Arman Jabbari | Sue Kim | Dana Lansigan
Stella Park | Rebekah Tang

PREDICTING THE TREATMENT COMPLETION FOR PSYCHOLOGICAL DISORDERS

GOAL: TOWARDS SUCCESSFUL TREATMENT COMPLETION

- Meta-analysis of 186 studies by Idaho University reports finds the average treatment-refusal rate of **8.2%** and average premature termination rate **21.9%**.
- Treatment completion not only contributes to **patient's recovery** but also **prevents the development of new mental illness**.
- Given this significance, it is important and necessary to understand determinants of completion.
- This project takes patient data and makes predictions that can help identify the factors that significantly affect the **likelihood of treatment completion**.

APPROACH

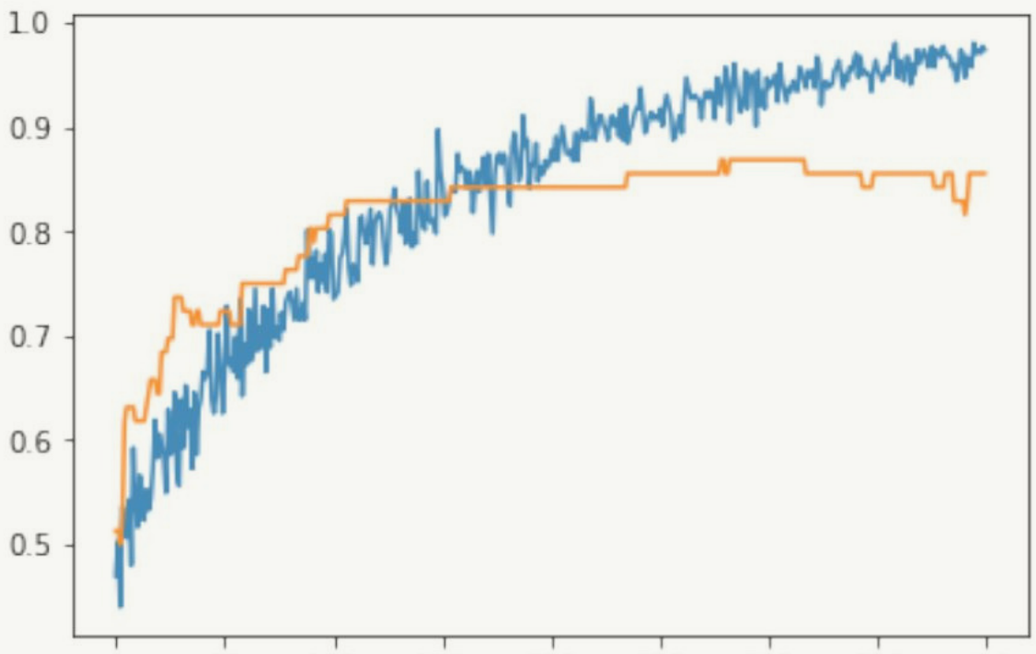
- Using data from the National Comorbidity Survey Replication (NCS-R), create a neural network to predict treatment completion for people with depression.
- Train network on NCS-R participants' sex, age, socioeconomic status, demographics, medication, treatment type, and symptoms.

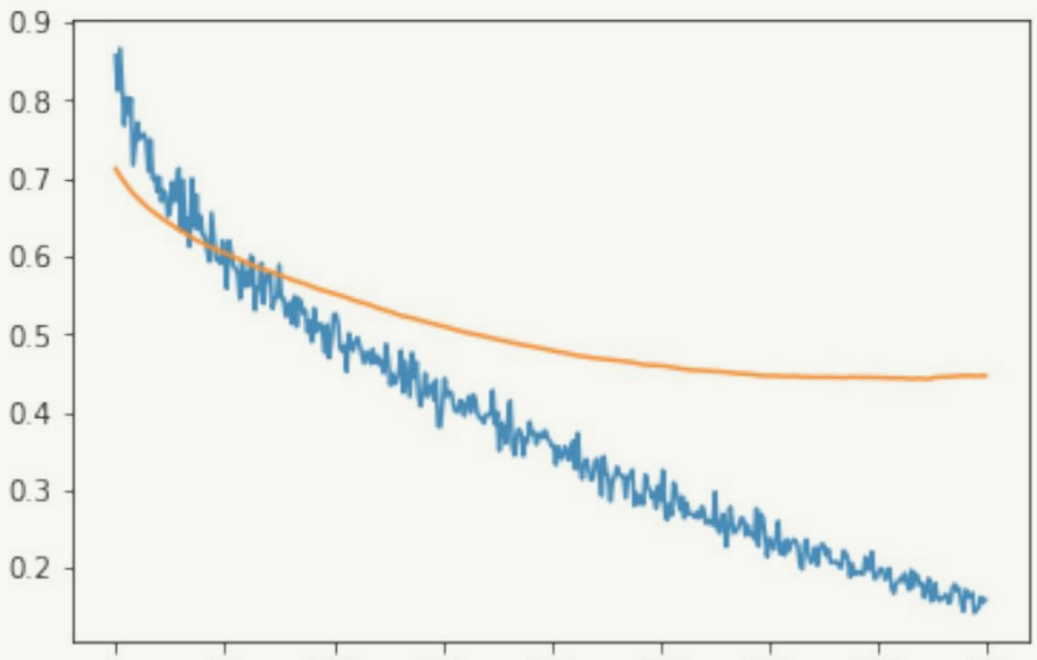
PREPROCESSING

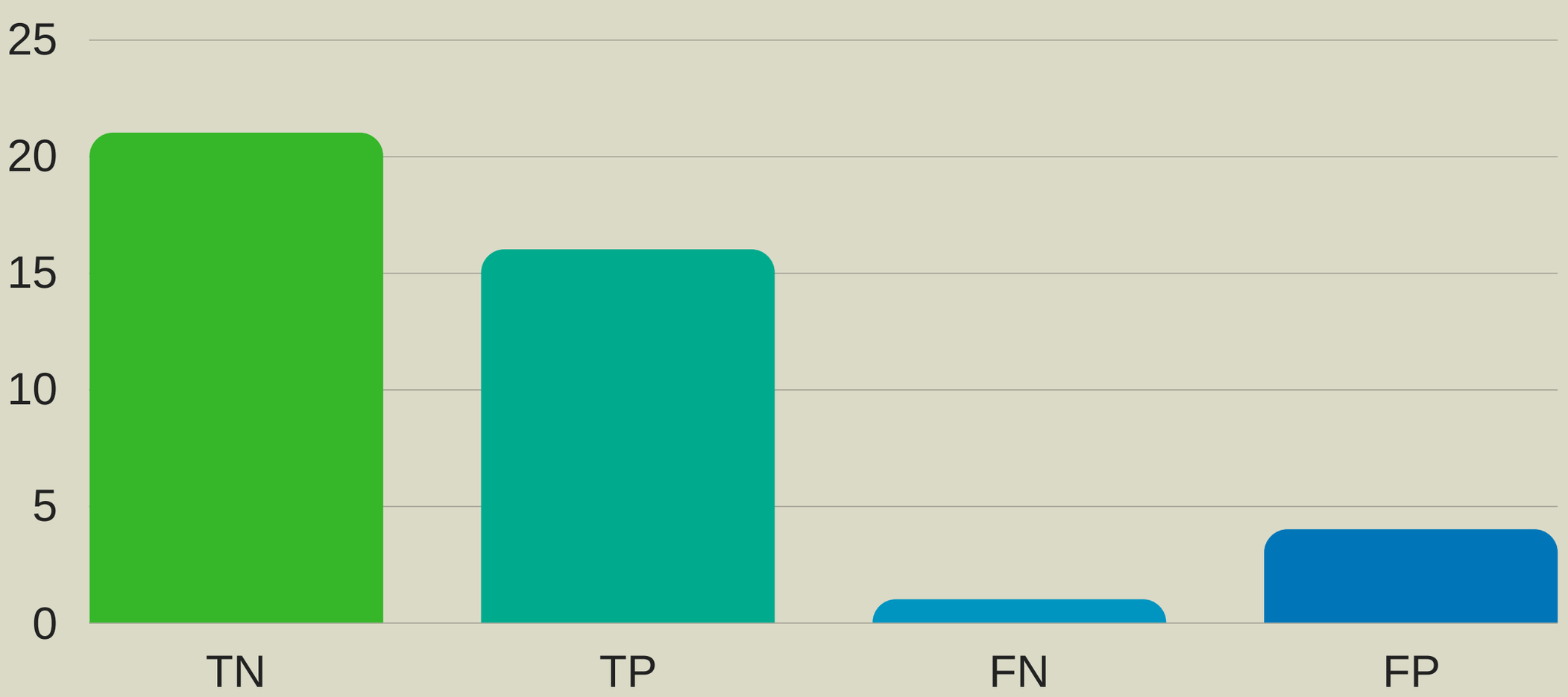
- Used data of respondents who were diagnosed with a major depressive disorder and underwent treatment.
- Handled missing data with various methods depending on each variable's statistical characteristics.
- After preprocessing, data size was 294.
- Balanced dataset by duplicating data with "no" responses.

FINAL STATE

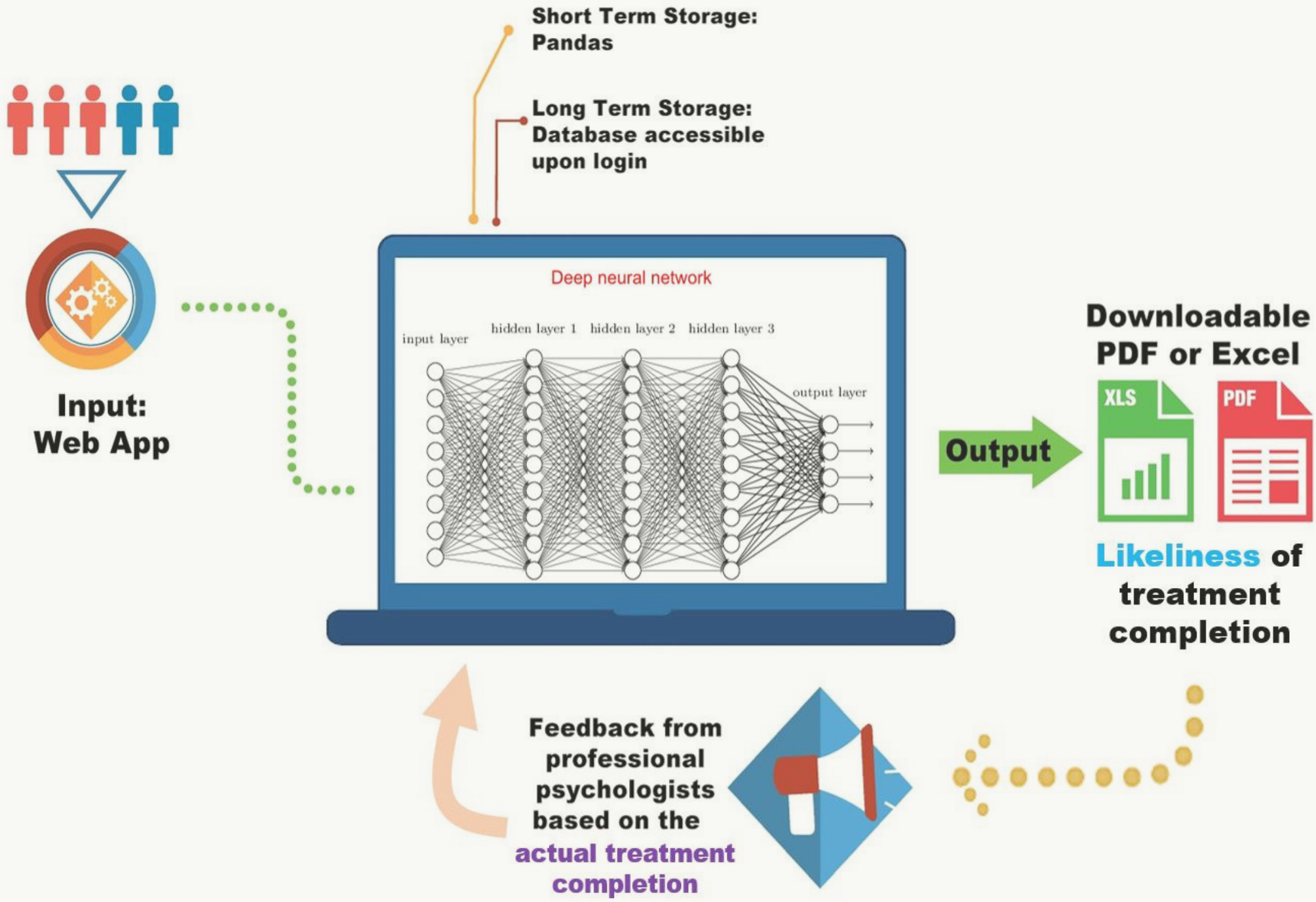
Achieved test accuracy of 88.09% and the F1 score of 86.48%.







ARCHITECTURE

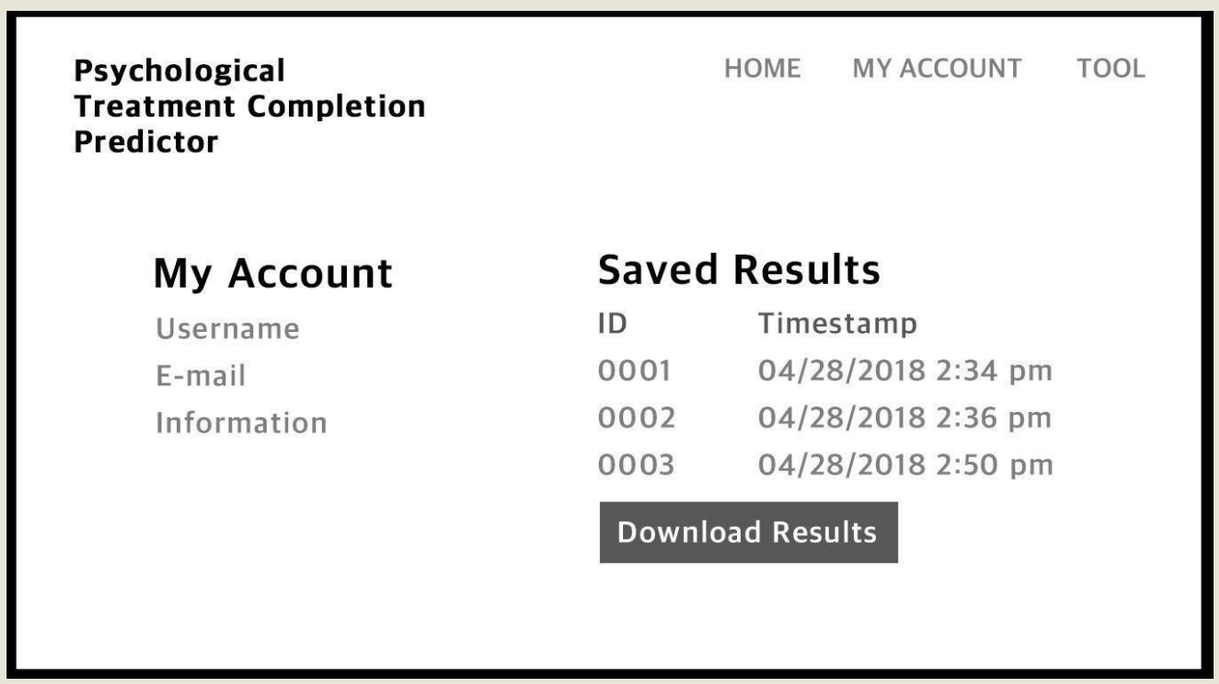


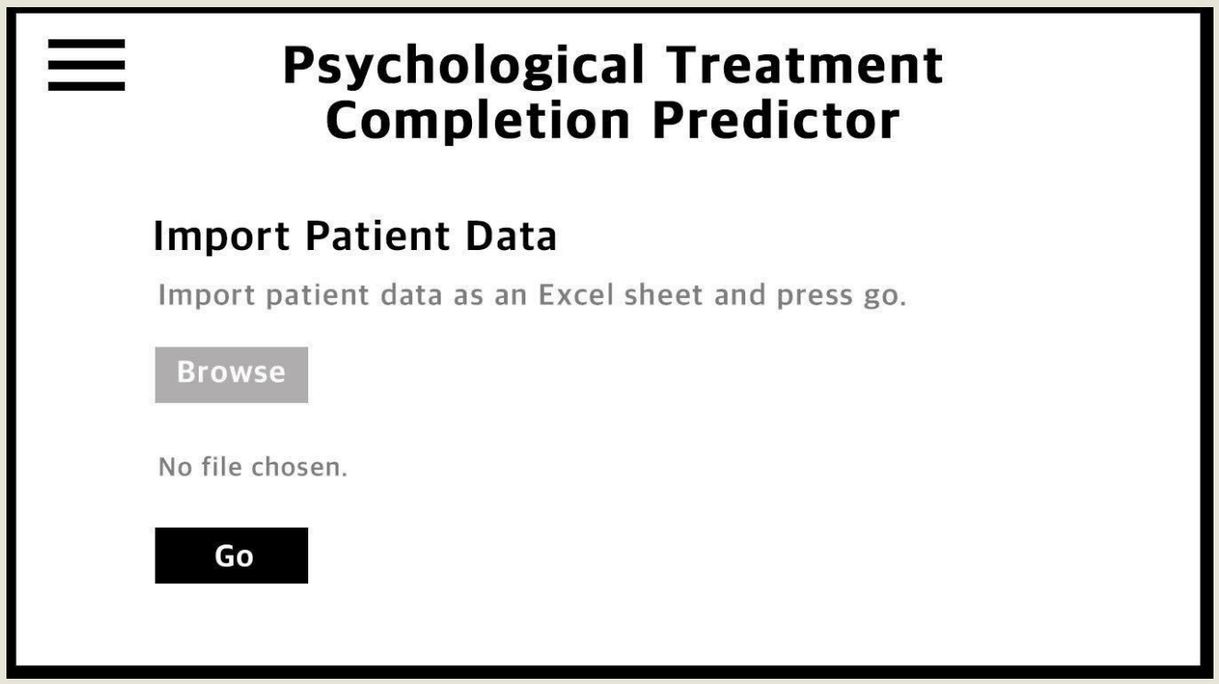
ARCHITECTURE

- Number of variables: 93
- Number of hidden layers: 3
    - Layer 1: 1000 nodes, activation: Relu
    - Layer 2: 500 nodes, activation: Relu
    - Layer 3: 200 nodes, activation: Relu
- Dropout rate: 90%
- Batch size: 40
- Number of epochs: 400

USER INTERFACE: WEBSITE

- Currently, we use a Jupiter notebook to take in the patient data, and output likelihood of treatment completion.
- We intend to develop a website that improves accessibility and user-friendliness.





LEARNING PATH: STARTING POINT

- Initially wanted to create a classifier that could diagnose and differentiate between psychological mood disorders.
- After consulting with professors Aaron Fisher and Sheri Johnson from the psychology department, we decided to predict treatment completion instead.
- The psychology community has been moving towards solving the problem of treatment completion as opposed to disease diagnosis.
- There has been some controversy for whether classifying and differentiating so many mood disorders is meaningful or productive.

LEARNING PATH: DATA COLLECTION

- Considered two datasets recommended by professors: **NCS-R** and **STEP-BD** (Systematic Treatment Enhancement Program for Bipolar Disorder, deals with which treatments are best to treat depression, mania, and bipolar disorder) datasets.
- We learned that is was very difficult to obtain psychological datasets--we were unable to obtain the STEP-BD dataset.
- Even though we were able to obtain the NCS-R data, which was overwhelmingly large, there was a lot of missing data, especially for sensitive topics--why our final dataset was the size of 294.
- Thousands of variables--needed to use both domain knowledge and statistical significance to narrow down to just over 50 features

  Good practical experience on how to deal with messy data.

LEARNING PATH: MODEL FINDING

- Attempted K-Nearest Neighbors, Random Forest, xgboost, neural networks, using cross-validation, feature selection by significance, hyperparameter tuning.
- The final model was a three-layer, dense neural network made with Keras, which had the best accuracy rates and F1 score.