Model Card for Logistic Regression

Siyuan Lu

November 27, 2024

Model Overview

- **Model Type**: Logistic Regression
- **Purpose**: The logistic regression model is used to predict the likelihood of binary outcomes based on several input features. In this case, the model could be used to predict whether women participate in a given role within a rebel group (e.g., frontline, noncombat, lead, etc.) based on various organizational and contextual features.

Intended Use

- The model aims to predict women's participation in rebel organizations. The model's predictions could be used by researchers or analysts in conflict studies or gender studies to understand patterns in female participation in armed conflicts and rebel movements.

Performance Metrics

- **Accuracy**: The accuracy of the model should be reported as the percentage of correct predictions (based on the test dataset).
- **Precision and Recall**: Precision (how many of the predicted positives were actually positives) and recall (how many actual positives were identified) can be calculated for each binary outcome (e.g., frontline, lead).
- **Confusion Matrix**: The confusion matrix can provide understanding into the model's true positives, false positives, true negatives, and false negatives.

Ethical Considerations

- **Bias**: The model could potentially reflect any biases present in the dataset, such as bias related to the underrepresentation of certain types of rebel organizations or gendered assumptions in the data collection process.
- **Fairness**: It is essential to ensure that the model does not unintentionally perpetuate stereotypes or biases about women's roles in conflict. The dataset should be analyzed for fairness before model deployment.

Limitations

- **Data Limitations**: The model relies on historical data, which may be incomplete or biased. It is crucial to account for potential underreporting or misclassification of women's roles in armed groups.
- **Generality**: The model's generalizability to different contexts or periods beyond the dataset (1946-2015) should be considered with caution.

Training Data

- The model is trained on the **WAAR Project Dataset v1.0**, which includes data on women's participation in 372 rebel organizations from 1946 to 2015. The dataset includes both ordinal and binary variables measuring women's participation in various roles (e.g., frontline, leadership, auxiliary).

Limitations and Future Work

- The current model may not account for all relevant features, such as qualitative narratives that provide context to the participation estimates. Future work could integrate this qualitative data to improve predictions.