

Model Card for 2020 US Cooperative Election Study*

Yuanyi (Leo) Liu

April 2, 2024

Model Card for the 2020 US Cooperative Election Study

Model Details:

- **Model date:** April 2024
- **Model version:** 1.0
- **Model type:** Supervised learning model predicting voter turnout and preferences.
- **Information about training algorithms, parameters:** Trained using a gradient boosting framework with parameters optimized for balanced accuracy across demographic groups.

Intended Use:

- **Primary intended uses:** This model is intended for academic research to understand voting behavior patterns
- **Primary intended users:** Researchers, policy makers, and civic engagement organizations.
- **Out-of-scope use cases:** Not intended for individual voter prediction or targeted political advertising.

Factors:

- **Relevant factors:** Model performance may vary by age, gender, race, geographic location, and previous voting history.

*The essay is available at: <https://github.com/leoyliu/STA302-Mini-Essay-MRP>

- **Evaluation factors:** Performance evaluated across demographic groups defined by race, gender, and age to ensure the model’s fairness and effectiveness across these dimensions.

Metrics:

- **Model performance measures:** Accuracy, F1 score, and area under the ROC curve (AUC), with results disaggregated by relevant demographic factors to assess fairness.
- **Decision thresholds:** Established to maximize F1 score uniformly across all demographic groups.
- **Variation approaches:** Performance metrics evaluated using cross-validation to ensure robustness of model results.

Evaluation Data:

- **Datasets:** Utilized the 2020 US Cooperative Election Study, supplemented with external validation datasets to ensure generalizability.
- **Motivation:** These datasets were chosen for their comprehensive coverage of the US voting population and richness in demographic details.
- **Preprocessing:** Data anonymized and preprocessed to normalize input features, handle missing values, and ensure privacy.

Training Data:

- Similar to evaluation data, the model was trained on a partition of the 2020 US Cooperative Election Study dataset, ensuring a split that reflects the diversity of the US population.

Quantitative Analyses:

- **Unitary results:** Showcases model performance across individual demographic groups.
- **Intersectional results:** Examines model performance at the intersection of demographic factors to identify any compounded biases.

Ethical Considerations:

- Emphasizes the model’s development with a focus on ethical practices, including fairness in model performance and avoiding misuse of the model for discriminatory purposes.

Caveats and Recommendations:

- Notes the importance of continuous monitoring for potential biases as the model is deployed in different contexts. Recommends periodic reevaluation of the model with newer data to ensure its relevance and fairness over time.