

# **Greenspace Team 3 - Therapeutic Alliance**

## **Week 7 Team Report**

Kohsin Chen, Zerui Zhang, Zheng Zhang, Bingshen Yang

### **- Team progress compared to the project plan and milestones**

In our planned timeline, between June 9th and June 15th, we aimed to complete the following tasks: Model Fine Tuning II and Confirmation of Factors. However, the previous model results did not successfully identify significant predictors for the therapeutic alliance score, and the model's performance was unsatisfactory. Consequently, we had to adjust our direction and timeline starting from June 2nd.

Last week, we explored recommendation algorithms for therapist matching and experimented with various methods. We employed Euclidean and cosine similarity to estimate patient similarity and tested several approaches for ranking therapists. Based on these rankings, we proposed the top 3 recommended therapists for new patients. Additionally, we utilized neural networks to investigate whether they could capture real-world matching behaviors.

### **Results:**

Based on the similarity results, initial recommendation models were built that prioritize matching based on closest patient similarities and assigned rankings according to match logic. However, it remains uncertain whether clinics in the real world actually assign therapists to new patients using this match logic. Therefore, further testing with real-world scenarios involving new patients may be necessary to validate these approaches.

For neural networks, we tried both CNN and DNN. The best model validation accuracy for CNN (i.e., whether the therapist predicted with the highest probability by the model is actually the therapist assigned to the patient) reached 40%. However, this is likely due to the input dataset being a combination of therapist, patient, and initial assessment code, where duplicate therapist and patient entries may cause the accuracy to be artificially high. Therefore, we tried cleaning the data again, expanding the columns to ensure each row represents a unique combination of patient and therapist. But the CNN accuracy dropped to about 0.1%. On the other hand, the DNN model seemed to perform better on this set of data. Validation with the existing test set data showed that the model could achieve an accuracy of 70%. However, the performance on new data still needs to be further verified with entirely new data.

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

<b>Group Member</b>	<b>Contribution</b>	<b>Challenges</b>
Kohsin	<ol style="list-style-type: none"> <li>1. Used Euclidean distance to calculate similarity for a new patient and identified the most similar patients in the dataset and then computed score rankings for therapists associated with these patients.</li> <li>2. Final data cleaning</li> </ol>	We are uncertain if this matching logic applies to the real world.
Zerui	<ol style="list-style-type: none"> <li>1. Data cleaning and merging for the recommendation model.</li> <li>2. Built matching model using a hybrid method, finding similar patients and using score improvement to rank therapists.</li> </ol>	This algorithm is easy to interpret but may have a lower accuracy, compared to the given patient-therapist match.
Zheng	<ol style="list-style-type: none"> <li>1. Built CNN model for therapist recommendation system.</li> <li>2. Data cleaning</li> <li>3. Built DNN model for therapist recommendation system.</li> </ol>	<ol style="list-style-type: none"> <li>1. The process of prediction is too computational complex causing the kernel dead for a large amount of data.</li> <li>2. Model is harder to interpret.</li> </ol>
Bingshen	<ol style="list-style-type: none"> <li>1. Evaluation of last week hybrid recommend algorithm</li> <li>2. Built SVM model</li> </ol>	Turns out that last week high accuracy prediction result from the hybrid recommendation algorithm was a huge misleading. The

		dataset contains several assessments under same therapist-patient pair which exists in both training and test dataset, resulting the unusual accuracy.
--	--	--

**- 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	Finalize algorithms. Validated results based on real world therapist-patient matching. Start working on the final report and presentation.
Zerui	Finalize the recommendation model and prepare to submit. Work on final report and presentation.
Zheng	Test DNN model accuracy with most recent data, fine tune the final model. Prepare final report and slides
Bingshen	Finalize and fine-tune the DNN model Prepare final report and slides