Hate-Crime Early-Warning Model for NYPD

Final Report
Author | Siddhant Anand Jadhav   Course | Machine Learning

---

1 Stakeholder

NYPD Hate-Crime Task Force and community-relations officers who decide where to send patrol cars and outreach teams.

---

2 Problem & Target

Hate crimes are rare but high-impact. NYPD wants to know **which precincts are likely to experience at least one hate-crime incident next month** so they can act early.

Target variable
`Hate_Crime_Occurred = 1` if a precinct logs ≥ 1 hate crime in the coming month; `0` otherwise.

---

3 Dataset

Item	Info
Source	NYC OpenData – "NYPD Hate Crime Incidents" (2019 – Mar 2024)
Raw size	3 325 incident rows
Reshaped	11 736 precinct-months (every precinct, every month)
Link	https://data.cityofnewyork.us/Public-Safety/NYPD-Hate-Crimes

---

4 Feature Engineering

Feature	Reason
Month (1-12)	captures seasonality (spring spike)
Season_Spring / Summer / Winter	smooths month noise
Lag 1 & Lag 2	"momentum" – recent crimes predict near-term risk
Crime Count (current month)	baseline hot-spot score

All zero-crime months are explicit so the model learns from quiet periods too.

5 Models Tried & Why

Model	Why include it	Key settings
Logistic Regression	Transparent baseline	`C=1`, max_iter 500
Random Forest	Handles non-linear patterns; feature importance	`n_estimators 200`, `max_depth 5`
XGBoost	SOTA gradient boosting; often best on tabular data	300 trees, depth 4, `scale_pos_weight` for imbalance

80 % train / 20 % test, **stratified** by target.

6 Evaluation Metrics & Results

Model	F1	Precision	Recall	AUC
Logistic Reg.	0.26	**0.72**	0.16	**0.65**
Random Forest	0.32	0.55	0.23	0.60
XGBoost	**0.47**	0.45	**0.49**	0.64

Why these metrics?

- **F1** balances missed crimes vs. false alarms.
- **Recall** matters most (better to over-patrol than miss a hate-crime month).
- **AUC** shows overall ranking quality.

7 Interpretation

Logistic Reg. ranks precincts best (AUC 0.65) but misses most hotspots (16 % recall).
Random Forest improves recall to 23 %.
XGBoost nearly triples recall to **49 %** while keeping AUC 0.64, giving commanders a far more useful early-warning list.

Feature importance confirms pattern: recent activity (`Lag 1`), `Month`, and `Season_Spring` drive risk.

8 Fairness Check

Precision by borough ranges **0.48 – 0.60** – no single borough is over-flagged. We will track fairness again after adding demographic features.

9 Recommendation

Adopt **XGBoost** for a monthly dashboard that lists the **top-5 risk precincts**.
Even if only half its alerts are correct, NYPD would still catch ~25 extra hate-crime months a year—well worth a few extra patrol shifts.

10 Deployment Plan

Step	Tool	Cost
Monthly cron pulls new CSV	AWS EventBridge + Lambda	$0
Lambda retrains XGBoost & writes `risk_latest.json`	Python + joblib	$0.20/mo
Static HTML dashboard reads JSON, shows table & map	S3 static site	$0.50/mo
Auto-email PDF to commanders	Lambda + SES	$0.10/mo

Zero servers, < $1 per month.

11 Limitations & Future Work

Gap	Next action
No demographic context	Join census income & race data
Recall 49 %	Tune thresholds & test weekly forecasts
Probability not calibrated	Apply isotonic calibration
Only borough-level fairness	Check by race/ethnicity once demographics added

12 Code & Reproducibility

- [GitHub repo]
- Environment: Python 3.9, packages listed in `requirements.txt`.