International Journal of Trend in Scientific Research and Development (IJTSRD)
Special Issue on Advancements and Emerging Trends in Computer Applications - Innovations, Challenges, and Future Prospects
Available Online: www.ijtsrd.com e-ISSN: 2456 – 6470

CRIME INSIGHTS: Visualization and Analysis with Python

Laxmit Kaur Chana

PG Student, Department of Computer Application, G. H. Raisoni University, Amravati, Maharashtra, India

ABSTRACT

The rise in crime rates across various regions has made it increasingly important to analyze and understand crime trends for effective law enforcement and policymaking. This project, "Crime Data Visualization and Analysis Using Python," aims to develop a system that enables the exploration, analysis, and visualization of crime-related data to identify patterns, trends, and anomalies. The system leverages Python libraries such as Pandas, Matplotlib, Seaborn, and Plotly to clean, process, and visualize crime datasets. It allows users to filter data by parameters like crime type, location, and time frame, enabling a deeper understanding of how crimes evolve over time and space. Advanced features like heatmaps, time-series graphs, and geographical mapping are included to provide a more intuitive and interactive user experience. The insights derived from this project can be used by law enforcement agencies, researchers, and policymakers to identify high-crime zones, predict future crime hotspots, and allocate resources more efficiently. Overall, this project demonstrates how data science and visualization can play a crucial role in enhancing public safety and strategic decision-making.

KEYWORDS: *Python (NumPy, Jupyter Notebook), HTML/CSS, JavaScript, React.js*

I. INTRODUCTION

Crime analysis plays a vital role in understanding the nature, frequency, and distribution of criminal activities within a particular region. With the increase in publicly available crime datasets, there is a growing opportunity to utilize data science and visualization techniques to derive meaningful insights from this information. Traditional methods of analyzing crime data are often time-consuming and lack interactivity, making it difficult for law enforcement agencies and policymakers to make quick and informed decisions.

This project, titled *Crime Data Visualization and Analysis Using Python*, is developed with the objective of analyzing and presenting crime data in a visually comprehensive manner. The system leverages Python programming along with data science libraries such as Pandas for data manipulation, Matplotlib and Seaborn for static visualization, and Folium for geographic mapping. These tools help to transform raw data into understandable and interactive visual formats such as bar charts, line graphs, heatmaps, and location-based maps.

The project primarily focuses on identifying patterns in crime occurrences based on parameters such as crime type, time, and geographical location. This allows users to detect trends, observe seasonal variations, and pinpoint high-crime areas. The insights generated can assist law enforcement agencies in effective decision-making, such as allocating resources to crime-prone zones or predicting future incidents based on historical data.

In addition to visualization, the project also introduces basic analytical techniques which could be extended in future work to include predictive modeling using machine learning. Overall, this project demonstrates power of data analytics in crime prevention and community safety, offering a foundation for more advanced crime Intelligence systems.

Abbreviations and Acronyms:

- **QoL**: Quality of Life
- **KPI**: Key Performance Indicator
- **NPS**: Net Promoter Score
- **CSAT**: Customer Satisfaction
- **CRM**: Customer Relationship Management
- **2FA**: Two-Factor Authentication
- **VPN**: Virtual Private Network
- **SSL**: Secure Sockets Layer

Units:

- **Number of Users**: Quantify the total users interacting with the portal.
- **Session Duration**: Measure the average time (in minutes or seconds) users spend per session.
- **Page Views**: Count the average number of pages viewed per session.
- **Bounce Rate**: Expressed as a percentage, this indicates the proportion of users who leave the portal after viewing only one page.
- **Response Time**: Average time (in milliseconds or seconds) the system takes to respond to user actions.
- **Response Rate**: Percentage of participants who complete the survey out of those invited.
- **Drop-off Rate**: Percentage of participants who start but do not complete the survey.
- **Average Response Time**: Mean time (in minutes or seconds) taken by participants to complete the survey.

II. RELATED WORK

Several research studies and projects have focused on the analysis and visualization of crime data to improve public safety and decision-making. Most of these works utilize publicly available datasets and apply statistical or machine learning techniques to identify crime trends, hotspots, and risk factors. One notable study is the use of **heatmaps and clustering algorithms** to identify crime-prone areas within large metropolitan cities.

These visual tools help law enforcement allocate resources more efficiently. Projects like **Crime Mapping using GIS** (Geographic Information Systems) have demonstrated the effectiveness of combining geographic data with crime statistics to create interactive dashboards for city administrators and police departments. Other studies have implemented **time-series analysis and prediction models**, using machine learning algorithms such as Linear Regression

and Random Forest to forecast future crime occurrences based on historical data.

In addition, platforms such as **Chicago Crime Data Portal** and **Kaggle Datasets** have made real-time crime data more accessible, enabling researchers to explore trends based on location, time, and crime category. However, these platforms primarily serve as data sources and do not provide customized analytical tools for specific research needs.

III. DATA AND SOURCES OF DATA

The primary dataset used for this project was obtained from publicly available sources such as the **Kaggle Crime Dataset**, **Chicago Data Portal**, and other open government platforms that provide real-time or historical crime records. These datasets typically include information such as the type of crime, date and time, location (latitude and longitude), and crime description. The data was cleaned and preprocessed using Python to remove missing values, inconsistencies, and duplicates before analysis.

The project was developed using the Python programming language, with the following libraries and tools:

- **Pandas**: For data loading, manipulation, and preprocessing.
- **NumPy**: For numerical operations and data handling.
- **Jupyter Notebook**: For coding, testing, and visual presentation of data analysis.

IV. RESEARCH METHODOLOGY

This research aims to analyze crime data and present it visually to uncover trends, patterns, and insights that can support decision-making for law enforcement agencies and policy makers. The methodology was developed through a series of structured steps, including data collection, preprocessing, exploratory data analysis (EDA), visualization, and interpretation.

1. **Data Collection:** The first step involved gathering publicly available crime datasets from sources such as the **Kaggle Crime Dataset**, the **Chicago Data Portal**, and other open government databases. These datasets contain detailed information about criminal activities, including crime types, geographical locations (latitude and longitude), and timestamps of the incidents. These datasets were selected for their comprehensive nature and frequent updates, ensuring the reliability and relevance of the data.
2. **Exploratory Data Analysis (EDA):** EDA was performed to understand the distribution of crime types, temporal trends, and geographical locations. Statistical methods were used to summarize the data, and initial visualizations were created to identify patterns in the data. This helped in detecting outliers, missing values, and understanding the structure of the dataset.
3. **Crime Data Visualization:** The main focus of the project was the visualization of crime data. Various visual representations were created to provide insights:
   - **Time-series analysis**: Line graphs were used to track crime occurrences over time, helping to identify trends and peak crime periods.
   - **Bar and pie charts**: Crime types were represented using bar charts, and the distribution of crimes by different categories was shown using pie charts.
4. **Limitations and Future Work:** One limitation of the project is the reliance on publicly available datasets, which may not be fully comprehensive or up to date. Future work could incorporate real-time crime data and apply machine learning models for predictive analysis of crime trends.

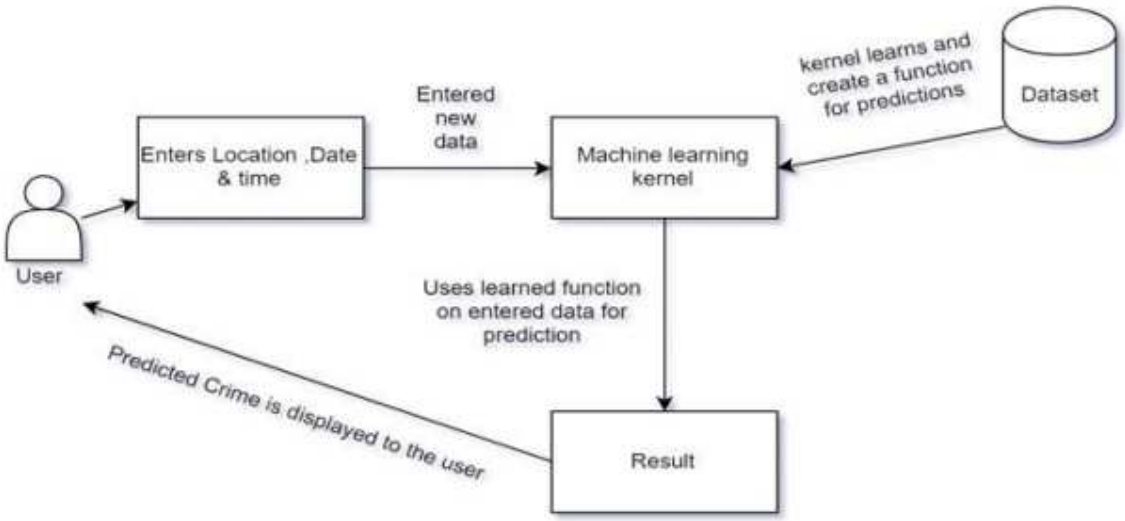

Fig.1 Architectural Diagram

There are many kinds of architecture diagrams, like software architecture diagram, system architecture diagram, application architecture diagram, security architecture diagram, etc. For system developers, they need system architecture diagrams to understand, clarify, and communicate ideas about the system structure and the user requirements that the system must support. It describes the overall features of the software and is concerned with defining the requirements and establishing the high level of the system. During architectural design, the various web pages and their interconnections are identified and designed. The major software components are identified and decomposed into processing modules and conceptual data structures and the interconnections among the modules are identified. The following modules are identified in the proposed system. The system architectural design is the design process for identifying the subsystems making up the system and framework for subsystem control and communication. The goal of architectural design is to establish the overall structure of software system.

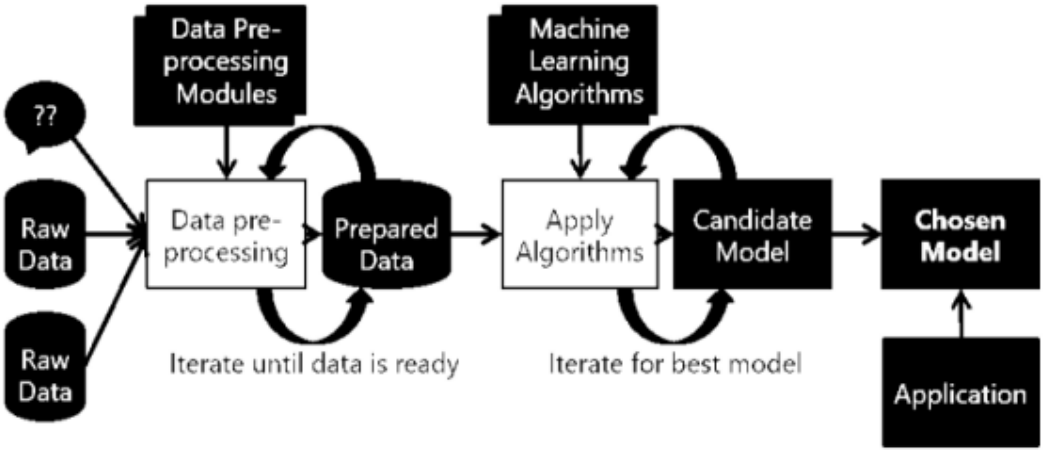

Fig.2 Data Flow Diagram

Steps Involved in Supervised Learning: • First Determine the type of training dataset • Collect/Gather the labelled training data. • Split the training dataset into training dataset, test dataset, and validation dataset. • Determine the input features of the training dataset, which should have enough knowledge so that the model can accurately predict the output. • Determine the suitable algorithm for the model, such as support vector machine, decision tree, etc. • Execute the algorithm on the training dataset. Sometimes we need validation sets as the control parameters, which are the subset of training datasets. • Evaluate the accuracy of the model by providing the test set.

V. RESULTS AND DISCUSSION

Results of Descriptive Statics of Study Variable

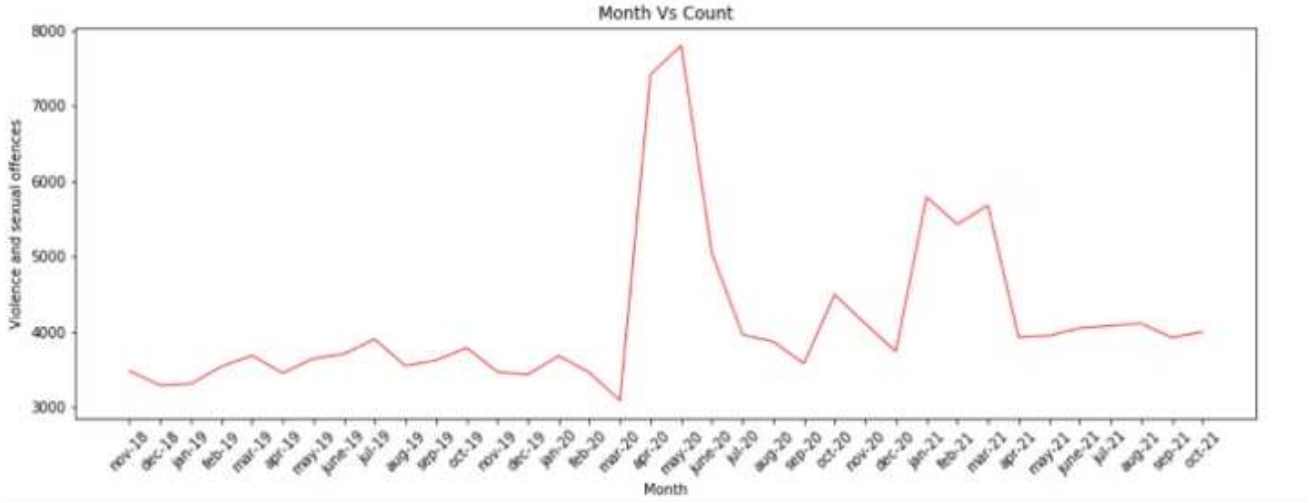

Fig 3: Month wise violence and sexual Offence

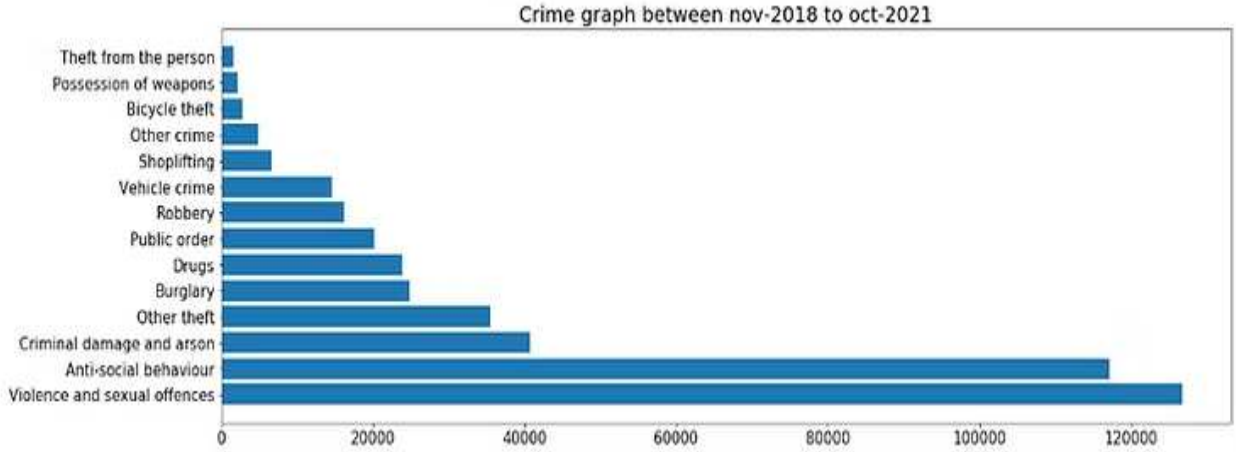

Fig 4: Crime graph between nov-2018 to oct-2021

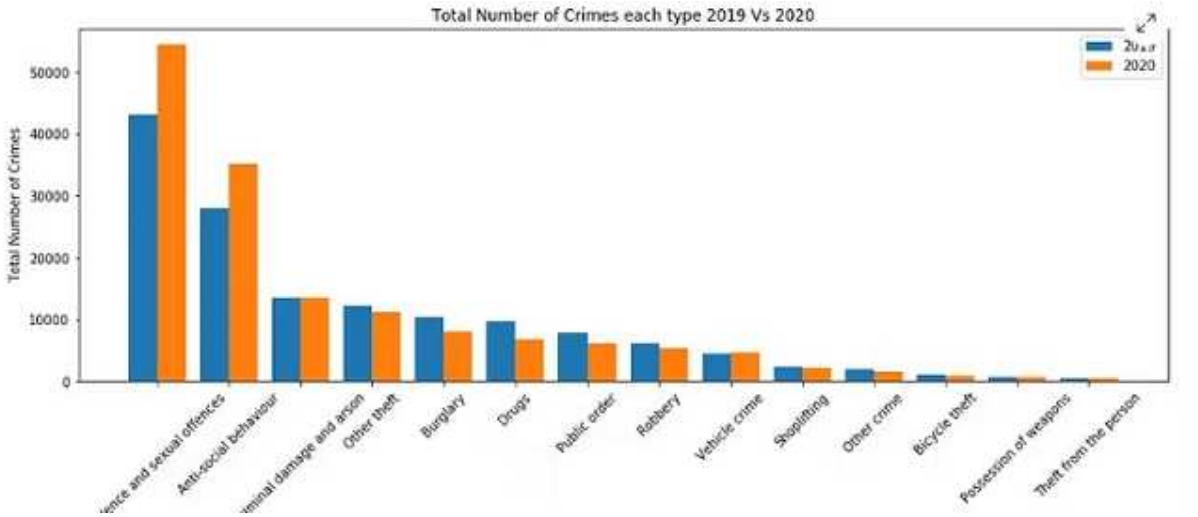

Fig 5: Total Number of crimes of each type

In addition to showcasing our expertise in crime data analysis and visualization, we at coder's arts are dedicated to providing comprehensive support and assistance to individuals and teams working on similar projects. Our team of experienced data scientists, analysts, and technical writers is ready to lend a helping hand to those facing challenges or seeking guidance in their projects. Whether you need assistance with data collection and preparation, exploratory data analysis, statistical analysis, visualization techniques, or any other aspect of your project, we are here to help. Feel free to reach out to us for personalized support and expert advice. Your success is our priority, and we look forward to collaborating with you on your data-driven endeavors.

The results highlight several socio-economic and behavioral patterns in crime distribution. For instance, the rise in cybercrime reflects broader societal digitization, while temporal trends suggest the need for enhanced night-time patrolling.

Urban crime concentration implies that resource allocation (e.g., police force deployment) should be optimized based on population density and economic indicators. Additionally, seasonal and daily time-based variations in crime rates can help law enforcement strategize more efficient shift planning and public safety campaigns.

Implications for Law Enforcement and Policy Makers

- **Predictive Policing**: The insights from this project can support the development of predictive models to proactively prevent crime in high-risk areas.
- **Public Awareness Campaigns**: Time-based crime trends suggest opportunities for targeted public awareness campaigns during peak hours and high-risk seasons.
- **Urban Planning and Social Services**: Identifying crime hotspots can inform better infrastructure planning, improved street lighting, and increased accessibility to social services in vulnerable areas.

Limitations and Future Work

- **Data Limitations**: The accuracy of the results is dependent on the completeness and quality of the crime dataset. Underreporting and data anonymization may affect outcomes.
- **Granularity**: Some datasets lacked sufficient demographic or geographic detail, limiting the depth of analysis.
- **Future Enhancements**: Incorporating machine learning for crime prediction, integrating real-time data sources, and developing interactive dashboards could greatly enhance the project's utility.

Detailed step-by-step implementation of the project

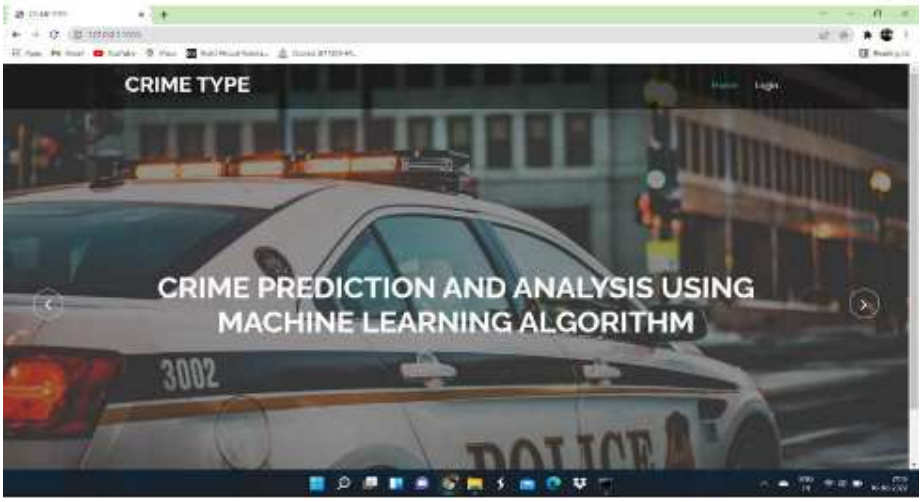

Fig 6: shows the homepage of our application of crime analysis and prediction.

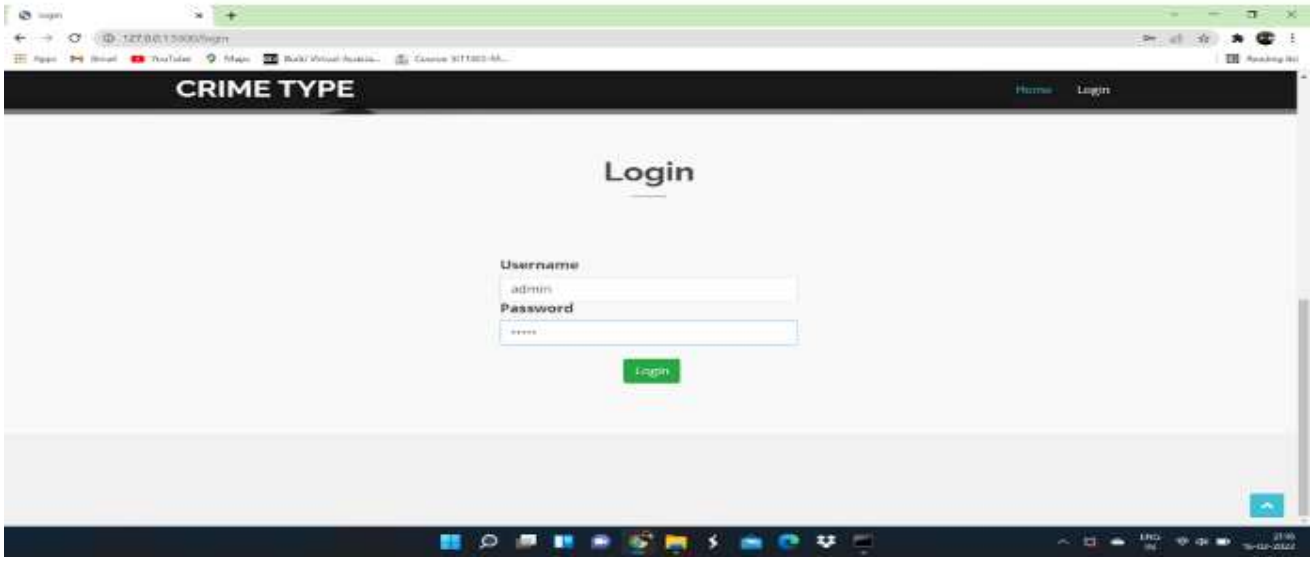

Fig 7: shows the login page for predicting the crime.

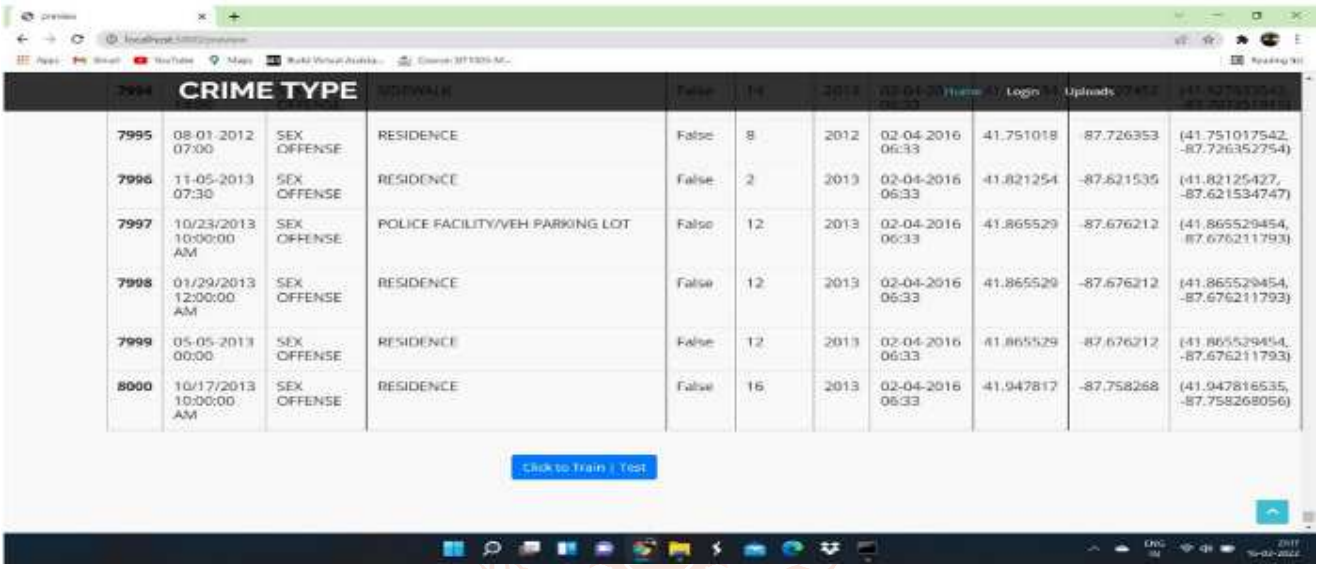

Fig 8: shows the dataset to train and test using machine learning algorithm

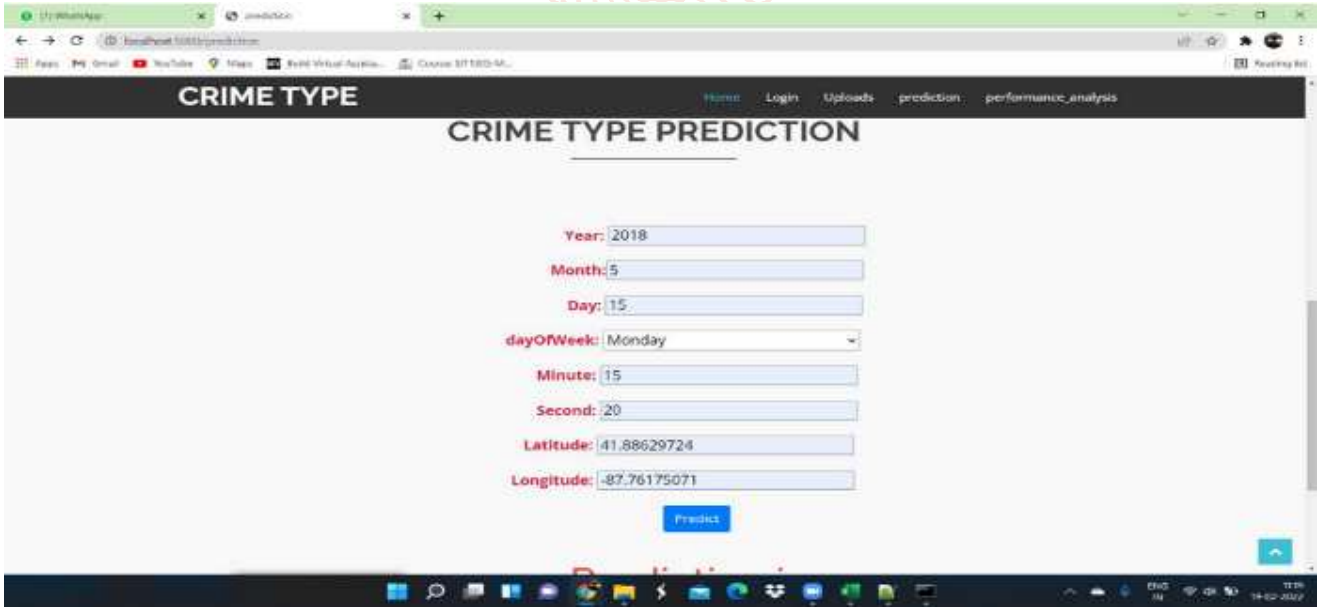

Fig 9: shows the web page filling the details about crime location, date and time to predict crime

Fig 10: shows the performance of crime occurrences with particular day of week.

VI. CONCLUSION

In this paper focused on building predictive models for crime frequencies per crime type per month. The crime rates in India are increasing day by day due to many factors such as increase in poverty, implementation, corruption, etc. The proposed model is very useful for both the investigating agencies and the police official in taking necessary steps to reduce crime. The project helps crime analysis to analysis these crime networks by means of various interactive visualization. Future enhancement of this research work on training bots to predict the crime prone areas by using machine learning techniques. Since, machine learning is similar to data mining advanced concept of machine learning can be used for better prediction. The data privacy, reliability, accuracy can be improved for enhanced prediction.

VII. REFERENCES

[1] **Treviranus, J. (2014).** "Leveraging the Web as a Platform for Economic Inclusion." Behavioral Sciences & the Law, 32(1), 39-51.

[2] **Aditya Nigam, Rhitvik Pasricha(2021)"**A Systematic Review on AI-based Proctoring Systems: Past, Present and Future".

[3] **KENRIE Hylton [a], Yair Levy [b], LaurieP. Dringus** "Utilizing webcam-based proctoring to detmisconduct in online exams"

[4] **Prakash Sinha, Dileshwari**" Remote Proctored Theory and Objective Online Examination" Int. J. Advanced Networking and Applications.

[5] H. **M. Alessio, N. J. Malay, K. Maurer, A. J. Bailer and B. Rubin,** "Examining the Effect of Proctoring on Online Test Scores," Online Learning, vol. 21, no. 1, 2017.

[6] **A. A. Alghamdi, M. A. Alanezi and F. Khan,** "Design and Implementation of a Computer Aided Intelligent Examination System," International Journal of Emerging Technologies (IJET), vol.15, no.01, pp.30-44, 2020.

[7] **G. Frankl, P. Schartner, and D. Jost,** "The 'Secure Exam Environment': E-Testing with Students' Own Devices," In IFIP World Conference on Computers in Education, pp. 179–188, 2017, https://doi.org/10.1007/978-3-319- 74310-3_20.

[8] **S. Khan and R. A. Khan**, "Online assessments: Exploring perspectives of university students," Education and Information Technologies, vol. 24, no. 1, pp. 661–677, 2019.