

Interpreting Black-box Model: Real-world evidence examples and application

David Olaleye, PhD (Author)

Jim Box – Principal Data Scientist (Presenter)



10th–13th November
RAI Amsterdam
The Netherlands

EU 2019

The Clinical Data
Science Conference



David Olaleye, PhD

David is Senior Manager and Principal Research Statistician at SAS in Cary, NC. He received his postgraduate training in statistics, demography, and clinical epidemiology from the University of Pennsylvania School of Arts and Sciences, and School of Medicine, Philadelphia. His areas of specialization include episode-of-care analytics, real world evidence, pharmacovigilance and drug safety, and statistical algorithms for mining healthcare and clinical trials databases. David has authored and published articles in peer-reviewed journals.

<https://www.linkedin.com/in/dave-olaleye-29005353/>



AGENDA

01

Introduction

02

Machine Learning Models

- Interpretability and Bias

03

Post-Predictive Model Features
Association Check Tool

04

Use Case Demonstration

- Hospital Readmission Risk

05

Conclusion

Introduction

- ❑ Machine learning model types
 - ❑ Black box and white box models
- ❑ Model interpretability methods
 - ❑ Pre- and post-hoc interpretability methods
 - ❑ Model-agnostic versus model-specific
 - ❑ Global versus local explanations
- ❑ Challenges and limitations

Machine Learning Models

❑ Black box models

- ❑ Examples: forest model, gradient boosting, deep neural networks
- ❑ Provide highly accurate predictions
- ❑ Ability to detect nonlinear relationship and complicated interactions
- ❑ Hard to interpret due to large number of parameters, association layers,

❑ White box models

- ❑ Examples: Regression, decision trees, etc.
- ❑ Simplicity and easy to understand
- ❑ Model parameters are interpretable

Machine Learning Models

Attributes

- ❑ Accuracy, fairness, transparency
 - ❑ Bias
 - ❑ Unintended/Unanticipated results
- ❑ Relevance and meaningfulness
 - ❑ Features selection and rules
 - ❑ Dynamics between model input features and model output predictions
 - ❑ How a single or combination of features partially influence model predictions
- ❑ Evidence generation
 - ❑ Model reliability
 - ❑ Model replicability and reproducibility
 - ❑ Model generalizability

Model Interpretability Methods

- ❑ Variable importance

- ❑ Examples: forest model, gradient boosting, deep neural networks

- ❑ Partial dependence (PD) plot

- ❑ Depict functional relationship between a model's predictions and model inputs
 - ❑ Show how model's predictions partially depend on input variables' values

- ❑ Individual conditional expectation (ICE) plot

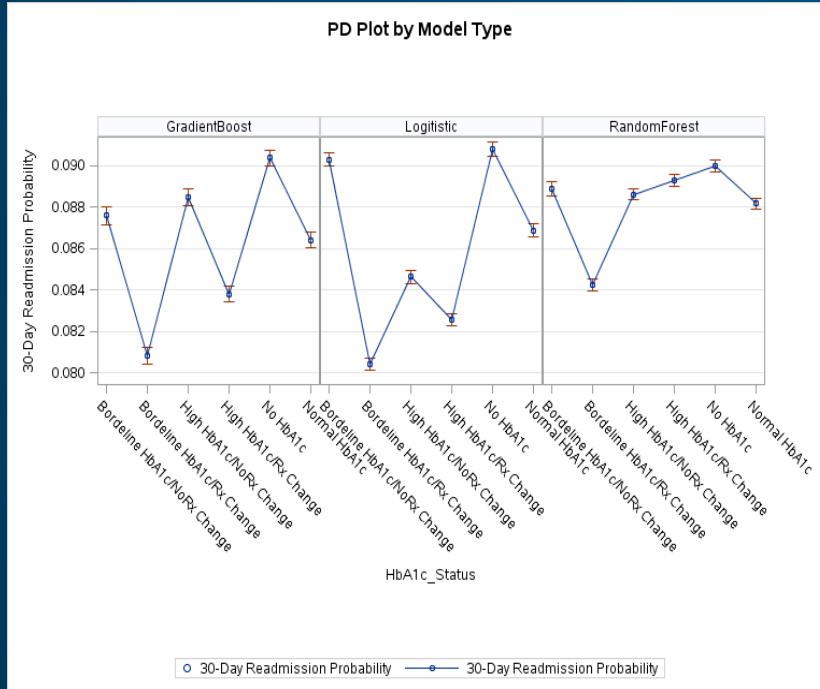
- ❑ Disaggregate PD function to reveal interactions and interesting subgroups
 - ❑ Ability to drill down to the level of individual observations

- ❑ Others: LIME, Shapley, etc.

Partial Dependence Plots

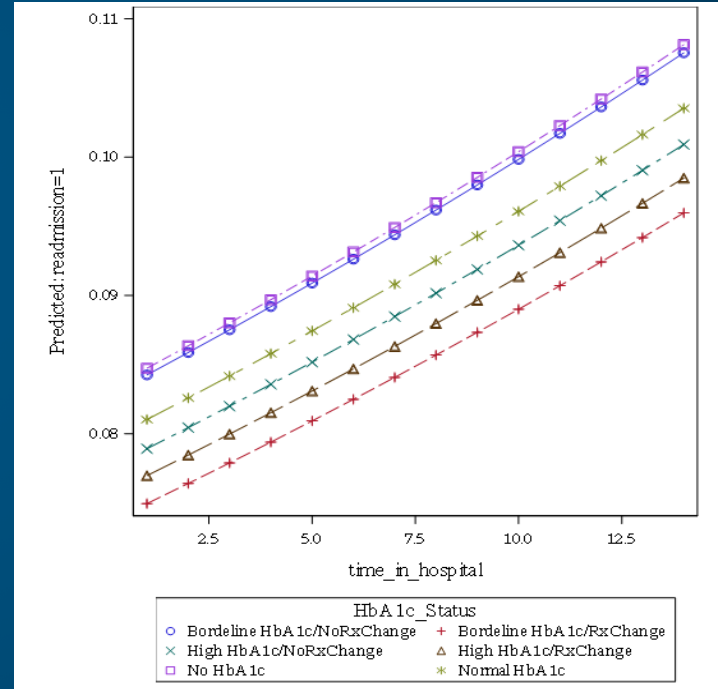
One-Ways and Two-Way Plots

One-Way Plot of HbA1c status



(ML models comparison)

Time spent in hospital by HbA1c status



(Logistic model)

Post-prediction Model Features Association Check

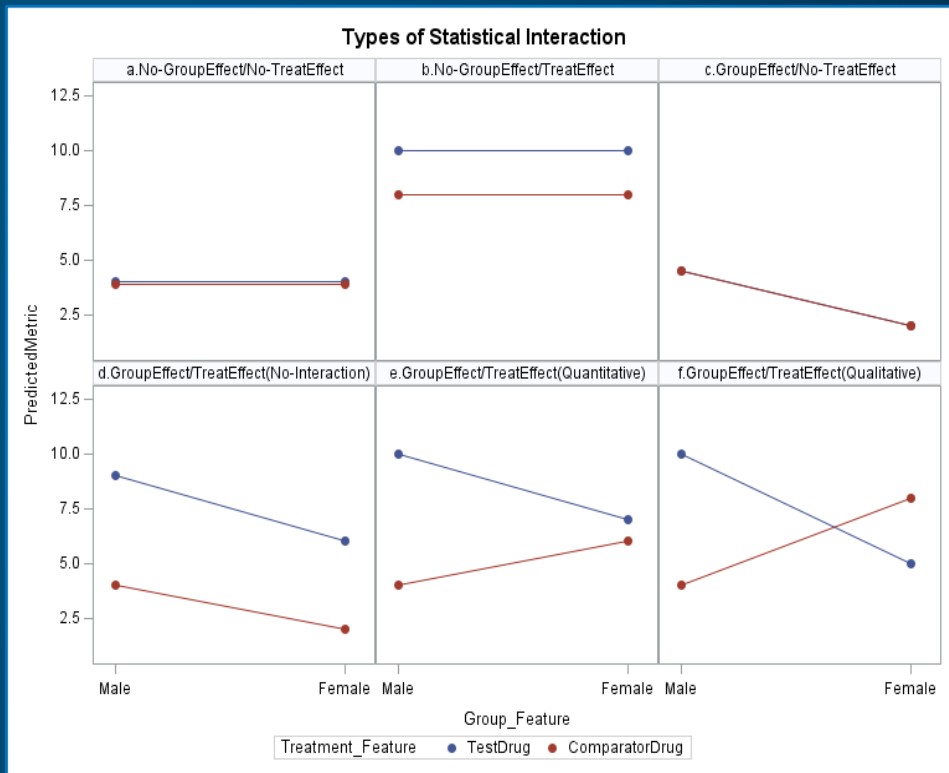
PMFAC Plot

- ❑ Post-hoc model-agnostic technique for checking post-predictive model predictions
- ❑ Sensitivity and discrepancy test check of observed test statistics from the training data compared with predicted test parameters from scored replicate samples of the data
- ❑ Post-hoc model-agnostic check of the model predictions and model inputs based on the observed data and scored replicates to detect variable interactions and association patterns
- ❑ Check model predictions and explore model fit for improvement

Model Features Interaction Types

Intuition behind PMFAC Plot

- (a) Group effect=No, Treatment effect=No
- (b) Group effect=No, Treatment effect=Yes
- (c) Group effect=Yes, Treatment effect=No
- (d) Group effect=Yes, Treatment effect=Yes (no interaction)
- (e) Group effect=Yes, Treatment effect=Yes (quantitative)
- (f) Group effect=Yes, Treatment effect=Yes (qualitative)



Use Case Demonstration

30-Day Hospital readmission risk prediction model

- **Objective:**
 - Use the PMFAC macro program for post-hoc model-fit check and generate model interpretability plots for predictive modeling of the impact of HbA1c measurement on 30-day hospital readmission risk
- **Data:**
 - Diabetes data set downloaded from the UCI Machine Learning Repository
 - Number of patients – 69,980
- **Solution:**
 - Fitted gradient boosting, forest, neural net, and logistic models.
 - PMFAC plots and PD function plots for model fit checks and features association interpretation

Research Article

Impact of HbA1c Measurement on Hospital Readmission Rates: Analysis of 70,000 Clinical Database Patient Records

Beata Strack,¹ Jonathan P. DeShazo,² Chris Gennings,³ Juan L. Olmo,⁴ Sebastian Ventura,⁴ Krzysztof J. Cios,^{1,5} and John N. Clore⁶

¹ Department of Computer Science, Virginia Commonwealth University, Richmond, VA 23284, USA

² Department of Health Administration, Virginia Commonwealth University, Richmond, VA 23298, USA

³ Department of Biostatistics, Virginia Commonwealth University, Richmond, VA 23298, USA

⁴ Department of Computer Science and Numerical Analysis, University of Cordoba, 14071 Cordoba, Spain

⁵ ITIS Polish Academy of Sciences, 44-100 Gliwice, Poland

⁶ Department of Medicine, Virginia Commonwealth University, Richmond, VA 23298, USA

Correspondence should be addressed to John N. Clore; jclore@mcvh-vcu.edu

Received 19 November 2013; Revised 25 February 2014; Accepted 7 March 2014; Published 3 April 2014

Academic Editor: Ali Rizvi

Copyright © 2014 Beata Strack et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

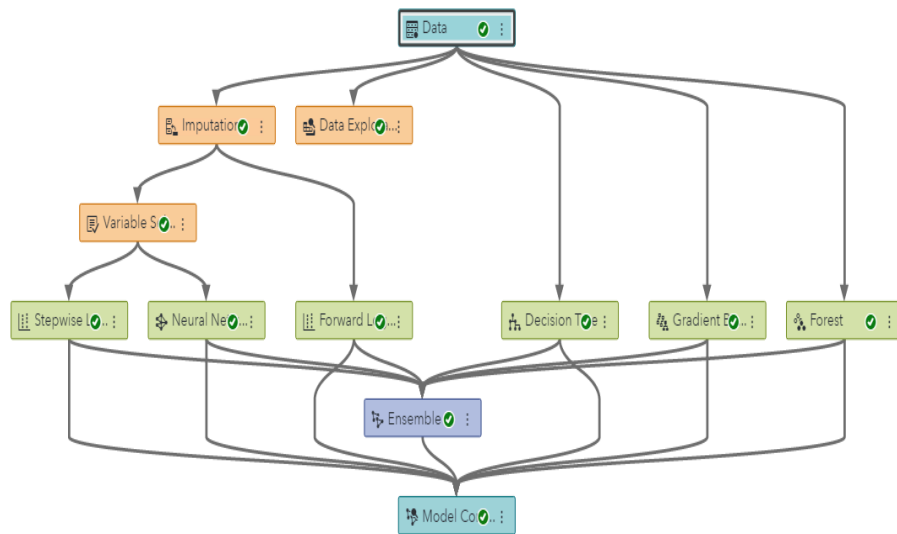
Management of hyperglycemia in hospitalized patients has a significant bearing on outcome, in terms of both morbidity and mortality. However, there are few national assessments of diabetes care during hospitalization which could serve as a baseline for change. This analysis of a large clinical database (74 million unique encounters corresponding to 17 million unique patients) was undertaken to provide such an assessment and to find future directions which might lead to improvements in patient safety. Almost 70,000 inpatient diabetes encounters were identified with sufficient detail for analysis. Multivariable logistic regression was used to fit the relationship between the measurement of HbA1c and early readmission while controlling for covariates such as demographics, severity and type of the disease, and type of admission. Results show that the measurement of HbA1c was performed infrequently (18.4%) in the inpatient setting. The statistical model suggests that the relationship between the probability of readmission and the HbA1c measurement depends on the primary diagnosis. The data suggest further that the greater attention to diabetes reflected in HbA1c determination may improve patient outcomes and lower cost of inpatient care.

Model Studio Pipeline

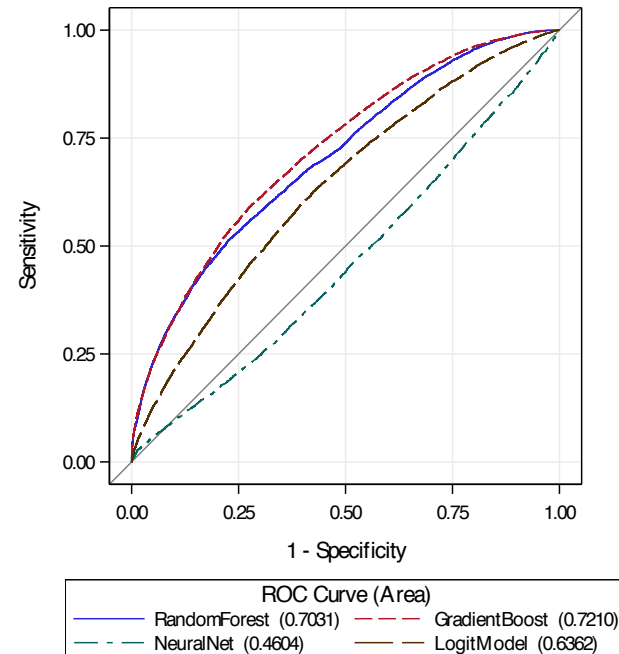
30-Day Hospital readmission risk prediction model

Diabetes Readmission Model Pipeline

Run Pipeline

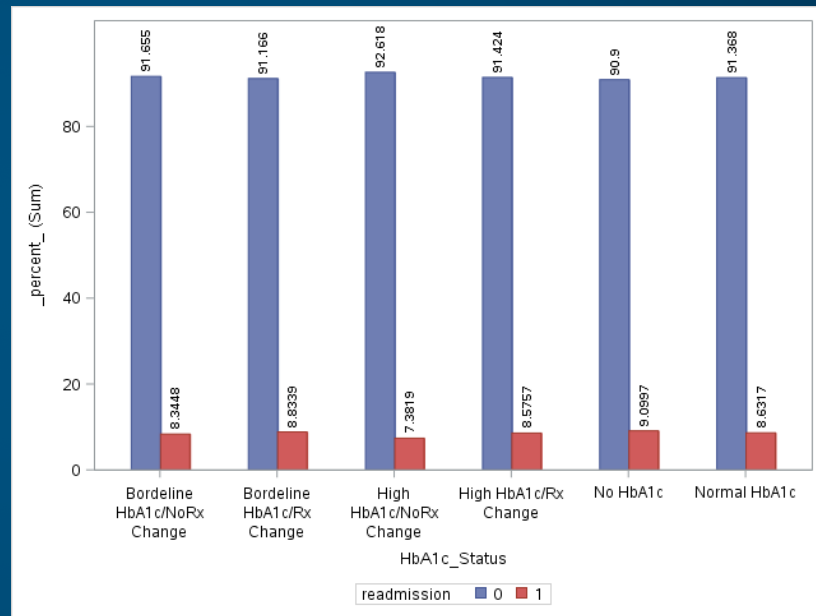
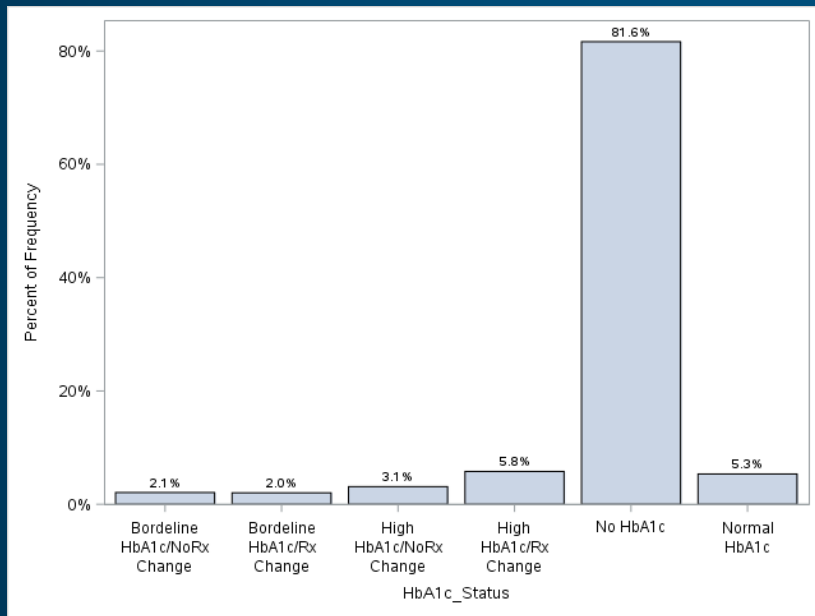


ROC Curves for Comparisons



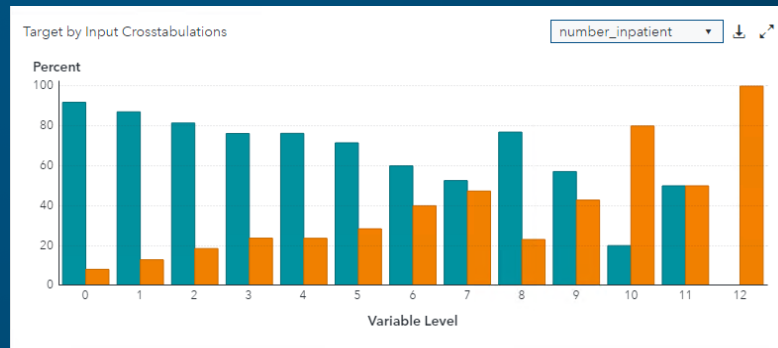
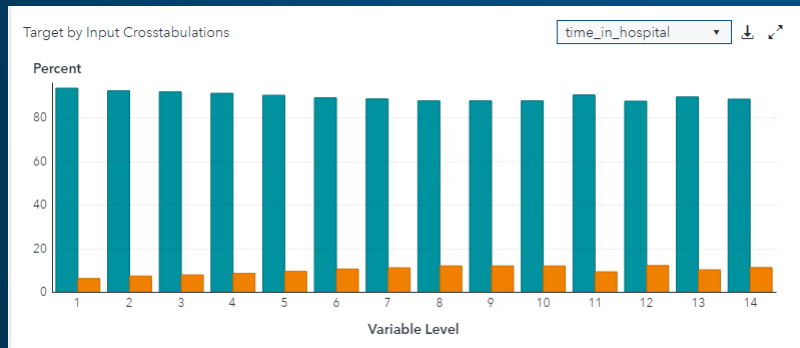
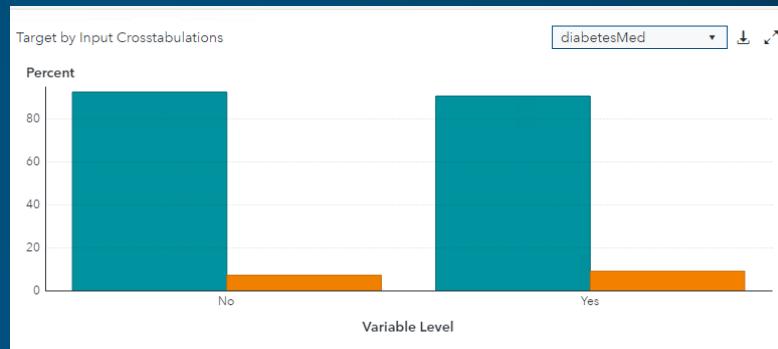
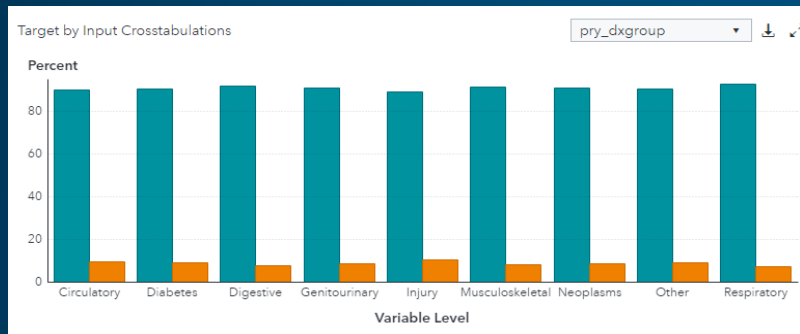
Key Variables Information

30-Day Hospital readmission risk prediction model



Key Variables Information

30-Day Hospital readmission risk prediction model



Model Interpretability (Variable Importance)

30-Day Hospital readmission risk prediction model

Model Variable Importance				
Variable Label	Variable Name	Training Import...	Importance Sta...	Relative Import...
	discharge_disposition	111.7307	0	1
	number_inpatient	105.5622	16.6416	0.9448
	sec1_dxgroup	23.7276	3.8261	0.2124
	pry_dxgroup	22.9711	1.6874	0.2056
	time_in_hospital	11.2107	0	0.1003
	rosiglitazone	10.7459	0	0.0962
	number_diagnoses	8.4724	0	0.0758

Local Model				
Effect	Binary Encoding	Estimate	Degrees of	Error
Intercept	No	0.1458	1	0.0014
discharge_disposition	Yes	-0.0283	1	0.0006
number_inpatient	Yes	-0.0384	1	0.0008
time_in_hospital	Yes	-0.0032	1	0.0007
number_diagnoses	Yes	0.0024	1	0.0005
rosiglitazone	Yes	-0.0044	1	0.0011

Gradient Boosting

Model Variable Importance				
Variable Label	Variable Name	Training Import...	Importance Sta...	Relative Import...
	number_inpatient	24.3776	4.4285	1
	time_in_hospital	10.3765	1.8850	0.4257
	discharge_disposition	6.8122	8.8151	0.2794
	pry_dxgroup	5.3797	1.8106	0.2207
	number_emergency	5.0559	1.7298	0.2074
	sec2_dxgroup	4.2310	1.8511	0.1736
	age	4.0479	1.8148	0.1660

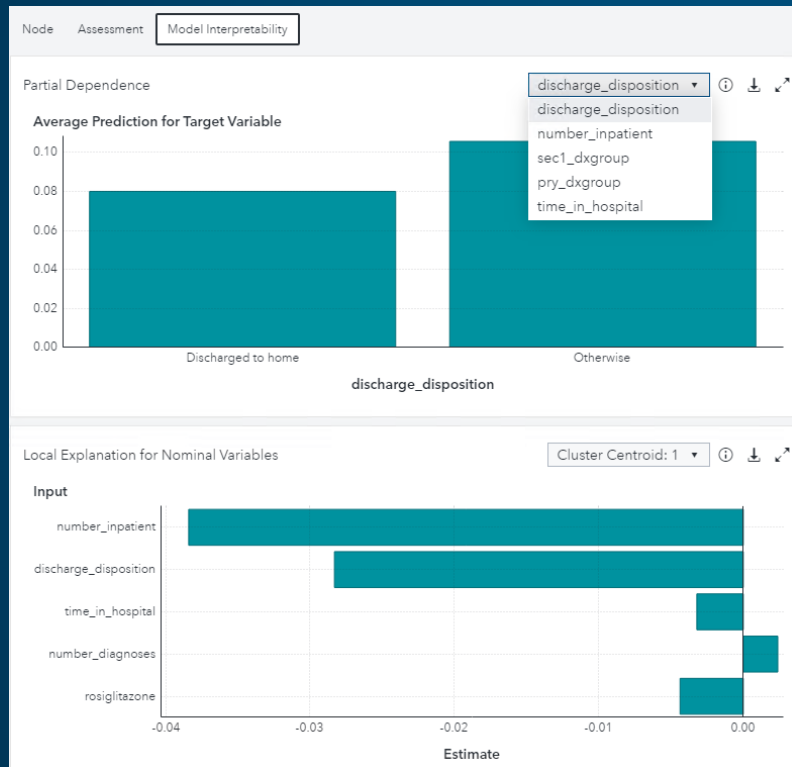
Local Model				
Effect	Binary Encoding	Estimate	Degrees of Free...	Standard Error
Intercept	No	0.1117	1	0.0092
number_inpatient	Yes	-0.0342	1	0.0007
discharge_disposition	Yes	-0.0170	1	0.0005
number_emergency	Yes	-0.0050	1	0.0008
repaglinide	Yes	-0.0066	1	0.0019
chlorpropamide	Yes	0.0299	1	0.0090

Forest Model

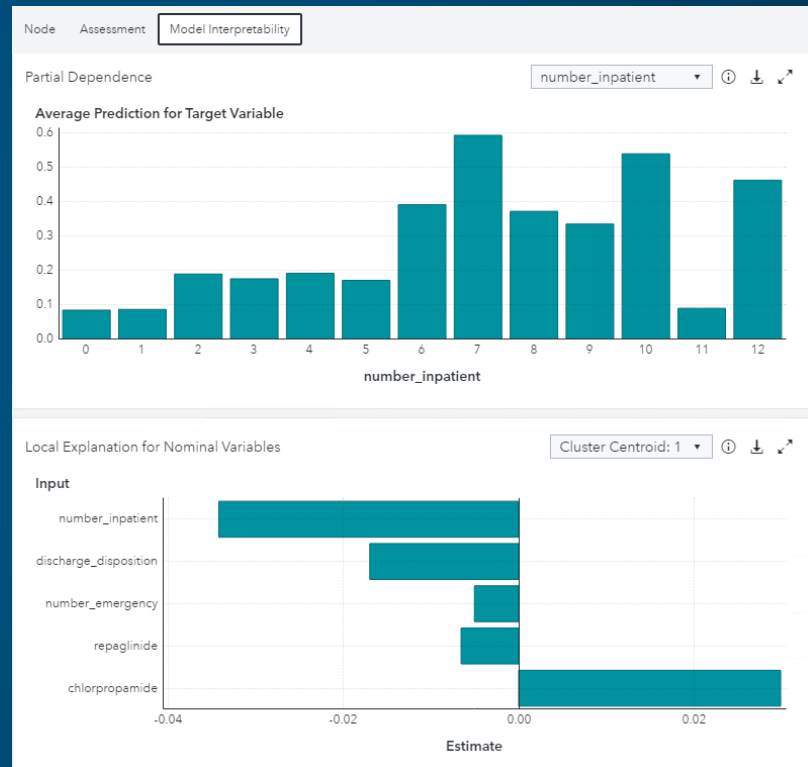


Model Interpretability (PD & ICE Plots)

30-Day Hospital readmission risk prediction model



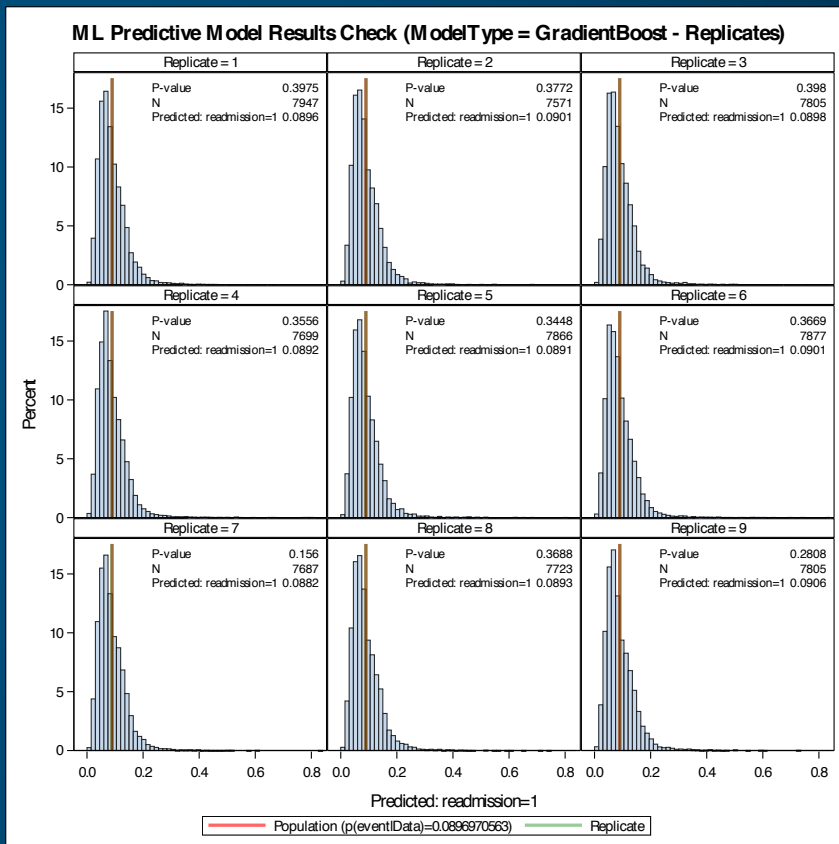
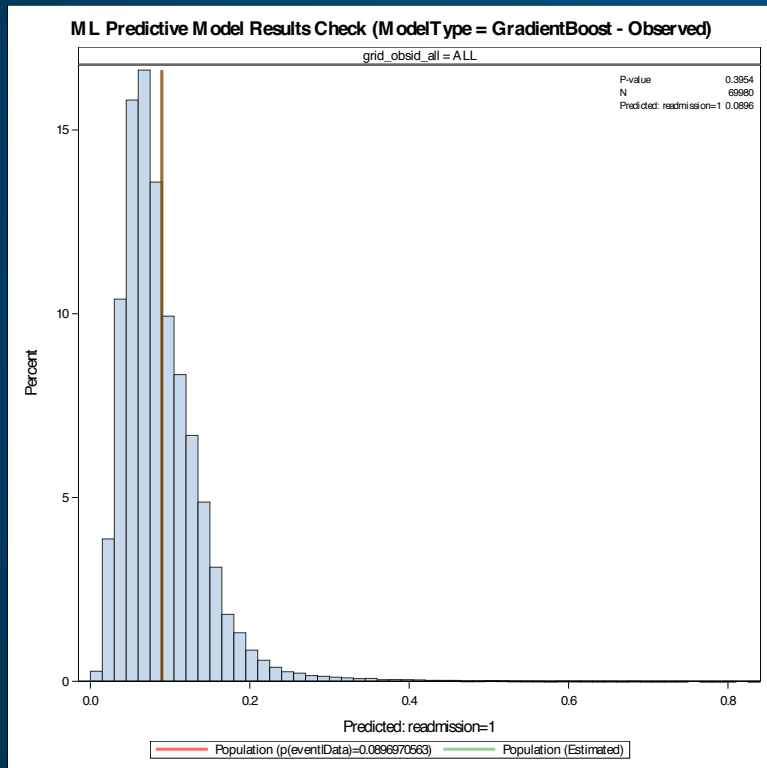
Gradient Boosting



Forest Model

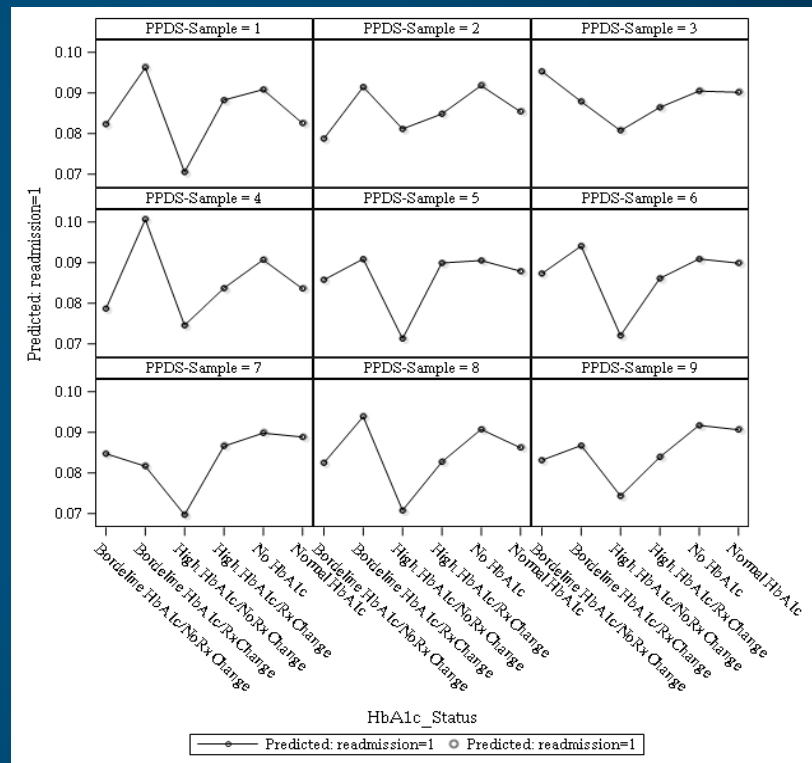
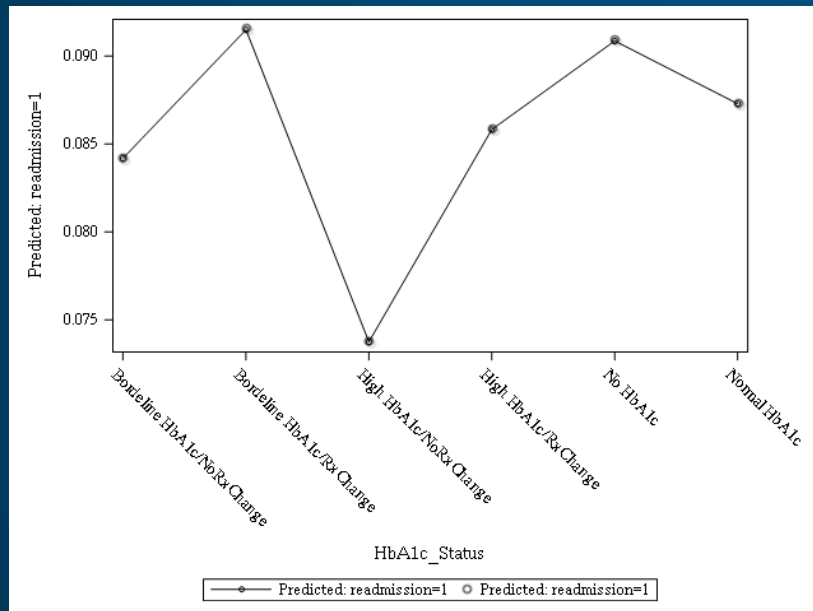
PMFAC Plot for Model Fit Check

30-Day Hospital readmission risk prediction model



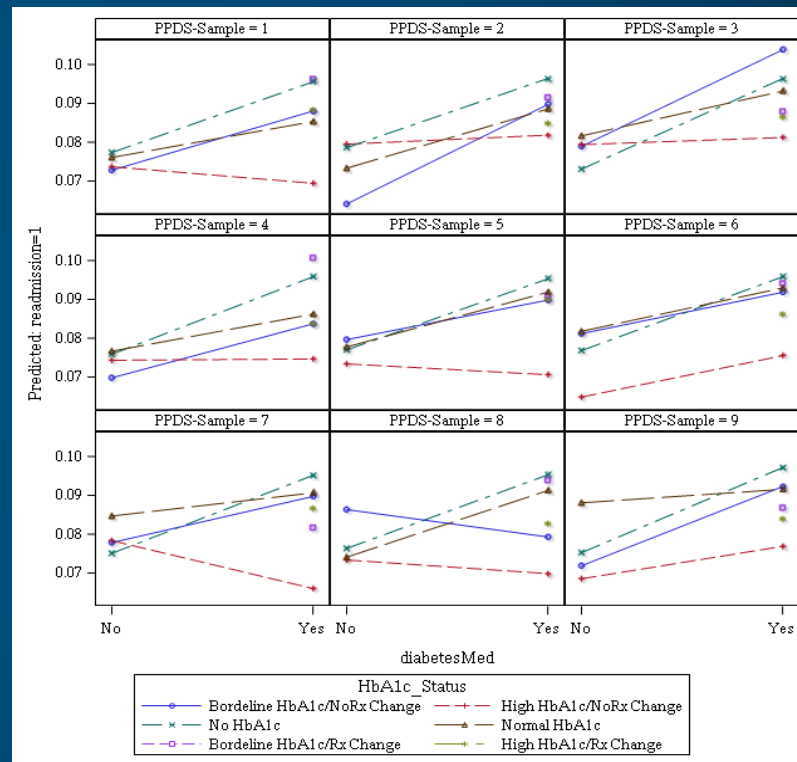
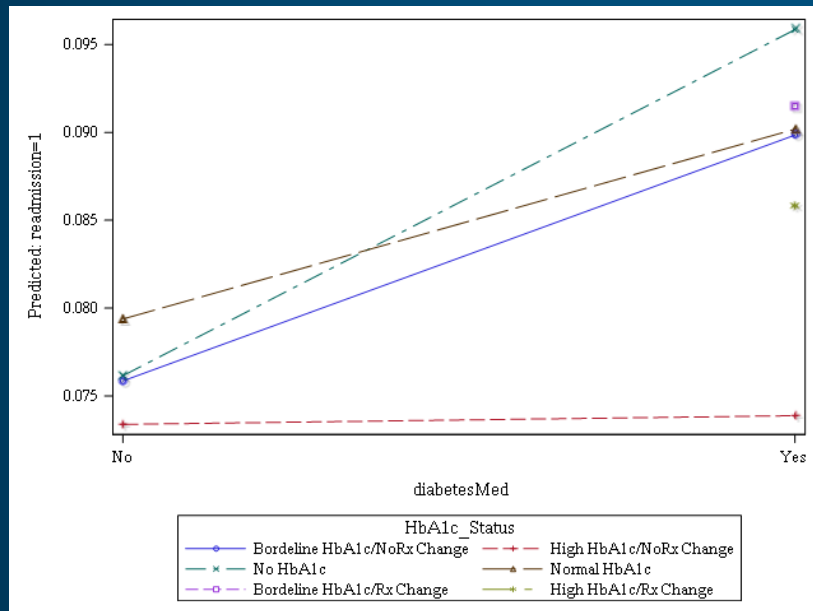
PMFAC Plot

HbA1c status



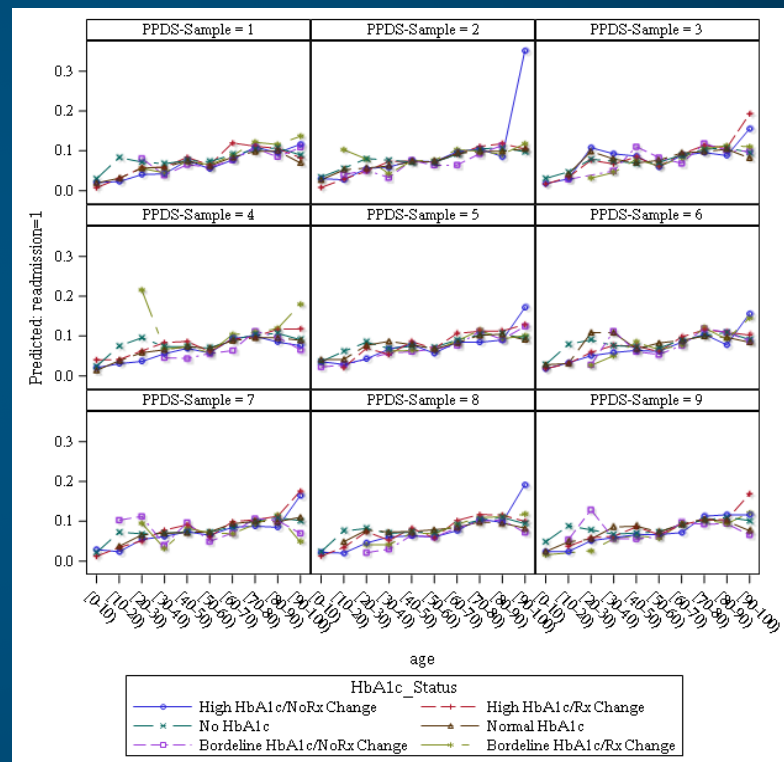
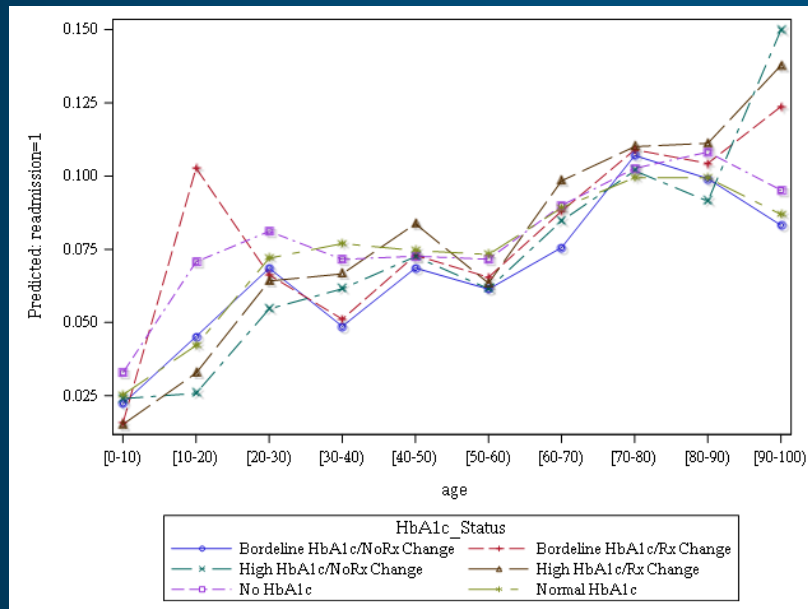
PMFAC Plot

Diabetes medication and HbA1c status



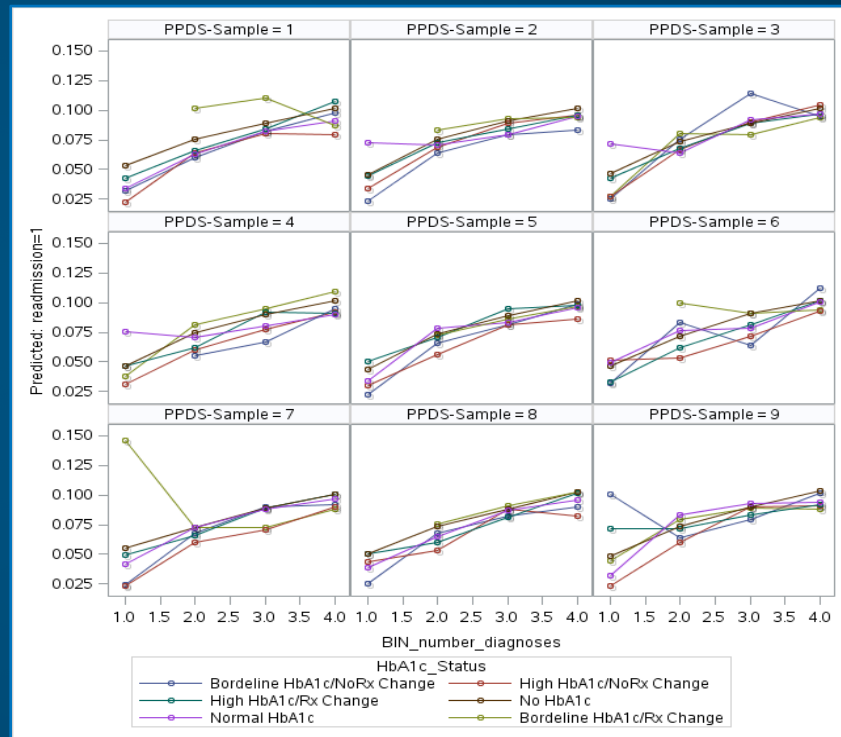
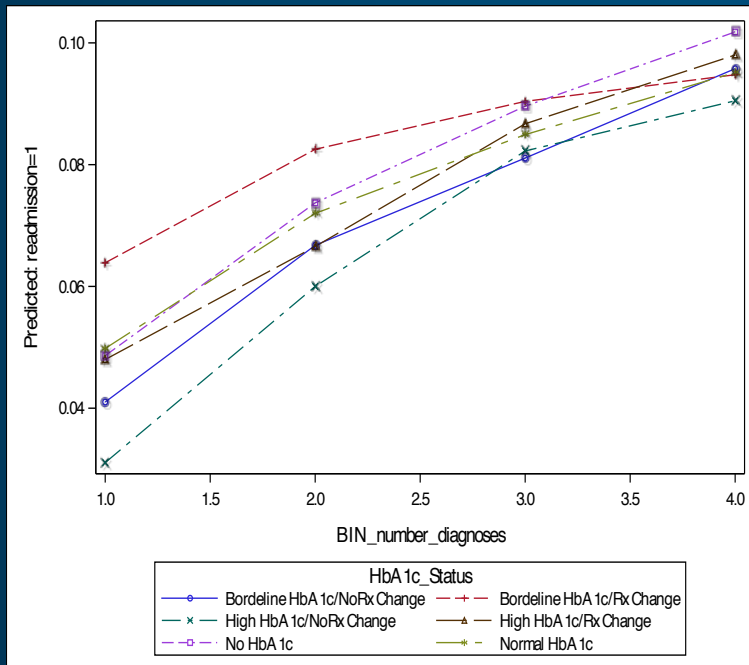
PMFAC Plot

Age group and HbA1c status



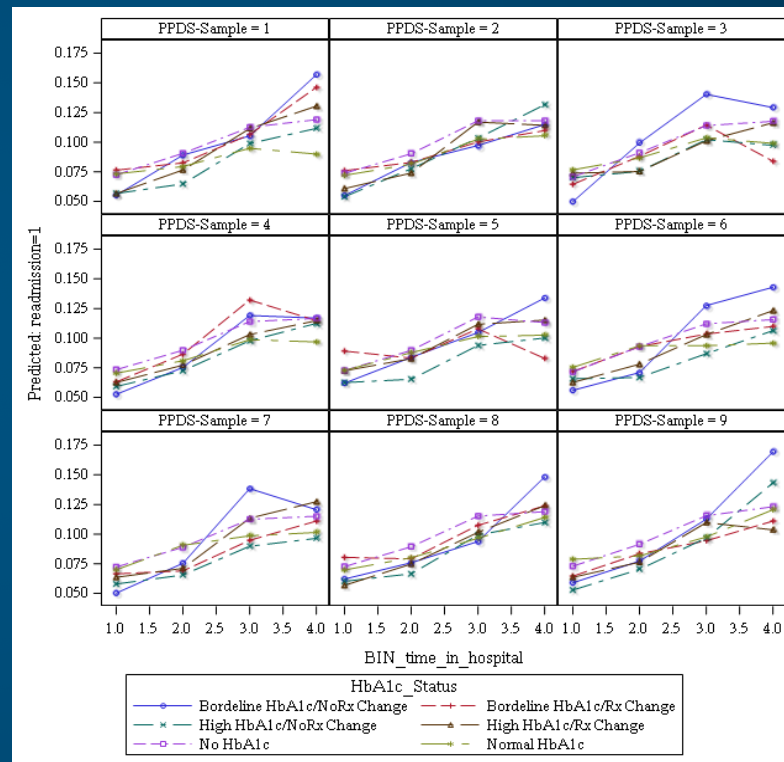
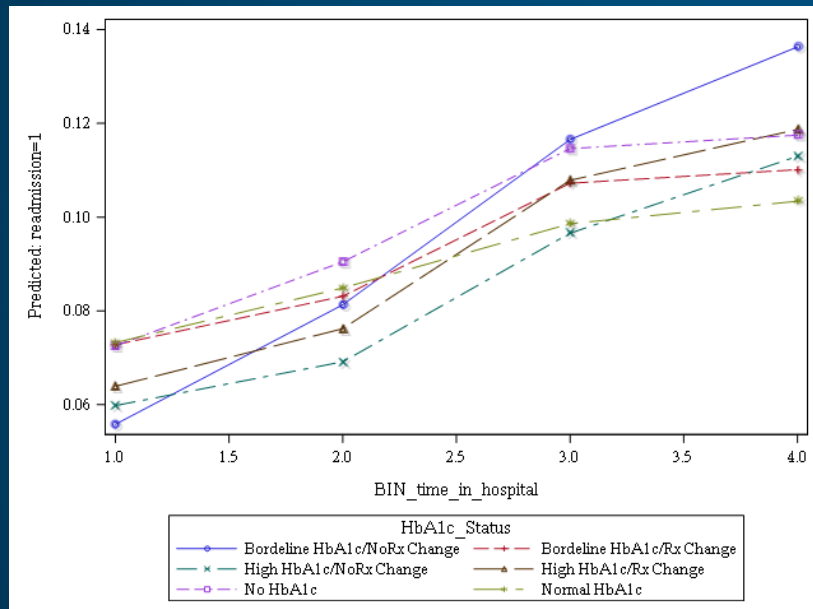
PMFAC Plot

Number of diagnoses (binned) and HbA1c status



PMFAC Plot

Length of stay (binned) and HbA1c status



Takeaways

- Machine learning interpretability tools
 - explain predictions from complex black box models
 - explain dynamic relationship between model inputs and output predictions
- PMFAC plots
 - check for consistency and reproducibility of ML model results
 - reveal input variables dependency and interactions
 - show how changes in model input variables can influence model predictions in different population samples



Questions?

David Olaleye

David.Olaleye@sas.com

<https://www.linkedin.com/in/dave-olaleye-29005353/>