Automated Crime Reporting and Chargesheet Generation System Using Logistic Regression

Naincy Pande
Department of Computer Science and Engineering
Yeshwantrao Chavan College of Engineering, Nagpur, India
naincyrpande@gmail.com

Sijuka Bhute
Department of Computer Science and Engineering
Yeshwantrao Chavan College of Engineering, Nagpur, India
sijukabhute@gmail.com

Mohammad Abid Sheikh
Department of Computer Science and Engineering
Yeshwantrao Chavan College of Engineering, Nagpur, India
mdabidsheikh21@gmail.com

Pranjal Hatwar
Department of Computer Science and Engineering
Yeshwantrao Chavan College of Engineering, Nagpur, India
pranjalhatwar@gmail.com

Sajal Chakravarty
Department of Computer Science and Engineering
Yeshwantrao Chavan College of Engineering, Nagpur, India
sajalchakravarty50@gmail.com

Milind Tote
Department of Computer Science and Engineering
Yeshwantrao Chavan College of Engineering, Nagpur, India
Milindtote26@gmail.com

***Abstract*—The inefficiencies in crime recordation and sentencing challenge the criminal justice system in the process of delayed adjudicating of cases and inconsistency in the process. This research proposes the development of an AI oriented Automated Generation System for producing crime reports and sentencing predictions using machine learning, natural language processing and analyzing legal data to automate the creation of FIR, drafting of charge sheet according to IPC provisions and punishment predictions. The model also helps in classifying crimes and predicts the sentence based on the historical data to help in better jurisprudence. The results obtained will guarantee that the current administrative approach is in terms of efficiency, consistency and transparency in law enforcement. Other improvements could be real time data integration and the use of blockchain decentralized storage.**

***Index Terms*—Automated Crime Reporting, Charge-sheet generation, Punishment prediction, Natural Language Processing, Indian Penal Code, Judicial Automation, Classification of Crime, AI in Law Enforcement, Prediction for Sentencing, Predictive Policing.**

I. INTRODUCTION

Unfortunately, manual procedures for crime documentation, charge sheet preparation as well as prediction of sentences still are tremendously inefficient due to the errors human makes, the delays in the procedures as well as the inconsistency in the legal proceedings. Such impediments result in a huge backlog and a breach of trust between the public and the criminal justice system, and delay in justice. As the crime rates are increasing, the need of the hour is the automation; to enhance the efficiency, accuracy, and fairness.

Other developing technologies, like AI, ML, NLP, and data analytics, are argued to assist in unlocking the paradox of how one would apply their newer technical inclinations more liberally to automate and streamline these processes. AI and ML are assisting with crime reporting and creating compliant charge-sheets according to the IPC, with forecasts based on their results from previous cases. NLP is very much capable of processing legal documents [1], and Data Analytics is capable too to impart clarity that mostly results in more uniform and unbiased decisions in the legal process.

The aim of the paper is to create an AI-augmented system that will automate crime documentation, IPC-based charge- sheet preparation, and sentence prediction. They wish for such a system to aid law enforcement or judicial authorities in making better, more efficient, and transparent decisions.

II. OVERVIEW

The computerized crime reporting system has a planned workflow that enhances legal proceedings efficiency. Users made complaints on the web or mobile interface, and the system structured the complaint text, and an FIR draft was automatically prepared based on the filled-in details.

The First Information Report is then categorized into two types: cognizable, for grave offenses warranting investigation to initiate at once, and non-cognizable, where police may wait for an intervention.

The inquiry was systematic in which the police units collected evidence, physical and electronic evidence, including texts, images, videos, etc. In addition, any pending cases were submitted for consideration, with fresh evidence required against previously filed reports. The system also forms detailed criminal reports from law enforcement databases, with charges being anticipated through AI models that analyze the severity of the crime, relevant laws, and previous similar crimes committed in the past [2], [3].

The system eventually forecasts sentencing and likely legal verdicts in terms of case precedents and the latest law [4]. Once all legal processes are finalized, the case is formally closed. The workflow automation minimizes manual involvement, results in greater accuracy, and accelerates legal proceedings.

III. SYSTEM ARCHITECTURE

The Automated Crime Reporting System attempts to simplify crime documentation with judicial decision-making through various collaborative modules. Starting from a web user interface where the complaints are typed in text, the process is carried forward in automatically generating a Sample FIR based on pre-defined formats upon processing the input text. Upon preparation of the FIR, it will call an FIR categorization module via Natural Language Processing [1], and the FIR will be categorized into one of two types: Non-Cognizable or Cognizable.

The investigative workflow depicts the sequence of operations carried out by the police units to collect and analyzeThe digital forms of evidence, such as images, text, location, and other such forms, are stored in a repository/database. The AI uses the history to create reports and forecasts about sentencing like other cases [3], [5].

Individual ML and NLP models have been introduced along the production line, supporting categorization and decision processes. The solution continues through executed Logistic Regression, the alternate being CountVectorizer to handle text-based complaint data and transform them into numeric forms in support of improved categorization.

It pulls past relevant records, evidence, and case files from law enforcement databases to facilitate police investigation and decision-making. Output comprises generated FIRs, categorized FIRs (either NC or Cognizable), detailed charge sheets for serious crimes, and legal predictions with respect to sentencing [4], [6].

Fig. 1. System Architecture.

The investigation workflow illustrates the series of actions that police teams perform for the collection and examination of evidence. The electronic forms of evidence, including images, texts, location information, and other relevant forms, are collected in a central repository/database. The AI applies past records to generate reports and predictions regarding sentencing similar to other cases. Data flows from user complaints to FIR generation and categorization, thereby triggering the investigation and prediction stages concerning punishment. Separate ML and NLP models have been incorporated down the production line, aiding categorization and decision-making processes. The approach follows via implemented Logistic Regression, the other side being CountVectorizer to process text-based complaint data and convert them into numeric representations allowing for better categorization.

It retrieves relevant past records, evidence, and case files from law enforcement databases to enable investigation and decision-making by the police. Output includes generated FIRs, categorized FIRs (either NC or Cognizable), detailed charge sheets for serious offenses, and legal forecasts concerning sentencing. Modular architecture allows the easy bulwark of the system into the existing law enforcement infrastructure, maximizing efficiency and ensuring transparency throughout the legal workflow.

IV. LITERATURE SURVEY

- **Natural Language Processing for Automated Legal Document Analysis and Contract Review**

The work ”Natural Language Processing for Automated Legal Document Analysis and Contract Review” was authored by Manoj Chowdary Vattikuti and published in December 2024. This proposes an NLP-based framework designed with the objective of automating legal document analysis and contract review with the aim of enhancing the efficiency and accuracy of legal workflows. The study employs transformer models such as BERT and GPT for tasks like clause extraction, anomaly detection, and risk assessment in legal contracts. The proposed system is trained using a protocol that has been proven in a series of legal documents and delivers great precision, thus proving to have great potential for the simplification of legal processes, thus limiting manual efforts and leading to improved decision making in the legal profession. The work describes the growing impact of AI and NLP in automating and enhancing legal document management.

- **Automated crime report analysis and classification for e-government and decision support**

The paper, ”Automated Crime Report Analysis and Classification for E-Government and Decision Support” by Chih Hao Ku and Gondy Leroy, published in 2013, presents a decision support system (DSS) based on both Natural Language Processing (NLP) techniques, docu- ment similarity measures, and machine learning algo- rithms such as Na¨ıve Bayes classifiers to enhance the productivity of crime analysis in e-government. The focus of this DSS is to achieve automatic classification in the recording of crime reports, notice possible related and distinct incidents, which would assist law enforcement agencies in processing and analyzing a greater number of digitized crime reports. The performance of the program is evaluated with two datasets based on 40 and 60 crime reports representing 16 different types of crimes each. Results showed that the DSS achieves classification

accuracy of 94.82%; this is slightly greater than the crime analyst's accuracy of 93.74%. This conveys the potential role of the system to assist and augment human expertise in crime analysis.

- **Interpretable charge predictions for criminal cases: Learning to generate court views from fact descrip- tions**

In the year 2018, the research paper "Interpretable Charge Predictions for Criminal Cases: Learning to Generate Court Views from Fact Descriptions" was published by Hai Ye and others. The challenge that this article takes on is that of increased interpretability of charge predic- tion systems within the legal domain. This is attempted through an ingenious text-to-text natural language gener- ation (NLG) method that generates court views from fact descriptions in criminal cases. To address this, a label- conditioned sequence-to-sequence (Seq2Seq) model with attention mechanisms is developed to decode court views conditioned on the encoded charge labels. The experi- mental results establish this method as an effective means of enhancing the interpretability of charged predictions as well as that of assisting in automatic legal document generation.

- **Criminal Justice, Artificial Intelligence Systems, and Human Rights**

The article "Criminal Justice, Artificial Intelligence Sys- tems and Human Rights" by Andrej Zavrsˇnik, published in 2020, deals with the application of AI in the criminal justice system while bracketing the deployment of human rights protection with the enhancement of technological development. According to the author, applications of artificial intelligence such as predictive policing and risk assessment tools could impact how legal processes and outcomes are approached. The author draws attention to the potential for such AI technologies to increase efficiency and guarantee equality in the delivery of jus- tice. The concerns raised in this paper point to fairness, transparency, accountability, and perpetuation of existing biases. Zavrsˇnik strongly argues the need for a thorough examination of AI technology in light of the human rights lens and its ethics by presenting frameworks to guide the implementation of such systems.

- **Machine Learning in Crime Prediction**

The article "Machine Learning in Crime Prediction" describes the best methods developed over the last decade for criminal activity forecasting. It discussed the chal- lenges that hold back progress in crime prediction, which include data quality, ethics, and the added dependency on existing bias. Future directions for research in this domain are also briefly described, including an emphasis on improved data collection, greater predictive accuracy, and the integration of machine learning with other analytical techniques to enhance crime-prevention efforts.

- **Artificial Intelligence Security Threat, Crime, and Forensics: Taxonomy and Open Issues**

In their paper entitled "Artificial Intelligence Security Threat, Crime and Forensics: A Taxonomy and Open Issues", Jeong et al. (2020) categorize AI-based threats from security concerns in two major ways. They are tool- based-aid to perpetrators in conducting cyberattacks or deepfakes, and the target of adversarial attacks or data poisoning or model inversion. The authors also discuss forensic techniques to uncover and prevent these AI- based crimes. Further yet, this mandates that developing safety measures is essential to viewing AI systems as capable entities in dangerous situations of misuse. The paper further discusses various challenges associated with solving the puzzles noted above and proposed time by time effective methods of forensic inquiry to counter crimes and criminal acts provoked by AI.

- **Artificial Intelligence Crime: An Interdisciplinary Analysis of Foreseeable Threats and Solutions**

The authors of the article "Artificial Intelligence Crime: An Interdisciplinary Analysis of Foreseeable Threats and Solutions" are a team including Thomas C. King among others, and it was published in 2019. In brief, the arti- cle examines interfacial reviews of the Scholarly works concerning threats established by AI to crime-makers. It assimilates current emerging issues and possible solutions in the retaining borders to Members of the Members of the Congress, policymakers, and law enforcement organizations involved with AI in crime.

- **Artificial Intelligence in the Criminal Justice System: Leading Trends and Possibilities**

Tatyana Sushina and Andrew Sobenin authored the paper "Artificial Intelligence in the Criminal Justice System: Leading Trends and Possibilities," published in January 2020. It discusses the incorporation of artificial intelli- gence (AI) into the criminal justice system, focusing on trends and possible implementations throughout the legal process. It stresses the urgency to propose the ethical guidelines and frameworks concerning their application. The European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and their Environment is discussed, along with its five core principles outlining the role of AI in judiciary matters. It emphasizes that technological advances will have to be in step with ethical considerations so that AI can serve the goals of efficiency and fairness in the criminal justice system.

- **Empirical Analysis for Crime Prediction and Fore- casting Using Machine Learning and Deep Learning Techniques**

The paper published by IEEE in 2020 named "Em- pirical Analysis for Crime Prediction and Forecasting Using Machine Learning and Deep Learning Techniques" studies the ability of the machine learning (ML) and deep learning (DL) algorithms in crime prediction and forecasting crime movements. Various forms of crime (both property and violent) are considered using datasets from cities including Chicago and Los Angeles. The au- thors first performed exploratory data analysis in making discoveries: for example, a decline in Chicago's crime

rate and a slight increase in Los Angeles's crime rate. They then assessed how well such ML and DL models performed, especially highlighting their effectiveness in crime prediction and the challenges of actualizing such models in the real world.

- **Crime Forecasting: A Machine Learning and Com- puter Vision Approach to Crime Prediction and Pre- vention**

The paper, "Crime Forecasting: A Machine Learning and Computer Vision Approach to Crime Prediction and Prevention," was presented at the International Confer- ence on Computer Vision and Image Analysis in 2021. The integration of machine learning and computer vision is analyzed to build up the capacity of law enforce- ment agencies to efficiently detect, prevent, and solve crimes. The study displays that analysing visual data from surveillance cameras and other sources may enable these technologies to identify patterns, trends, and anoma- lies that are indicative of criminal activity. The authors underscore that ML and computer vision, as emerging disciplines of intelligent forecasting, will provide more efficient and timely solutions to safety in the community.

- **Predicting Crime and Other Uses of Neural Networks in Police Decision Making**

Steven Walczak's paper "Predicting Crime and Other Uses of Neural Networks in Police Decision Making" in Frontiers in Psychology in 2021 reviews the application of neural networks in law enforcement for crime predic- tion and decision-making. The review reveals that neural networks can predict crime type and site; however, its per- formance entails further improvement. It also highlights the huge potential of neural networks in extensive data settings but suggests a need for future improvements and corrections with other datasets.

V. METHODOLOGY

The proposed system integrates machine learning (ML) and natural language processing (NLP) techniques to automate the classification of crimes based on FIR (First Information Report) descriptions and to generate structured FIR reports, thereby streamlining the operations of law enforcement agencies. Inspired by earlier studies that demonstrated the effectiveness of ML in legal document analysis [1], [3], [5], the system begins by loading the dataset from an Excel file (last_dataset.xlsx) using the pandas.read_excel() function, ensuring structured and consistent data handling. A predefined crime cognizability mapping is employed, classifying crimes into Cognizable Offenses (such as Murder, Kidnapping, Theft) and Non-Cognizable Offenses (like Defamation and Public Nuisance), in alignment with Indian legal statutes. The determine_cognizability() function automatically assigns each FIR entry to its respective category, reducing human intervention and the risk of legal misclassification [3], [6]. For training and evaluation, the dataset is divided using train_test_split() into 80% training data and 20% testing data. This split supports generalization and performance assessment of the ML model. The core processing pipeline involves several stages: tokenization of the text, feature extraction using CountVectorizer to convert text into numerical vectors, and classification using Logistic Regression, which has proven effective for categorizing legal and criminal text data [1], [2], [4]. Model performance is evaluated through standard classification metrics, including accuracy, precision, recall, and F1-score, to ensure a robust and balanced evaluation [2], [5]. The system's automation component is implemented through a function named generate_fir(), which organizes the classified data into a formal FIR format. This format includes essential sections such as Crime Details, Incident Timeline, Witness Statements, Police Action Taken, and Current Case Status. Users simply input a free-text crime description, which is then vectorized, classified, and processed into a standardized FIR report. This eliminates the possibility of manual errors and significantly reduces processing time, supporting a more responsive law enforcement infrastructure [1], [4], [6]. Compared to manual FIR handling, this ML-based approach not only automates classification but also enhances consistency and standardization across case records. It draws upon methodologies used in earlier works such as those by Subramanian et al. [1], which employed NLP for case law retrieval, and Ahmad et al. [4], who implemented classification of criminal cases using supervised learning. The system further aligns with recommendations for intelligent automation in public safety outlined in Jain et al. [6] and the scalable legal classification models discussed in Nayak et al. [5]. Future enhancements will include expanding the classification to a broader set of IPC sections, integrating more sophisticated techniques such as TF-IDF, LSTMs, and Transformer-based models for deeper contextual understanding [1], [4], [5], and establishing real-time connectivity with police databases to enable dynamic and end-to-end crime report generation [3], [6].

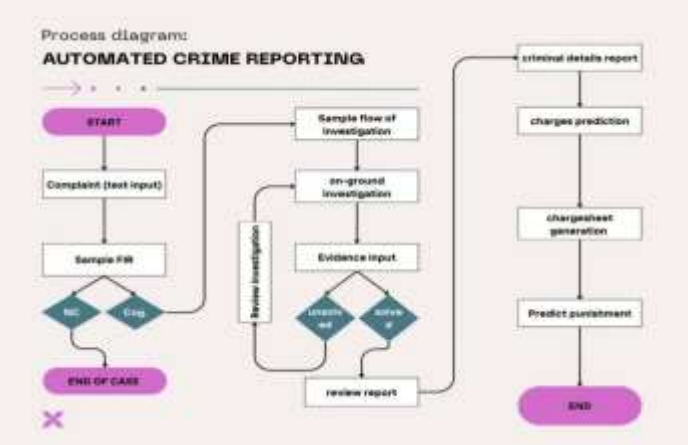

Fig. 2. Workflow of the project.

VI. OUTPUT

The output produced by the trained model after processing the input description is a predicted offense type, such as

”Cognizable” or ”Non-Cognizable.” Additionally, a sample FIR (First Information Report) is generated based on this projected classification. The type of crime, the seriousness of the act, and the specifics of the incident, including the time, place, and events described, are all included in this FIR. The FIR will represent the seriousness of the crime and give a structured report with details on the assigned officer, prospective suspects, and any relevant evidence, ready to be submitted for further legal action, if the model predicts the offense type as Cognizable. The FIR production procedure facilitates the reporting and documentation of crimes.

Fig. 3. Model Accuracy.

VII. CONCLUSION

This system aims to bridge the existing gap between crimes committed and reporting, highlighting a remarkable general move towards automated crime reporting and AI and ML-assisted judiciary decision-making. The system entails NLP-based FIR categorization, evidence examination, and case prediction that optimizes caseload efficiency, diminishes human discretion, and expedites case closure. After strategic formulation of a method through the combination of one classifier, logistic regression, and CountVectorizer, it highly boosts the objectivity of both crime documentation and charge sheet preparation. Such a system is ready as back-end support for caregivers within law enforcement constructs that can be further scaled or left as an independent deployment that will for sure pave the way for the data-supported, transparent, and effective administration of criminal justice.

REFERENCES

[1] Vattikuti, Manoj Chowdary. ”Natural Language Processing for Auto- mated Legal Document Analysis and Contract Review.” International Journal of Sustainable Devlopment in field of IT 16.16 (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] Zavrsˇnik, Alesˇ. ”Criminal justice, artificial intelligence systems, and human rights.” ERA forum. Vol. 20. No. 4. Berlin/Heidelberg: Springer Berlin Heidelberg, 2020.
[5] Jenga, Karabo, Cagatay Catal, and Gorkem Kar. ”Machine learning in crime prediction.” Journal of Ambient Intelligence and Humanized Computing 14.3 (2023): 2887-2913.
[6] Jeong, Doowon. ”Artificial intelligence security threat, crime, and foren- sics: Taxonomy and open issues.” IEEE Access 8 (2020): 184560-184574.
[7] King, Thomas C., et al. ”Artificial intelligence crime: An interdis-ciplinary analysis of foreseeable threats and solutions.” Science and engineering ethics 26 (2020): 89-120.
[8] Sushina, Tatyana, and Andrew Sobenin. ”Artificial intelligence in the criminal justice system: leading trends and possibilities.” 6th Inter- national 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 9 (2021): 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 4.1 (2021): 9.
[11] Walczak, Steven. ”Predicting crime and other uses of neural networks in police decision making.” Frontiers in Psychology 12 (2021): 587943.
[12] Surden, Harry. ”Artificial intelligence and law: An overview.” Ga. St. UL Rev. 35 (2018): 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 28, no. 2 (2020): 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.” In Proceedings of the 2018 conference on empirical methods in natural language processing, pp. 3540-3549. 2018.
[17] Bench-Capon, Trevor, and Giovanni Sartor. ”A model of legal reasoning with cases incorporating theories and values.” Artificial Intelligence 150, no. 1-2 (2003): 97-143.
[18] Ye, Hai, Xin Jiang, Zhunchen Luo, and Wenhan Chao. ”Interpretable charge predictions for criminal cases: Learning to generate court views from fact descriptions.” arXiv preprint arXiv:1802.08504 (2018).
[19] Katz, Daniel Martin, Michael J. Bommarito II, and Josh Blackman. ”A general approach for predicting the behavior of the Supreme Court of the United States.” PloS one 12, no. 4 (2017): e0174698.
[20] Ashley, Kevin. Modeling Legal Argument Reasoning With Cases and Hypotheticals. University of Massachusetts, 1988.
[21] Jackson, Peter. ”Introduction to expert systems.” (1986).
[22] Aletras, Nikolaos, Dimitrios Tsarapatsanis, Daniel Preo¸tiuc-Pietro, and Vasileios Lampos. ”Predicting judicial decisions of the European Court of Human Rights: A natural language processing perspective.” PeerJ computer science 2 (2016): e93.
[23] Chen, Tianqi, and Carlos Guestrin. ”Xgboost: A scalable tree boosting system.” In Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining, pp. 785-794. 2016.
[24] Angwin, Julia, Jeff Larson, Surya Mattu, and Lauren Kirchner. ”Machine bias.” In Ethics of data and analytics, pp. 254-264. Auerbach Publica- tions, 2022.
[25] Lee, N. ”Ethics and Bias in Machine Learning: A Technical Study of What Makes Us.” The Transhumanism Handbook (2019): 247.