ANALYZING THE IMPACT OF SOCIOECONOMIC FACTORS ON HATE-CRIME RATES POST-2016 ELECTION

Aarin Amin, Sophia Kurz, Kevin Love, Aswathama Shanmugam Marimuthu, Andrew Normandin, Achintya Srikanth
Overview of Statistical Learning 36600 - A

Background & Introduction

Socio-economic and demographic factors play a significant role in the examination of hate crimes across the United States. This study looks at hate crimes across the nation, exploring correlations with variables such as median household income, unemployment rates, educational attainment, and voting patterns.

By identifying trends and geographic disparities, this analysis provides insights into potential drivers of hate crimes, with the aim of reducing such offences.

Data Pre-Processing

Overview:

- The dataset includes state-level variables such as median household income, share of unemployed population, urbanization, educational attainment, racial demographics, and hate crime rates per 100,000 people (from SPLC and FBI sources).
- Data spans all 50 states with details on socio-economic and political indicators.

Steps Taken:

- **Handling Missing Data**: States not included in the analysis (e.g., Maine, Mississippi, South Dakota, and Hawaii) were excluded from specific visualizations.
- **Variable Transformation**: Scaled continuous variables for consistent interpretation in modeling and visualizations.
- **Region Encoding**: Categorical variables (e.g., state names) were transformed to lowercase to align with mapping datasets.

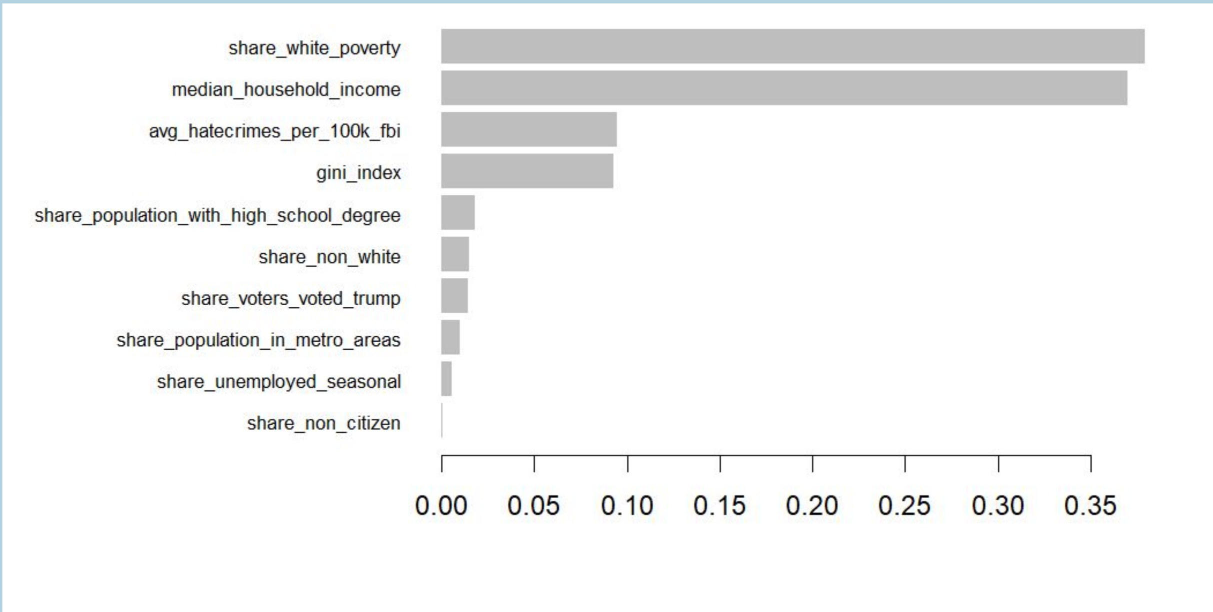
Methods & Analysis

Visualization:

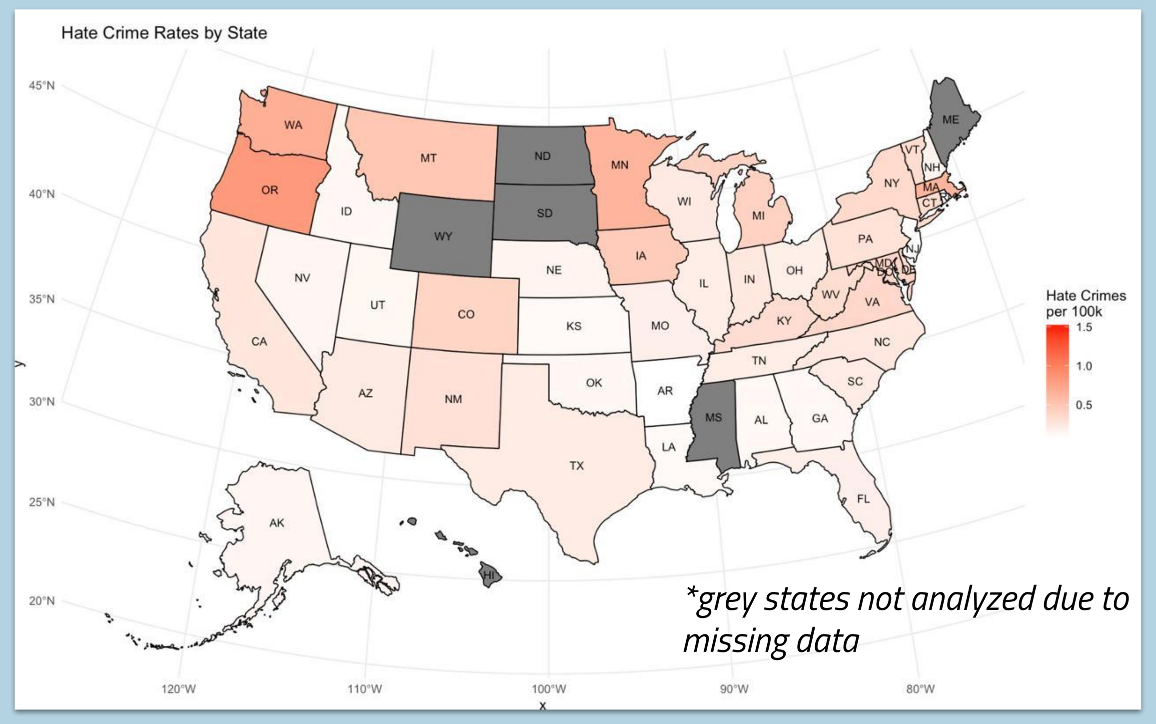
- Geographic heat maps to highlight state-by-state analysis.
- Overlayed state abbreviations for interpretability.
- Scaled indicators of hate crime severity for clearer geographic comparisons.
- Geographic mapping of hate crimes using usmap and ggplot2 in R to visualize spatial disparities.

Modeling:

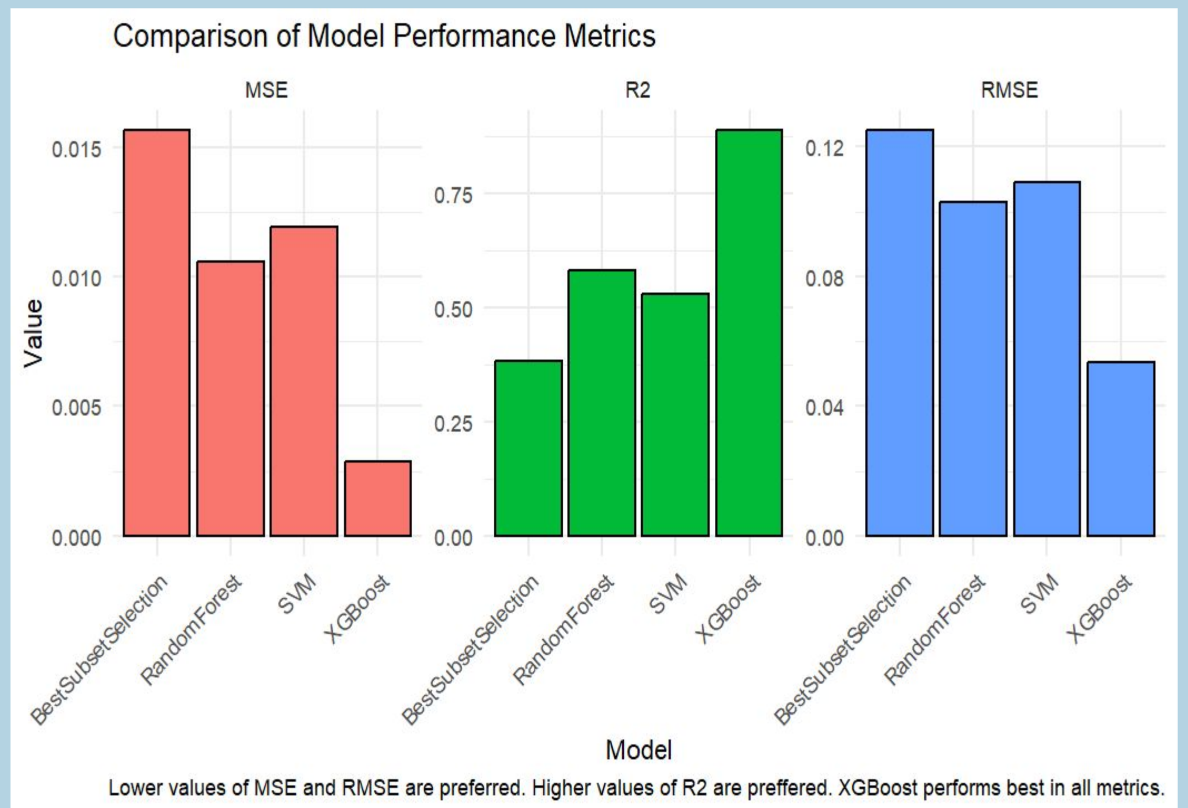
- Regression analyses to assess the influence of socio-economic variables on hate crime rates.
- Best Subset Linear Regression as determined by the Akaike Information Criterion was used as a baseline for evaluating other modeling techniques
- Models evaluated:
  a. Support Vector Machine
  b. Random Forest Regression
  c. Extreme Gradient Boosting (XGBoost)
- The residuals of the models indicate that the nonlinear nature of Extreme Gradient Boosting was best able to accommodate the distribution of the response when training.
- Extreme Gradient Boosting provided the best prediction by all metrics



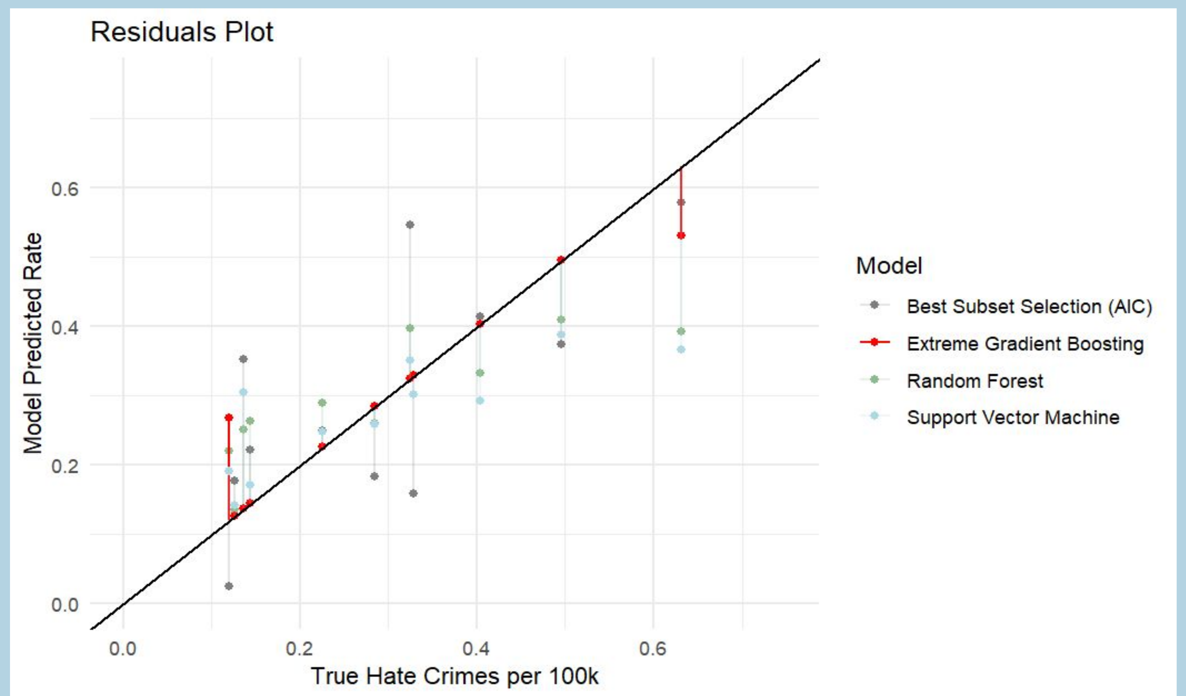
Comparison of Explained Variance Among Features (XGBoost)



Geographical Distribution of Rates of Hate Crimes



Comparison of Mean Squared Error, R-squared Value and Root Mean Squared Error Across Evaluated Models



Plot of Residuals of Evaluated Models

Results

Extreme Gradient Boosting was found to return the best predictive results, followed by Random Forest regression, Support Vector Machine, and Best Subset Selection for Logistic regression, respectively. With Extreme Gradient Boosting, we were able to predict the rate of hate crimes per 100,000 people after the 2016 presidential election with a mean squared error of .0029. The most impactful factors were found to be the share of white residents who are living in poverty in 2015 and the median household income in 2016. While these features were identified as the most impactful predictors for the Extreme Gradient Boosting model, Random Forest found the share of 2016 U.S. presidential voters who voted for Donald Trump and the average annual hate crimes per 100,000 population as reported by the FBI from 2010-2015 to be the most predictive variables. Extreme Gradient Boosting was able to produce an R-squared statistic of .887, indicating a strong fit for the data and a strong predictive ability. The overall accuracy and fit of Extreme Gradient Boosting make it likely that the share of white residents who are living in poverty in 2015 and the median household income in 2016 do truly carry the best predictive weight. The states with the highest reported hate crime incidents after Trump's election were found to be Massachusetts and Oregon with a rate of around 1 incident for every 100,000 people.

References

- R Packages: ggplot2, usmap, tidyverse, bestglm, randomForest, e1071, and xgboost
- Reinhart, A. (2017, February 24). *Hate crimes after the 2016 presidential election*. Hate crimes after the 2016 presidential election. http://rosmarus.refsmmat.com/datasets/datasets/hate-crimes/