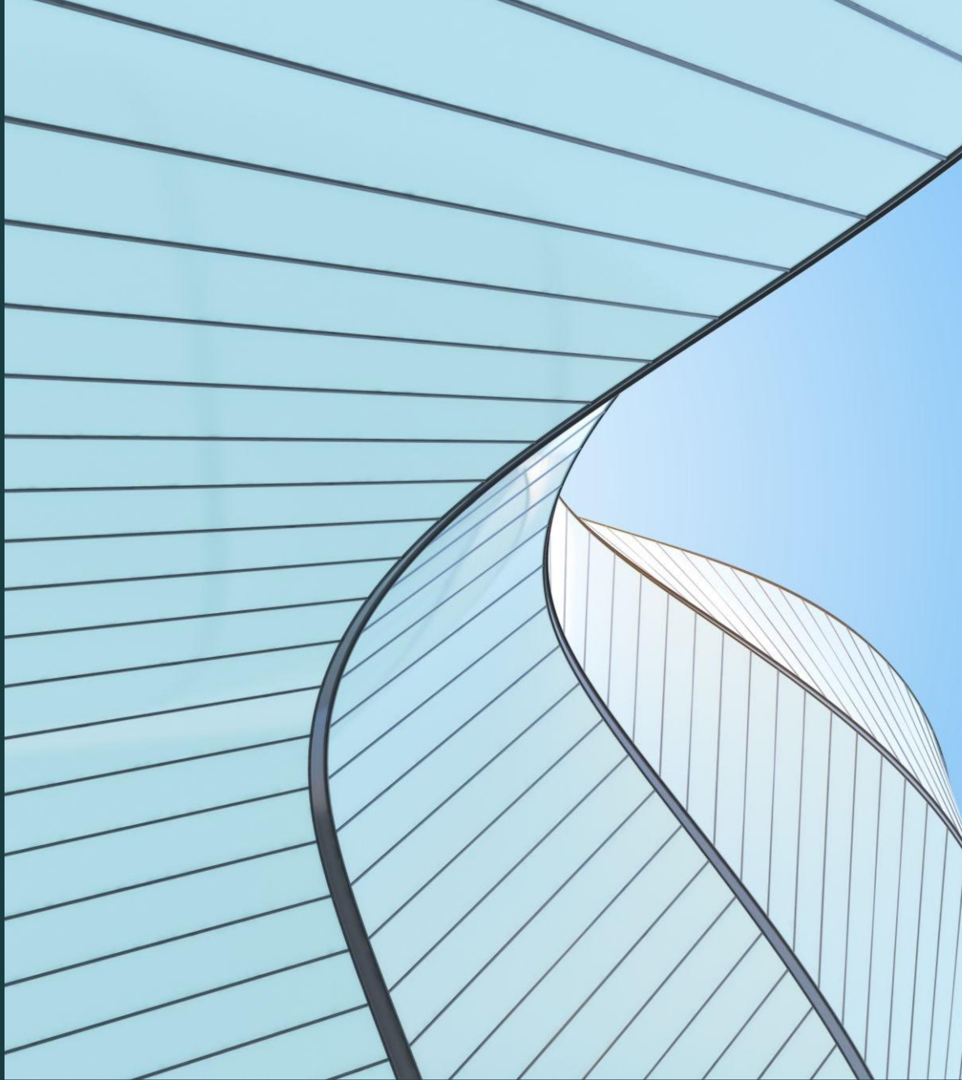
Modeling User Adherence for MyYouthspan

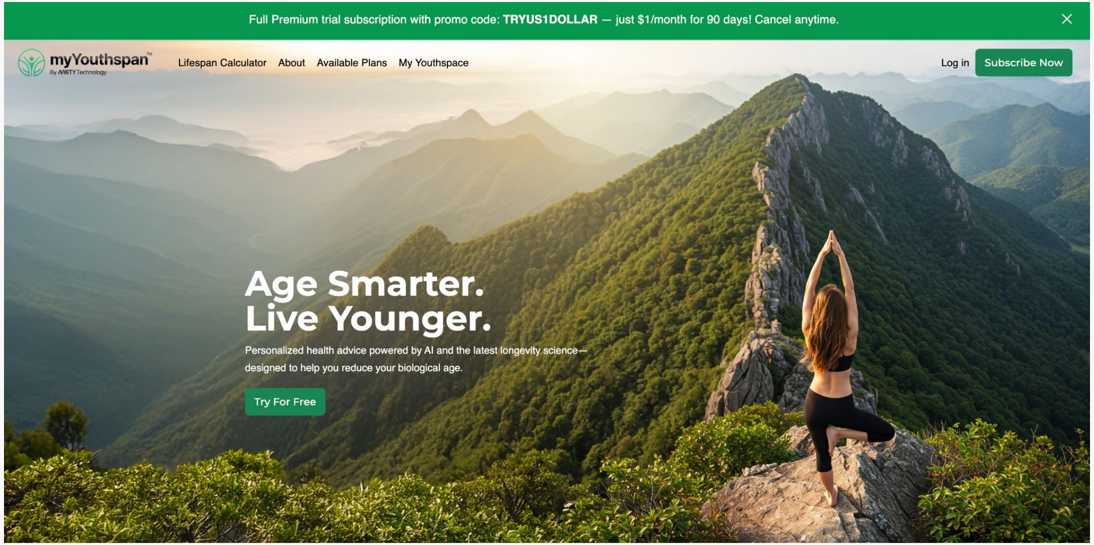
By:

Karunya Iyappan
Sam Knisely
Alec Pixton
Trenton Ribbens



Project Overview

- **MyYouthSpan**: Personalized wellness platform for tracking daily health behaviors
- **Project goal**: Model and understand adherence to user-specific lifestyle plans
- **Outcome**: Support more personalized, achievable recommendations
- **Long-term goal**: Increase user engagement and retention through tailored guidance



UVA SCHOOL of DATA SCIENCE

Our Research Questions

1. How can we define and measure user adherence?
2. What are the key predictors of user adherence?
3. Can users be segmented into behavioral personas?
4. How can modeling adherence improve platform personalization?

Overview of the Data

myYouthspan™
By METY Technology

Lifespan Calculator | My Profile | My Plan | My Log | My Progress | Resources | About | Available Plans | My Youthspace | Karunya | Logout

Make Your Plan

Interventions with * are still being analyzed by our team.

Interventions with ! may provide additional benefits when exceeding the optimal amounts, however these additional benefits are small.

Diet

		Current	Optimal	New
Processed Meat	Servings / Week	1	0	
Red Meat	Servings / Week	0	0	
Poultry	Servings / Week	1	0	
Fish*	Servings / Week	1		

Benefits	Current	Optimal	New
Lifespan (in years)	88.13	104.32	88.13
Cancer	-38%	-53%	-38%
Type 2 Diabetes	+40%	-49%	+40%
Cardiovascular Disease	-47%	-80%	-47%
Stroke	-53%	-87%	-53%
Depression	0%	-28%	0%

Note: Values represent the percentage change. For example, +25% means the risk is 25% higher than the average, and -25% means the risk is 25% lower than the average.

Refresh

Initial exploration into behaviors: PCA Analysis

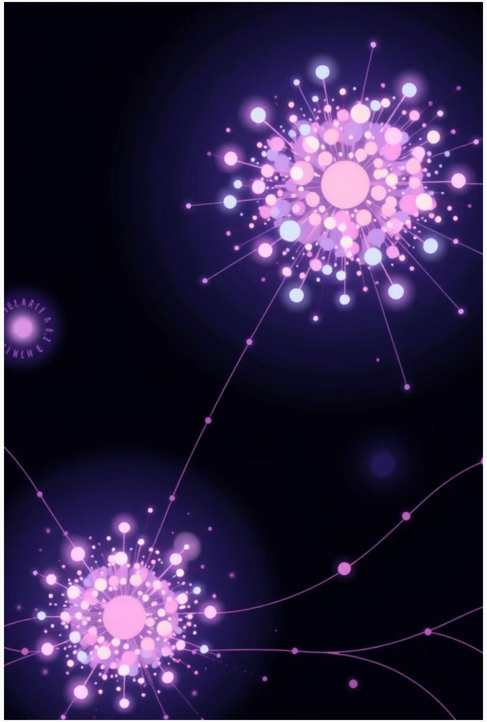
PC1	High adherence to fitness, nutrition, and social goals; indicates general engagement with lifestyle plans
PC2	Variation in consistency of logging activity
PC3	Differentiates users who exceed goals from those who just meet them

PCA confirmed the existence of distinct user types, providing a clear foundation for clustering.

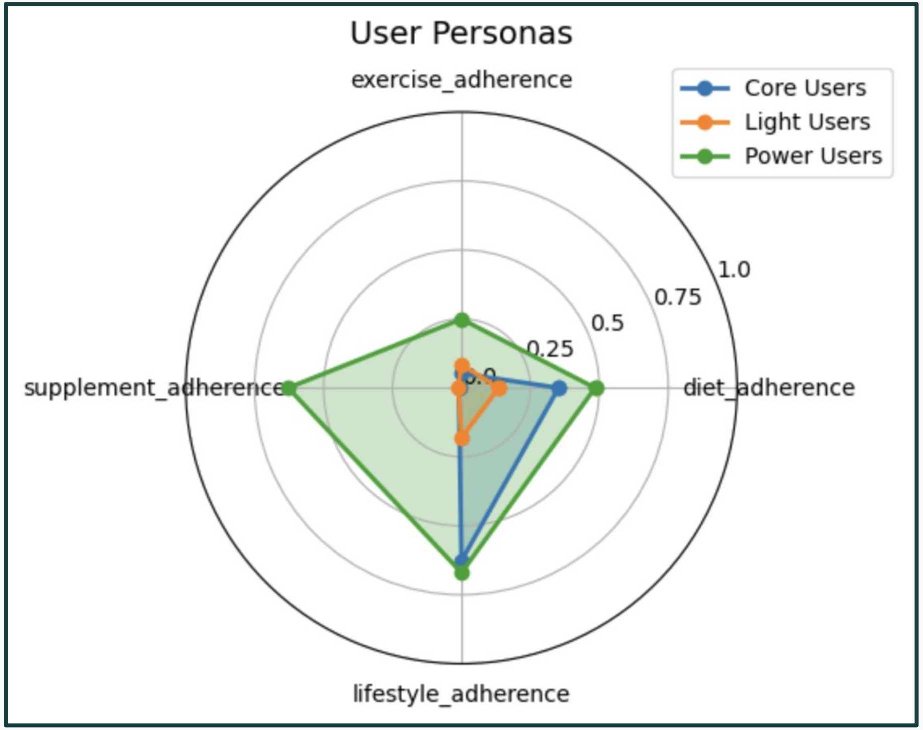
Clustering Analysis Using K-Means

- **Methodology**
    - Grouped variables into Diet, Lifestyle, Exercise, and Supplements
    - Defined adherence for each group of variables
    - Aggregated by user and used these four variables to perform clustering along with number of log entries

- **Results**
    - 3 distinct clusters
    - Best results with only group adherence variables
    - Silhouette score = **0.314**



Clustering Analysis Results



Light User

- Overall Adherence = **0.103**
- Row Count = **13.2**
- User Count = **20**

Core User

- Overall Adherence = **0.260**
- Row Count = **28.1**
- User Count = **14**

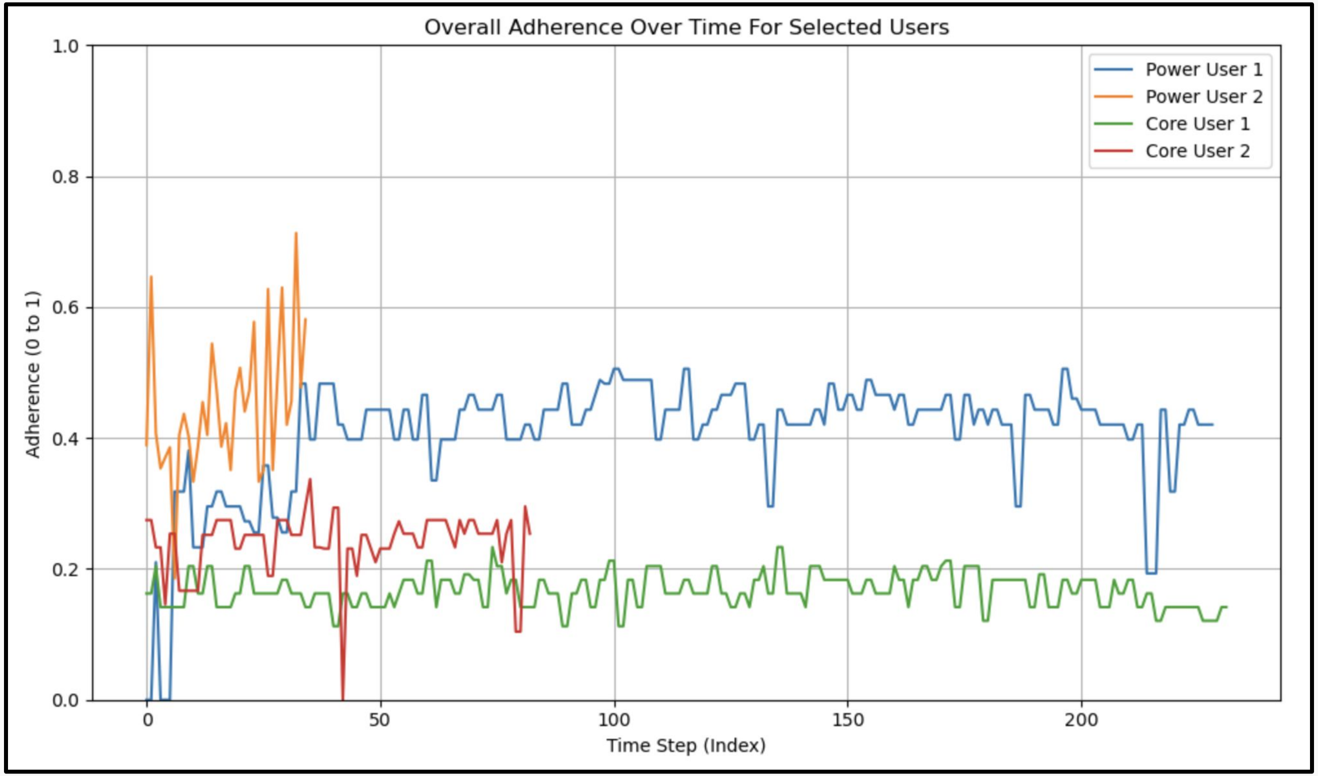
Power User

- Overall Adherence = **0.508**
- Row Count = **29.5**
- User Count = **10**

Icons by Gamma

UVA
SCHOOL of DATA SCIENCE

Adherence Over Time of Selected Users

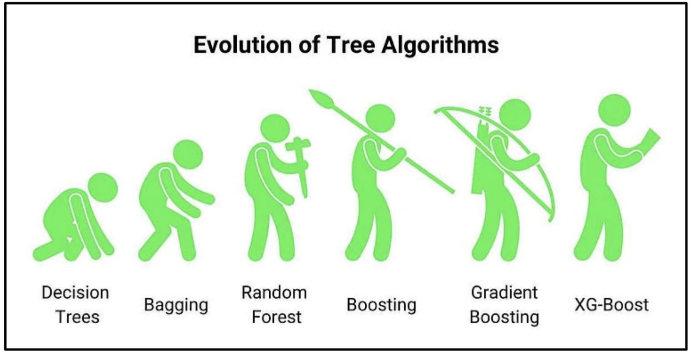


UVA
SCHOOL of DATA SCIENCE

XGBoost Regression Model of Adherence

- **Methodology**
  - Defined adherence as the proportion of goals achieved in the log data over goals set in the plan data.
  - Only variables that have non-null/non-zero values in the plan data are utilized, and adherence is updated directionally if going below the plan is beneficial (i.e., drinking less alcohol)
  - Three models tested: all parameters, median importance pruned model, SHAP-pruned model

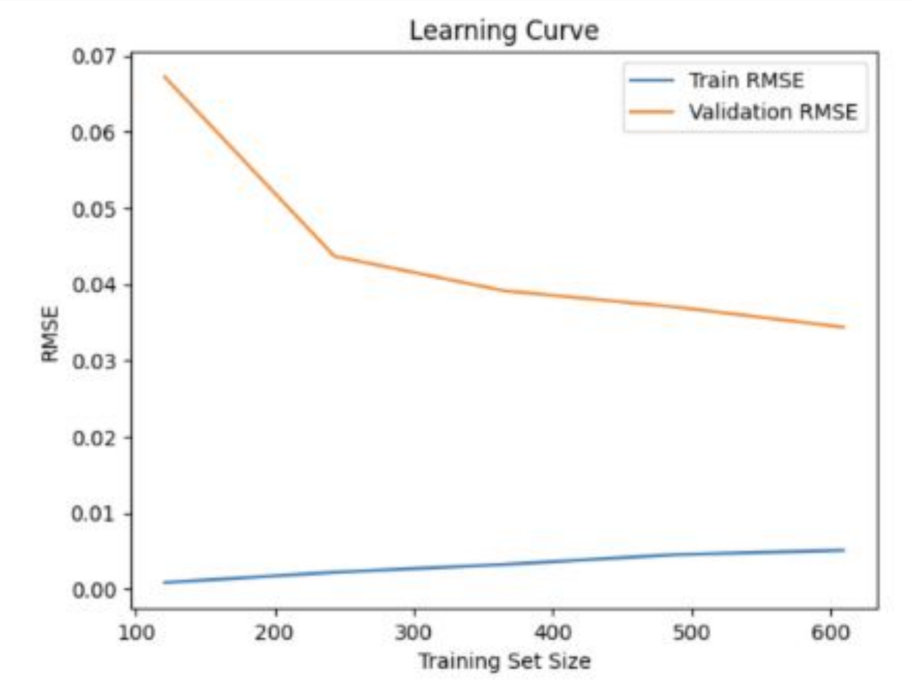
- **Results**
  - Top performing model included all parameters
  - 5-fold cross-validation during training showed model stability
  - No significant signs of overfitting

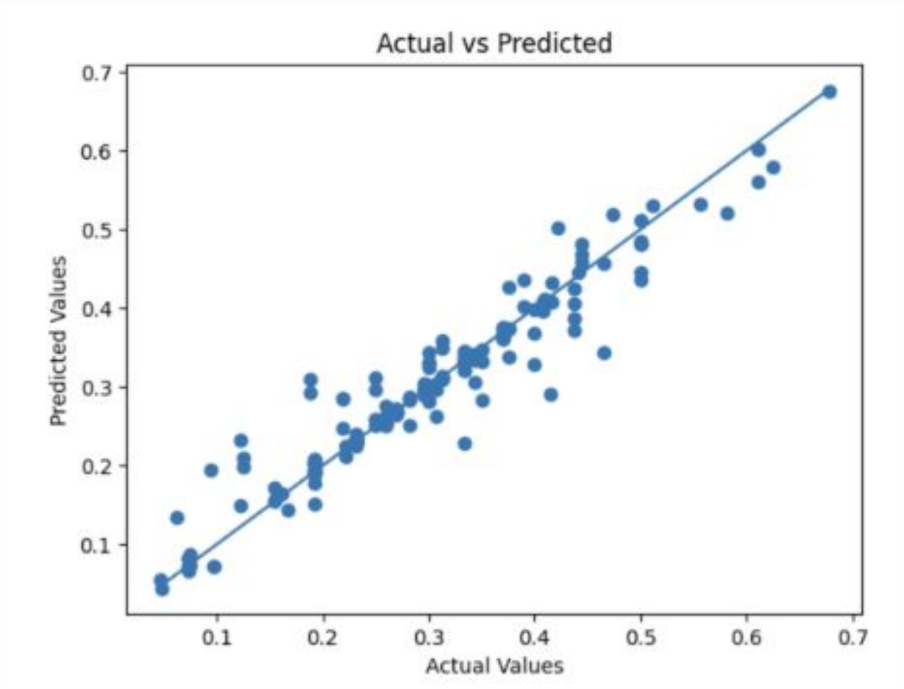


XGBoost Model Diagnostics

Model	**RMSE**	**R^2**
All Predictors	0.0322	92.95%
Median Importance Pruning	0.0401	89.09%
SHAP Pruning	0.0694	67.32%

XGBoost Plots





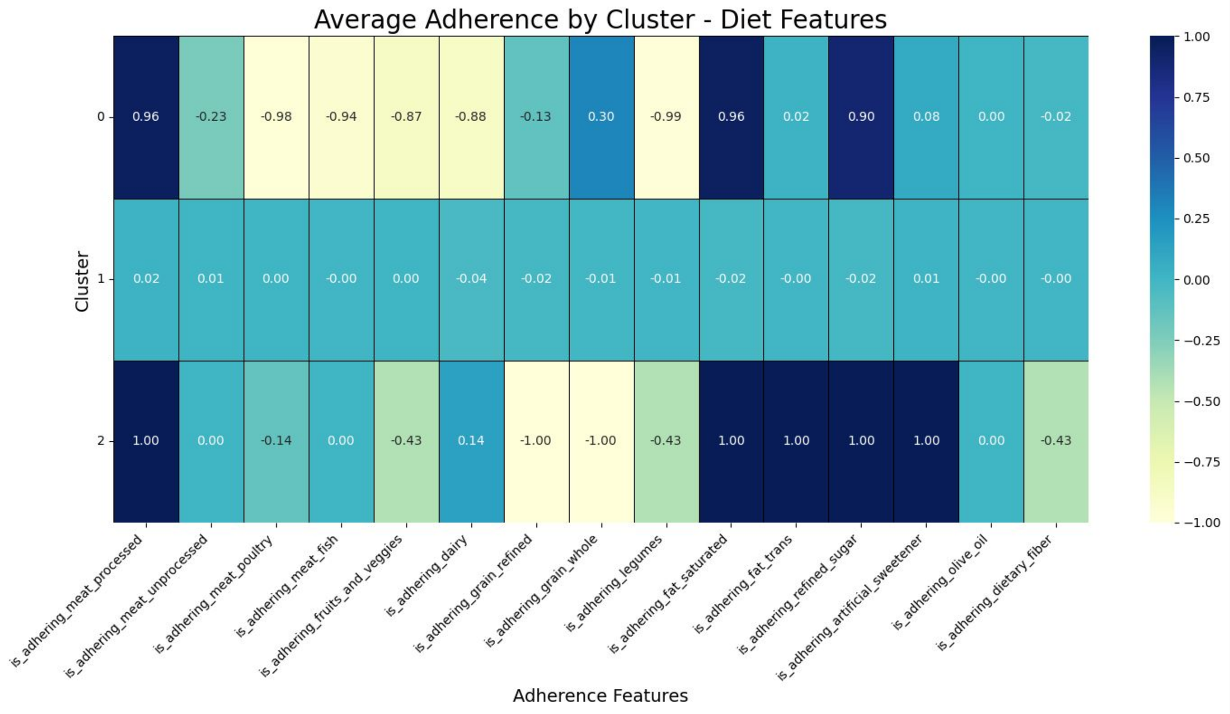
UVA
SCHOOL *of* DATA SCIENCE

Recurrent Neural Networks: Cluster Summary

- Cluster 0: Core Users - Low Adherence
  - Tracking major diet features(meat, dairy, fat, sugar)
  - Tracking intense exercise
  - Tracking basic lifestyle(sleep, alcohol, water)
  - No supplement tracking
- Cluster 1: Light Users - Needs more planning
  - Most features are not tracked, those that are have too few data points to determine adherence
  - Trying to track basic items such as meat, dairy, sleep and water
- Cluster 2: Power Users - Moderate Adherence
  - Tracking all diet except meat_unprocessed, meat_fish, and olive oil
  - Tracking all exercise except cardio_high
  - Tracking basic lifestyle(sleep, cigarettes, alcohol, water)
  - Tracking some supplements

Recurrent Neural Networks: Diet Features

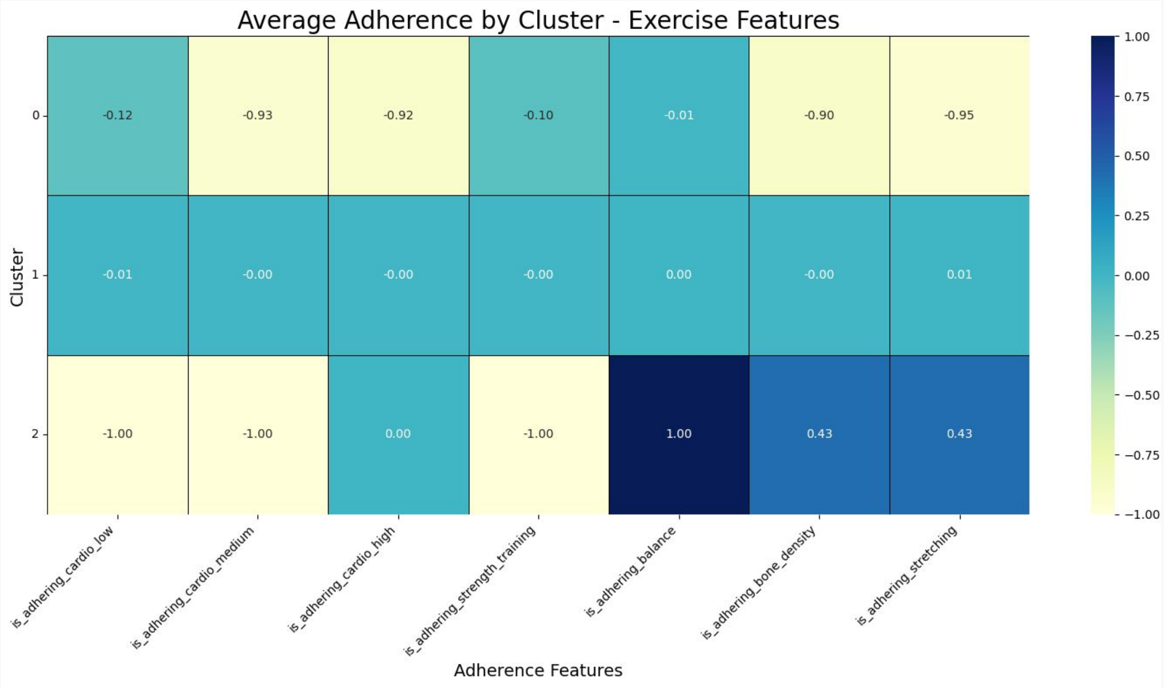
- Cluster 0: Tracking many variables but failing to adhere most of the time
- Cluster 1: Users are not tracking features or submitting logs
- Cluster 2: Users are tracking a majority of features with moderate adherence



UVA
SCHOOL *of* DATA SCIENCE

Recurrent Neural Networks: Exercise Features

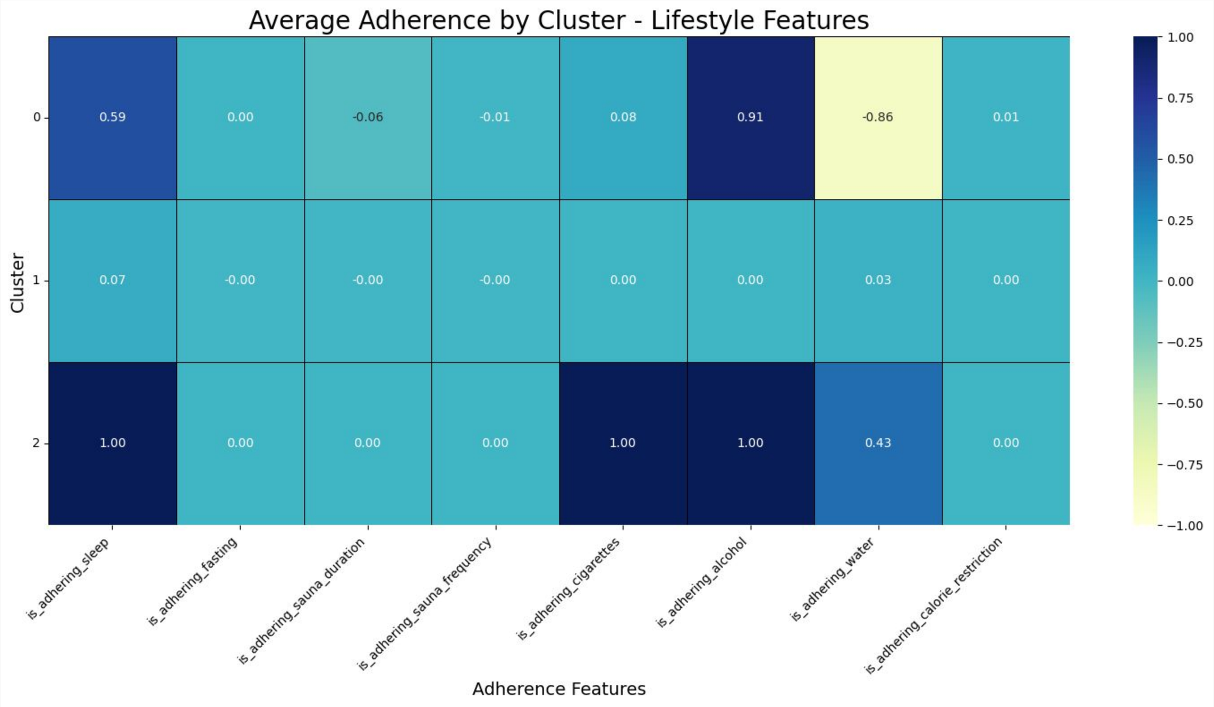
- Cluster 0: Tracking half of variables but failing to adhere entirely
- Cluster 1: Users are not tracking features or submitting logs
- Cluster 2: Users are tracking a majority of features with unsuccessful adherence



UVA
SCHOOL of DATA SCIENCE

Recurrent Neural Networks: Lifestyle Features

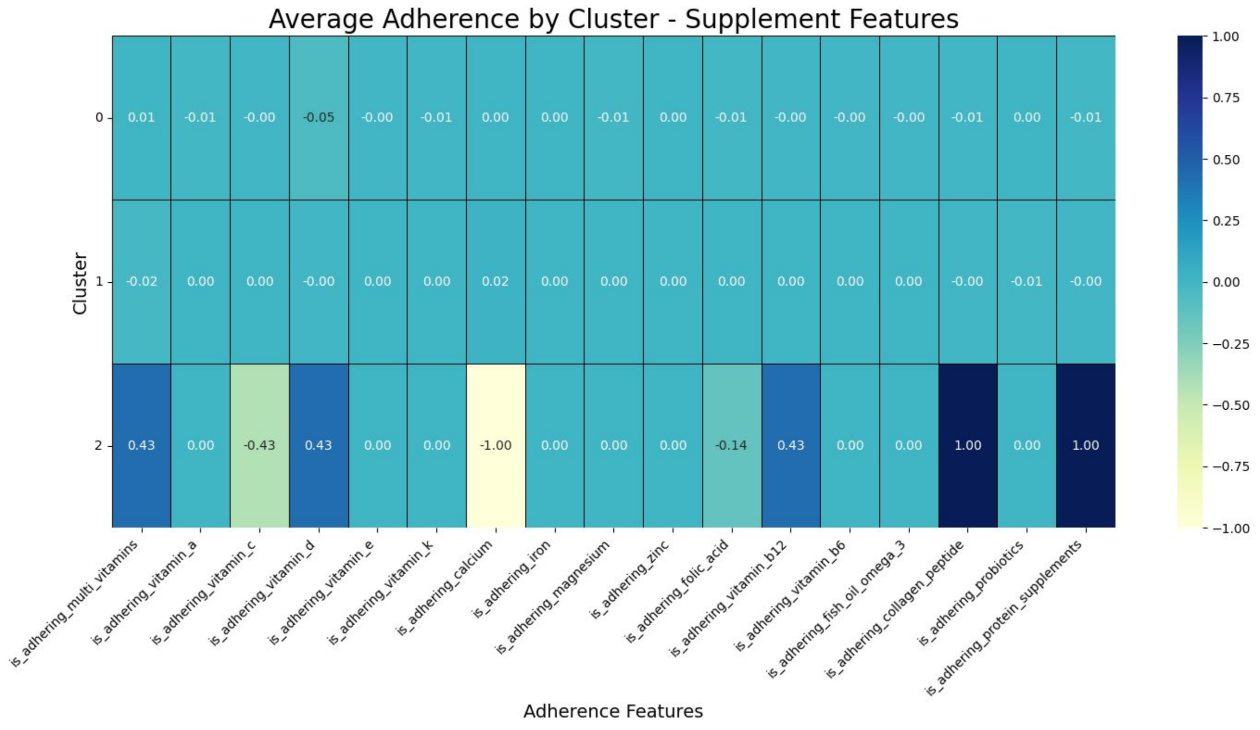
- Cluster 0: Tracking a few variables adhering half the time
- Cluster 1: Users are not tracking features or submitting logs
- Cluster 2: Users are successfully tracking and adhering to half of features



UVA
SCHOOL *of* DATA SCIENCE

Recurrent Neural Networks: Supplement Features

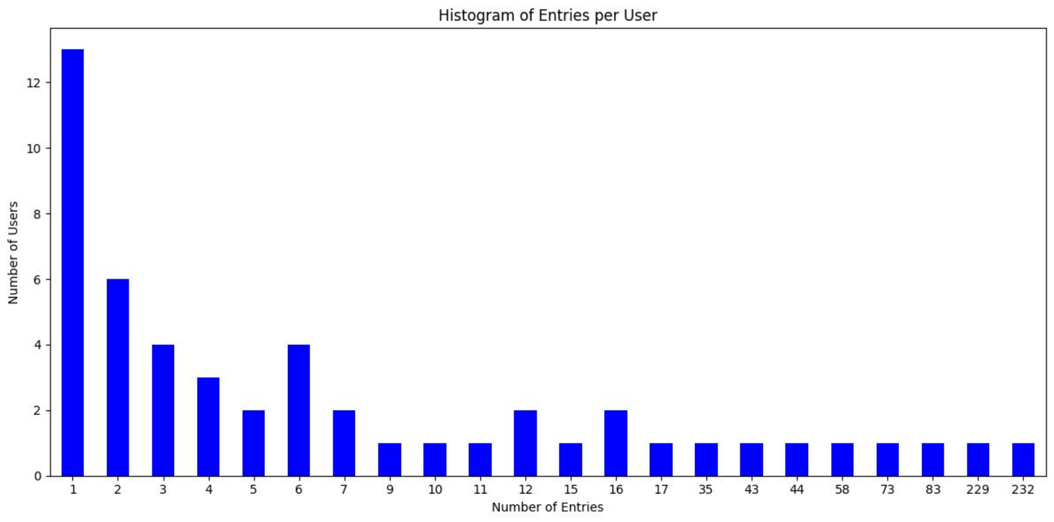
- Cluster 0: Users are not tracking features or submitting logs
- Cluster 1: Users are not tracking features or submitting logs
- Cluster 2: Users are successfully tracking a few features with mixed adherence success



UVA
SCHOOL of DATA SCIENCE

Recurrent Neural Networks: Autoencoder

- Goal: Use RNNs to gain a feature specific understanding of user adherence patterns
- Challenges:
  - Defining adherence
  - Variable Sequence Length
- Components:
  - Encoder: Converts time series data into a user embedding
  - Decoder: Reconstructs original time series from embeddings
  - Loss Function: Mean-Square Error used to measure reconstruction loss
- As loss decreases, the encoder's ability to capture relevant user information improves
- Encoder alone is used to create user embeddings



UVA
SCHOOL *of* DATA SCIENCE

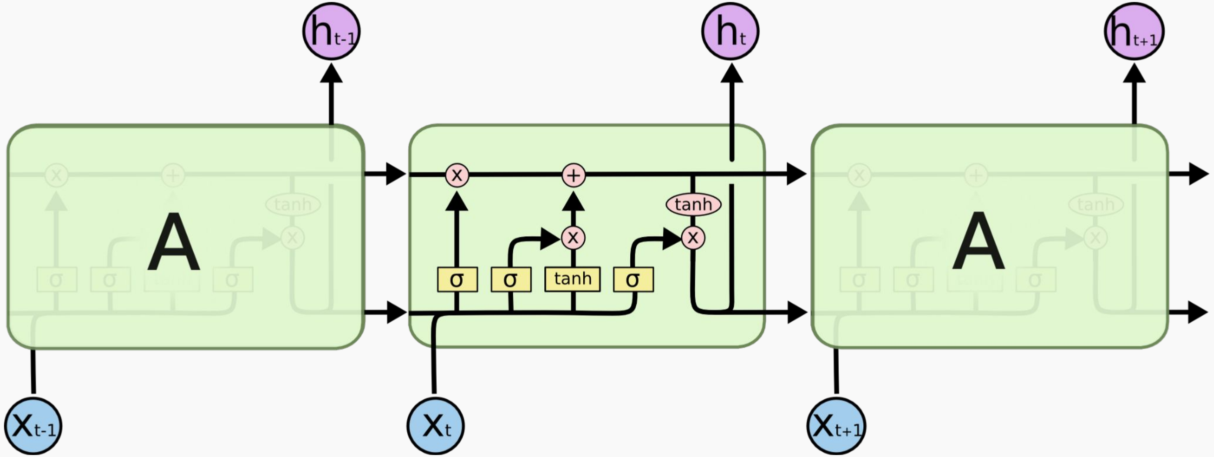
Recurrent Neural Networks: Defining Adherence

- Defined 4 rules for determining adherence(Not the same as feature categories)
  - Ratio_range
  - Greater_than_goal
  - Less_than_goal
  - Exact_goal
- Successful adherence: 1
- Unsuccessful adherence: -1
- Untracked goal: 0
- Padded values: 99

user_id	end_date	is_adhering_meat_unprocessed	is_adhering_meat_poultry	is_adhering_meat_fish	is_adhering_fruits_and_veggies	is_adhering_dairy
07f86f6c976848c1916b7c23f3743db5	Sat Jan 11 2025	0.0	-1.0	0.0	1.0	1.0
08d54e3402a84950b9d1ffc12020ce89	Mon Jan 13 2025	0.0	-1.0	0.0	1.0	1.0
08d54e3402a84950b9d1ffc12020ce89	Sat Apr 05 2025	0.0	1.0	0.0	-1.0	-1.0
08d54e3402a84950b9d1ffc12020ce89	Sat Feb 15 2025	0.0	1.0	0.0	-1.0	1.0
094a27f6082943e5af4ed1e883bd33c3	Sat Sep 21 2024	-1.0	-1.0	0.0	-1.0	0.0

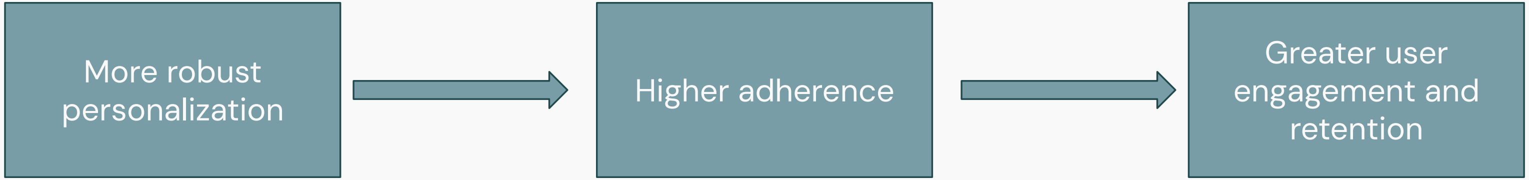
Recurrent Neural Networks: Model Architecture

- Single LSTM: A single LSTM layer for both the encoder and decoder.
- Stacked LSTM: Two stacked LSTM layers for both the encoder and decoder.
- Bidirectional LSTM: A single bidirectional LSTM encoder with a standard LSTM decoder.
- Stacked Bidirectional LSTM: Two stacked bidirectional LSTM encoders with two standard LSTM decoders.
- TimeDistributed Dense layer: Converts decoder output to feature variables.
- Hyperparameter search
  - Latent dimensions
  - Dropout rate



Recommendation & Impact

- **Enhance Data Collection**: Increase user volume, reduce sparsity, and extend logging history.
- **Deploy Predictive Insights**: Integrate XGBoost adherence predictions into MyYouthSpan to proactively target at-risk users.
- **Refine Personas**: Leverage expanded RNN analysis for more granular segmentation and personalized recommendations.
- **Personalize Interventions**: Tailor reminders, goals, and feedback to each persona to improve retention.



Acknowledgements

- **Sponsor**: John Leddo, PhD - Co-founder MyYouthspan
- **Client Technical Support:** Ansh Riyal - Technical Liaison

- **Faculty Mentor**: Adam Tashman, PhD

UVA

SCHOOL *of* DATA SCIENCE