

Predicting Treatment Adherence Using Machine Learning

Peyton Hopkins
January 8, 2026

Leading Source for Neuropsychiatry Real-World Data

36,500,000+

Patients

40+

Geographically-diverse health systems

20+

Years of longitudinal data



Linkability

Connects patient data across time and care settings to enable longitudinal analysis and richer clinical insights.



Longitudinal

Over 20 years of neurology, behavioral health, mental health, and neuropsychiatry data from diverse care environments including inpatient, outpatient, emergency, crisis, and community settings.



Comprehensive

Combines structured data with NLP-enriched clinical notes to support multidimensional cohort development and outcomes research.



Depth

Offers validated neuropsychiatry variables—assessments, symptoms, testing, and interventions—for advanced analytics and decision-making.

<https://www.neuroblu.ai/>

Structured Data

 Diagnosis Codes (ICD-9, ICD-10)		Outcome Measures	Patient Counts
		CGI-S	423K
 Patient Demographics		GAF	308K
 Prescription Data		PHQ-9	48K
 ED, inpatient and outpatient data across the same patients in 39 of 44 clinics		CSSRS	57K
		QIDS	23K
		BDSS	20K
		PSRS	10K

Unstructured Data



Clinical notes (intake notes, progress notes etc.) generated at every visit with rich information about the patient



Mental Status Examination (MSE)

Categorized notes on patient's function, appearance and mood



External Stressors

Social, relational and occupational events that may affect the patient's mental health



Side Effects

Clinician-observed side effects from medication

NeuroBlu Platform

We used Holmusk's NeuroBlu Analytics to:

- Access their database
- Perform data preprocessing, cleaning, and exploratory analysis.
- Develop and test our prediction model



Medical Adherence

Definition: Taking medications as prescribed (correct dose, time, and duration).

Non-adherence Damages

- Non-adherence to medical treatment leads to:
 - Estimated \$100 billion - \$300 billion in avoidable healthcare costs
 - 3%–10% of total U.S. healthcare costs
 - Poor health outcomes
 - Increased hospitalizations

8,000,000+

Patient records with Major Depressive Disorder (MDD)
in NeuroBlu Analytics

Why Focus on MDD?



In 2021, **8.3%** of U.S. adults, or **21 million** people, had a major depressive episode in the past year.



Nonadherence rates for patients with MDD range from **28%** to **52%**.

Julius et al., 2009, Journal of Psychiatric Practice.

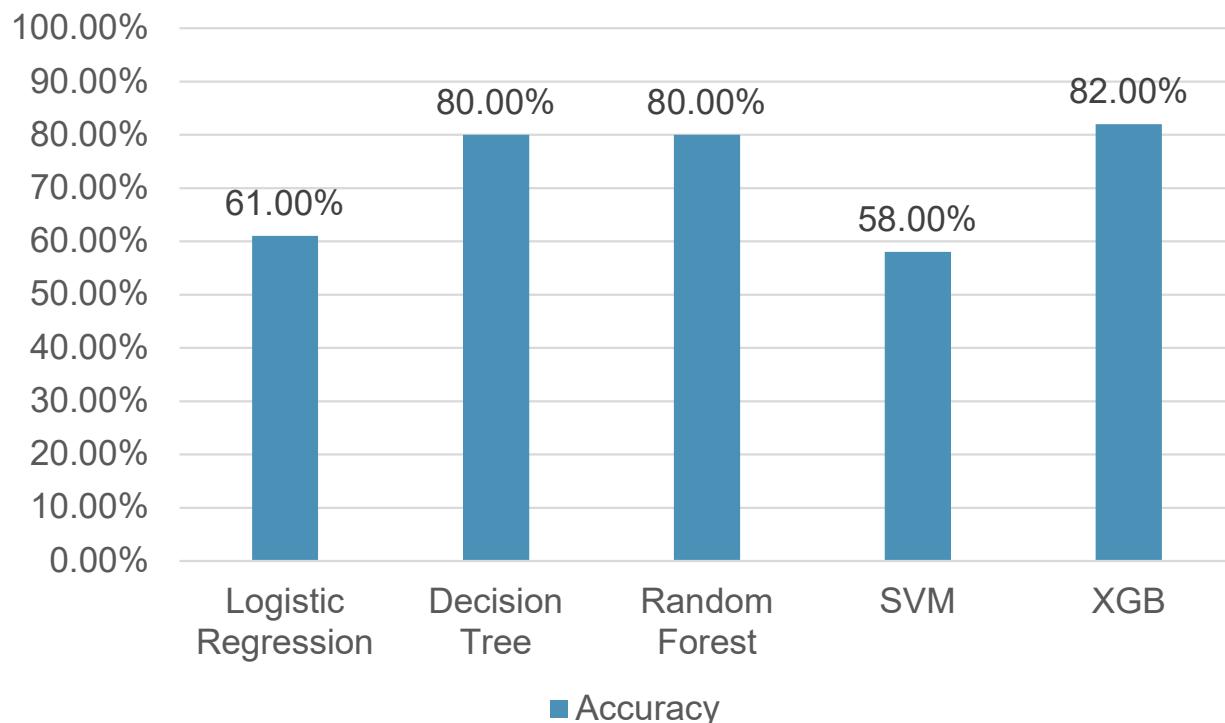
Data Extraction and Preprocessing



- Explored Neuroblu's data
- Learned OMOP and OHDSI standards
- Identified key features for adherence modeling
- Used SQL to extract data from NeuroBlu
- Cleaned data by addressing missing values
- Removed duplicate records from drug_exposure
- Created clean datasets for modeling
- Verified data quality with initial analysis

Machine Learning Model Selection

Comparison of Models Accuracy



- We tested several Machine Learning Models.
- Out of all models the Gradient Boosting Model (XGB) performed the best with an accuracy score of **82%**.

Interpretability

- Builds trust and transparency.
- Explains how features impact prediction.
- Identifies key features that should be focused on.

Impact in Healthcare



Avoid cost of nonadherence intervention (\$300 billion per year).



Healthcare providers being able to take action to educate patients early on.



Provide better care.

Personal Impact

- Experience working with an industry-grade data warehouse
- Learned how Computer Science work can bring about positive change
- Developed a deeper understanding of how a machine learning model "learns"

Thanks!