Social Pressure and Voting

Overview

The Problem

1

Low Voter Turnout

2

Poor Political Representation

How might we use machine learning to understand the role of **social pressure** *in determining* **voting behaviour** *such that we can design* **interventions** *to* **increase voter turnout?**

Identify

subgroups

Dataset Overview

Target column is **'voted'.**

G2000 P2000 DEM	SEX YOB % < 5	MEDIAN $ UNEMPLOY	>HS >BACH < 9-12	ZIP HH_ID
Past Voting Behaviour	Age/Gender	Income & Employment	Education	Location & Household Info.

Feature Engineering

1 **One-Hot Encoding**

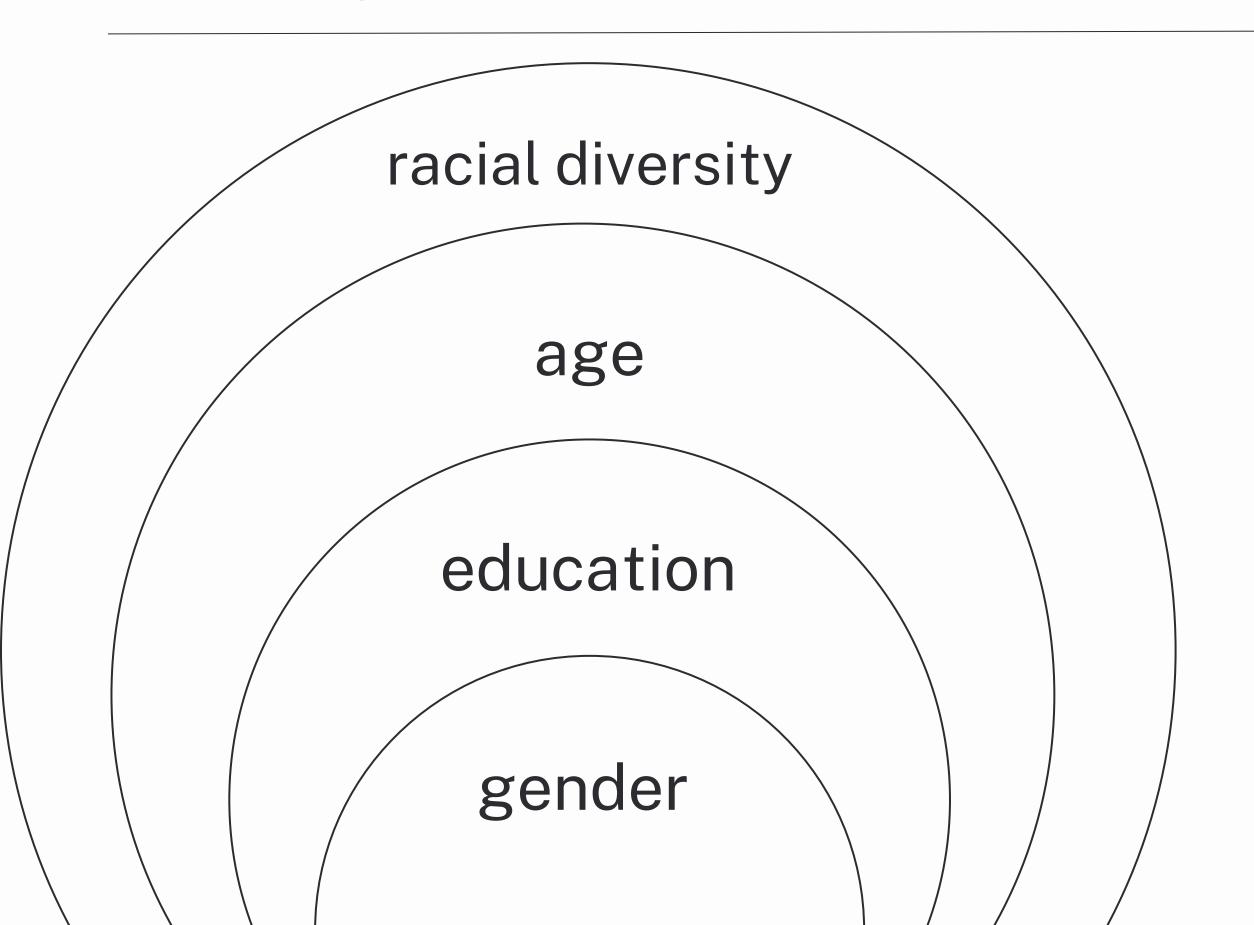Turned treatment columns into dummy columns.

2 **Dropping Columns**

Removed columns that were redundant, had low variance or were highly correlated with other variables.

ADVANCED EDA

racial diversity

age

education

gender

CLUSTERING

Using MiniBatchKMeans

SCORING

LUsed Inertia and Silhouette Score.

MODELLING

Yet to be integrated into modelling but promising.

MODELLING APPROACHES

Logistic
Regression

- 7 models
- Generally, accuracy ~69%
- Good precision and recall for the majority class (not voted)
- Poor precision and recall for minority class
- Used SMOTE

Decision
Trees

- Hyperparameter tuning using RandomizedSearchCV and GridSearchCV

Random
Forest

- 2 models
- Better recall for 'first pass' model with no drop in accuracy
- Much better recall for class-balanced model

Gradient
Boosting

- Slight improvement in accuracy and precision

Neural
Network

- 3 layers, funnel-shaped
- Used dropout layers for generalisability

Evaluation
Metrics

Accuracy

F1 Score

NEXT STEPS

Clustering
Integrate into modelling

Class-Balancing
Try with more models

Causal Forests
Estimate CATE

Shapley Values
To interpret models better