

Single Day Event

Predict & Explain: Voluntary Patient Dropout
for Efficient Execution of Clinical Trials

Amisha Kaswankar,
Johnson & Johnson



Disclaimer

This presentation is for informational purposes only and does not represent professional guidance or advice. Any views and opinions expressed during this presentation are those of the presenters and do not necessarily reflect the views or policies of Johnson & Johnson, or any other company in the Johnson & Johnson Family of Companies.

The data used for this case study has been simulated to depict real life patterns.

This presentation may not be reproduced, modified, or copied, in whole or in part, without the express written permission of Johnson & Johnson.



Agenda

1

**Problem Statement
Overview**

3

**Model &
Explainability**

5

Results & Insights

7

Future Scope

2

Data - Domains

4

**Data Pipeline
Journey**

6

**Interactive
Visualization**



Spotting the Silent Exit: Voluntary Patient Dropouts



How to Tackle this issue?

Using an innovative approach of Modern ML and Explainable AI with potential to tackle the challenges of patient retention & compliance

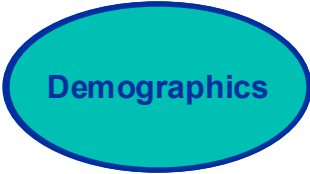


Understanding the Data





Understanding the Data

A teal-colored oval with a dark blue border, containing the word "Demographics" in dark blue text.

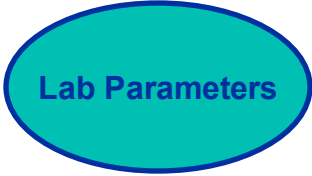
Demographics



Understanding the Data

A teal-colored oval with a dark blue border, containing the text "Adverse Events" in dark blue.

Adverse Events



Lab Parameters

Understanding the Data



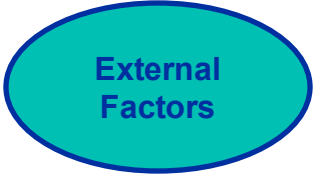
**Site
Characteristics**



Understanding the Data



Patient Reported
Outcomes



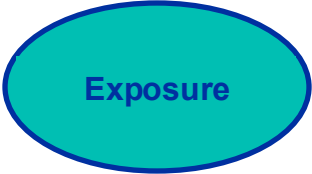
**External
Factors**



Understanding the Data

Subject Visits

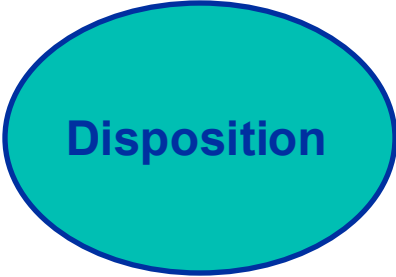
Medical History



Exposure



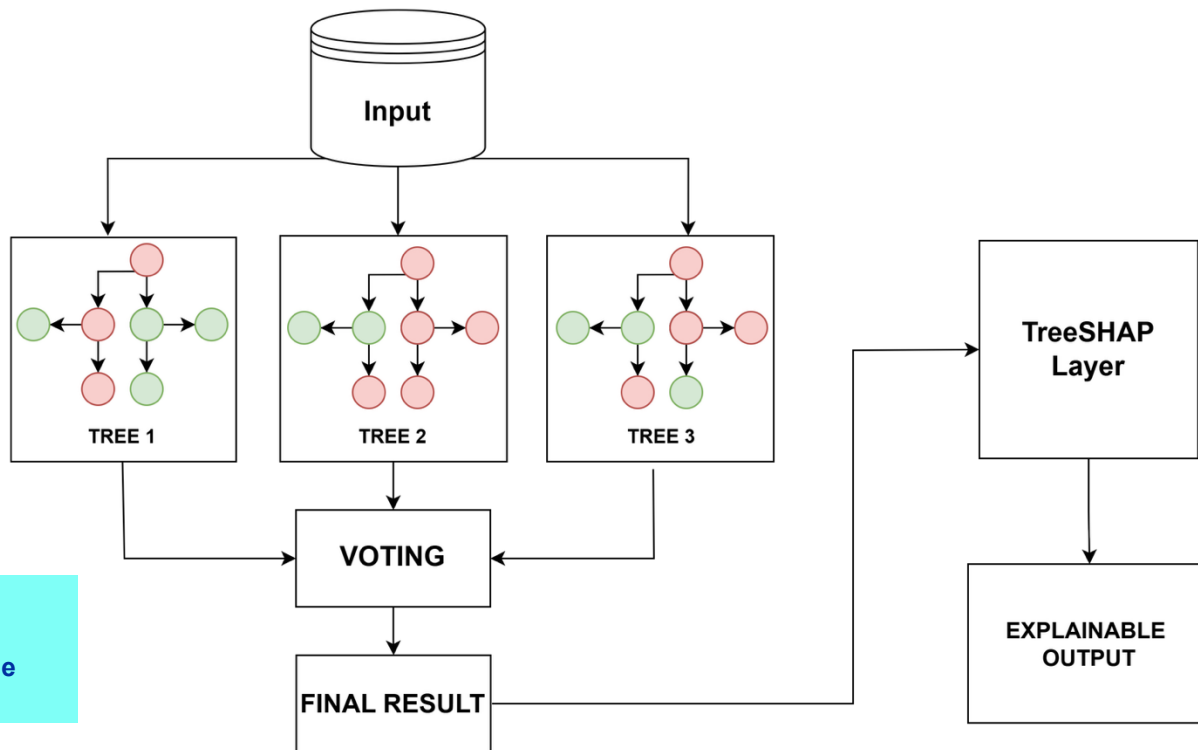
Understanding the Data

A teal-colored oval with a dark blue border, containing the word "Disposition" in dark blue text.

Disposition



Model Framework

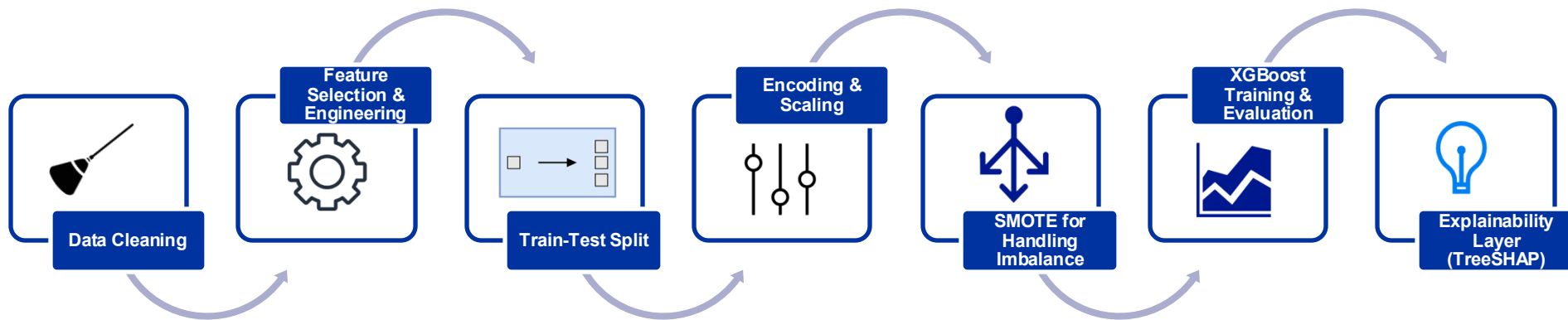


XGBoost: a Tree Based, Gradient Boosting Machine Learning Model

TreeSHAP: a Tree Based, Explainable AI (XAI) Technique

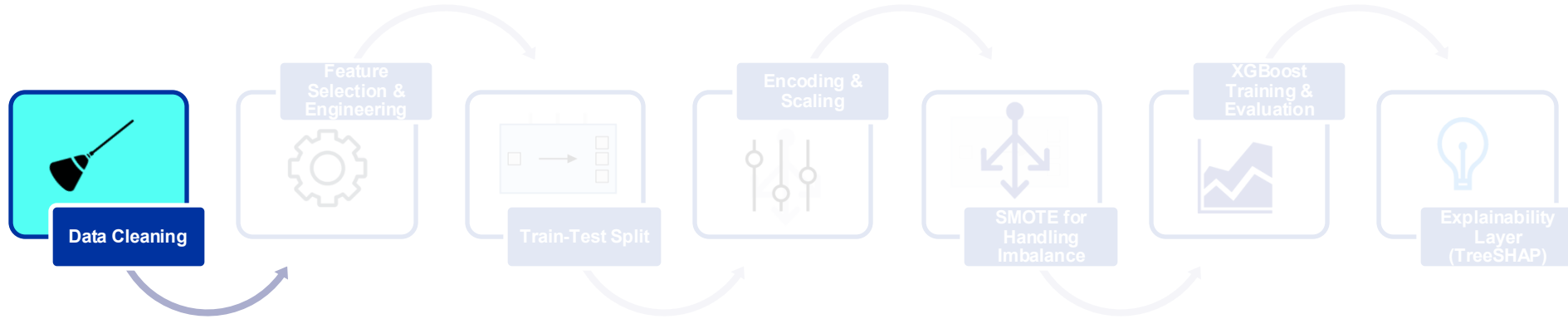


Data Pipeline Journey





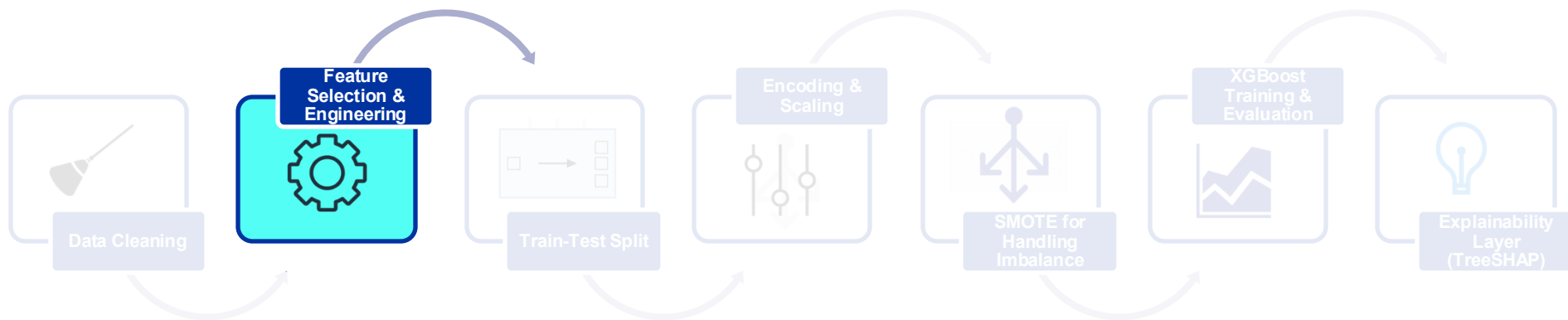
Data Pipeline Journey



**Data gets separated into two parts:
Predictor dataset (X) & Target dataset (y)**



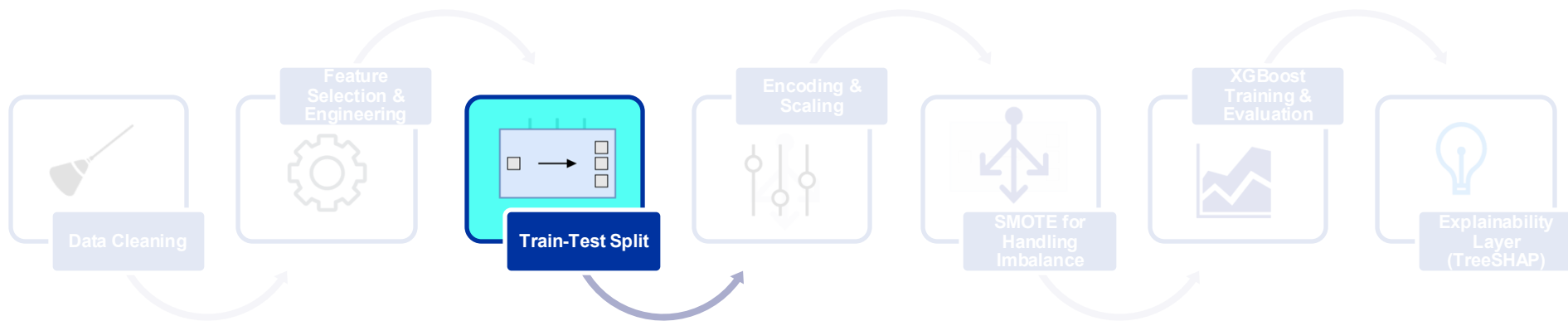
Data Pipeline Journey



Identification of relevant variables through correlation checks and domain knowledge



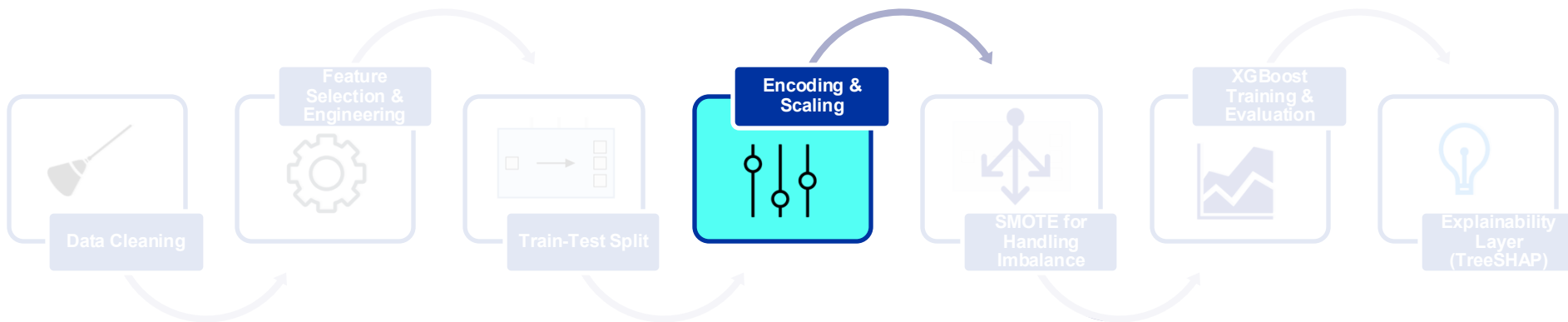
Data Pipeline Journey



Data gets separated into two parts: Training set (70%) for Model Training and Testing set (30%) for Model Testing



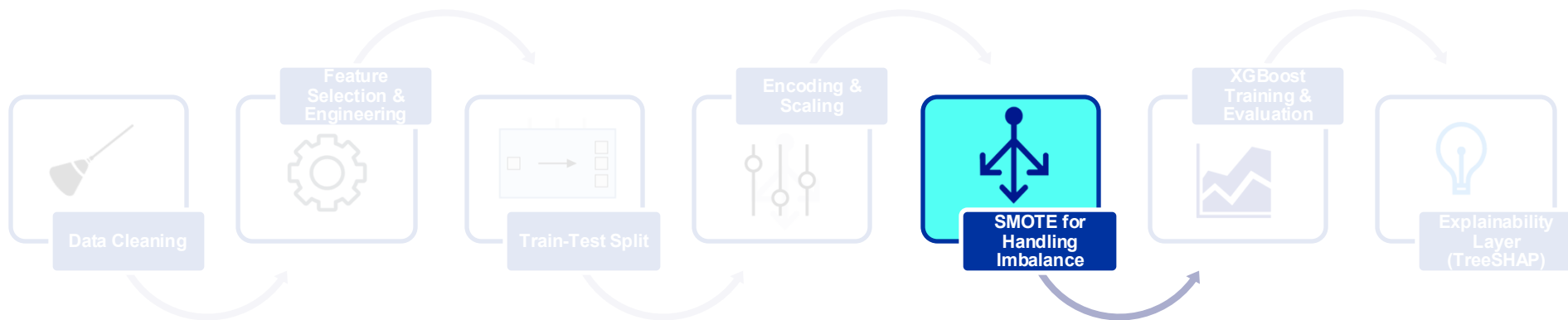
Data Pipeline Journey



Encoding: all categorical variables get converted to numeric form
Scaling: all the values get normalized and converted to a similar range



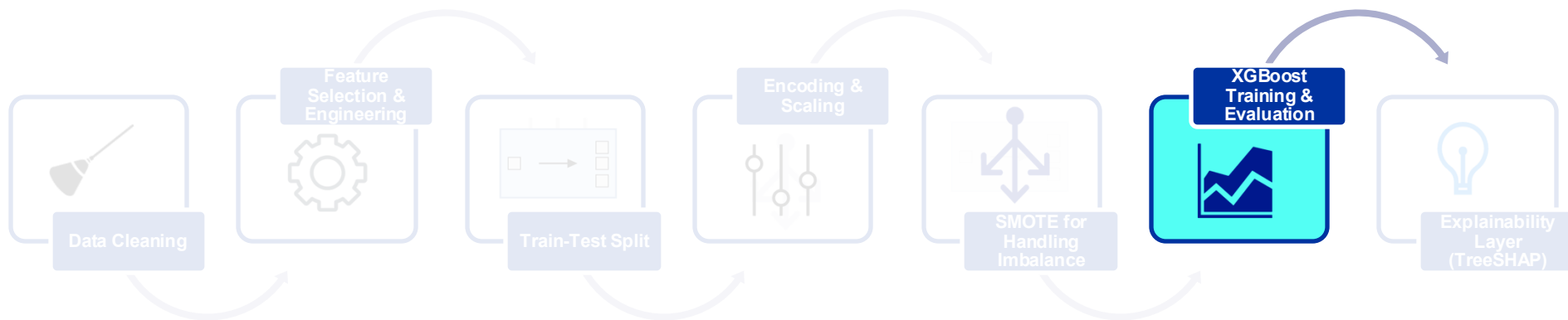
Data Pipeline Journey



SMOTE (Synthetic Minority Oversampling Technique) helps to balance the target variable by synthetically generating new records



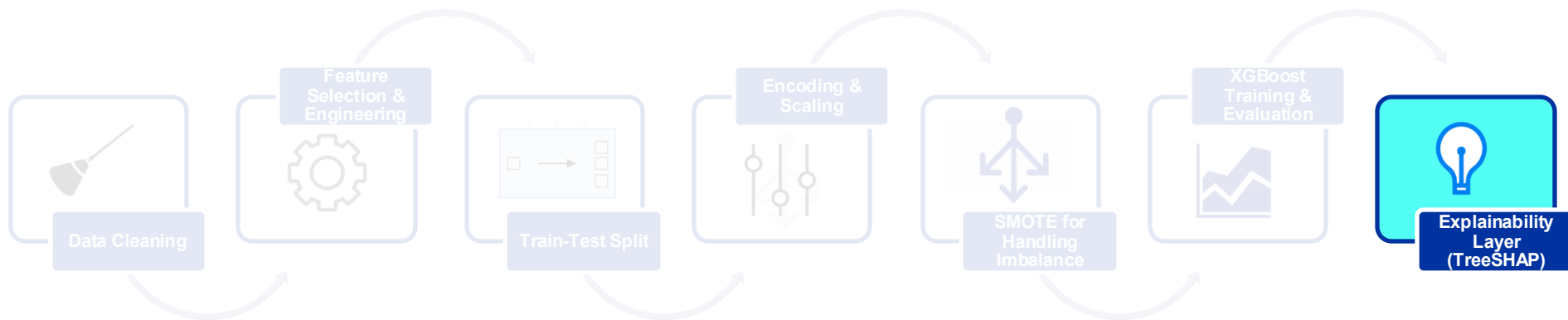
Data Pipeline Journey



Model gets trained on pre-processed data and learns the best working (parameters) through hyperparameter tuning



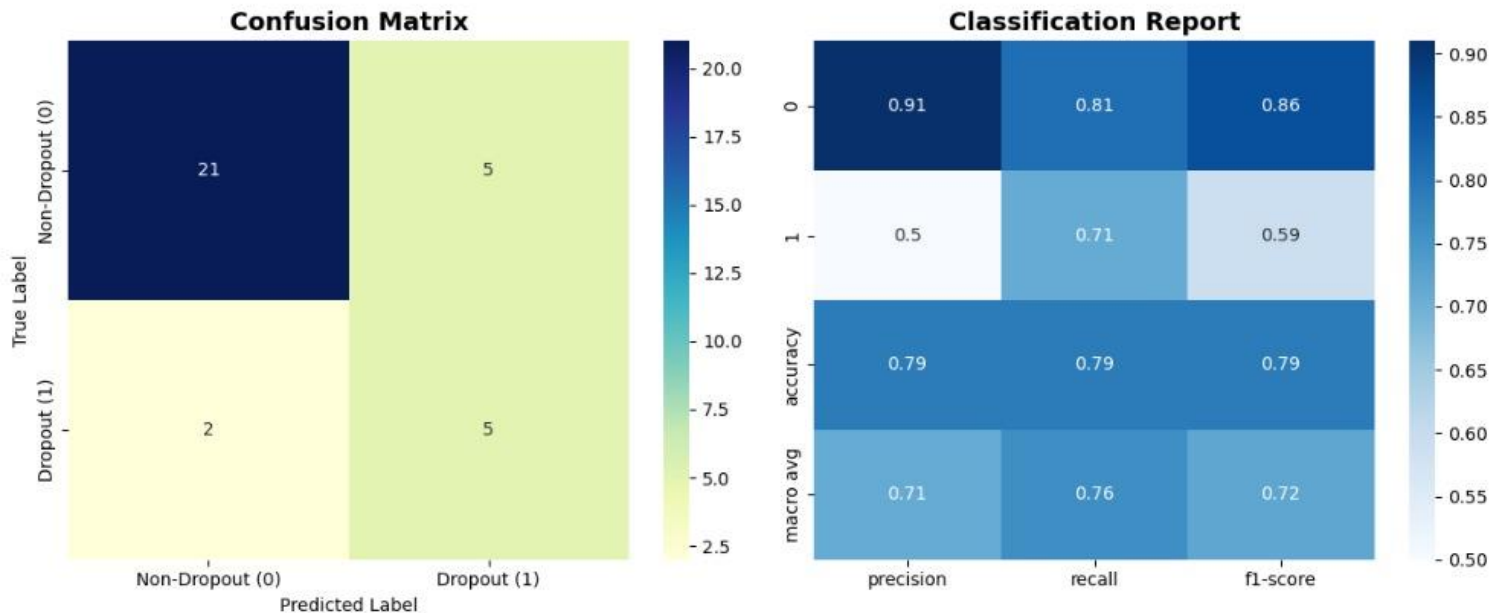
Data Pipeline Journey



XGBoost Output is passed to the TreeSHAP engine for calculation of contribution scores for variables



Results & Insights



Model Performance:
ROC-AUC = 0.831

Indicative of Moderate Predictive Power, scope for improvement with richer data



Results & Insights

Patient Level Explainability

Top 3 features driving each dropout prediction

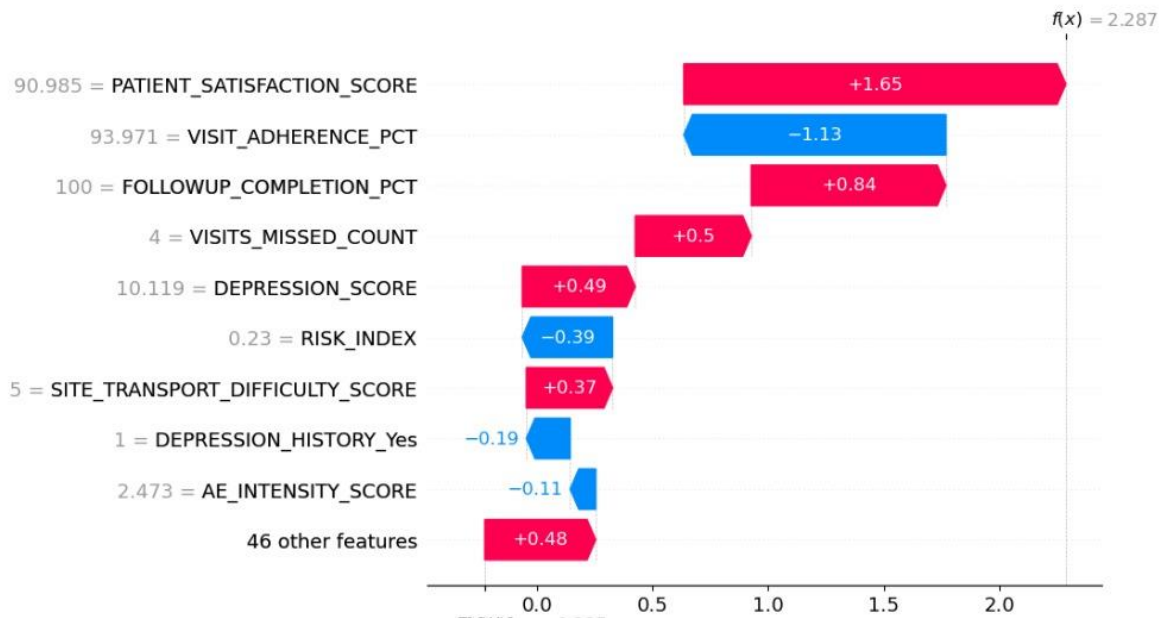
PATIENT_ID	Top3_Features	Top3_SHAP_Values	Actual_Dropout_Flag	Dropout_Probability	Dropout	Prediction_Interpretation
P079	[PATIENT_SATISFACTION_SCORE, FOLLOWUP_COMPLETION_PCT, VISITS_MISSED_COUNT]	{'PATIENT_SATISFACTION_SCORE': 1.12090003490448, 'FOLLOWUP_COMPLETION_PCT': -1.0121999979019165, 'VISITS_MISSED_COUNT': -0.9081000089645386}	0	0.083	0	Likely Retained

PATIENT_ID	Top3_Features	Top3_SHAP_Values	Actual_Dropout_Flag	Dropout_Probability	Dropout	Prediction_Interpretation
P026	[PATIENT_SATISFACTION_SCORE, VISIT_ADHERENCE_PCT, FOLLOWUP_COMPLETION_PCT]	{'PATIENT_SATISFACTION_SCORE': 1.6519999504089355, 'VISIT_ADHERENCE_PCT': -1.1335999965667725, 'FOLLOWUP_COMPLETION_PCT': 0.8417999744415283}	1	0.908	1	Likely Dropout



Results & Insights

Prediction for Patient ID P026:



Patient Level Explainability

Visual Representation of the factors influencing for a Patient Dropout Prediction.

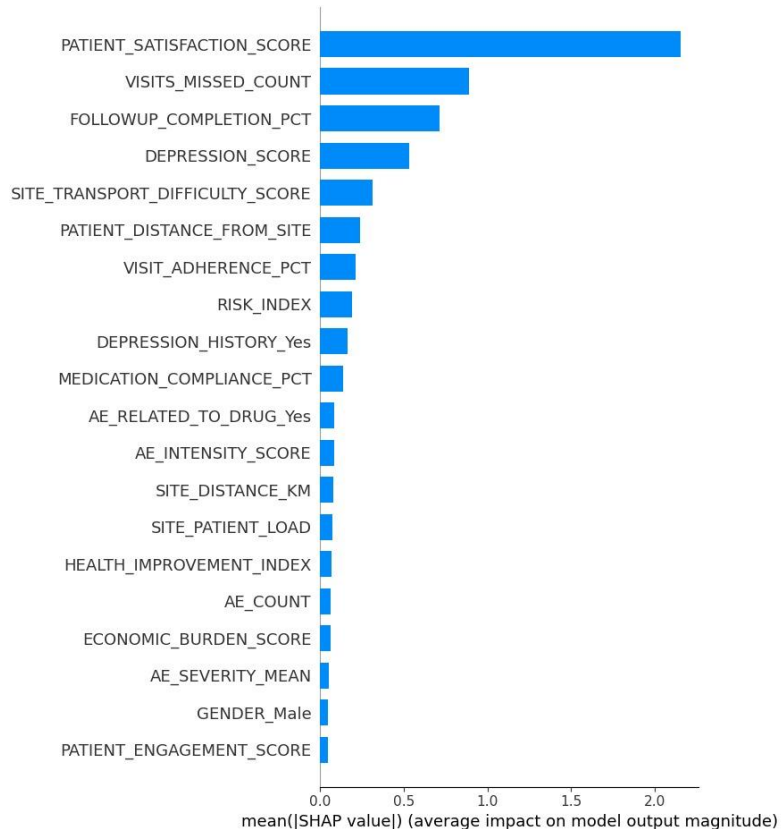
Positive value of $f(x)$ denotes that patient will likely dropout



Results & Insights

TreeSHAP Global Feature Importance Bar Plot

Top factors influencing Dropout
Predictions that should be prioritized
while designing retention strategies





Results & Insights

TreeSHAP Global Feature Importance Beeswarm Plot

Top factors influencing Dropout Predictions, showcasing their effect and consistency





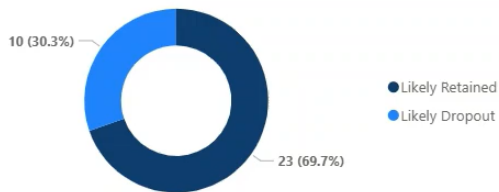
Interactive Visualization

Patient Dropout Prediction Dashboard

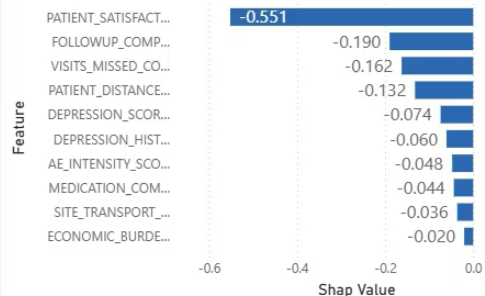
Select

All

Dropout vs Retention Distribution



Features by Shap Value

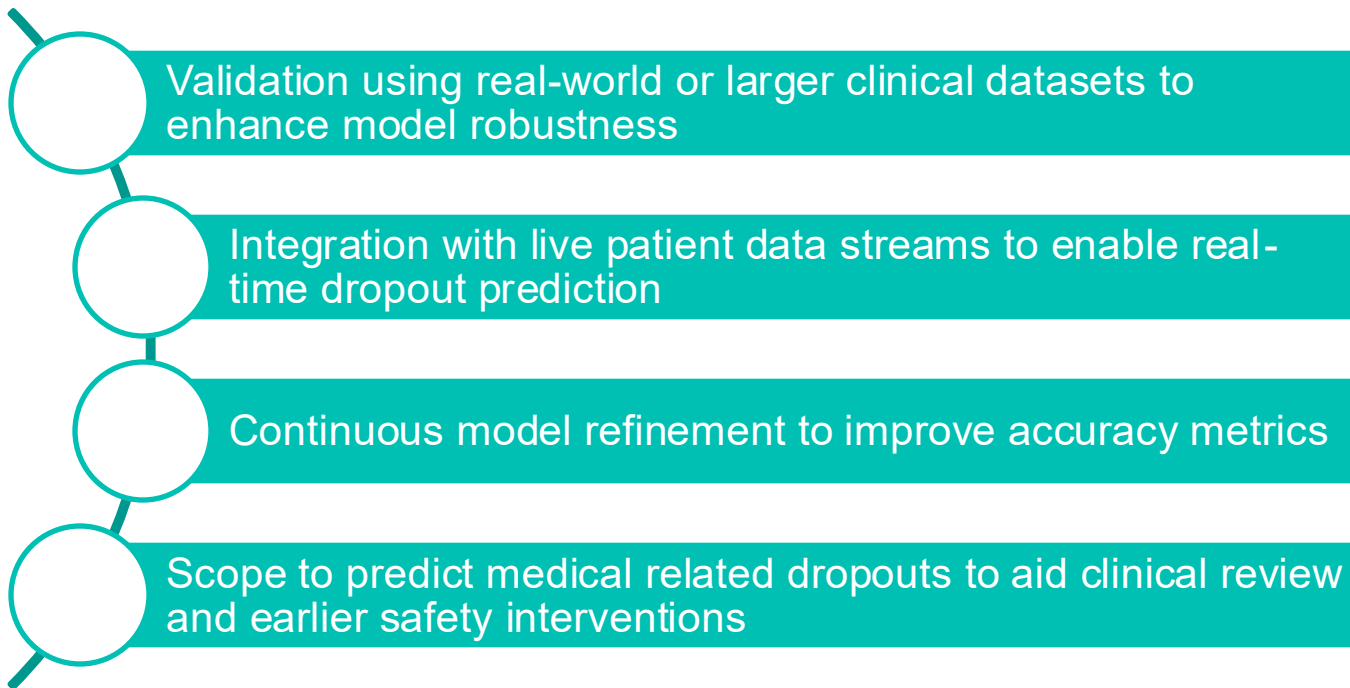


Patient Level Predictions and Key Features

patient Id	feature 1	feature 2	feature 3	Dropout Probability
P002	'PATIENT_SATISFACTION_S CORE': 1.68139	'VISITS_MISSED_COUNT': -1.5477	'FOLLOWUP_COMPLETIO N_PCT': -0.8296	0.24
P003	'PATIENT_SATISFACTION_S CORE': 1.50250	'VISITS_MISSED_COUNT': -1.1425	'FOLLOWUP_COMPLETIO N_PCT': 0.9147	0.64
P006	'PATIENT_SATISFACTION_S CORE': 1.584599	'VISITS_MISSED_COUNT': -1.4871	'FOLLOWUP_COMPLETIO N_PCT': -0.9262	0.23
P017	'PATIENT_SATISFACTION_S CORE': -1.136600	'VISITS_MISSED_COUNT': 1.037	'FOLLOWUP_COMPLETIO N_PCT': -0.6340	0.34
P020	'PATIENT_SATISFACTION_S CORE': -3.23359	'VISITS_MISSED_COUNT': 0.870	'DEPRESSION_SCORE': 0.4794	0.11
P026	'PATIENT_SATISFACTION_S CORE': 1.651999	'VISIT_ADHERENCE_PCT': -1.1335	'FOLLOWUP_COMPLETIO N_PCT': 0.8417	0.91
P028	'PATIENT_SATISFACTION_S CORE': 1.730100	'PATIENT_DISTANCE_FROM SITE': -0.7253	'FOLLOWUP_COMPLETIO N_PCT': 0.6658	0.88
P030	'VISITS_MISSED_COUNT': -1.347900	'FOLLOWUP_COMPLETION _PCT': -1.0657	'PATIENT_SATISFACTION_S CORE': 1.0520	0.03
P032	'VISITS_MISSED_COUNT': 1.048900	'PATIENT_SATISFACTION_S CORE': -0.92	'DEPRESSION_SCORE': -0.8532	0.53
P033	'PATIENT_SATISFACTION_S CORE': -3.751600	'FOLLOWUP_COMPLETION _PCT': 0.4986	'DEPRESSION_SCORE': 0.4174	0.05
P034	'PATIENT_SATISFACTION_S CORE': -2.57890	'VISITS_MISSED_COUNT': -1.1805	'FOLLOWUP_COMPLETIO N_PCT': -0.5512	0.01
P036	'PATIENT_SATISFACTION_S CORE': -3.475899	'VISITS_MISSED_COUNT': 0.7706	'SITE_TRANSPORT_DIFFIC ULTY_SCORE': -0.5569	0.03



Future Scope



THANK YOU!



**The Global Healthcare
Data Science Community**

Contact Channels

📞 UK +44 1843 609600
✉ office@phuse.global
🌐 phuse.global

Social Media

🐦 [@phusetwitta](https://twitter.com/phusetwitta)
📘 [/phusebook](https://facebook.com/phusebook)
▶ [/phusetube](https://youtube.com/phusetube)
🌐 [/company/phuse](https://company.phuse.com)

Amisha Kaswankar
Associate Stat Programmer
akaswank@its.jnj.com
[LinkedIn Profile](#)