Modeling and Predicting Conspiracy Belief

Capstone Sprint 2

Elliot Carter

The Problem Area

- Who is most susceptible to conspiracy belief and misinformation?
- Big problem, big consequences

How Data Science Can Help

- Much research on conspiracy theories has small sample sizes

- Recent survey by Imhoff et al produced a large dataset (~100k respondents).

- My project: explore and model their data

Conspiracy mentality and political orientation across 26 countries

Roland Imhoff[1✉], Felix Zimmer[1], Olivier Klein[2], João H. C. António[3], Maria Babinska[4], Adrian Bangerter[5], Michal Bilewicz[4], Nebojša Blanuša[6], Kosta Bovan[6], Rumena Bužarovska[7], Aleksandra Cichocka[8], Sylvain Delouvée[9], Karen M. Douglas[8], Asbjørn Dyrendal[10], Tom Etienne[11], Biljana Gjoneska[12], Sylvie Graf[13,14], Estrella Gualda[15], Gilad Hirschberger[16], Anna Kende[17], Yordan Kutiyski[11], Peter Krekó[17], Andre Krouwel[18], Silvia Mari[19], Jasna Milošević Đorđević[20], Maria Serena Panasiti[21], Myrto Pantazi[22], Ljupcho Petkovski[23], Giuseppina Porciello[21], André Rabelo[24], Raluca Nicoleta Radu[25], Florin A. Sava[26], Michael Schepisi[21], Robbie M. Sutton[8], Viren Swami[27,28], Hulda Thórisdóttir[29], Vladimir Turjačanin[30], Pascal Wagner-Egger[31], Iris Žeželj[32] and Jan-Willem van Prooijen[18]

Introduction to the Dataset

- Target: scores on *Conspiracy Mentality Questionnaire (CMQ)*

- Predictors: demographic features, political orientation measures

- Feature engineering of country-level features:
  - GDP and unemployment % from previous year
 - Economist Democracy Index

Statistical Analysis

- Strongest positive correlations with target: unemployment %, preference for *socially* right-wing parties, lack of high school diploma

- Strongest negative correlations: university degree, country-level democracy index, favoured political party being in power

- Political orientation and conspiracy mentality related, but not linearly…

Visualizing the Relationship Between Political Orientation and Conspiracy Mentality
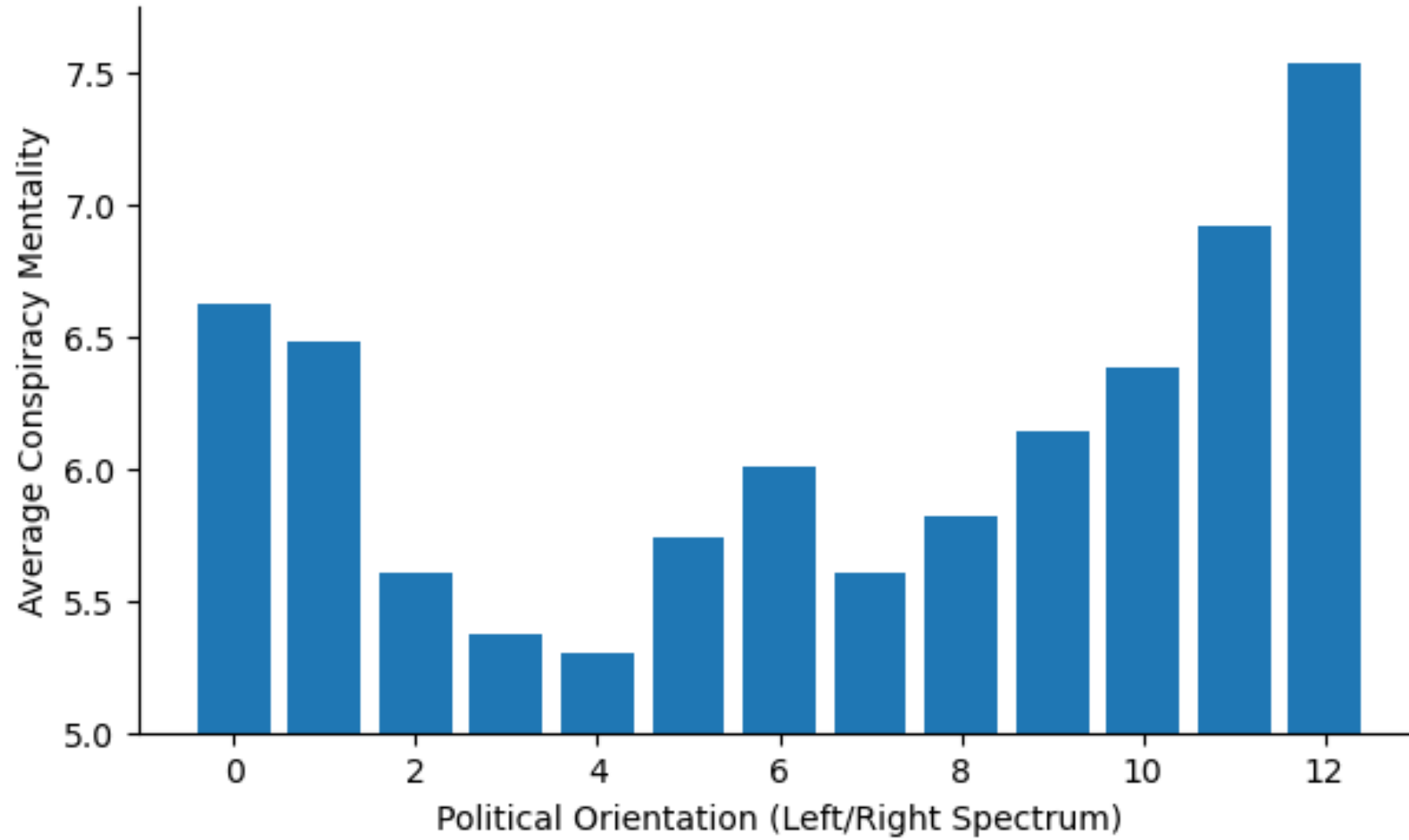
Mean Conspiracy Mentality and Unemployment Rate by Country
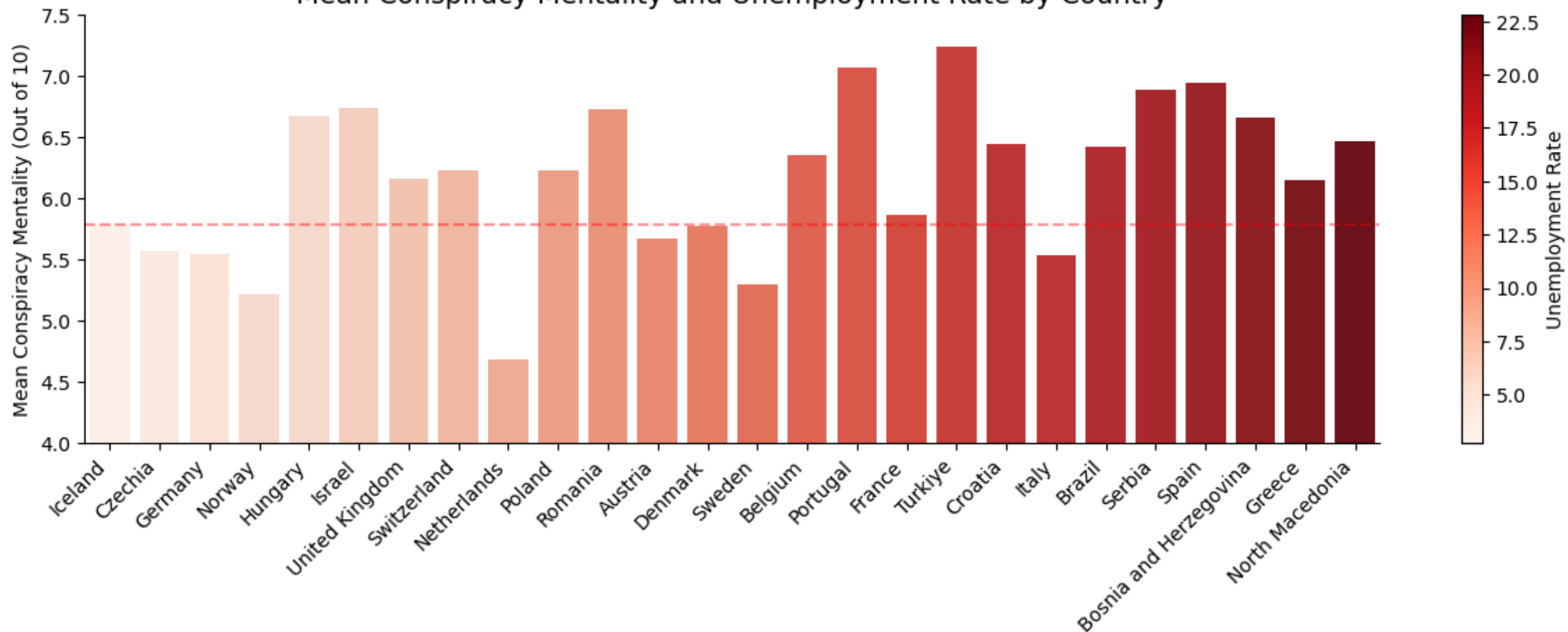
Baseline Modelling Results

Type of Model	Scoring	Notes
Linear Regression	R^2: 0.21	Many statistically significant features; low R^2
Decision Tree Regressor	R^2: 0.23	Not much improvement over linear regression
Gradient Boosting Regressor	R^2: 0.27	Best R^2 among regressors; hard to interpret
Logistic Regression	Accuracy: .68 Recall: .71 for class 1; .64 for class 0	Class 1 = above average CM; Class 0 = below average Similar coefficients to linear regression
KNN	Accuracy: .67	Worse results than logistic regression
Gradient Boosting Classifier	Accuracy: .69 Recall: .74 for class 1, .63 for class 0	Slightly better accuracy than logistic regression but slightly larger imbalance in recall

Next Steps

- Hyperparameter optimization with cross-validation for best models
- Clustering to try to find profiles of typical conspiracy believers
- Further investigating non-linear relationships between predictors and target