Improved Predictive Models for Readmission of Patients with Diabetes

Chathurangi Heshani Karunapala Pathiravasan(chathurangi@siu.edu),
R. W. M. A. Madushani, PhD(anusha.wasalamudiyanselage@medicine.ufl.edu), *Gabrielle LaRosa* (gal41@pitt.edu)
Southern Illinois University, University of Florida, University of Pittsburgh

Problem Statement

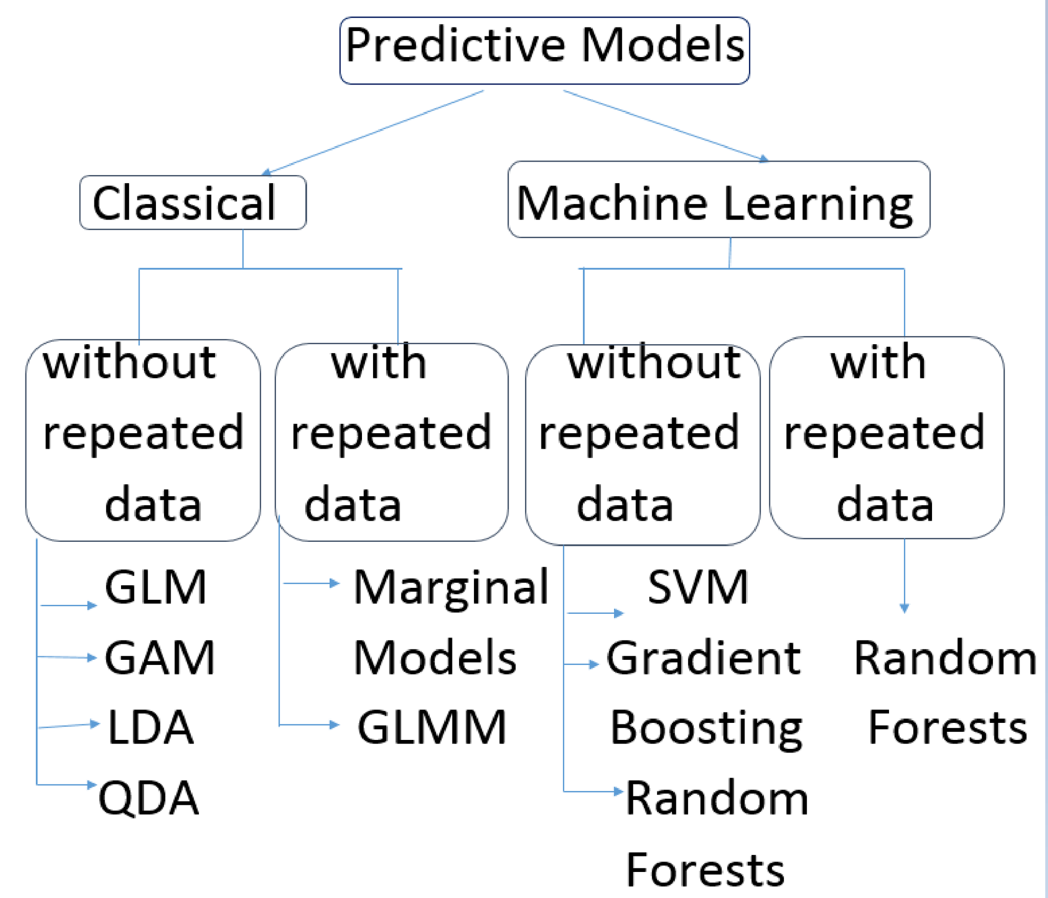
- Examining the historical patterns of diabetics care is very essential which leads to improvements in patient safety and prevent future readmissions.

- This improves the quality of health care and reduces the medical expenses on readmission.

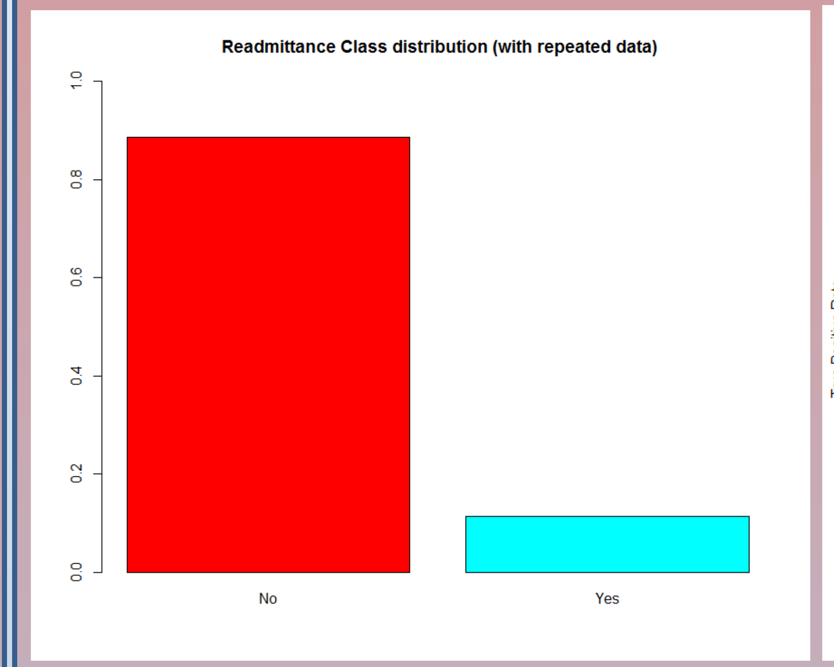
Goals and Objectives

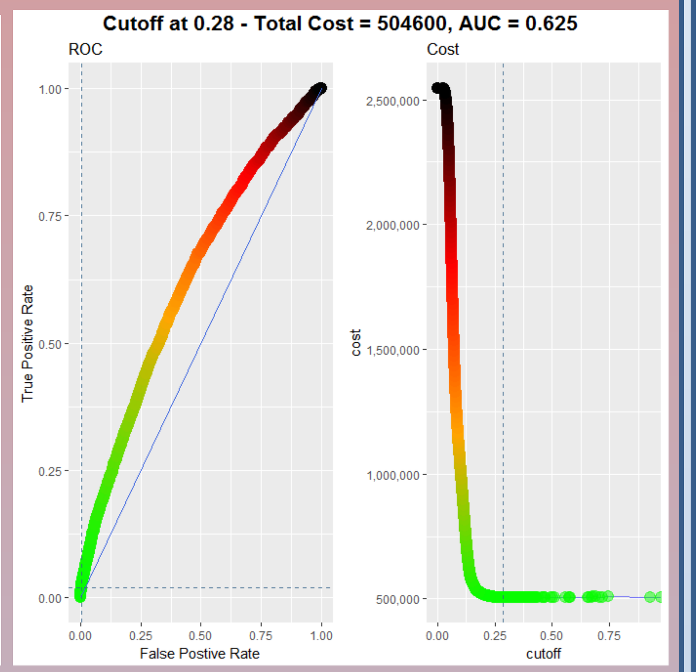
- To build accurate predictive models for hospital readmission of patients with diabetes.
- Identifying key contributing factors of readmission.

Methods



- The data set is obtained from UCI machine learning repository and it is highly imbalanced and contains multiple patient visits. We used ROC and AUC to compare models.





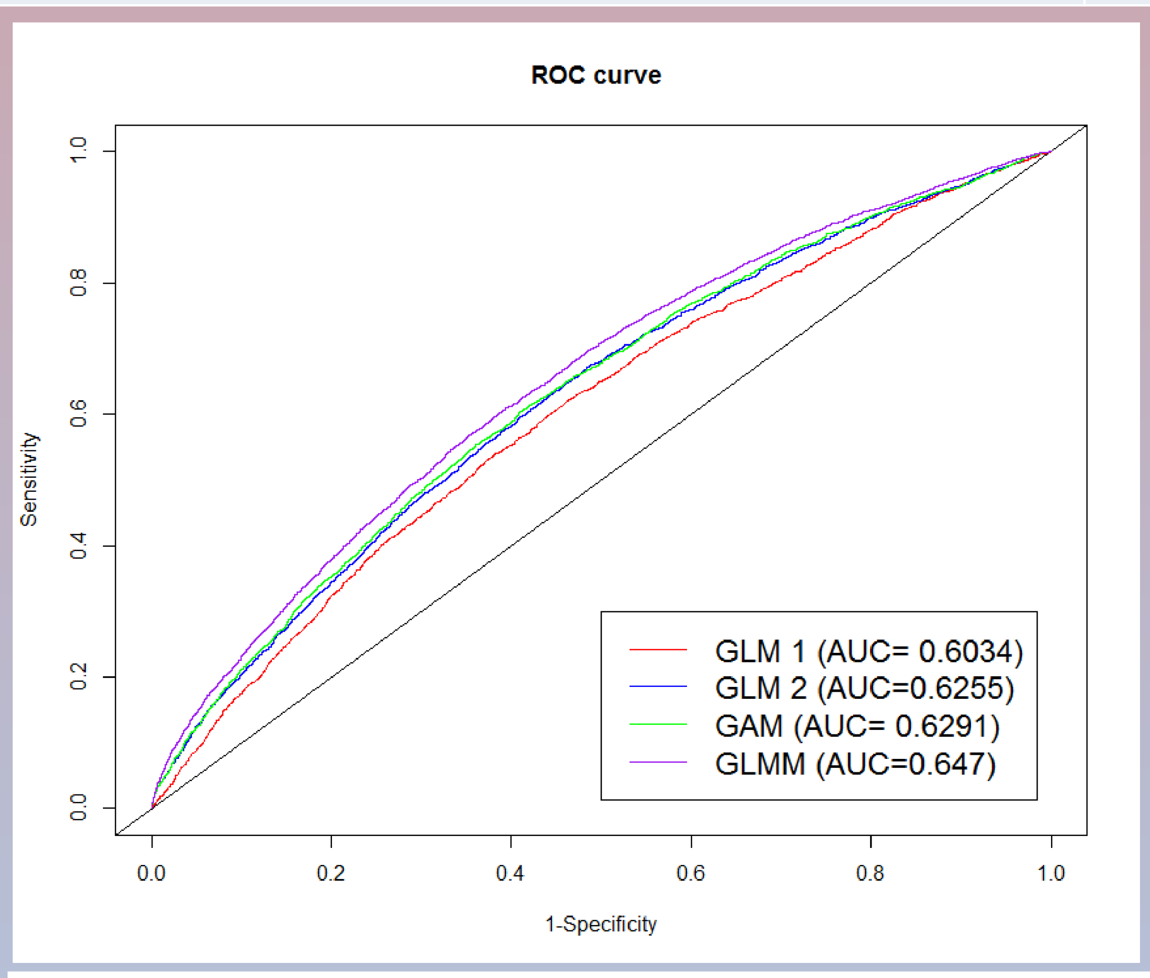
- We use 19 predictors (8 continuous predictors and 11 categorical predictors).

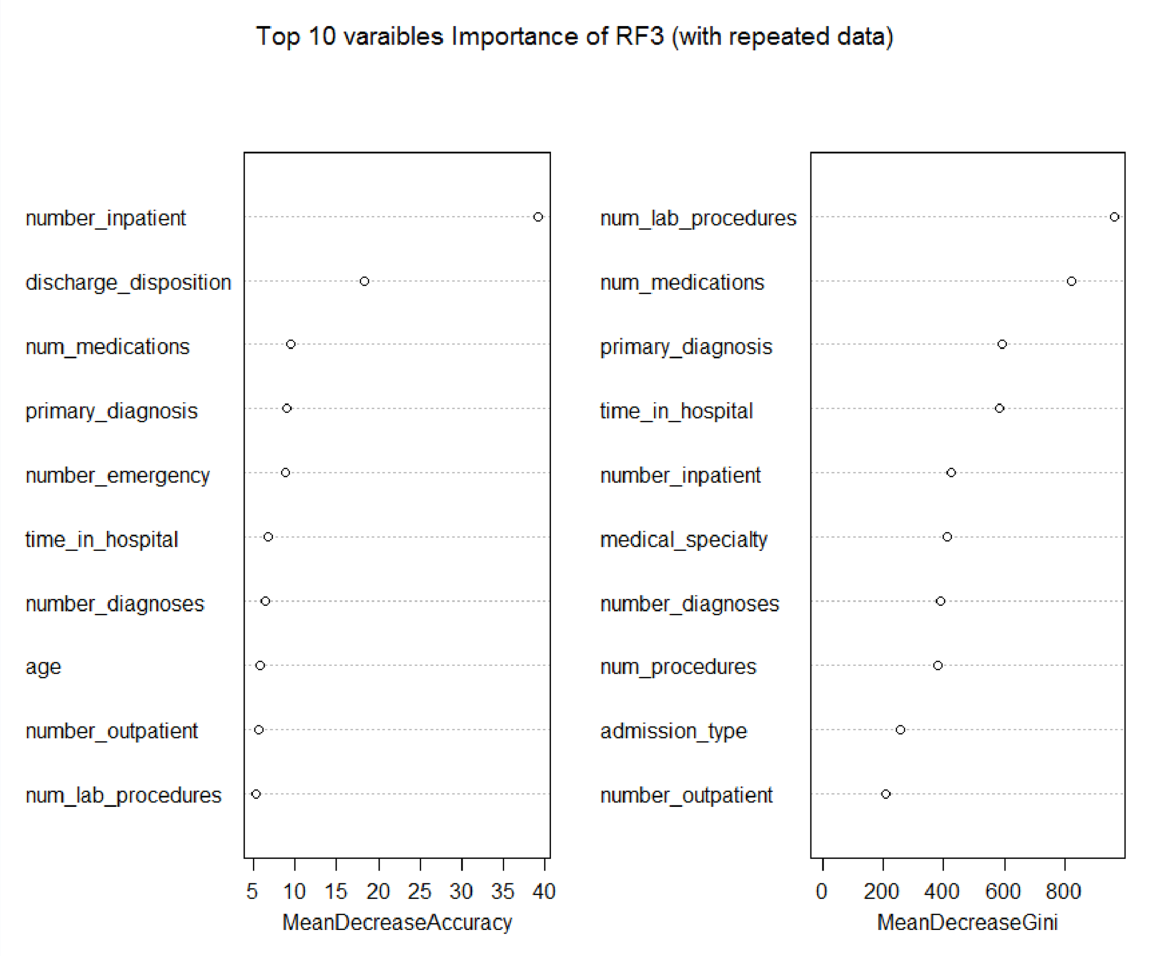
Results

Model	AIC	AUC
GLM 1 (Base Model)	24733.7	0.6034
GLM 2 (with 19 predictors)	24509.1	0.6255
GLM 3 (backward elimination sub model)	24491.3	0.6239
GAM (all excluding number of procedure)	24482.4	0.6291

Model	AUC
LDA (Linear Discriminant Analysis)	0.6016
QDA (Quadratic Discriminant Analysis)	0.6018
MM (Marginal Models with repeated data)	0.6480
GLMM (Generalized Linear Mixed Models with repeated data)	0.6470

Model	AUC
NB (Naïve Bayes)	0.6249
RF1 (Random Forest without repeated data)	0.6360
RF2:Dealing with imbalanced (without repeated)	
Over Sampling	0.6547
Under Sampling	0.6527
RF3 (Random Forest with repeated data)	0.6560
SVM (Support vector machine: 19 Predictors)	0.5416
EGB(Extreme gradient boosting)	0.5835





Conclusions

- Compared to the base model GLM model 2 has more predictability.
- GAM model is a better fit (AIC lower) and it has more predictive power than GLM models.
- Naïve Bayes model has better predictability compared to GLM.
- Substantial improvement in the predictability, when we consider marginal and mixed effect models.
- Random Forests performs well compared to other machine learning approaches.
- Number of lab procedures, number of medications, time in hospital, number of diagnoses, number of procedure, primary diagnosis and medical specialty are the main contributing factors of readmission.

References

[1] B. Strack, et al.. Impact of hba1c measurement on hospital readmission rates: analysis of 70,000 clinical database patient records. BioMed research international, 2014.
[2] A. Frank and A. Asuncion, UCI Machine Learning Repository, University of California, School of Information and Computer Science, 2010.

Acknowledgement

This project is under the grant NSF CCF0939370: Center for Science of Information (https://www.soihub.org)