NATIONAL CENTER FOR HEALTH STATISTICS

# Using the National Health Interview Survey Linked Mortality Files to Train and Assess the Performance of Machine Learning Prediction Models to Predict All-Cause Mortality and Interpret Model Predictions using Explainable AI

Orlando Davy, MPH; Frances McCarty, PhD; Yulei He, PhD; Cordell Golden, MPS

## Background

- NCHS Data Linkage Program produces high-quality linked data files that can be used for a wide-range of health-related research topics and data science applications
- Machine learning (ML) prediction models can be used to predict outcomes with high predictive accuracy, but require quality and accurate data for training
- Explainable AI (XAI) involves visual global and local methods used to enhance the understanding of complex ML models overall and make model predictions interpretable

## Objectives

- Evaluate selected ML prediction models using linked data as the training data and validation source to assess model performance for predicting all-cause mortality
- Use XAI to interpret model results

## Methods

### Data Source

- Public-use 2000 - 2001 National Health Interview Survey (NHIS) Linked Mortality Files (LMF) with mortality information through 2019
- IPUMS NHIS harmonized data used for the analysis
- Linkage eligible NHIS sample adult respondents with complete information for selected predictor variables (n = 46,949)

### Data Analysis

- Analysis performed using R v 4.3.1, package included: caret v6.0-94, pdp v0.8.1, lime v0.5.3, and yardstick v1.2.0
- Selected ML prediction models: Random Forest (RF) , Gradient Boosting Machine (GBM), Support Vector Machine (SVM), Naïve Bayes (NB)
- XAI methods used: Variable Importance Plots (VIP); Partial dependence plots (PDP), LIME Local Features Plot
- Initial feature set included 20 predictor variables plus one target (outcome) variable
- Only records with complete information for all predictor variables were included
  - Training set: 2000 NHIS LMF (n = 23,210)
  - Validation set: 2001 NHIS LMF (n = 23,739)

Steps:

1. VIP used to identify ten most important variables (2000 NHIS LMF)
2. ML models trained using most important variables (2000 NHIS LMF)
3. Trained models were re-run using 2001 NHIS LMF
4. Model results compared to outcome from 2001 NHIS LMF
5. Precision-Recall curves (PR-AUC) used to visually assess model performance
6. Additional metrics evaluated to assess performance: Misclassification Error, Precision, Recall, Balanced Accuracy, F1-Score, and Processing Time

## Results

- Older age (70+ years), hypertension, diabetes, widow/divorced/separated, and inactivity were consistent top 10 predictors for all-cause mortality across all models
- The GBM model had the lowest error rate, highest F1-score and PR-AUC estimate
- SVM precision was greater compared to the other models, but had the lowest recall
- Processing time was reasonable across models, the NB model was the most efficient
- Partial dependence shows older age groups tend to have a more positive impact on predicting the correct class (Mortality = "Yes")

## CONCLUSIONS

- The ML models performed reasonably well when predicting all-cause mortality when trained with high-quality data

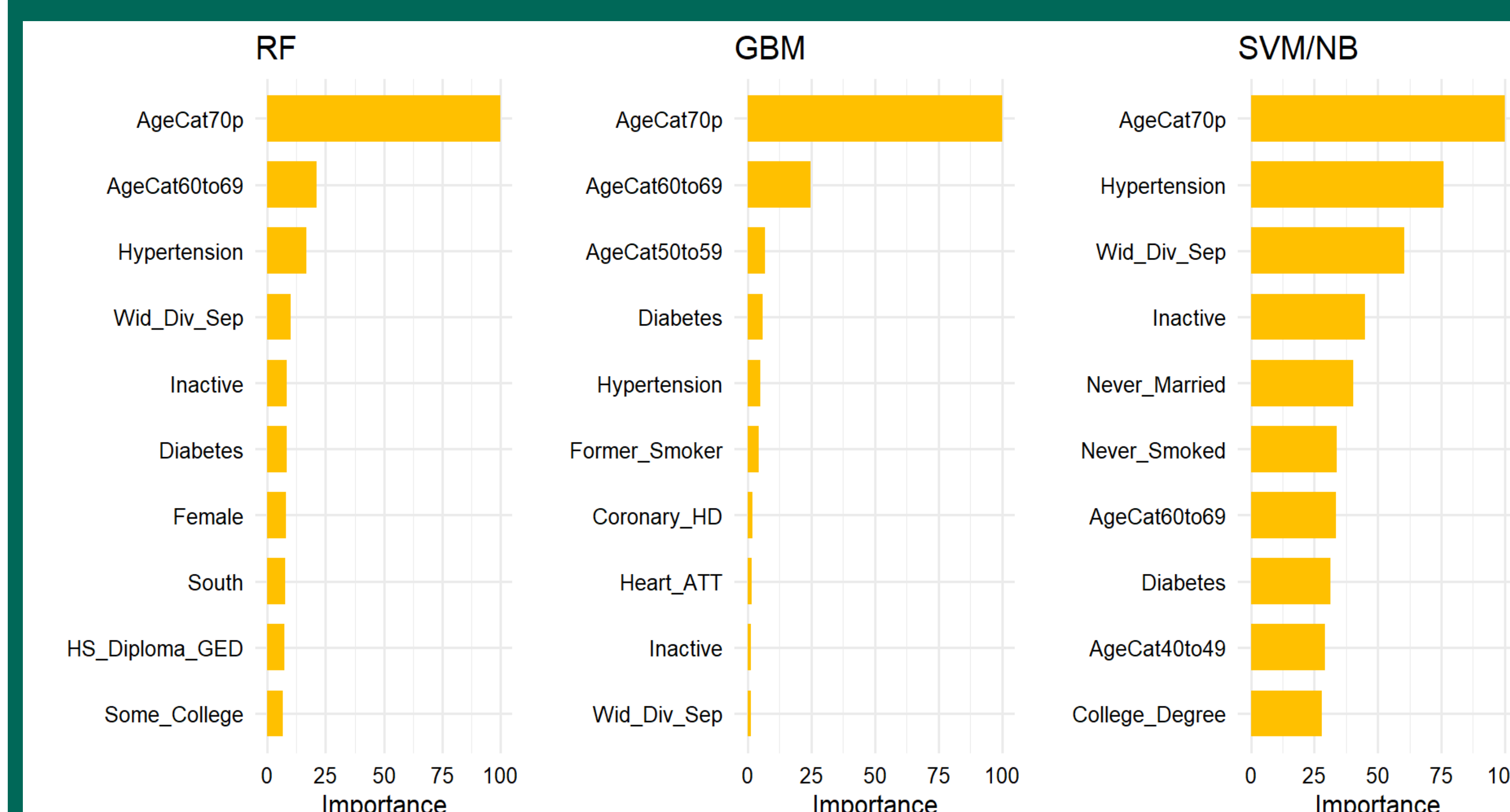
## Limitations

- Only complete cases used for analysis
- Information for predictor variables derived from self-reports

## Target Variable* and Feature Set

1. Age
2. Sex
3. Race and Ethnicity
4. Education
5. Marital Status
6. Poverty to Income Ratio
7. Health Insurance
8. Inactivity
9. Smoking Status
10. Excessive Alcohol Consumption
11. Body Mass Index
12. Hypertension
13. Diabetes
14. Coronary Heart Disease
15. Heart Condition
16. Heart Attack
17. Place for Care
18. Barrier to health Care: Cost
19. Psychological Distress
20. Region
21. **Mortality***

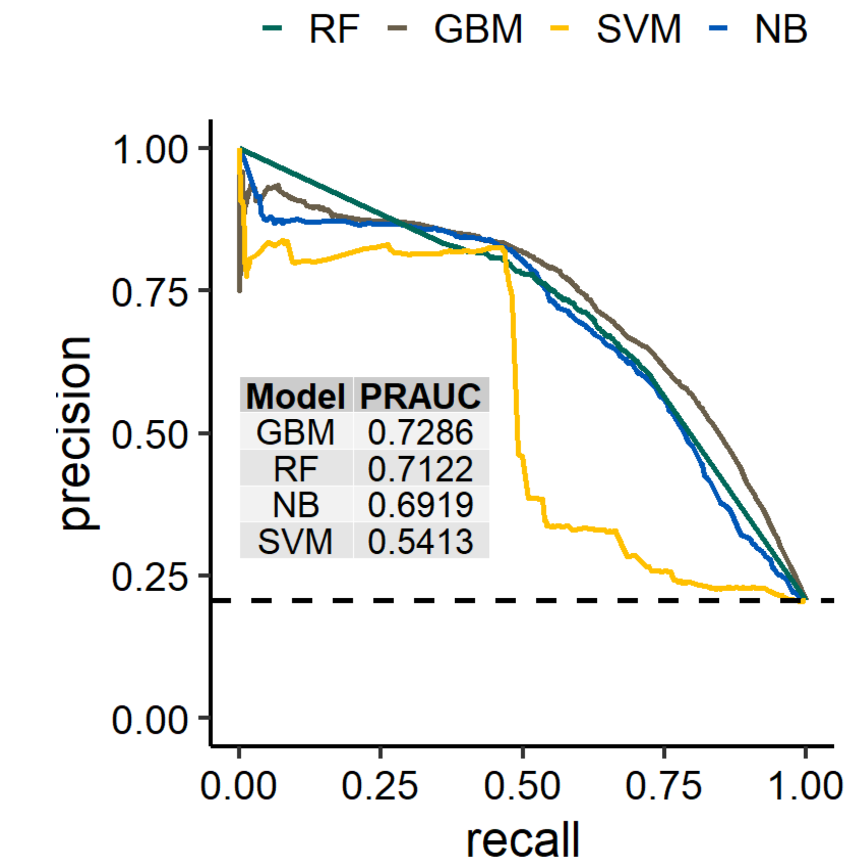
## Variable Importance

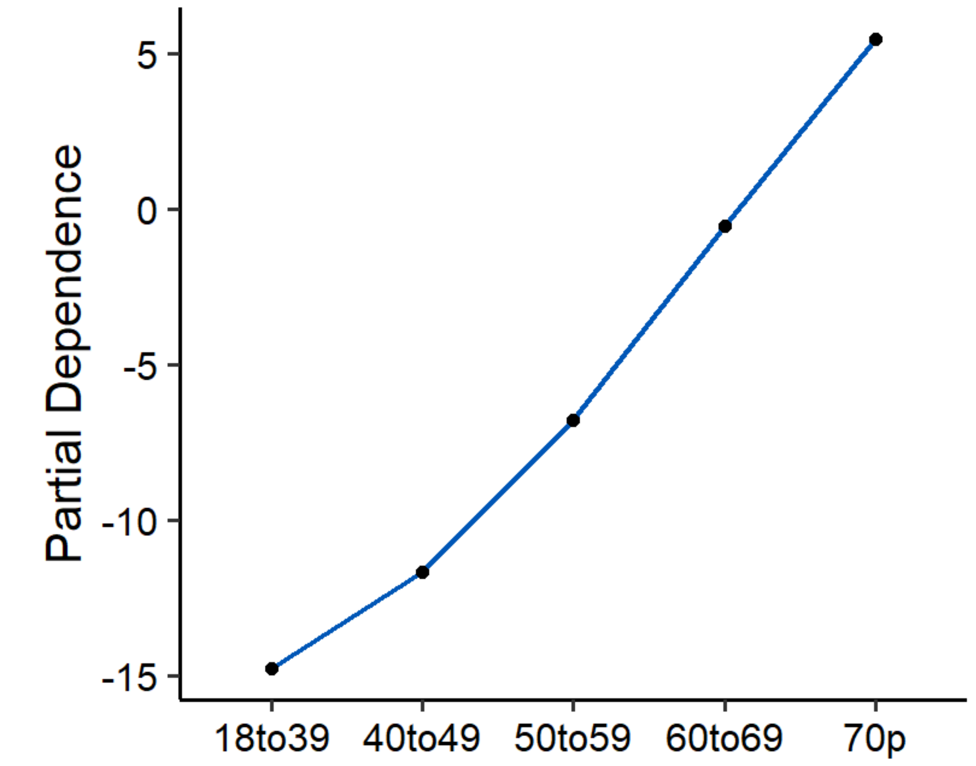



## Performance Measures

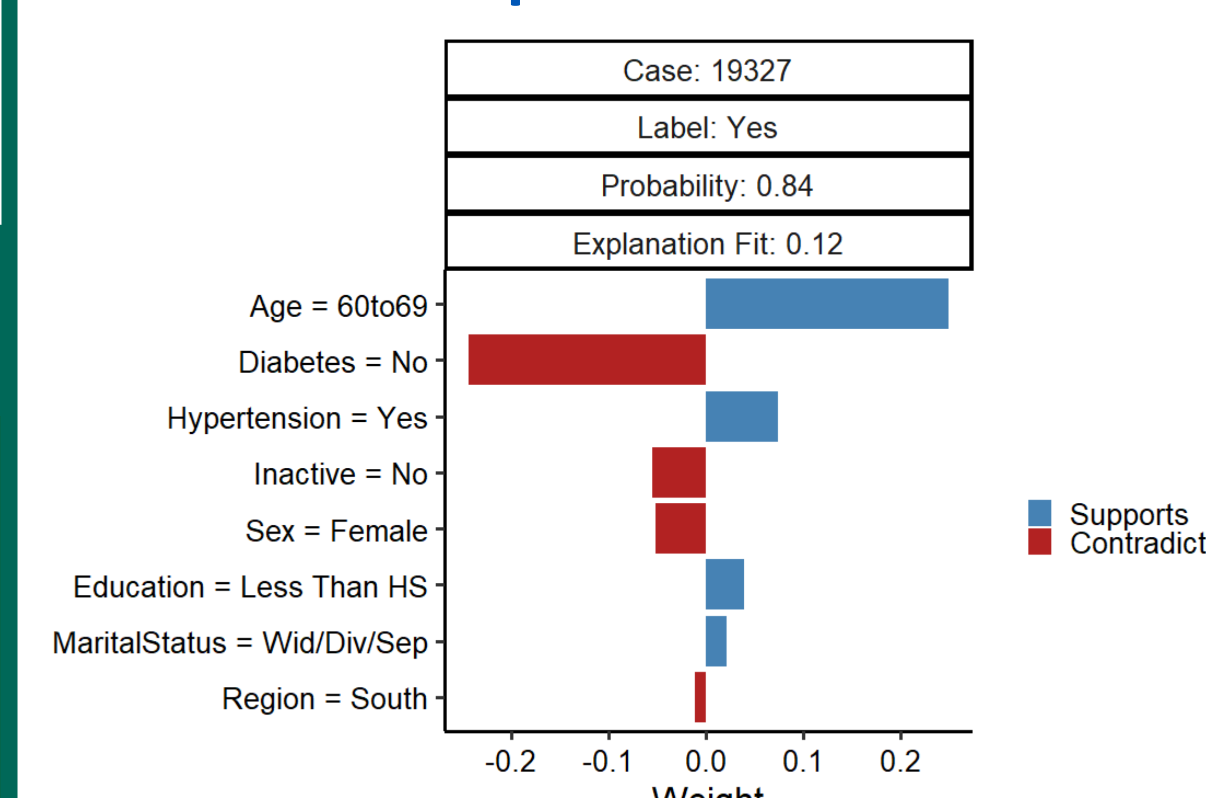
| | RF | GBM | SVM | NB |
|---|---|---|---|---|
| Misclassification Error | 0.1305 | 0.1239 | 0.1320 | 0.1560 |
| Precision | 0.7304 | 0.7752 | 0.8262 | 0.6044 |
| Recall | 0.5826 | 0.5957 | 0.4562 | 0.7051 |
| Balanced Accuracy | 0.7633 | 0.7723 | 0.7156 | 0.7926 |
| F1-Score | 0.6481 | 0.6648 | 0.5878 | 0.6509 |
| Process Time(min) | 3.20 | 0.58 | 1.00 | 0.02 |

## Global Procedures




## An Example of a Local Procedure




CONTACT INFO: odavy@cdc.gov