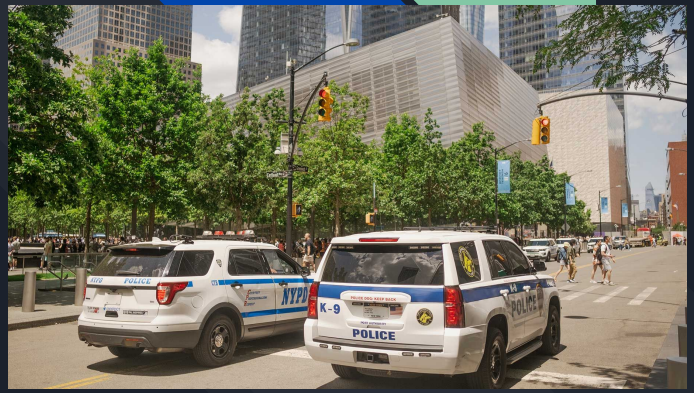
Erdos Data Science Bootcamp Project



Perceptions of Police Interactions

Members: Alistair Fernandez, Danyh Tolah, Duke Ochichi, Kazi Tawhid, Pingchuan Ma

Motivation

- Research Question: “What is the public perception of the police after contact with law enforcement personnel based on socioeconomic features?”
- Thesis Statement: Public perception of the police following any encounter with law enforcement may vary significantly depending on socioeconomic background.

This project helps law enforcement, local and federal government, and community leaders better understand how the general public perceives police behavior, especially across different socioeconomic groups.

- Provides insight into public trust and perceived fairness
- Enables data-driven decision-making in training, oversight, and community engagement
- Promotes the use of interpretable, calibrated models for transparency and accountability

Goal

- Identify which groups are **more likely to report improper police behavior**
- Build models that **predict and calibrate the risk** of perceived misconduct
- Focus on **probability-based outputs**, not just classification, to reflect **real-world risk**

Data Overview

Source: ICPSR 38872 — Police Public Contact Survey

Population: 130,000+ survey responses.

Selected Features: AGE (categorical), SEX, HISP (ethnicity variable), and INCOME.

Target Variable: survey indicator of proper vs. improper police behavior (V347).

Final Sample Size: 17,535 records (after filtering & cleaning).

Preprocessing:

- Removed invalid/missing values
- One-hot encoded categorical variables
- Balanced training data using SMOTE

Methods and Models

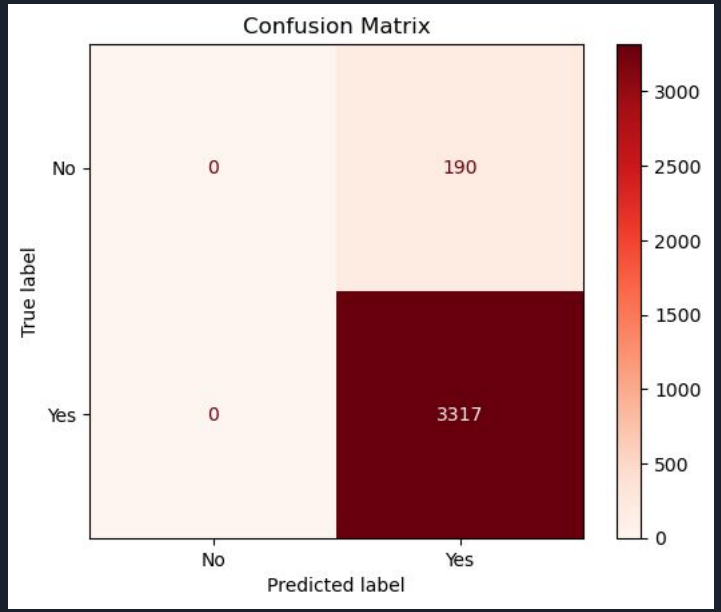
Baseline Model: Logistic Regression:

- Logistic Regression
- Logistic Regression with SMOTE

More Advanced Model:

- XGBoost
- AdaBoost

Logistic Regression Result



Accuracy: 0.95

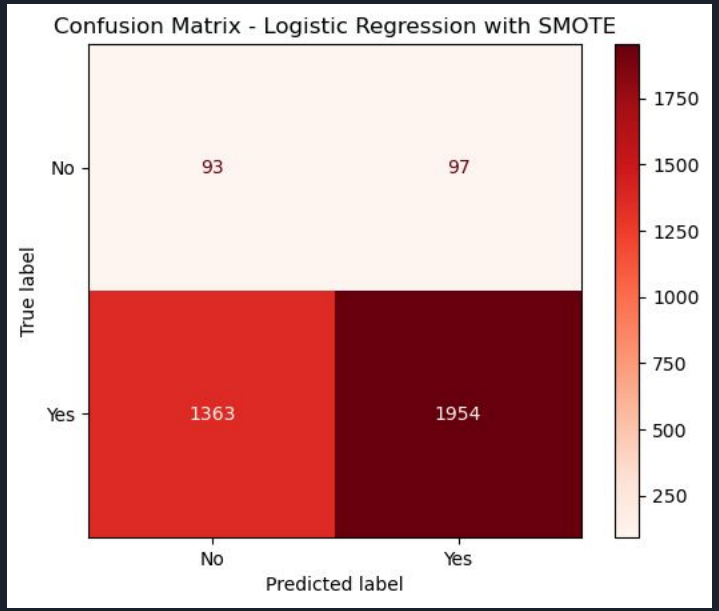
Problem:

- The model is biased towards predicting "Yes" for every case.
- It never identifies "No" cases correctly.

Why?

- Imbalance data: majority of the training data voted “Yes” to the survey question.
- A way to balance data is needed!

Logistic Regression with SMOTE



Accuracy: 0.58
Threshold: 0.5 (default) or adjusted to 0.4test set.

Improvement:

- The model **started identifying "No"** cases (minority class).
- **Recall for "No"** increased from **0.00 → 0.49**
- It now considers **both classes** in its predictions.

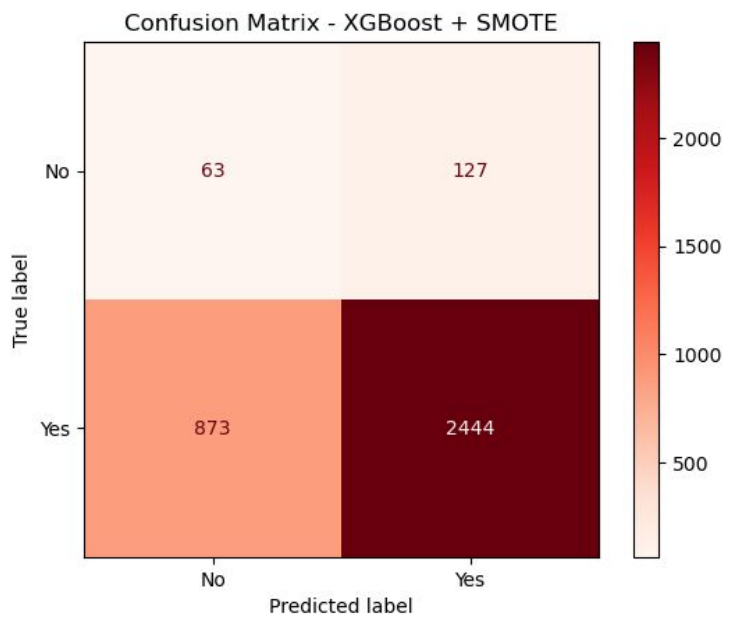
Problem (Still Exists):

- **Precision for "No" is very low** (as low as 0.06).
- Many **false alarms**: predicting "No" when it's actually "Yes".

Why?

- SMOTE helped balance the training data,
  but the test data remains imbalanced.
- Logistic Regression is still **limited in modeling complexity**.
- **Minority class patterns are harder to learn** and generalize.

Calibrated XGBoost



Accuracy: 0.72

Class 0 ("No") :

- **Precision** (truly No) ≈ **6.7%**
- **Recall** (No were correctly predicted)≈ **33.2%**

Class 1 ("Yes"):

- **Precision** ≈0.95
- Recall ≈0.737

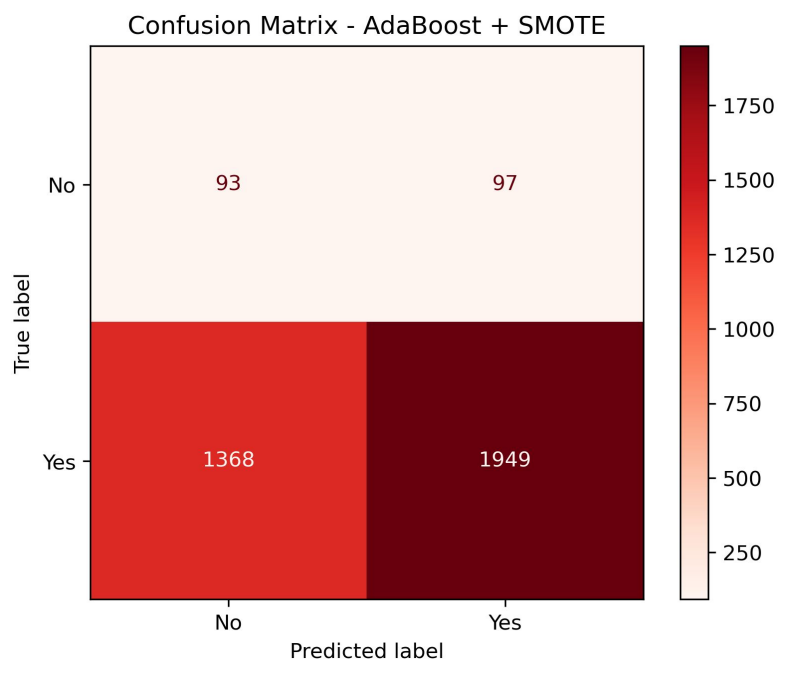
Improvement:

- Model now **detects more "No" cases** (Recall = 33%).
- **Overall performance improved** compared to Logistic Regression.
- **Class 1 ("Yes") performance remains strong.**

Problem:

- **Precision for “No” remains low (6.7%)**

AdaBoost Result



Accuracy: 0.58

Class 0 ("No") :

- **Precision** (truly No) ≈ **6%**
- **Recall** (No were correctly predicted)≈ **49%**

Class 1 ("Yes"):

- **Precision** ≈ 95%
- Recall ≈ 59%

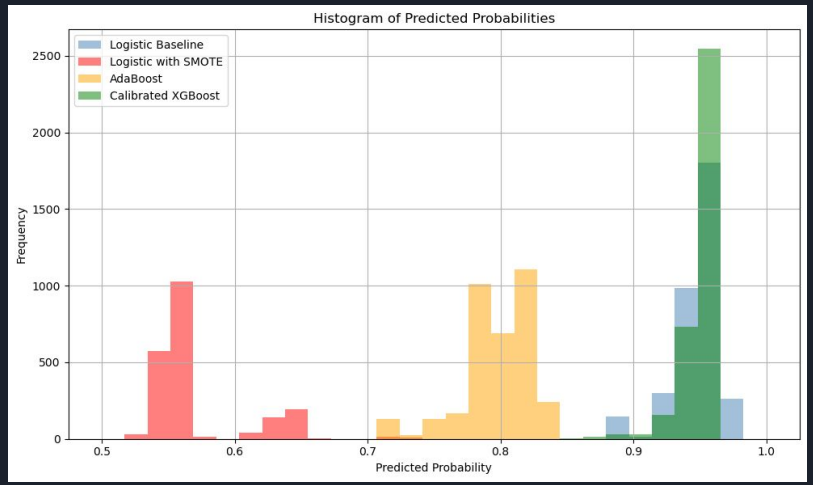
Improvement:

- Model now **detects more "No" cases** (Recall = 49%).
- **Overall performance improved** compared to Logistic Regression.
- **Class 1 ("Yes") performance remains strong.**

Problem:

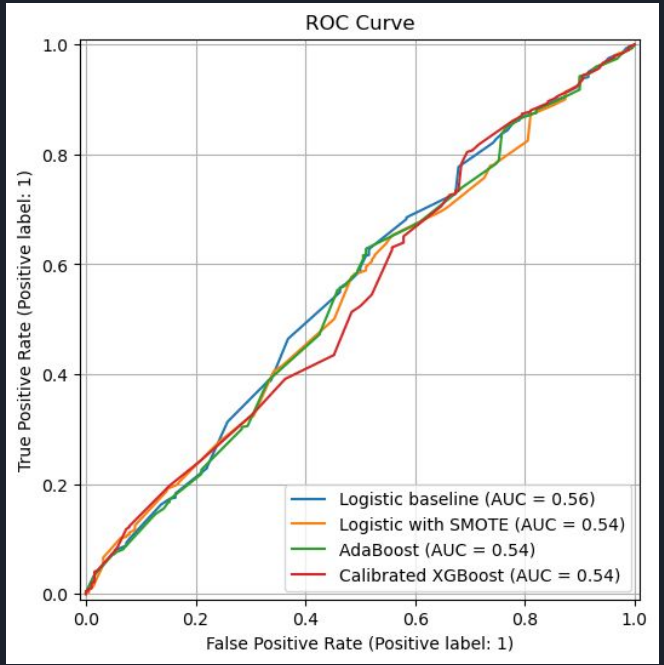
- **Precision for “No” remains low (6.7%)**
- **Overall accuracy is 58%**, lower than XGBoost (71%) due to increased false alarms.

Comparing the results: probabilities



1. Logistic regression: 0.93 ~ 0.97.
2. SMOTE logistic: around 0.5~0.7.
3. AdaBoost: 0.75~0.85.
4. Calibrated XGBoost: ~0.95.

Comparing results from all models



ROC Analysis:
All models have AUC scores near **0.5**, which indicates **low discriminative power** (i.e., they're close to random guessing).
While recall for "No" improved through SMOTE and AdaBoost, **overall model ability to distinguish between classes remains poor.**

Conclusion

- This imbalance in the survey data results in 95% results in poor performance of the models.
- The Models are not sufficient to analyze survey data containing categorical variables.
- Public agencies should collect more survey data information.

Future directions

- Test on additional features in the data set.
  - Previous interaction with the police.
  - If the interviewees' contact was due to violations.
- Building a similar project on data sets that are more insightful.
  - Current Dataset: Imbalance data gives low accuracy.
- Study the theoretical background of the predicted probabilities.
  - Data findings could be used in academia and research.
  - Socioeconomic and Demographic background could impact on the public perception of police.