International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 12 Issue: 12 | Dec 2025 www.irjet.net p-ISSN: 2395-0072

Automated Crime Reporting and Charge sheet Generation System Using Reinforcement

Vikash Mallick[1], Tanmayi Jaiswal[2], Sujal Dhurve[3], Arpita Bankar[4], Rakesh Moharle[5], Sushama Telrandhe[6]

[1,2,3,4]Dept of CSE, Gurunanak Institute of Engineering and Technology, Nagpur, Maharashtra, India
[5]Assistant Professor, Dept of CSE, Gurunanak Institute of Engineering and Technology, Nagpur, Maharashtra, India
[6]Associate Professor, Dept of CSE, Gurunanak Institue of Engineering and Technology, Nagpur, Maharashtra, India

---

Abstract - *Manual preparation of crime records, First Information Reports (FIRs), and charge-sheets remains a significant bottleneck in policing due to human error, procedural delays, and inconsistencies in documentation. To address these limitations, this research introduces an AI-driven framework that automates FIR drafting, crime categorization, IPC-based charge-sheet preparation, and penalty estimation. The system integrates machine learning, natural language processing, and legal-data analytics to classify offense types, map them to appropriate IPC sections, and generate structured legal documents. Logistic Regression combined with text-vectorization techniques enables reliable crime classification, while predictive analytics supports sentencing estimation based on historical case data. The approach enhances operational efficiency, reduces manual dependency, and ensures uniformity and transparency in law-enforcement workflows. Future extensions may include block chain-based data security and real-time integration with national crime databases.*

Key Words: **Automated Crime Reporting, FIR Generation, Charge-Sheet Automation, Crime Classification, Logistic Regression, Natural Language Processing, Indian Penal Code (IPC), Judicial Automation, Sentencing Prediction, Predictive Analytics, Machine Learning in Law Enforcement, Legal Document Processing, Crime Data Analysis, AI-Assisted Policing, Digital Forensics Integration.**

1. INTRODUCTION

Despite major digital governance initiatives, core components of India's criminal justice system still rely heavily on manual workflows such as FIR drafting, charge-sheet preparation, evidence logging, and IPC-based classification. Such manual dependency increases the likelihood of human error, procedural delays, and inconsistent interpretation of legal guidelines [4]. As police departments process thousands of incident reports daily, these inefficiencies contribute to overburdened courts, case backlogs, and delayed justice delivery.

AI-driven automation offers transformative potential to address these gaps. NLP methods provide accurate extraction of semantic information from textual complaints, enabling automated processing of narrative descriptions [1], [12]. ML classification algorithms, especially Logistic Regression, Naïve Bayes, and XG Boost, have been widely used to classify legal documents and crime types with high accuracy [2], [5]. Similarly, predictive systems have shown promising results in identifying crime trends, forecasting potential criminal behaviour, and estimating judicial outcomes [9].

By integrating these capabilities, the proposed system enables automatic FIR drafting, offense categorization, IPC-mapping, and sentencing prediction. Such automation aligns with global trends where legal AI systems are increasingly used for document screening, case-law retrieval, fraud detection, and judicial decision support [12], [15]. The system thus enhances efficiency, transparency, and fairness within law-enforcement procedures.

2. LITERATURE SURVEY

AI and NLP have become integral tools in modern legal informatics due to their ability to interpret, categorize, and analyse complex textual information. Several major research works highlight their application in criminal justice:

[1] NLP for Legal Documentation

Studies such as Vattikuti (2024) demonstrate that transformer-based models like BERT and GPT significantly improve clause extraction, anomaly detection, and document review accuracy in legal workflows [1]. Such systems reduce reliance on manual reading and improve consistency.

[2] Automated Crime Classification

Ku & Leroy (2013) developed a decision-support system using NLP and Naïve Bayes classifiers to categorize crime reports. Their system achieved an accuracy of 94.82%, surpassing human analysts [2]. This demonstrates the reliability of ML in automated crime categorization.

[3] Charge Prediction and Legal Reasoning

Ye et al. (2018) introduced a text-to-text model generating judicial reasoning ("court views") from case descriptions.

Their label-conditioned Seq2Seq architecture improved interpretability and showed strong applicability for charge prediction [3].

[4] Ethical & Human Rights Concerns
Završnik (2020) emphasized that AI in criminal justice must consider fairness, transparency, and risk of bias to avoid reinforcing systemic inequalities [4]. These concerns underline the importance of explainable and accountable AI systems.

[5] Neural Networks for Crime Forecasting
Walczak (2021) and Shah et al. (2021) highlighted how neural networks and computer-vision-based systems can predict crime hotspots, detect anomalies, and classify incidents using multimodal data sources [10], [11].

[6] Deep Learning and Crime Prediction
Safety et al. (2021) used deep learning models to analyse large datasets from Chicago and Los Angeles, demonstrating strong performance in crime forecasting but also highlighting challenges such as dataset bias and data imbalance [9].

Across literature, it is evident that combining NLP and ML with legal-domain knowledge can significantly enhance the automation of crime-reporting processes and judicial documentation.

3. SYSTEM ARCHITECTURE

The architecture follows a multi-module design where each subsystem contributes to automation and decision support.

1. Complaint Processing Interface
Users submit text-based descriptions of incidents. NLP techniques such as tokenization, NER, and dependency parsing help extract essential details. Similar approaches have been used for legal information retrieval and document classification.

2. NLP-Based FIR Generator
NLP models structure free-text input into formal FIR fields (crime nature, incident timeline, location, involved persons). Prior research confirms the effectiveness of NLP in legal entity extraction and structured document generation.

3. Crime Classification Module
Count Vectorizer transforms text into numerical features, and Logistic Regression classifies the offense into cognizable or non-cognizable categories. Logistic models are commonly used for law-related text classification due to their interpretability.

4. Evidence Repository
Digital evidence—images, videos, documents—is stored in a centralized database. Such repositories support efficient retrieval and cross-case referencing as suggested by forensic-AI research.

5. Charge-Sheet Generation Engine
Based on the offense type, the system maps the incident to relevant IPC sections using pre-programmed legal rules and ML inferences. Similar hybrid systems combining rule-based and ML approaches have been proposed in legal-analytics research.

6. Sentencing Prediction Module
Using historical case records, the system predicts sentencing severity with ML models. Predictive systems have been used in various legal contexts, including ECHR judgment prediction, US Supreme Court predictions, and legal-judgment modelling.

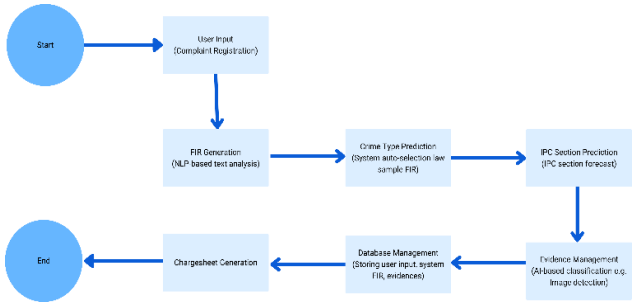

Fig . 1. System Architecture

4. FIR GENERATION

The FIR is generated automatically using a structured template. NLP facilitates extraction of:

- Crime description and contextual narrative
- Legal classification (cognizable / non-cognizable)
- Spatio-temporal attributes
- Suspect/witness identification (NER)
- Evidence summary extracted from uploaded documents
- Officer assignment and case status

5. OUTPUT AND INTERPRETATION

The system outputs:

- Crime category (cognizable or non-cognizable)
- Structured FIR generated through NLP
- IPC-mapped charges (for cognizable offenses)
- Recommended sentencing ranges based on ML predictions

ML-based classification reliability is evaluated using accuracy, precision, recall, and F1-score, similar to metrics used in crime-forecasting studies [9], [10].

6. CONCLUSION

This research confirms that integrating NLP and ML can significantly automate essential components of criminal-justice workflows. FIR drafting, offense classification, IPC mapping, and sentencing prediction—traditionally manual and time-consuming tasks—can be reliably automated. The proposed system draws on proven methodologies such as Logistic Regression, Seq2Seq reasoning, crime-prediction models, and forensic-AI frameworks [1]–[14], [16]. Future improvements include adoption of transformer-based models, block chain-backed audit trails, and real-time connectivity with national crime databases for end-to-end automation. These advancements mark a major step toward transparent, efficient, and technologically empowered criminal-justice administration.

7. REFERENCES

[1] Vattikuti, Manoj Chowdary. "Natural Language Processing for Automated Legal Document Analysis and Contract Review." International Journal of Sustainable Development in Field of IT, vol. 16, no. 16, 2024.

[2] Ku, Chih-Hao, and Gondy Leroy. "Automated Crime Report Analysis and Classification for E-Government and Decision Support." Proceedings of the 14th Annual International Conference on Digital Government Research, 2013.

[3] Ye, Hai, et al. "Interpretable Charge Predictions for Criminal Cases: Learning to Generate Court Views from Fact Descriptions." arXiv preprint arXiv:1802. 08504, 2018.

[4] Završnik, Ales. "Criminal Justice, Artificial Intelligence Systems, and Human Rights." ERA Forum, vol. 20, no. 4, 2020.

[5] Jenga, Karabo, Cagatay Catal, and Gorkem Kar. "Machine Learning in Crime Prediction." Journal of Ambient Intelligence and Humanized Computing, vol. 14, no. 3, 2023, pp. 2887–2913.

[6] Jeong, Doowon. "Artificial Intelligence Security Threat, Crime, and Forensics: Taxonomy and Open Issues." IEEE Access, vol. 8, 2020, pp. 184560–184574.

[7] King, Thomas C., et al. "Artificial Intelligence Crime: An Interdisciplinary Analysis of Foreseeable Threats and Solutions." Science and Engineering Ethics, vol. 26, 2020, pp. 89–120.

[8] Sushina, Tatyana, and Andrew Sobenin. "Artificial Intelligence in the Criminal Justice System: Leading Trends and Possibilities." 6th International Conference on Social, Economic, and Academic Leadership (ICSEAL-6-2019), Atlantis Press, 2020.

[9] Safat, Wajiha, Sohail Asghar, and Saira Andleeb Gillani. "Empirical Analysis for Crime Prediction and Forecasting Using Machine Learning and Deep Learning Techniques." IEEE Access, vol. 9, 2021, pp. 70080–70094.

[10] Shah, Neil, Nandish Bhagat, and Manan Shah. "Crime Forecasting: A Machine Learning and Computer Vision Approach to Crime Prediction and Prevention." Visual Computing for Industry, Biomedicine, and Art, vol. 4, no. 1, 2021.

[11] Walczak, Steven. "Predicting Crime and Other Uses of Neural Networks in Police Decision Making." Frontiers in Psychology, vol. 12, 2021, 587943.

[12] Surden, Harry. "Artificial Intelligence and Law: An Overview." Georgia State University Law Review, vol. 35, 2018, p. 1305.

[13] Yang, Wenmian, Weijia Jia, Xiaojie Zhou, and Yutao Luo. "Legal Judgment Prediction via Multi-Perspective Bi-Feedback Network." arXiv preprint arXiv:1905 .03969, 2019.

[14] Medvedeva, Masha, Michel Vols, and Martijn Wieling. "Using Machine Learning to Predict Decisions of the European Court of Human Rights." Artificial Intelligence and Law, vol. 28, no. 2, 2020, pp. 237–266.

[15] Ashley, Kevin D. Artificial Intelligence and Legal Analytics: New Tools for Law Practice in the Digital Age. Cambridge University Press, 2017.

[16] Zhong, Haoxi, Zhipeng Guo, Cunchao Tu, Chaojun Xiao, Zhiyuan Liu, and Maosong Sun. "Legal Judgment Prediction via Topological Learning." Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, pp. 3540–3549, 2018.