Chronic Kidney Disease Prediction

Project #2

Christine Orosco

2/4/2023

Agenda

- Introduction
- Methods
- Analysis
- Conclusion

Introduction

Objective	• Chronic Kidney Disease Prediction • Binary Classification
3 Models	• Random Forest Classifier • AdaBoost • XgBoost
Dataset	• Study conducted *(Dua and Graff, 2019)* • 400 observations
Validation	• Validation dataset (Atul, n.d.) • CURE-CKD - *(Tuttle et al., 2019)* • CRIC (*Chronic-Kidney-Disease Research Chronic Renal Insufficiency Cohort Study Kidney Disease*, n.d.)

Methods

EDA

- Correlations
- Linear Relationship

Feature Selection

- RFE
- RFECV

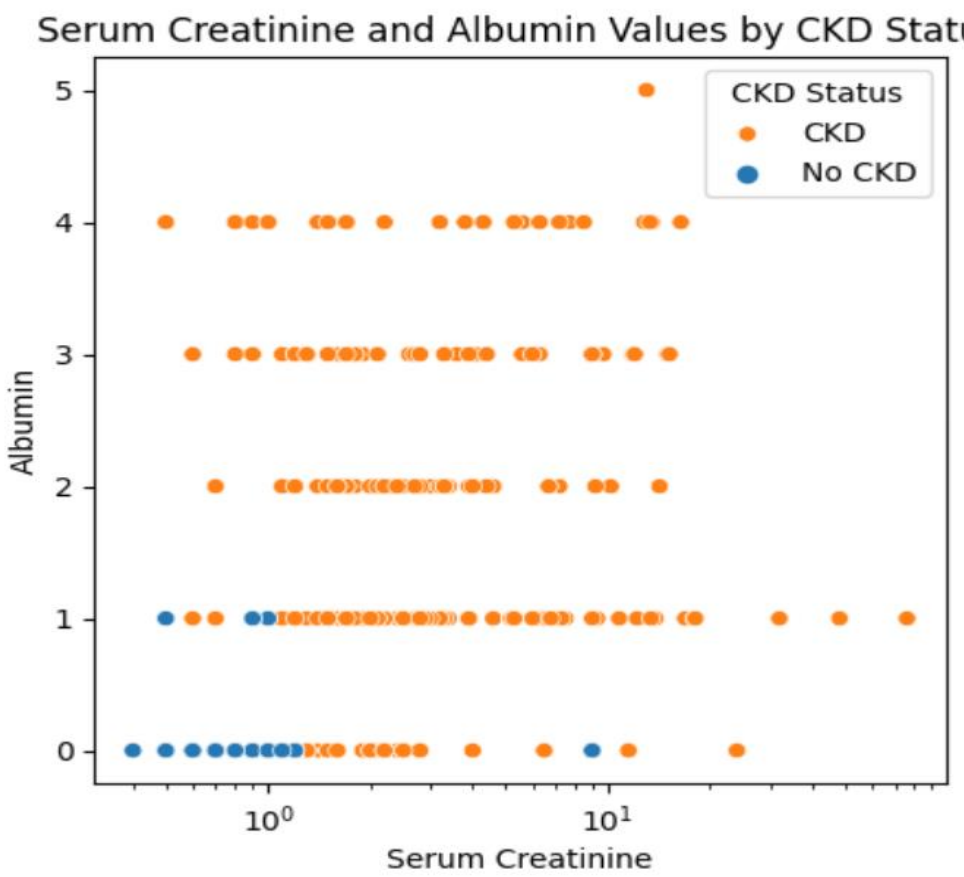
Build and Fit Models

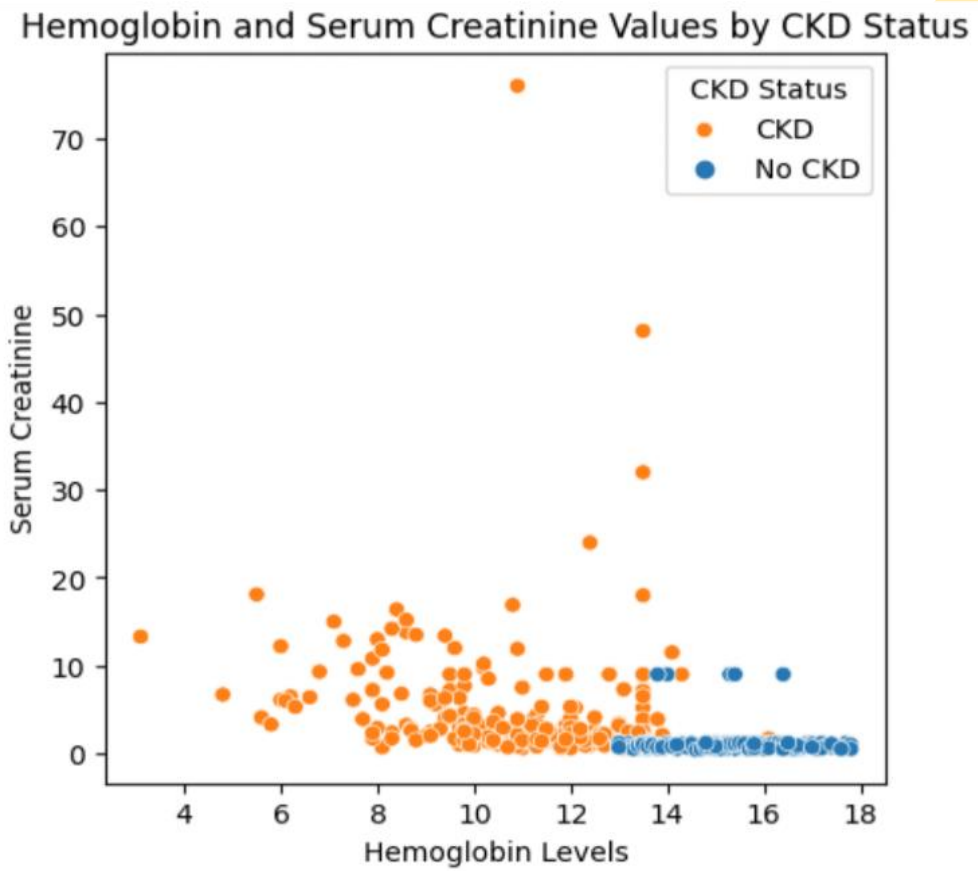
- Cross-Validation
- GridSearch
- Compute predictions

Compare results

- Validation dataset
- Published studies

Data Exploration

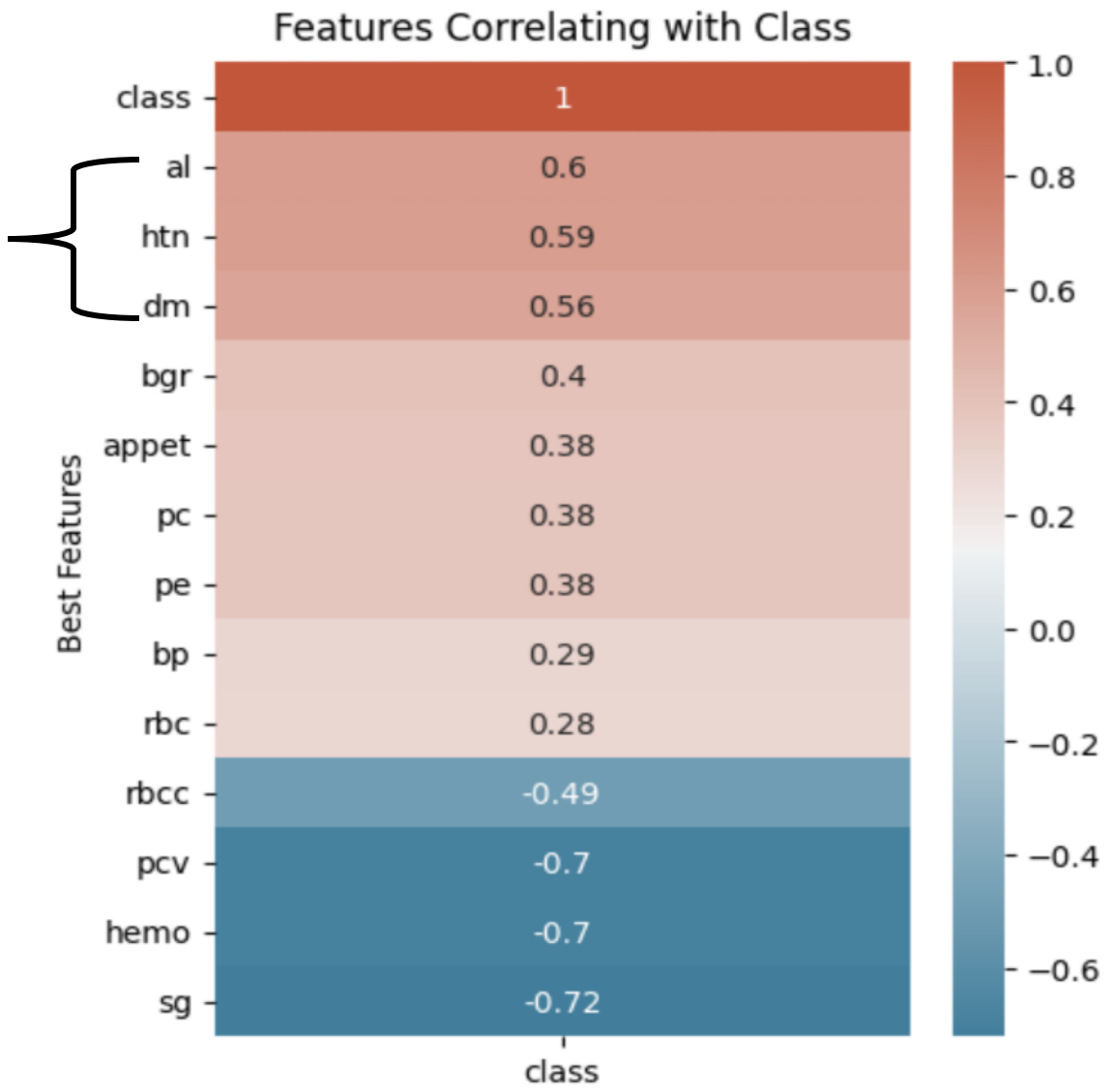




CRIC and CURE-CKD confirm:

- Diabetes and hypertension 2 leading cause of CKD
- Albumin to Creatine levels are elevated in CKD

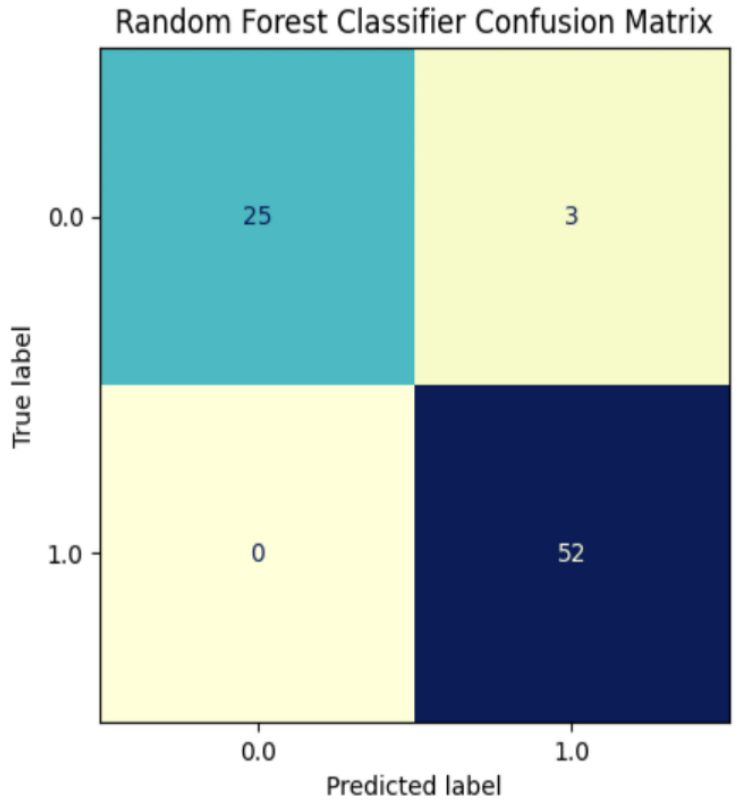
Feature Selection

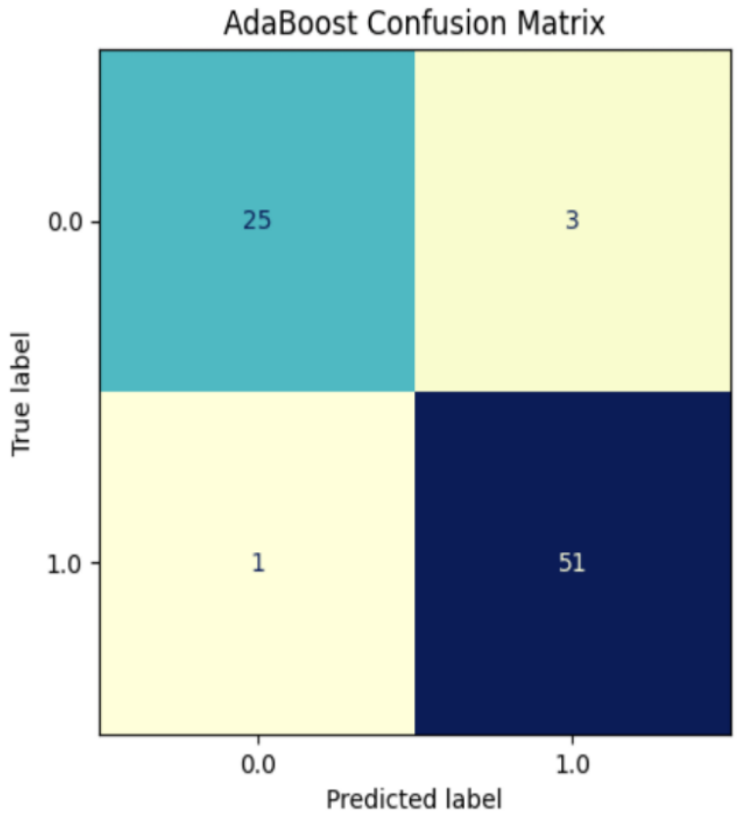


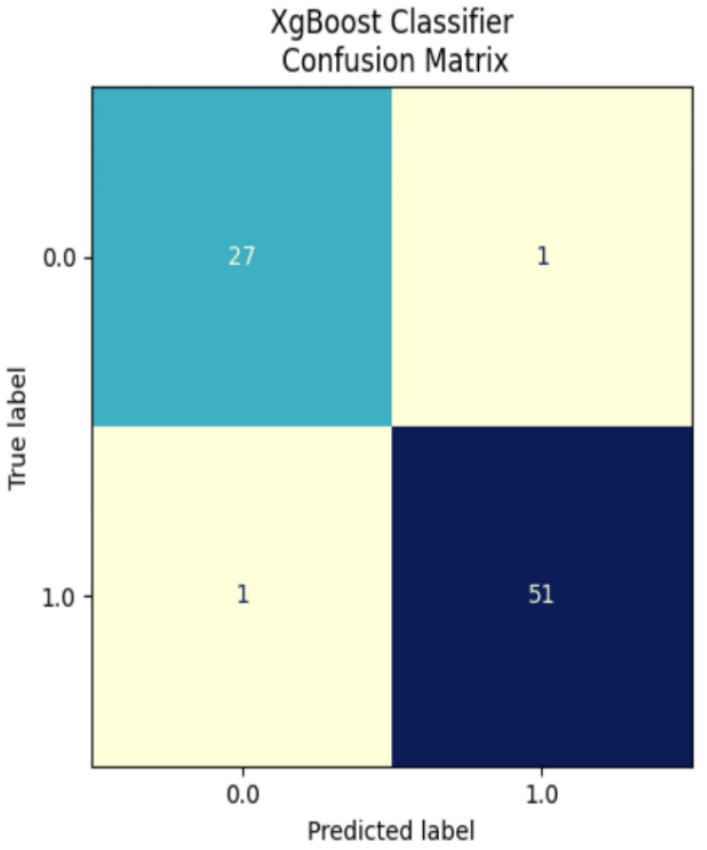
Results

MODEL	TEST ACCURACY SCORE	1ST VALIDATION ACCURACY SCORE	2ND VALIDATION ACCURACY SCORE
Random Forest	0.96	0.85	.85
AdaBoost	0.95	0.57	.5
XgBoost	0.98	0.93	.92

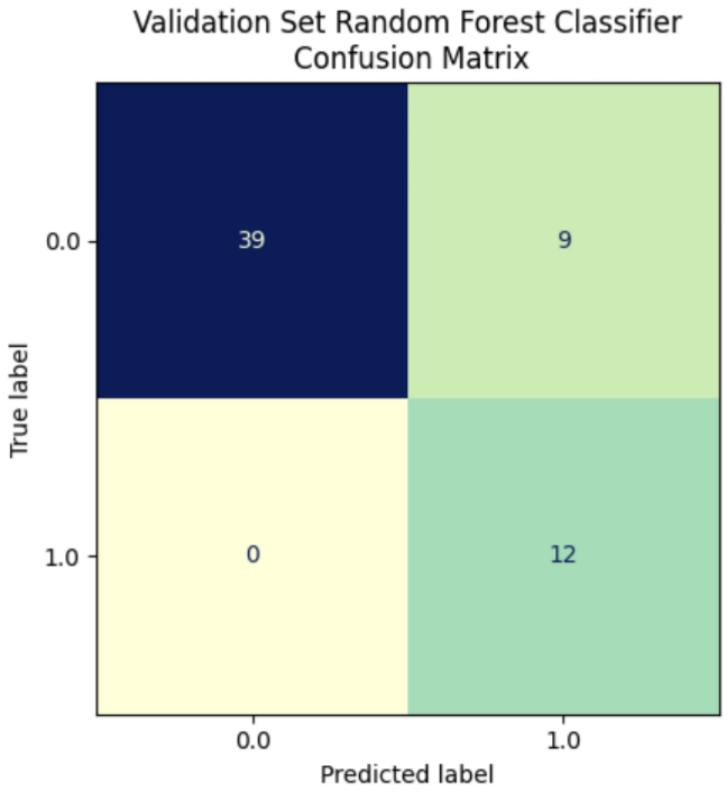
Prediction Results – Test Data

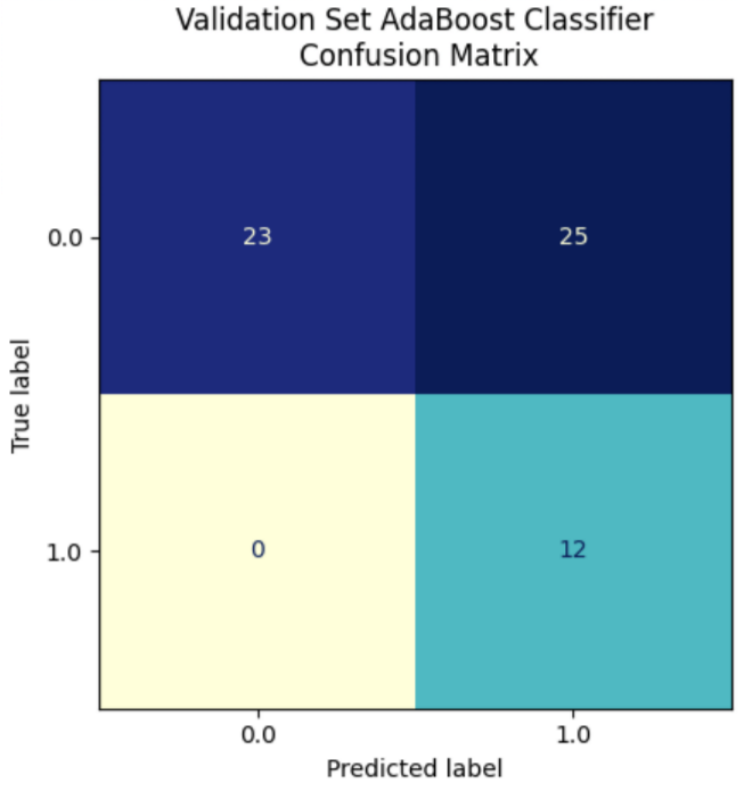


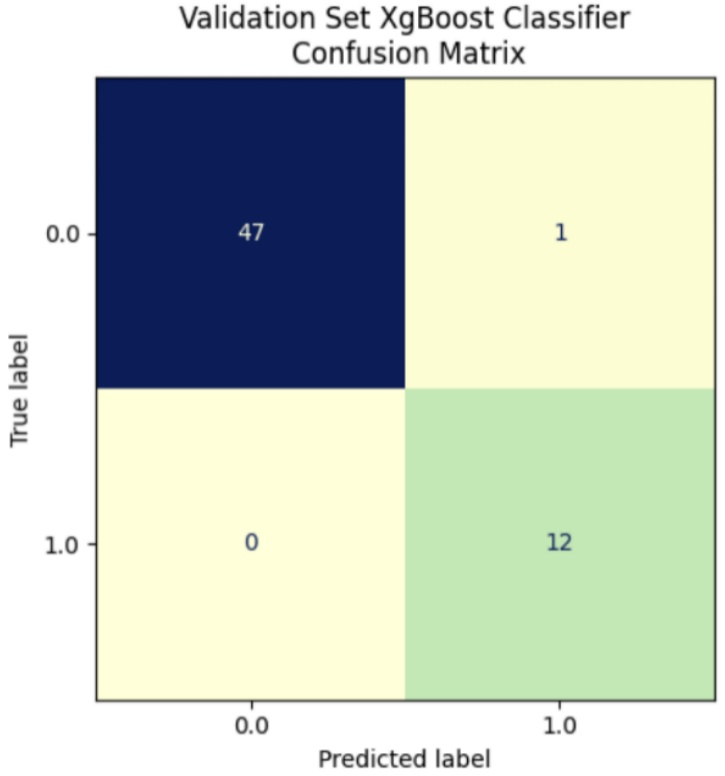




Prediction Results – Validation Data







Conclusion

- XgBoost has the best accuracy and consistent across the test and validation datasets
- Random Forest is a potential
- AdaBoost not a viable model
  - Prone to Overfitting
  - Lags behind XgBoost in speed and CPU memory usage

References

1. Tuttle, K. R., Alicic, R. Z., Duru, O. K., Jones, C. R., Daratha, K. B., Nicholas, S. B., McPherson, S. M., Neumiller, J. J., Bell, D. S., Mangione, C. M., & Norris, K. C. (2019). Clinical Characteristics of and Risk Factors for Chronic Kidney Disease Among Adults and Children. *JAMA Network Open*, *2*(12), e1918169. https://doi.org/10.1001/jamanetworkopen.2019.18169

2. *Chronic-Kidney-Disease - Research Chronic Renal Insufficiency Cohort Study Kidney Disease*. (n.d.). http://www.cristudy.org/Chronic-Kidney-Disease/Chronic-Renal-Insufficiency-Cohort-Study/CRIC-DataView

3. Dua, D. and Graff, C. (2019). UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science.

4. Atul, A. (n.d.). *Kidney Disease.csv*. https://raw.githubusercontent.com/AP-Atul/Chronic-Kidney-Disease/master/dataset/train.csv

5. *Kidney Disease*. (2022, December 17). National Institute of Diabetes and Digestive and Kidney Diseases. https://www.niddk.nih.gov/health-information/kidney-disease