Networks

7.2.2.1 Convolutional Neural Network (CNN)

The below graphs illustrate the performance of a CNN model in a therapist recommendation task, which is validated to determine whether it accurately predicts the correct therapist for a patient by learning the relationships between therapists and patients. The training loss decreases steadily while the validation loss plateaus and slightly rises after 10 epochs, suggesting overfitting. Both training and validation accuracies improve but remain low, peaking just above 0.14, indicating that the CNN is not highly effective for this classification task. For therapist recommendation, which involves understanding complex, non-spatial relationships across diverse features in order to classificate the therapists for patients, DNNs may be more suitable.

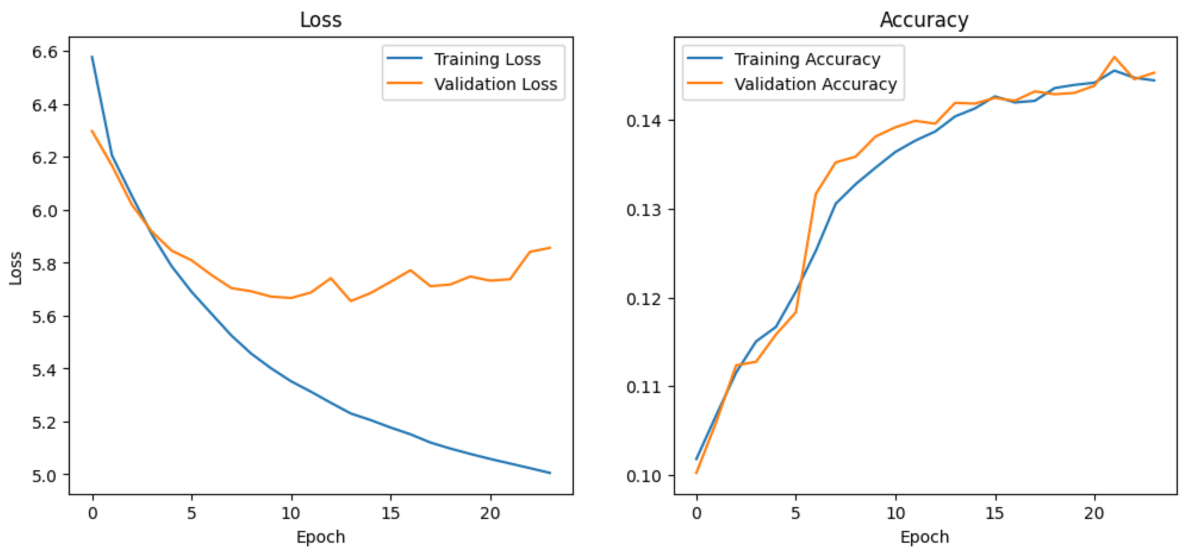
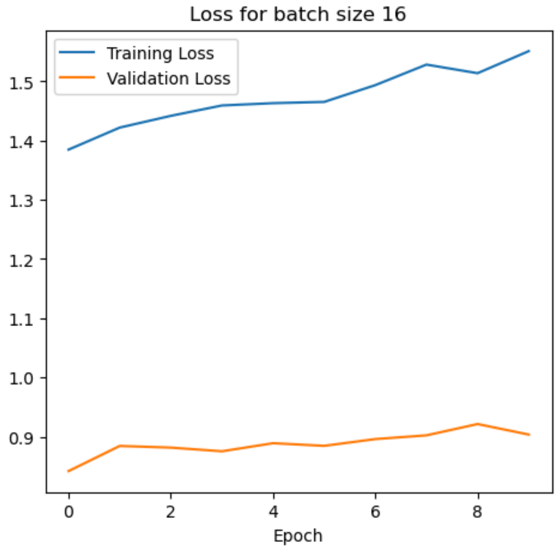
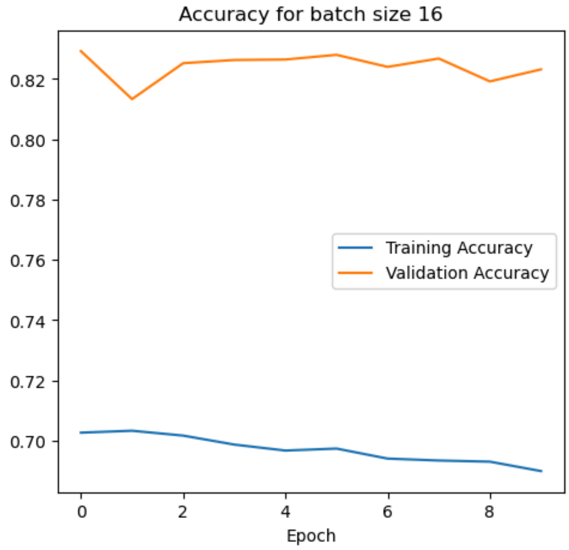


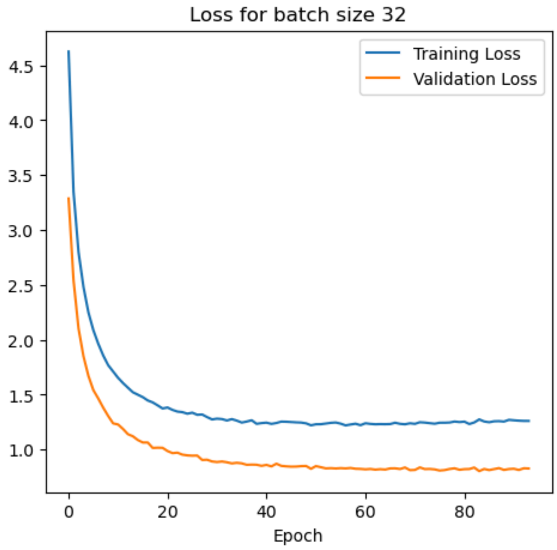
Figure 8: CNN Model Performance

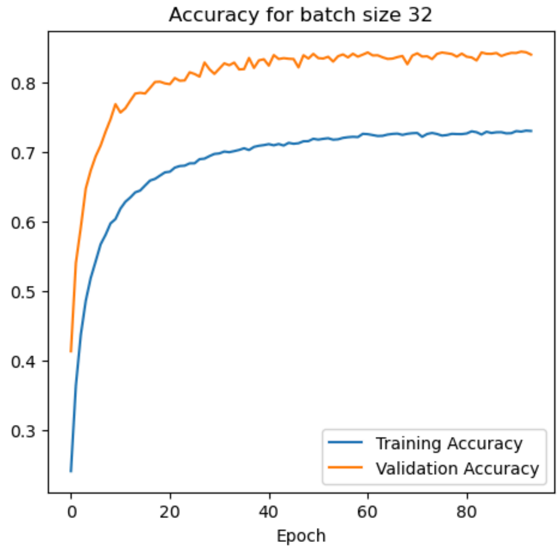
7.2.2.2 Dense Neural Networks (DNN) Fine-Tune

Based on the batch size comparisons, larger batch sizes (64) show more stable and generalized learning, evident from the smoother loss reduction and more consistent accuracy performance. In contrast, smaller batch sizes (16 and 32) exhibit higher variability and potential overfitting as indicated by fluctuations in loss and accuracy graphs.









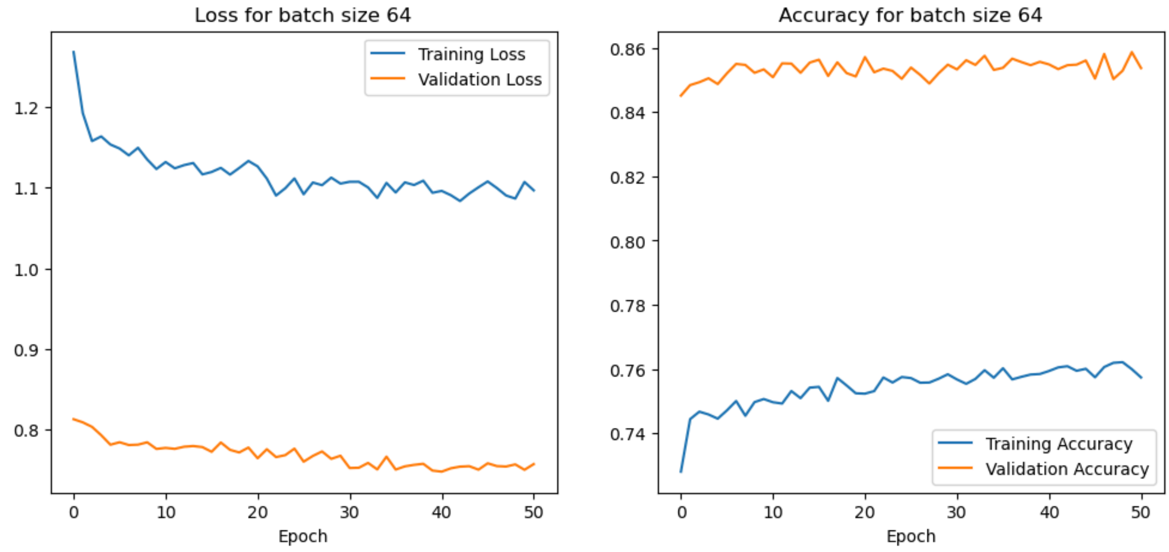
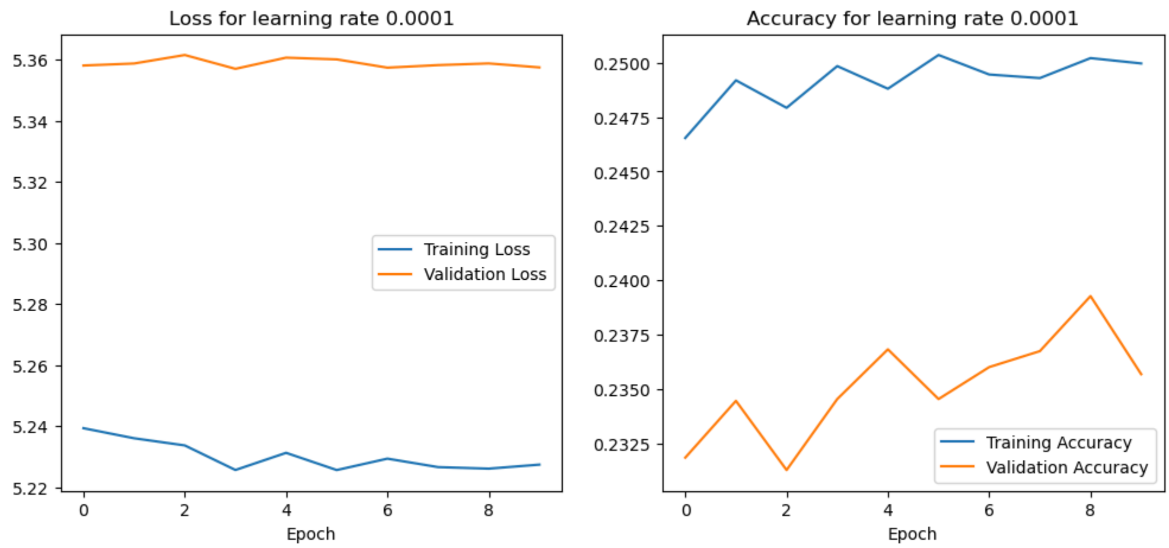


Figure 9: DNN Model Performance with Batch Sizes

Optimal learning rate selection is essential for balancing learning speed with the model's ability to generalize well on unseen data. A lower learning rate (0.0001) ensures stability but slow convergence, a moderate rate (0.001) offers a balance between convergence speed and model accuracy, while higher rates (0.01) lead to rapid learning but with significant overfitting risks.



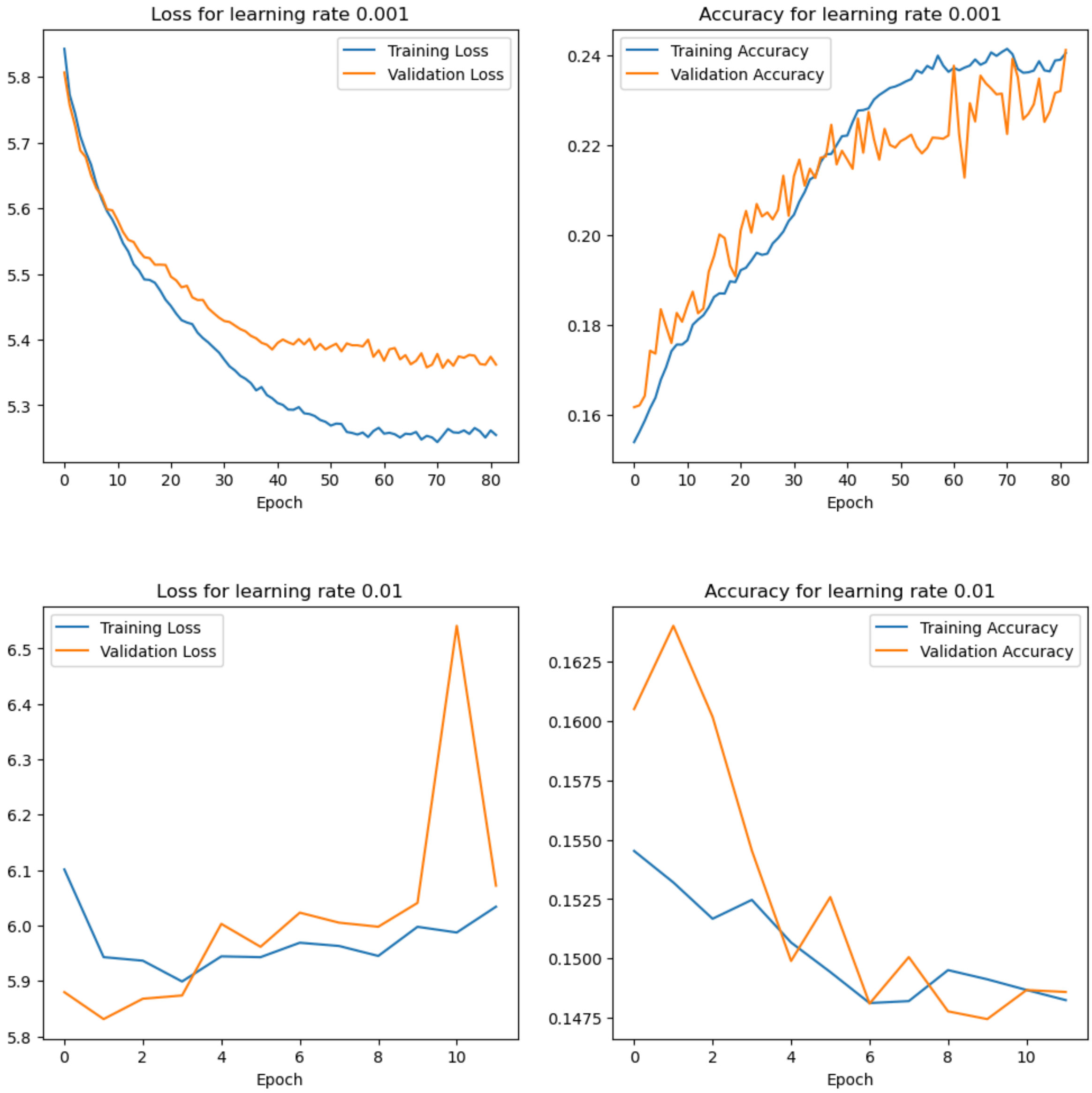


Figure 10: DNN Model Performance with Learning Rates

7.2.2.3 Dense Neural Networks (DNN)

The diagrams show the training and validation loss and accuracy of a DNN model with best parameters over 70 epochs. Both training and validation losses decrease sharply at the beginning and then stabilize. With the training loss consistently lower than the validation loss, it indicates good model convergence. The accuracy graph shows significant improvements in the early epochs, with training accuracy stabilizing around 80% and validation accuracy reaching approximately 90%. This suggests that the model generalizes well on the validation set, despite slightly lower performance during training.

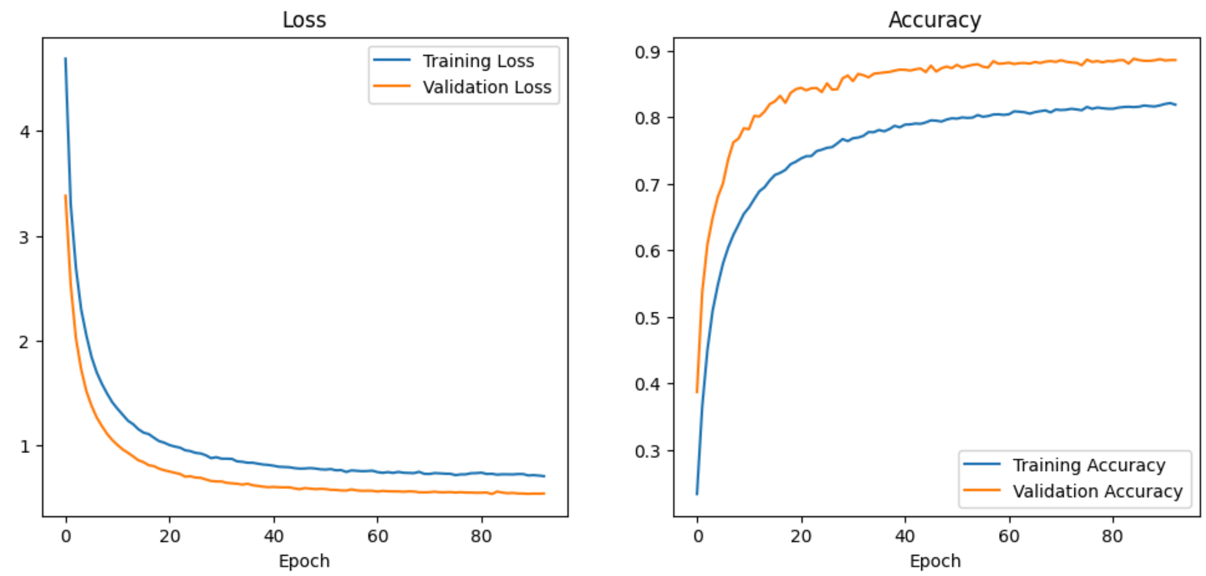


Figure 12: DNN Model Performance

Although these promising results in training and validation phases, the model's performance on completely new data is less satisfactory. When tested on 23 recent new patients, only 4 were correctly predicted by the DNN model. This substantial drop in performance on new, unseen data suggests potential issues such as overfitting to the training and validation data or the model's inability to generalize beyond the types of data seen during training.

7.2.2.4 Challenges

The challenges faced when implementing DNNs in environments characterized by class imbalances. Class imbalance occurs when certain therapist categories have significantly more patients than others, leading to a model bias towards more frequently represented classes. DNNs are highly sensitive to such disparities because they depend on extensive data sets to capture complex patterns. When data is skewed, the model may struggle to learn the diverse behaviors of therapists with fewer patients, affecting predictive accuracy. Besides that, therapists with fewer patients often receive disproportionately low matching probabilities because the DNN lacks sufficient behavioral data to create reliable probability. This can lead to underrepresentation in recommendations, reducing the likelihood of these therapists being matched with new patients, even though there is potential suitability. Another challenge is that the model might frequently recommend therapists from outside a patient's clinic if it cannot fully assess the suitability of in-clinic therapists with smaller data samples.

7.2.3 Discussion

7.2.3.1 Comparison Between Recommendation Systems and Neural Networks

	Recommendation Systems	Neural Networks
Approach	Build on match logic, assume TA score and improvement performance of each therapist serve an important role in matching	May capture real-world matching behaviors
Interpretability	Straightforward and easy to interpret	Black box - hard to interpret
Outcome	Recommend theoretically suitable therapists, but may differ from real world matching	Assigned therapists based on historical data, which may not always indicate the best match
Limitation	Hard to validate	Class imbalance

Table 4: Recommendation Systems Approach Result Comparison

Both approaches have their strengths and weaknesses. Recommendation algorithms offer transparency and straightforward interpretation but may not always capture the complexities of real-world behaviors. The DNN model, while potentially more accurate in real-world applications, is harder to interpret.

7.2.3.2 How To Choose

DNN is better suited for applications where the focus is on leveraging vast amounts of data to uncover hidden patterns and improve matching accuracy. Currently, DNN recommendations tend to be less restricted to specific clinics. Therefore, if a patient is not required to go to a specific clinic, this method might be suitable. However, this algorithm may face issues of overfitting.

Recommendation algorithms are currently limited to specific clinics because they require the input of a clinic ID before use. However, these algorithms offer greater flexibility and can be easily adjusted to meet user needs. For Approach 1, the algorithm can be modified to search across all clinics' datasets without being restricted to a specific clinic. For Approach 2, if not restricted to a specific clinic, calculating cosine similarity might require substantial time and resources, making it computationally expensive.

7.2.3.3 Societal and Ethical Considerations

Both the recommendation system and the DNN model introduce significant societal and ethical considerations. The deployment of these systems could change how therapists are evaluated and assigned, potentially impacting job satisfaction and professional dynamics. There's a risk that therapists might be pigeonholed into seeing only certain types of patients based on algorithmic assignments, which might not fully reflect their skills or professional development interests.

Ethically, handling sensitive patient and therapist data demands stringent protection to prevent breaches and unauthorized use. The opacity of the DNN model complicates transparency and accountability, necessitating audit and explanation mechanisms for its decisions. Additionally, it's important that both patients and therapists are fully informed about how their data is used and have control over whether to participate in algorithm-driven assignments, ensuring respect for their consent and autonomy. Addressing these concerns is essential for the ethical deployment of these systems in healthcare environments.

8. Conclusion

We are committed to finding predictors of therapeutic alliance. However, due to the complexity of TA and the limitations of our database and data types, even though we have tried various supervised and unsupervised models, the variables identified from the dataset are not enough to provide significant predictive power or insights into the factors influencing TA.

Regarding therapist matching, we have provided two algorithms: recommendation systems and a Deep Neural Network (DNN). Each model has its strengths and weaknesses. While recommendation systems could be a good approach to finding suitable therapists for a new patient, DNN might capture some behaviors in real-life matching scenarios.
Understanding the nuances and limitations of these models is crucial for their effective application. Future research should aim to improve data quality and explore advanced modeling techniques to enhance the accuracy and reliability of therapeutic alliance predictions and therapist matching.

9. Future Work

Future work on the recommendation system for therapist matching could focus on several key areas to enhance its accuracy, usability, and overall effectiveness:

- Conduct a Survey in the Clinic
  A survey can be administered to understand the various factors influencing therapist assignments for patients. This will include gathering insights from both therapists and patients regarding their preferences and experiences, which can inform improvements to the matching algorithm.
- Test the Recommendation System with Real-World Patient Data
  The algorithm can be tested using more actual patient data to validate its accuracy and effectiveness. This will help identify any discrepancies between the algorithm's recommendations and real-world outcomes, allowing for further refinement.
- Conduct Trials Comparing Outcomes
  We suggest conducting trials to compare patient outcomes between those who were randomly assigned therapists and those matched by the algorithm. This will involve evaluating metrics such as patient satisfaction, improvement rates, and therapeutic alliance scores.
- Enhance the Algorithm with Additional Variables
  The algorithm might be enhanced by incorporating additional variables, such as therapist specialties, patient demographics, and treatment preferences. Machine learning techniques will be employed to identify further factors that could influence successful matching.
- Cross-Clinic Comparisons
  Comparisons of the recommendation system's performance across different clinics can be conducted. This will help identify best practices and standardize successful strategies, ensuring the system's effectiveness in diverse clinical environments.

By addressing these areas, the recommendation system can be refined and enhanced to better serve patients and therapists, ultimately improving the quality of mental health care.

10. References

1. Ardito,R.B.,&Rabellino,D.(2011).Therapeuticallianceandoutcomeof psychotherapy: historical excursus, measurements, and prospects for research. Frontiers in Psychology, 2, 270.
2. Martin,D.J.,Garske,J.P.,&Davis,M.K.(2000).Relationofthetherapeutic alliance with outcome and other variables: A meta-analytic review. Journal of Consulting and Clinical Psychology, 68(3), 438.
3. Schwartz,B.,Cohen,Z.D.,Rubel,J.A.,Zimmermann,D.,Wittmann,W.W., & Lutz, W. (2021). Personalized treatment selection in routine care: Integrating machine learning and statistical algorithms to recommend cognitive behavioral or psychodynamic therapy. Psychotherapy Research, 31(1), 33-51.
4. Constantino,M.J.,Boswell,J.F.,Coyne,A.E.,Swales,T.P.,&Kraus,D.R. (2021). Effect of matching therapists to patients vs assignment as usual on adult psychotherapy outcomes: A randomized clinical trial. JAMA Psychiatry, 78(9), 960-969.
5. Valant. (n.d.). Best practices for matching new patients with your therapists. Retrieved June 21, 2024, from https://www.valant.io/resources/blog/best-practices-for-matching-new-patients-with-your-therapists/
6. GreenspaceHealth.(n.d.).Home.RetrievedMay4,2024,from https://greenspacehealth.com/en-ca/
7. LinkedIn. (n.d.). How do you choose between collaborative filtering and content-based filtering for recommendation systems?. Retrieved June 21, 2024, from https://www.linkedin.com/advice/0/how-do-you-choose-between-collaborative-content-based
8. De O. Melo, P., Prudêncio, R. B. C., & Gama, J. (2019). *Collaborative Filtering vs. Content-Based Filtering: Differences and Similarities*. arXiv. https://arxiv.org/abs/1912.08932
9. James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Learning*. Springer.
10. Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1), 267-288.

11. Appendix

Appendix A.

T-test results for comparing high TA score group vs. low TA score group

Variables	**P-value**
COMPLETED_RATIO	2.7621387442505534e-25
Gender_Male	6.59395915458162e-10
DIAGNOSIS_PTSD	1.787455541365882e-06
Occupation_Unemployed	2.1942212849703487e-05
Age	0.00025623151359061796
SUICIDALITY_FLAG_True	0.00032204099057011384
Gender_Other	0.0007475619604149788
Occupation_Health Care	0.0011150652428821694
DIAGNOSIS_Anxiety_Depression	0.0042590074226893
Occupation_Student	0.012928391347566633
Occupation_Education	0.02297643214914339
DIAGNOSIS_Addiction	0.02745008694247103
DIAGNOSIS_Eating_Disorder	0.031505281390338834
AVERAGE_INITIAL_SCORE	0.045809733598244855
INOFFICE_True	0.05706557716675558
DIAGNOSIS_Postpartum	0.08520561632282861
DIAGNOSIS_Personality_Disorder	0.10230409105299306

Occupation_Other	0.2589672246084003
DIAGNOSIS_Abuse	0.3159796658504715
DIAGNOSIS_Dissociation	0.31870627927876216
Occupation_Professional Services	0.36684429207163627
DIAGNOSIS_Depression	0.6147511294702661
DIAGNOSIS_Anxiety	0.6859545531649707
DIAGNOSIS_OCD	0.7431124373328641
Occupation_Manufacturing and Construction	0.8325403654018448

Appendix B.

Tools, Libraries and Techniques

The following are the list of tools used as part of our project:

- GitHub: https://guides.github.com/
- Snowflake: https://www.snowflake.com/en/
- Python: https://www.python.org/about/

The following are the list of Python libraries used as part of our project:

- Pandas - https://pandas.pydata.org/pandas-docs/stable/#
- Scikit-learn - https://scikit-learn.org/stable/index.html
- SciPy - https://docs.scipy.org/doc/scipy/reference/#
- Tensorflow - https://github.com/tensorflow/tensorflow
- Keras - https://github.com/keras-team/keras
- Seaborn - https://seaborn.pydata.org/
- Matplotlib - https://matplotlib.org/stable/tutorials/pyplot.html