MACHINE LEARNING PROJECT

Ohav Zach, Gal Peleg

Link to the Github: https://github.com/Galpelegq/Course-Project-Data-Mining-2025

Introduction

Hate crimes are offenses motivated by prejudice against specific groups in the population. In this project, we analyze data from the Austin Police Department (Texas) between the years 2017-2025. This topic is significant because understanding this data can help law enforcement and the community identify where, when, and against whom these incidents occur. These insights can help with better prevention strategies and help make the city a safer place for everyone.

The primary goal of this project is to build a machine learning model capable of predicting the motive behind a crime, such as racism, religion, or sexual orientation, based on the specific details of the incident.

Dataset Description

Each record in the dataset represents a single crime incident and includes a variety of attributes:

-Date of incident
-Zip code and offense location
-Number of victims and offenders,-disaggregated by age groups
-Bias description and classification
-Demographic characteristics of offenders (if reported)

This dataset provides both numerical and categorical information, suitable for supervised multi-class classification tasks, capturing temporal, spatial, and behavioral aspects of hate crimes.

Handling Missing and Invalid Values

Columns representing counts of victims and offenders were safely converted to numeric values. Invalid or missing entries were replaced with zero. The incident date column was converted to a standardized datetime format, enabling consistent temporal feature extraction.

Victim and Offender Feature Engineering

Aggregate features were created to summarize each incident:

-Total victims = victims under 18 + victims over 18
-Total offenders = offenders under 18 + offenders over 18
-Victim–offender ratio, with a small smoothing value to avoid division by zero

These features help the model understand the severity and structure of each event.

Temporal Features

Temporal patterns were captured from the incident date, including the day of the week, a binary weekend indicator, and the season (Winter, Spring, Summer, Fall). These features allow the model to identify weekly and seasonal trends in hate crime occurrences.

Frequency Features

Zip code frequency was calculated to reflect how often incidents occur in each area. Rare offense locations (fewer than five occurrences) were grouped as Other_Location to reduce sparsity and prevent overfitting.

Bias Processing and Target Variable Definition

The primary bias motivation was extracted from the original Bias field and mapped into broader super-classes: Race/Ethnicity, Sexual Orientation/Gender, Religion, and Other. This reduces label noise and improves class balance. The resulting super-class was used as the target variable (Y) and label-encoded numerically.

Offender Information Encoding

A binary feature indicates whether offender information is available: 1 = reported, 0 = unknown or missing. This captures the completeness of offender data, which can be informative for classification.

Feature Selection and Categorical Encoding

Non-predictive or leakage-prone columns, such as incident IDs, raw Bias text, and the original date, were removed. Categorical features (month, day of the week, police sector, offense location group, season, and council district) were one-hot encoded to be used in machine learning models.

Methodology:

The goal of this project is to predict the type of bias behind hate crime incidents using features such as the age of victims and offenders, their counts, offense type, day of the incident, and location.

In the original dataset, there were dozens of unique and rare bias categories. To simplify the task and improve model performance, these were consolidated into **four main super-classes (Super-Classes): Race/Ethnicity, Sexual Orientation/Gender, Religion, and Other**. This consolidation was critical to reduce noise in the data and significantly enhance the predictive power and accuracy of the models.

Supervised Models:

-Logistic Regression: Simple, interpretable; shows each feature's effect on bias type.
-Decision Tree: Captures non-linear feature relationships, creating rules to separate bias types.
-Random Forest: Ensemble of trees; reduces overfitting and improves classification accuracy.
-Gradient Boosting: Detects subtle patterns to differentiate similar bias types.
-SVM: Separates bias types in complex, high-dimensional feature space.

Unsupervised Techniques:

-Clustering: Finds groups of similar incidents, revealing hidden patterns.
-Anomaly Detection: Identifies unusual or high-risk cases for further investigation.

Experiments and Results:

Logistic Regression Model Analysis

The model achieved 48% accuracy, nearly doubling the 25% random baseline, indicating a meaningful predictive signal. However, performance varies by class: the model

performs well on frequent classes (1 and 2) but struggles with rare ones (0) due to data sparsity, limiting balanced classification.

Logistic Regression

	precision	recall	f1-score	support
0	0.00	0.00	0.00	1
1	0.59	0.50	0.54	26
2	0.50	0.67	0.57	12
3	0.44	0.38	0.41	21
accuracy			0.48	60
macro avg	0.38	0.39	0.38	60
weighted avg	0.51	0.48	0.49	60

Decision Tree Model Analysis

The Decision Tree model achieved 57% accuracy, outperforming Logistic Regression. It shows more balanced performance across the main classes (1, 2, and 3), with F1-scores up to 0.60, indicating that non-linear patterns improve bias detection. However, Class 0 remains unclassified due to extreme data sparsity, highlighting a persistent structural limitation.

Decision Tree

	precision	recall	f1-score	support
0	0.00	0.00	0.00	1
1	0.58	0.58	0.58	26
2	0.55	0.50	0.52	12
3	0.59	0.62	0.60	21
accuracy			0.57	60
macro avg	0.43	0.42	0.43	60
weighted avg	0.57	0.57	0.57	60

Tuned Random Forest Model Analysis

The tuned Random Forest achieved the highest accuracy at 63%, demonstrating improved ability to capture complex patterns. It performed especially well on Class 2, reaching 82% precision, and showed balanced results across the main classes (1, 2, and 3). However, Class 0 remained unclassified, underscoring that even advanced models cannot overcome extreme data scarcity.

Tuned Random Forest

	precision	recall	f1-score	support
0	0.00	0.00	0.00	1
1	0.60	0.69	0.64	26
2	0.82	0.75	0.78	12
3	0.58	0.52	0.55	21
accuracy			0.63	60
macro avg	0.50	0.49	0.49	60
weighted avg	0.63	0.63	0.63	60

Gradient Boosting Model Analysis

The Gradient Boosting model achieved 57% accuracy, slightly below the tuned Random Forest. It performed strongly on Class 2 with 80% precision and showed reasonable balance between Classes 1 and 3, though with more sensitivity to noise. Overall, the results indicate that Gradient Boosting is more affected by small dataset size and minority-class scarcity in this setting.

Gradient Boosting

	precision	recall	f1-score	support
0	0.00	0.00	0.00	1
1	0.56	0.54	0.55	26
2	0.80	0.67	0.73	12
3	0.50	0.57	0.53	21
accuracy			0.57	60
macro avg	0.47	0.44	0.45	60
weighted avg	0.58	0.57	0.57	60

Support Vector Machine (SVM) Model Analysis

The SVM model achieved 52% accuracy, outperforming Logistic Regression but underperforming compared to tree-based models. It showed its strongest results in Class 1, with a 62% F1-score, indicating effective detection of frequent bias types. However, lower precision in Class 3 suggests overlapping or complex class boundaries that the current SVM configuration struggles to separate.

SVM

	precision	recall	f1-score	support
0	0.00	0.00	0.00	1
1	0.62	0.62	0.62	26
2	0.54	0.58	0.56	12
3	0.42	0.38	0.40	21
accuracy			0.52	60
macro avg	0.39	0.39	0.39	60
weighted avg	0.52	0.52	0.52	60

Model Configuration & Parameters

-**Logistic Regression:** max_iter=1000, class_weight='balanced' – A baseline linear model; balanced weights were crucial due to class frequency differences.
-**Decision Tree:** max_depth=15, class_weight='balanced' – Captures non-linear rules but is prone to high variance (overfitting).
-Random Forest (Tuned): n_estimators=500, max_depth=20, min_samples_split=5, min_samples_leaf=2, class_weight='balanced' – Our top performer; uses an ensemble of 500 trees to reduce error and handle data imbalance.
-**Gradient Boosting:** n_estimators=200, learning_rate=0.05 – Iteratively corrects errors from previous trees, detecting subtle patterns.
-**SVM:** kernel='rbf', class_weight='balanced' – Uses non-linear boundaries to separate complex classes in high-dimensional space.

Evaluation Metrics

-Accuracy: Measures the overall percentage of correct predictions.
-Weighted F1-score: Our primary metric, as it balances **Precision** and **Recall** while accounting for the disproportionate class sizes.

Final Analysis & Insights

-The best Model: Tuned Random Forest achieved the best results (**63% Accuracy**), effectively managing the complex interactions between location, time, and bias types.
-Impact of Class Weighting: Using class_weight='balanced' was essential across all models; without it, minority classes (like Class 0) would have been entirely ignored by the algorithms.
-Data Constraints: Performance was consistently limited by **Data Sparsity**. Even high-performing models struggled with categories containing very few samples, indicating that more data or oversampling (like SMOTE) could further boost accuracy.
-Complexity vs. Interpretability: While Logistic Regression offered the clearest interpretation, it couldn't match the predictive power of ensemble methods (Random Forest/Gradient Boosting).

Conclusion: The **Tuned Random Forest** is the most reliable model for this dataset. It successfully captures non-linear patterns and remains robust against noise, providing a strong baseline for predicting hate crime characteristics.

Unsupervised Learning – Experiments and Insights

Selected features for clustering: Total Victims, Total Offenders, Victim Offender Ratio,

Zip Code Frequency, Is Weekend, Is Offender Known

Choosing the Number of Clusters
-Used the **Elbow Method** to test k values from 2 to 8.
-Inertia stabilized around **k=5**, indicating this is the optimal number of clusters.

-This value balances capturing distinct patterns while avoiding too many small or overlapping clusters.

We chose features that capture **victim/offender dynamics, temporal patterns, and spatial frequency**, because they are most relevant for identifying underlying crime patterns.
-These features allow clusters to reflect meaningful differences between incidents without using the target label.

K-Means Clustering
-Applied K-Means with n_clusters=5 and n_init=10.
-Assigned cluster labels to all data points.

Cluster 0: Known-Offender Weekday Incidents

- Weekdays only, offenders always known (Is_Offender_Known = 100%)
- Low–moderate geographic concentration (Zip_Code_Frequency ≈ 13.8)
-Interpretation: Mostly single victim/offender, routine interpersonal incidents

Cluster 1: Anonymous Weekday Routine

- Almost exclusively weekdays, offenders rarely identified (Is_Offender_Known = 0%)
- Low geographic concentration (Zip_Code_Frequency ≈ 11.4)
- Interpretation: Typically single victim/offender, minor or low-priority cases

Cluster 2: Non-Personal & Anomalous Reports

- No recorded offenders (Total_Offenders = 0), extreme victim–offender ratio.
- Mixed temporal pattern (Is_Weekend ≈ 0.34)
-Interpretation: Moderate geographic spread (Zip_Code_Frequency ≈ 17), property/report-driven incidents

Cluster 3: High-Density Weekend-Leaning Hotspots

- Weekend-leaning (Is_Weekend ≈ 0.58), high offender identification (≈ 86%).
- Highest geographic concentration (Zip_Code_Frequency ≈ 47.4)
-Interpretation: Slightly more offenders per incident, hotspot with frequent confrontations

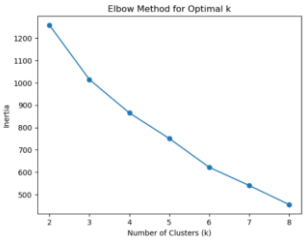
Cluster 4: Weekend Confrontational Incidents

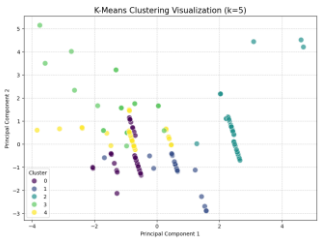
- Weekends only, high offender identification (≈ 89%)
- Moderate geographic concentration (Zip_Code_Frequency ≈ 12.7)
-Interpretation: Slightly more offenders per incident; linked to social/leisure activities.

hidden patterns

-The features were carefully chosen to capture **who, when, and where**, enabling meaningful separation of incidents.
-The clustering analysis shows that hate crimes are not random:
Time: Weekday vs. weekend patterns.
Location: High-crime areas vs. quieter areas.
Complexity: Anonymous/vandalism vs. group attacks.

-This unsupervised analysis complements the supervised models by uncovering **hidden patterns** in hate crime data.





Outlier Detection – Isolation Forest

We applied **Isolation Forest** to detect outliers in the hate crime dataset.
-n_estimators=200 – number of isolation trees.
-contamination=0.05 – assumes 5% of the data are anomalies.
-random_state=42 – for reproducibility.

The model classified each data point in the scaled feature space (X_unsup_scaled) as:
1 → Normal point **-1** → Outlier

Results: **Normal points:** 283, **Outliers:** 15

Examining the outliers revealed the following patterns:

- Some events had unusually high numbers of offenders (e.g., 4 offenders), which is rare in hate crimes.
- Some events had 0 offenders but multiple victims, resulting in extremely high Victim_Offender_Ratio values (e.g., 2,000,000 and 3,000,000). These cases likely correspond to anonymous acts, such as vandalism or graffiti on public property.
- One outlier had 0 victims, indicating a hate crime without a direct target, again suggesting acts like graffiti or property damage.

Insights for Outlier Detection

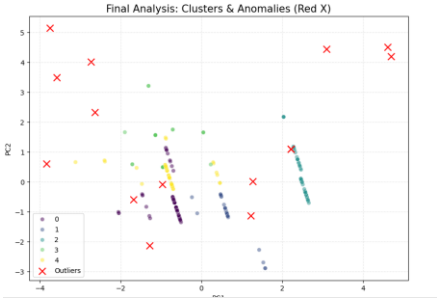
Isolation Forest successfully highlighted **extreme or atypical events** that deviate from the general clustering patterns. These outliers provide important insights into the **diversity and complexity** of hate crimes beyond the main clusters identified with K-Means.

- The presence of 0 victims or 0 known offenders shows that some crimes are non-interactive or anonymous, differentiating them from the majority of incidents.

- Events with multiple offenders or unusually high ratios stand out as rare but significant incidents that may require separate attention in analysis and policy considerations.

]:

	Total_Victims	Total_Offenders	Victim_Offender_Ratio	Super_Bias_Class
21	1	4	0.25	RACE_ETHNICITY
22	0	2	0.00	RACE_ETHNICITY
37	2	4	0.50	SEXUAL_GENDER
80	3	0	3000000.00	RACE_ETHNICITY
159	2	0	2000000.00	SEXUAL_GENDER



Conclusion and Discussion

General Findings: The analysis revealed several key insights about hate crime incidents:

Predictive Performance: The highest supervised model accuracy achieved was **0.63**. While this may appear modest, it reflects the **sparse and variable nature of the dataset**, rather than poor model quality. It is important to note that **even after merging the data into 4 consolidated categories**, the dataset size remained relatively small for such complex patterns. Reaching this level of prediction under these constraints is **meaningful and significantly better than random chance**, proving that the model successfully captured the underlying signals.

Patterns in Hate Crimes: Unsupervised clustering uncovered distinct behavioral profiles:

- Crimes vary by **time** (weekday vs. weekend) and **location** (hotspots vs. low-density areas).
-Some incidents are **anonymous or involve unusually large groups**, creating extreme cases that were identified as outliers.
-Religious bias crimes often involve no direct offender contact, suggesting acts like vandalism or graffiti.
-Racial and gender-related crimes occur more in busy urban areas and during weekends, highlighting environmental and temporal influence.

Outliers and Edge Cases: Isolation Forest identified rare events, such as incidents with **zero victims** (likely vandalism) or unusually high victim/offender ratios. These cases emphasize **the diversity and complexity** of hate crime incidents.

Importance and Implications

- Hate crimes are **not random**:they follow identifiable spatial, temporal, and social patterns.
- This topic is extremely important and **highly educational**, as understanding these patterns can directly help **prevent future incidents**.
- Predictive models, even with current data limitations, provide **useful insights** for anticipating and mitigating hate crimes.

Team Roles and Contributions

We, Ohav and Gal, worked together in full collaboration on all parts of the project. We jointly contributed to the data preprocessing and feature engineering, the implementation of supervised and unsupervised models, the analysis of results, and the interpretation of insights. All decisions and analyses were made collaboratively.

Future Directions

Collecting more data is crucial in order to allow the models to learn from a wider range of cases and improve prediction accuracy. Expanding the dataset and adding **additional contextual features** could significantly enhance model performance.
In addition, focusing on **urban hotspots and weekend-related incidents** may help guide targeted prevention strategies. Future work could also explore deeper relationships between different bias types and environmental or temporal factors to better understand and prevent future hate crime incidents.