

Greenspace Team 3 - Therapeutic Alliance

Week 3 Team Report

Kohsin Chen, Zerui Zhang, Zheng Zhang, Bingshen Yang

- Team progress compared to the project plan and milestones

According to the project plan, Kohsin, Zerui, Bingshen and Zheng are involved in dataset exploring and loading tasks, scheduled to be completed by May 12, 2024. This part of the work is progressing as planned. The discussion and setup of the model type are planned to be completed by May 15, 2024. Our team has already experimented with two types of models and identified the need for adjustments and optimizations from the results.

Specifically, for the model part, our team has tried both Random Forest Regressor and Linear Regression models to predict patients' therapeutic alliance scores (THERAPEUTIC_ALLIANCE_SCORE) with predictor normalized scores from assessments assigned to patients (SCORE_NORMALIZED). However, both models showed high Mean Squared Error (MSE) and low R-squared values, indicating worse predictive performance.

Due to the poor performance of the initial models, our team has decided to undertake the following division of tasks to further explore and optimize the model:

Kohsin is tasked with determining the main diagnosis based on the types of assessments assigned to each patient.

Zerui is responsible for identifying patient grouping information in the tags. This will help consider the characteristics of different patient groups in the model.

Bingshen and Zheng are in charge of filtering through the questions and responses to identify characteristics of patients that may impact their therapeutic alliance scores.

- The individual contributions on what they have done in previous week

Group Member	Contribution	Challenges
Kohsin	Utilized SQL queries to download assessment data and built random forest and regression models to test the relationship between initial	Currently, the calculated counts for completed assessments and missing

	<p>assessment intake scores and therapeutic alliance scores.</p> <p>After discussing with the Greenspace team, they suggested several factors we could investigate. One suggestion was to extract the counts of completed and missing assessments to see if these could predict therapeutic alliance. As a result, I was working on extracting this information.</p>	<p>assessments do not seem to match the number of assessments assigned to each person.</p>
Zerui	<p>Used random forest regressor model to investigate the initial assessment score as a predictor. Categorized and analyzed tags attached to patient records. The tags were normalized and grouped into categories for common mental health conditions. Additionally, tags were categorized based on treatment methods, such as CBT, DBT, mindfulness, and so on. Recognized tags that are unclear in meaning, which will not be included in further analysis.</p>	<p>A significant challenge was ensuring that the tags were accurately grouped despite variations in wording and terminology. We are unable to confirm the use of many tags, since they are fully customizable by clinics, and different therapists/clinics may use them in different ways.</p>
Zheng	<p>Built regression model and random forest model to predict therapeutic alliance scores using assessment scores. Utilized SQL queries to efficiently download all relevant data encompassing questions, responses, and patient information directly from the database. Classify and assign response labels to patients.</p>	<p>It's difficult to sort all potential features related to psychological assessments that should be extracted from the data beyond the basic demographic. The responses to questions are various, hard to categorize.</p>
Bingshen	<p>Developed methods to extract key patient characteristics from the dataset based on questions and responses, including identifying patients' age, gender, and</p>	<p>Identifying and categorizing questions due to the diverse ways the same question can be phrased. Different language expressions for</p>

	occupation from the responses data. Classify and assign response labels to patients.	similar questions lead to difficulties in accurately capturing all relevant data points.
--	--	--

- Team communication and collaboration

1. We schedule regular team meetings every Wednesday and Saturday, with occasional additional meetings on Thursday, to discuss progress, challenges, and next steps.
2. We share regular updates on everyone's progress during these meetings.
3. We discuss insights from our meetings with Greenspace and split our tasks accordingly.

- Clear work plan with tasks assigned to each person for the next week

Group Member	Next Week Tasks Assigned
Kohsin	Try to fix the SQL query to ensure that the sum of the completed assessment count and the missing assessment count equals the total number of assessments. Attempt to identify the main diagnosis based on the types of assessments assigned to each patient. Greenspace suggested that since identifying the main diagnosis might be challenging, we could try to determine it based on the most frequently assigned assessment.
Zerui	Split complex tags into smaller tags, ensuring each tag contains only one keyword. This process will focus on data from the three largest clinics. The grouping strategy developed will then be applied to evaluate its effectiveness. Finally, the tags will be used as predictors in building models to investigate the relationship between tags and therapeutic alliance.
Zheng	Following the suggestions from Greenspace, focus on extracting and organizing the big client's questions according to the frequency of occurrence to better understand the data.

	Identification and selection of useful questions to be integrated as features across the entire dataset with Bingshen, help to build predictive models.
Bingshen	<p>Focus on the modeling aspect of the project, especially integrating the refined features into predictive models to assess their impact on patient outcomes.</p> <p>Incorporate the newly identified and cleansed features with Zheng from the broader dataset to enhance model accuracy and reliability.</p>