



AI-Based Crime Prediction and FIR Automation

A Project Presentation for the Degree of
Bachelor of Technology in Computer Science & Engineering

Team Members

PUSHKAR PAN 14200122139	RANJITA DE 14200122146
SUKANYA GHOSH 14200122164	SRIJON BANERJEE 14200122144
DEBMALYA GHOSH 14200122080	ABHISHEK MAITRA 14200122173

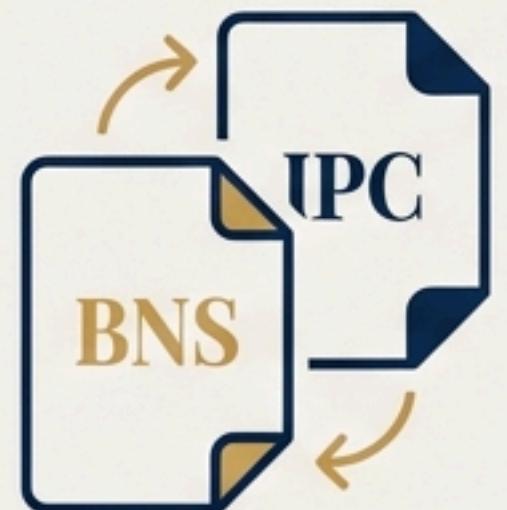
Supervisor

Md. Hamid Islam, Assistant Professor



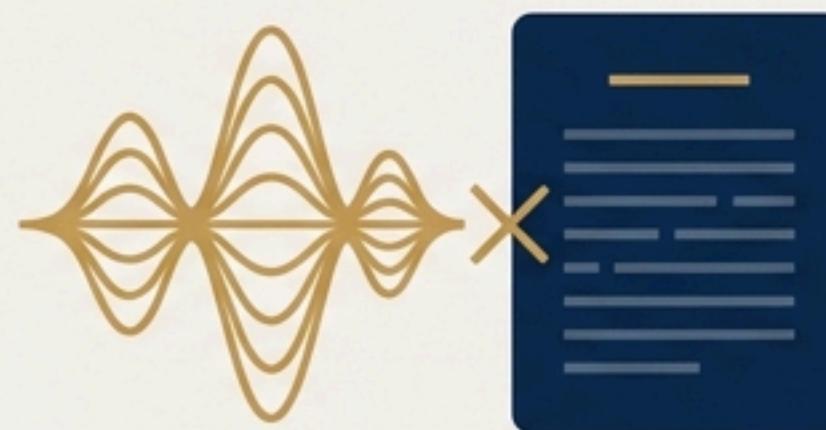
A Justice System at a Critical Juncture

The Indian criminal justice system faces a confluence of three major challenges that strain resources and impede efficiency.



Legislative Disruption

The transition from the Indian Penal Code (IPC) to the Bharatiya Nyaya Sanhita (BNS) creates a massive administrative and cognitive burden, risking widespread clerical error and confusion.



Semantic Ambiguity

A significant disparity exists between colloquial language used in citizen complaints and the formal language of legal statutes. This gap leads to incorrect FIR classifications and delays.



Reactive Policing & Analytical Bias

Traditional policing methods are often reactive. Furthermore, analysis based on absolute crime counts can be misleading, resulting in inefficient resource allocation.

Our Vision: An Intelligent Platform for Proactive and Precise Justice

To engineer an intelligent, integrated platform that enables a proactive, precise, and citizen-centric justice system.

Integrated Platform



BNS Search System

Automating FIR legal mapping with semantic understanding.

Bias-Free Crime Analysis

Uncovering true crime patterns with relative frequency analysis.

Proactive Crime Prediction

Providing actionable foresight for law enforcement.

Pillar 1: From Chaos to Clarity - Automating FIR Legal Mapping

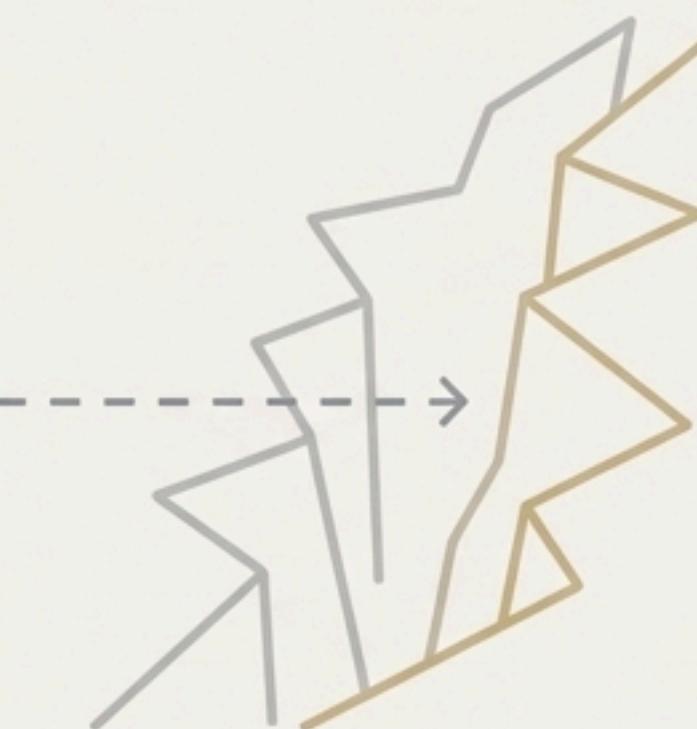
The Core Problem: Lexical Search Fails in Legal Contexts

Keyword-based search is inadequate for legal work. It cannot interpret the conceptual similarity between a citizen's colloquial description and the formal language of a statute.

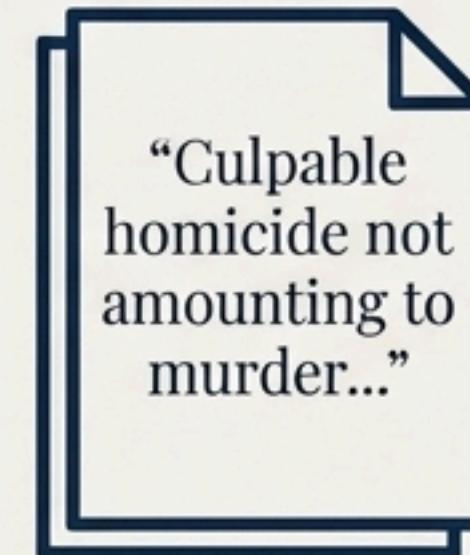
FIR Narrative (Colloquial)



driving



BNS Statute (Formal)



This semantic gap is a primary source of misclassification, especially during the BNS transition where established statutory numbers are obsolete.

Our Solution: A semantic search engine that understands the **meaning** and **intent** of an FIR narrative to retrieve the most relevant BNS sections.

A Two-Phase Workflow for BNS Semantic Search

Phase 1: Index Creation (Offline Process)



- 1. Ingest Data:** The system processes the complete corpus of 358 BNS sections.



- 2. Generate Embeddings:** Each BNS section is converted into a 384-dimensional vector using the `all-MiniLM-L6-v2` model.



- 3. Build Index:** The generated vectors are indexed into a FAISS database, creating a searchable legal knowledge base.

Phase 2: Runtime Retrieval (Live Process)



- 1. User Input:** An officer inputs a new FIR narrative as a query.



- 2. Generate Query Vector:** The FIR text is converted into an equivalent 384-d vector.



- 3. FAISS Search:** The system performs a high-speed search to find the 5 BNS vectors with the minimum L2 Distance to the query vector.

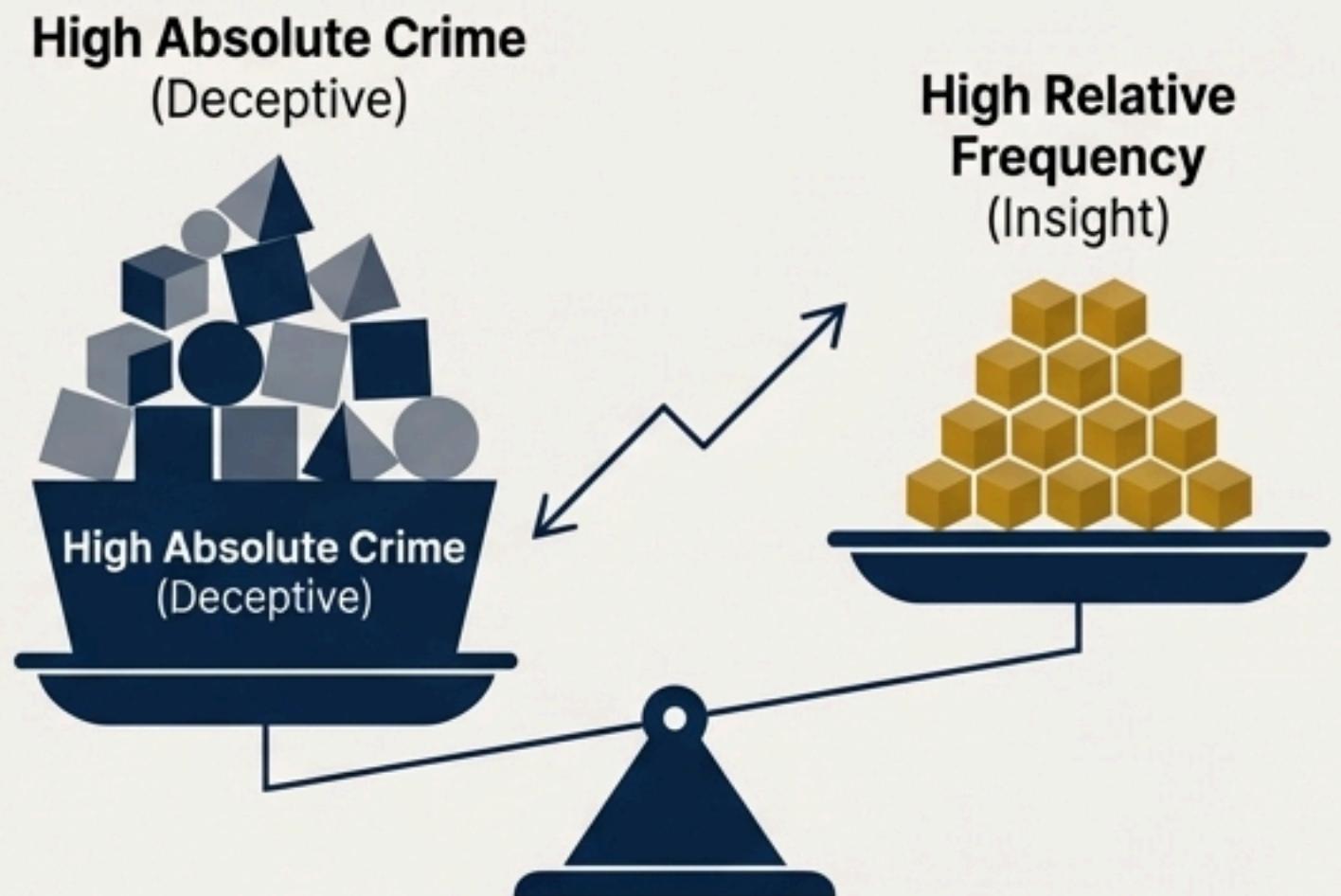


- 4. Return Results:** The Top 5 most semantically relevant BNS sections are presented to the user.

Pillar 2: From Bias to Insight - Redefining Crime Analysis

The Problem: Absolute Crime Counts are Deceptive

Analyzing crime using raw counts creates a biased view. A ward with a high total crime volume may overshadow a smaller ward where a specific crime type is disproportionately concentrated. This leads to misinterpreting crime patterns and misallocating resources.



Our Solution: Relative Crime Frequency Analysis

We calculate crime frequency as a percentage of the total crimes within a given ward. This normalized approach reveals the dominant, localized crime types and allows for meaningful, unbiased comparisons between different areas.

Crime Percentage = (Number of Crimes of a Specific Type in a Ward / Total Number of Crimes in that Ward) * 100

The Power of Relative Analysis: A Ward 18 Case Study

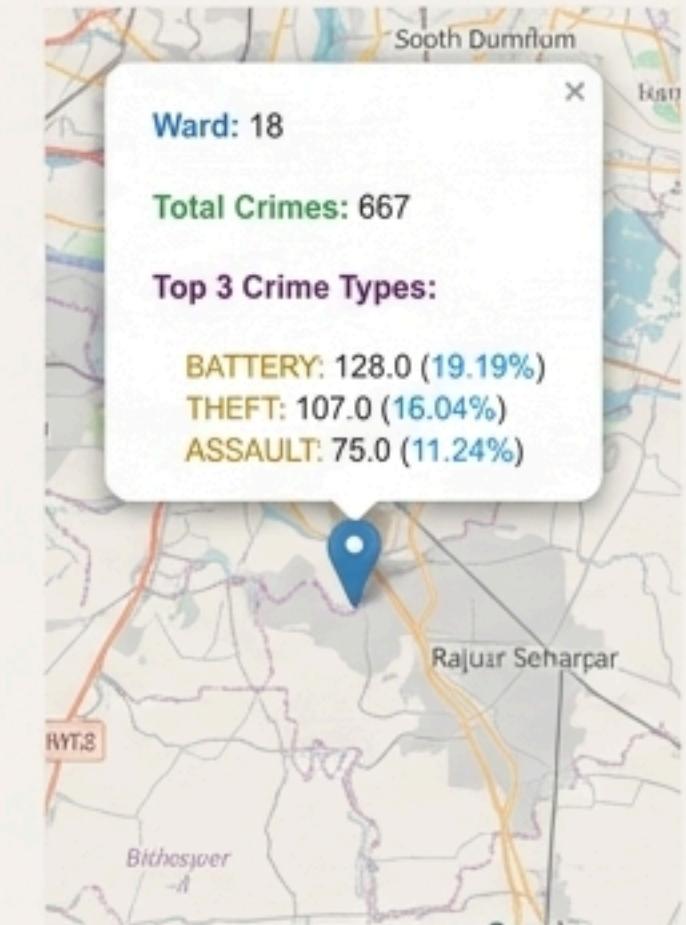
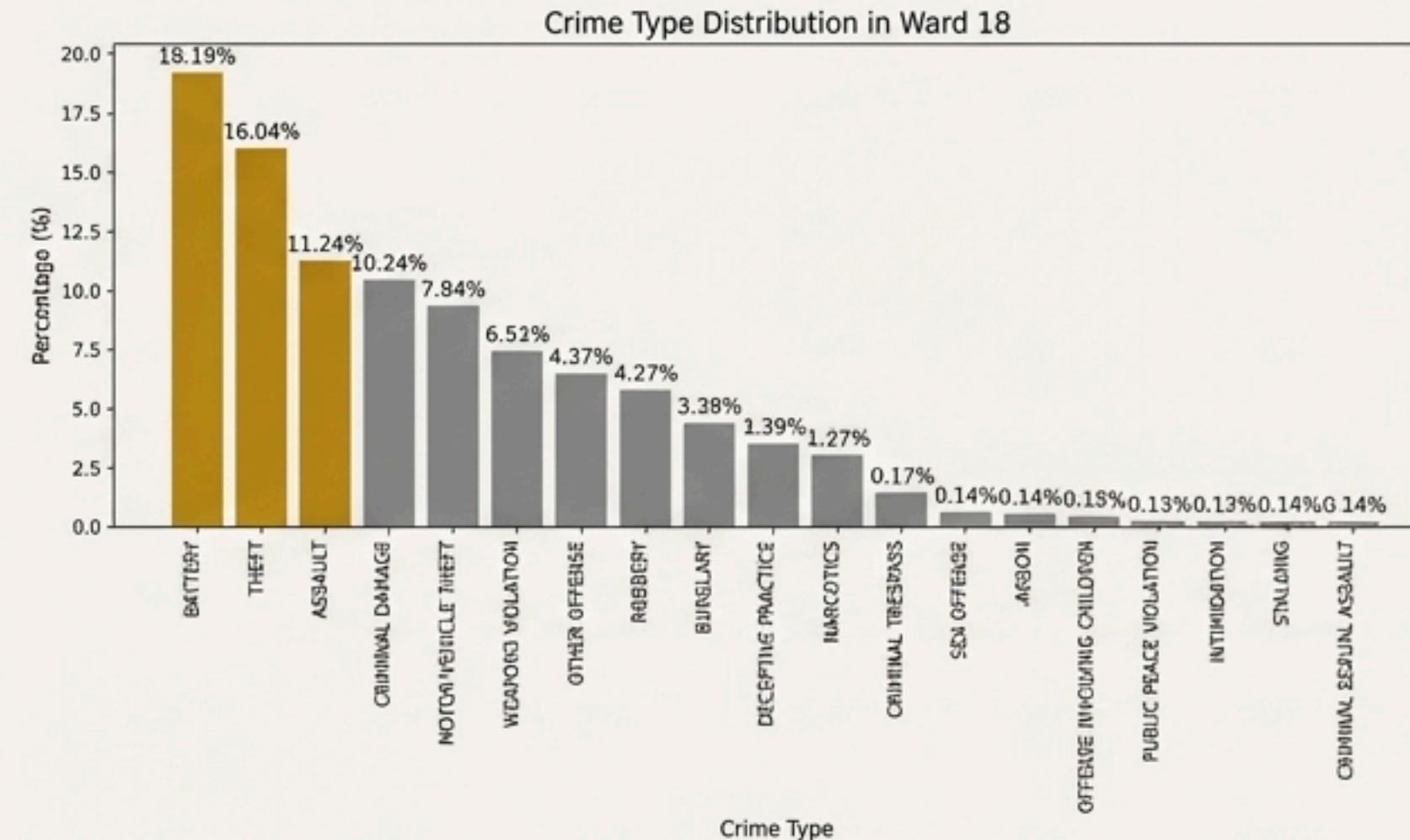
Misleading Absolute View vs. Insightful Relative View

While one ward might have more total thefts, another may have a higher **proportion** of thefts, indicating a more specialized problem.

Misleading Absolute View



Project Output for Ward 18

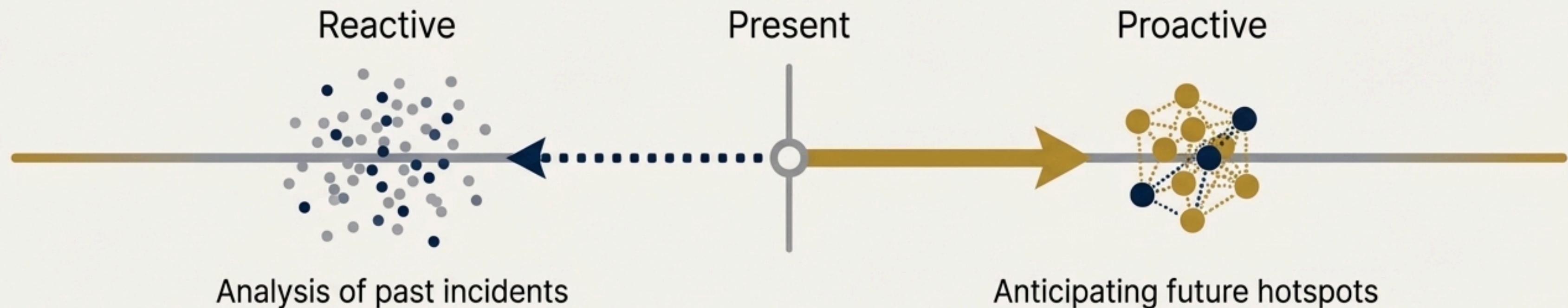


Our analysis identifies the true crime distribution, showing 'Battery', 'Theft', and 'Assault' as the Top 3 most prevalent crimes by percentage. This provides actionable intelligence for localized policing strategies.

Pillar 3: From Reaction to Foresight – Predictive Policing

The Problem: Law Enforcement is Overwhelmingly Reactive

Traditional policing responds to incidents after they have occurred. To effectively prevent crime and optimize resource deployment, law enforcement agencies require data-driven tools that provide proactive foresight into future risks.

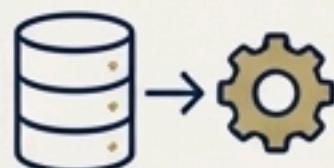


Our Solution: A Machine Learning Model for Crime Prediction

We developed a regression model that leverages historical data to predict future crime counts at the ward level. This allows authorities to anticipate hotspots and allocate patrols strategically before crimes happen.

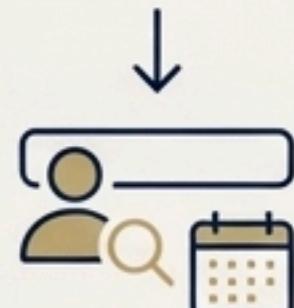
From Data to Decision: Prediction Workflow and Risk Mapping

Prediction Workflow



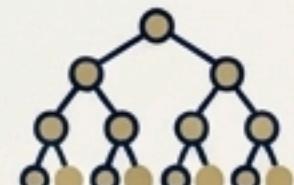
Historical Data Input

The model is trained on 16,800 aggregated crime records.



User Query

The user specifies a 'Ward', 'Year', and 'Month'.



Predict Crime Count

The trained Random Forest model predicts the expected number of crimes.



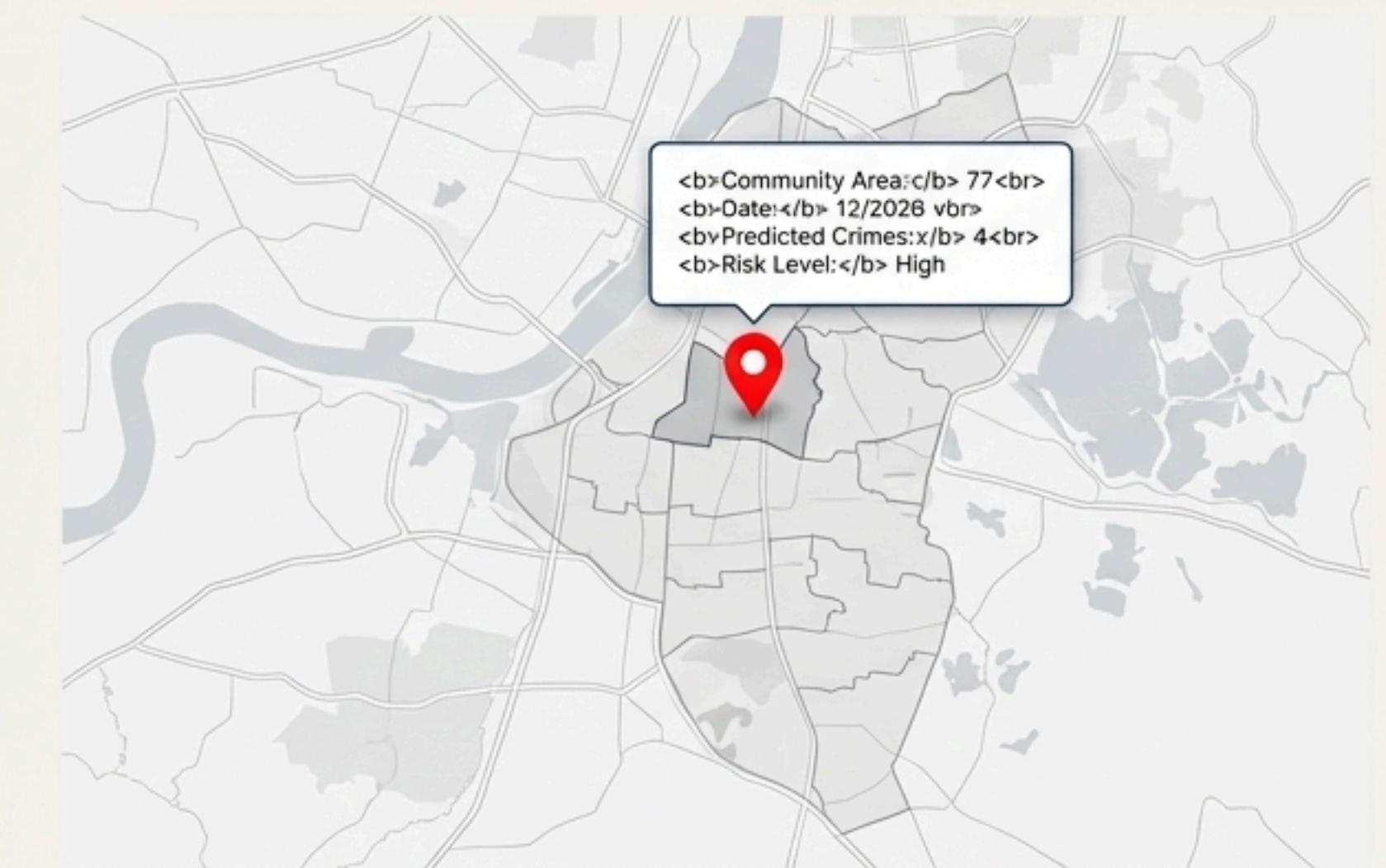
Classify Risk Level

The predicted count is categorized into a risk level.

Low Med High

Actionable Risk Classification

- **Low Risk**(<= 1 Predicted Crime)
- **Medium Risk** (2 Predicted Crimes)
- **High Risk**(> 2 Predicted Crimes)



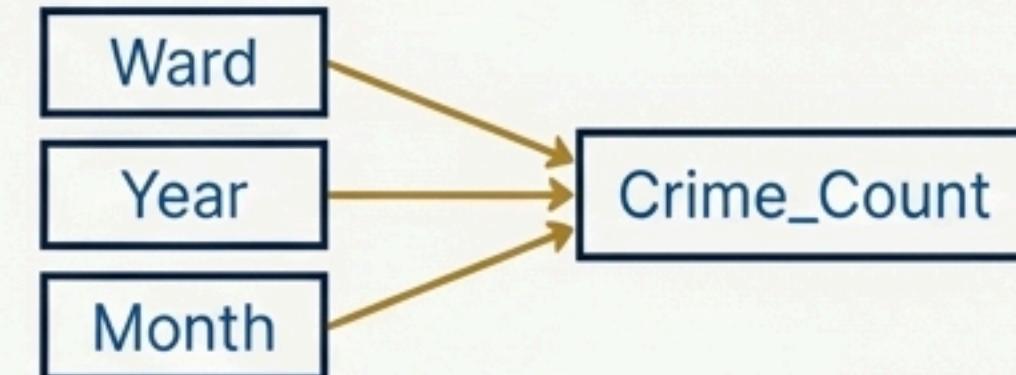
The Prediction Model: Random Forest in Action

Dataset



A comprehensive dataset of **16,800 crime records** from Kolkata, spanning the years 2001 to 2025.

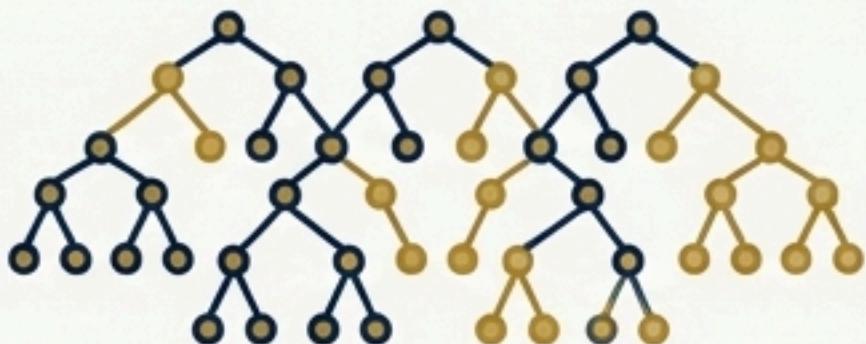
Features and Target



Input Features: 'Ward', 'Year', 'Month'

Target Variable: 'Crime_Count'

Model



Random Forest Regressor: An ensemble learning method chosen for its robustness and its ability to capture complex, non-linear patterns inherent in crime data.

Performance



The model achieved an **R² score of 0.67**. This indicates a strong predictive capability, demonstrating that it effectively explains 67% of the variance in the real-world crime data.

A Unified System: The Synergy of Three Pillars

Our platform integrates prediction, automation, and analysis into a continuous improvement loop for law enforcement.



Fostering Justice: Our Contribution to SDG 16

This project directly contributes to the United Nations Sustainable Development Goal 16: **Peace, Justice and Strong Institutions.**



How Our Project Supports SDG 16



Promotes Rule of Law

By ensuring accurate and consistent application of the new BNS legal code.



Builds Effective Institutions

By providing law enforcement with intelligent, data-driven tools to streamline workflows and enhance efficiency.



Ensures Equal Access to Justice

By reducing delays and errors in the critical first step of the justice process—the FIR.

Future Scopes: Engineering the Next Generation of Legal Tech

Our architecture is designed for future evolution. Key enhancements include:



Expansion to Regional Languages

Integrate multilingual sentence transformer models to process FIRs filed in major Indian languages like Hindi, Bengali, and Tamil.



Dynamic Confidence Thresholding

Implement a machine learning classifier to dynamically predict the accuracy of the top-ranked BNS section, providing an intelligent 'low confidence' warning in ambiguous cases.



Adversarial Training for Resilience

Introduce a component to generate synthetic, challenging crime scenarios. This will continuously train the model on edge cases, improving its ability to differentiate between highly similar statutes.

Redefining Law Enforcement for a New Era

Our integrated system moves policing beyond traditional methods. By unifying predictive analytics, semantic FIR automation, and unbiased crime analysis, this platform streamlines workflows and supports India's critical transition toward automated, data-driven, and citizen-centric law enforcement.



Thank You.