INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)
(Int Peer Reviewed Journal)
Vol. 05, Issue 09, September 2025, pp : 194-198
www.ijprems.com
editor@ijprems.com
e-ISSN : 2583-1062
Impact Factor : 7.001

CRIME PREDICTION AND DETECTION USING MACHINE LEARNING

Chaitra PN[1], Dr. Md. Irshad Hussain B[2]

[1]Student, Department Of MCA, UBDT College Of Engineering, Davanagere (Affiliated To VTU University Belagavi), India.

[2]Assistant Professor, Department Of MCA, UBDT College Of Engineering, Davanagere (Affiliated To VTU University Belagavi), India.

ABSTRACT

Crime prediction and analysis has proven critical in helping people stay safe and assist law enforcement agencies. The emergence of big data and machine learning means that predictive models can detect previously unnoticed trends in crime data to assist in predicting the occurrence of a crime. In this project, machine learning algorithms, namely Random Forest, Decision Tree and Support Vector machine are applied to a crime dataset in order to obtain both prediction and analysis. The system includes a web-based frontend developed on ReactJS, Vite, and Tailwind CSS and provides an opportunity to view patterns of criminal activity and analyze them in real-time. It has been experimentally tested that the accuracy of Random Forest is highest of the models tested. The system indicates that ML-enabled systems could potentially help law enforcement to make decisions and allocate resources.

1. INTRODUCTION

One of the major issues of society, defining the conditions of economic growth and population safety, is crime. Most criminal activity that is investigated in the old system much relies on manual enquiries and utilization of previous trends that in the majority of cases are not a help in providing forecasts of forthcoming activities that may occur as far as crime is concerned. The development of the Artificial Intelligence (AI) concept and the field of the Machine Learning (ML) made it possible to analyze the data on crime properly and extract the hidden associations.The goal of the present study is to develop a crime prediction and analysis system based on the concept of machine learning algorithm that will be applied to analyzing and classifying the crime patterns. An adequate project also incorporates the effective visualization and interactive analytical interface through a friendly web interface.

2. LITERATURE SURVEY

[1] The first is Crime Analytics and Predictive Policing: Basic studies in the field of criminology show that crime is not randomly distributed, but tends to cluster around certain geographic areas commonly known as hot spots. Research conducted by Sherman and coworkers (1989, 1995) confirmed that a small percentage of destinations always produce high proportion of incidents. This concept is known as predictive policing and is developed by using statistical and machine learning models on the history of crimes. Such systems are learning spatial-temporal patterns (e.g., seasonal changes or crime-specific neighborhood trends) to predict risk in the future time window. With the ability to predict when and where crimes are most probable to be committed, police agencies can effectively plan the deployment of patrol forces, engage in community outreach, and take proactive action. The expediency of the data-driven decision-making in its working definition , has been proved by the trial-and-error of the cities like Los Angeles and Chicago that in the predictive sense, when properly put into practice, will prevent more incidents than a uniform patrol. [2] Spatial Hot-Spot Detection :Crime analytics continue to rely on Geographic Information System (GIS)-based approaches. Kernel density estimation (KDE) and spatial scan statistics to locate and map high-risk areas have been extensively used. Those methods measure the level of crime activity in neighborhoods and generate a heat map that is used to enable tactical use. Clustering algorithms (K-Means and DBSCAN) and other such algorithms have been embraced because of their interpretative nature and their efficiency when it comes to computation. They are particularly applied to operational dashboards since they can summarize live concentrations of crime. Dimensionality reduction algorithms such as Principal Component Analysis (PCA) are also typically used in conjunction with clustering to remove noise, operate on multi-feature data and provide more clean visual representations of crime-prone areas. The integrated approaches not only can identify hot spots but can turn the findings into more practice-friendly information. [3] Repeat and Near-Repeat Patterns: Another empirical finding that has been well documented is repeat and near-repeat victimization. Criminology research has shown that once a place is victimized (e.g. burglary or theft) the place is at high risk of being victimized again in a short time. This effect is usually transferred to geographically close regions forming a near-repeat pattern. These features when incorporated in predictive models help a great deal in improving forecasting. Strong predictors (indicatively) are the number of crimes in the preceding few days, the spatial location (or vicinity) of the last occurrence, and the land-use environment of a neighborhood (e.g. commercial, residential). Spatio-temporal predictors are considered effective in models which predict short-term risks because they

outperform those that are based on long-term averages.[4] Crime Prediction by Machine Learning : In the recent progress, machine learning algorithms have been found to be superior to classical statistical methods in crime prediction. Random Forests and Gradient Boosting Machines (GBMs) have become popular due to their ability to process non-linear relationships, heterogeneous data, and feature spaces of high dimensions with few assumptions. Such models are highly effective at categorizing geographic units or periods based on risk levels (high, medium, low). They also inherently prioritize the importance of features, which contributes to interpretability and can help policy makers make sense of the drivers of crime risk. In addition, the exploratory analysis could also be used so as to identify hidden clusters in the unlabeled data by using non-supervised algorithms such as K-Means clustering that would give a clue of new crime patterns, or new hot spots. A combination of the supervised and unsupervised methods offers complementary features: prediction and pattern discovery.[5] Decision Support and Visualization : Finally, the impact of predictive systems is often dependent on decision support tools that communicate the results in an easy-to-understand way to the end-user. Visualization dashboards are extremely important, as they help identify hot spots and temporal trends (e.g., day of week, time of day) and also provide information on the status of cases in an intuitive way. Effective interfaces bring together maps, charts and tables to support evidence-based deployment. In practice, dashboards not only show crime risk but also recommend steps for action, such as shifting patrol footprints to emerging hot spots, timing outreach to high-risk communities, or tracking the impact of interventions. Outcomes over time to enable feedback loops in which agencies can refine models and determine prevention effectiveness In this sense, visualization is the gateway between the complex machine learning predictions and real-world decision making.

3. METHODOLOGY

The proposed system is based on a systematic approach of data collection, preprocessing, model training and evaluation:

- **Dataset:** We have a crime dataset (upload.csv) with crime records that include date, location, type of crime, victim information and suspect information.
- **Pre-processing:** Missing values were handled, categorical variables were factorised and feature scaling was applied.
- **Algorithm:** we used the Decision Tree and Random Forest algorithms and SVM algorithm to compare prediction accuracy.
- **Model Evaluation:** Accuracy, precision, recall and F1 score of the models were calculated to determine which is the best.

4. IMPLEMENTATION

The system was developed as a complete stack application with the following parts:

- **Frontend:** Created with ReactJS, Vite and Tailwind CSS for a responsive UI.
- **Backend:** Trained and deployed ML models for classification of crimes.
- **Workflow**: The user uploads datasets through the interface, the system processes the input and produces predictions and visualizations Graphs show the distribution of the crime by year, month, type and location.
- **Model Integration:** Best performing Random Forest is integrated into the application.

5. RESULT AND DISCUSSION

The system was tested for accuracy and other performance metrics:

Decision Tree: ~82% accuracy

Support Vector Machine (SVM): ~84% accuracy

Random Forest: ~88% accuracy

The Random Forest method was found to be the best among other classifiers in terms of accuracy in predicting the crime categories. Visualization modules offered further information, such as which crimes are more common by time of day, location hotspots, and by type of crime. These insights can be used by law enforcement agencies to better allocate their resources.

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)
(Int Peer Reviewed Journal)
Vol. 05, Issue 09, September 2025, pp : 194-198
www.ijprems.com
editor@ijprems.com
e-ISSN : 2583-1062
Impact Factor : 7.001

USER LOGIN

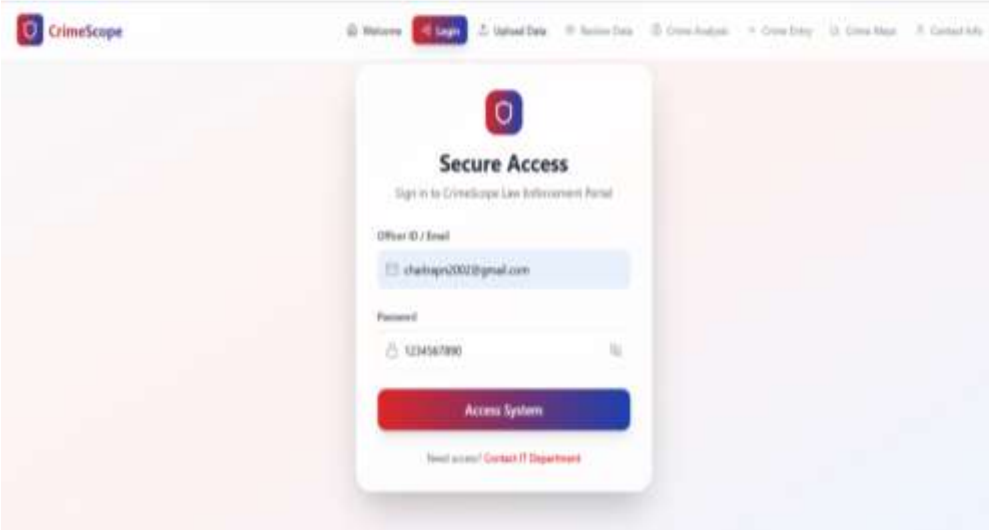

Figure 1: Login Page

UPLOADING FILE

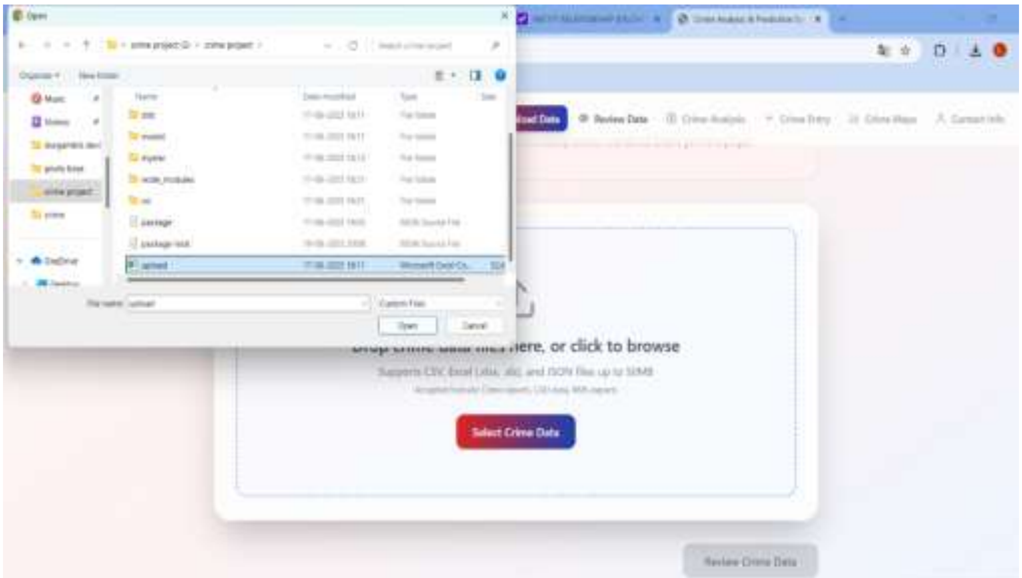

Figure 2: Upload Page

PREVIEW PAGE

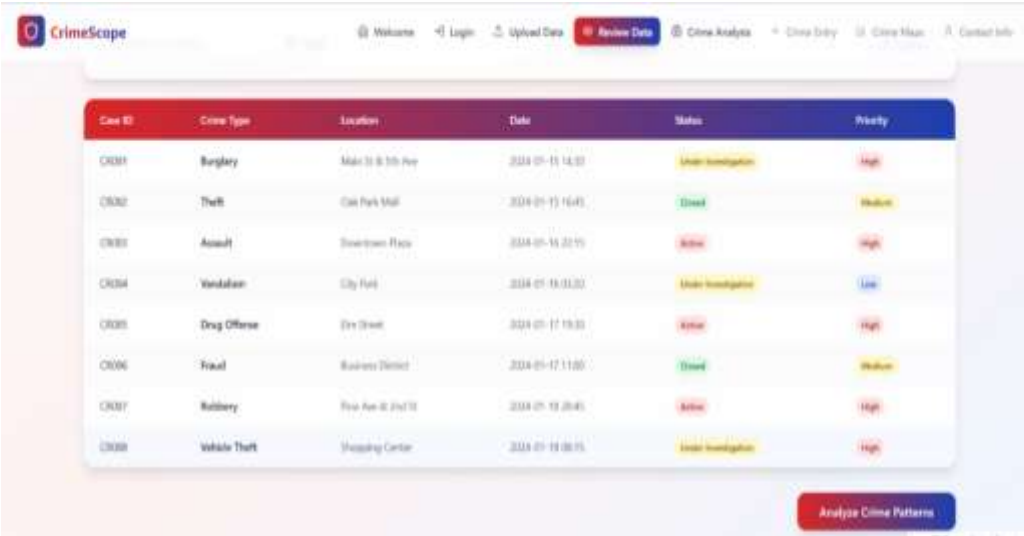

Figure 3: Preview Page

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)
(Int Peer Reviewed Journal)
Vol. 05, Issue 09, September 2025, pp : 194-198
www.ijprems.com
editor@ijprems.com
e-ISSN : 2583-1062
Impact Factor : 7.001

CRIME ENTRY

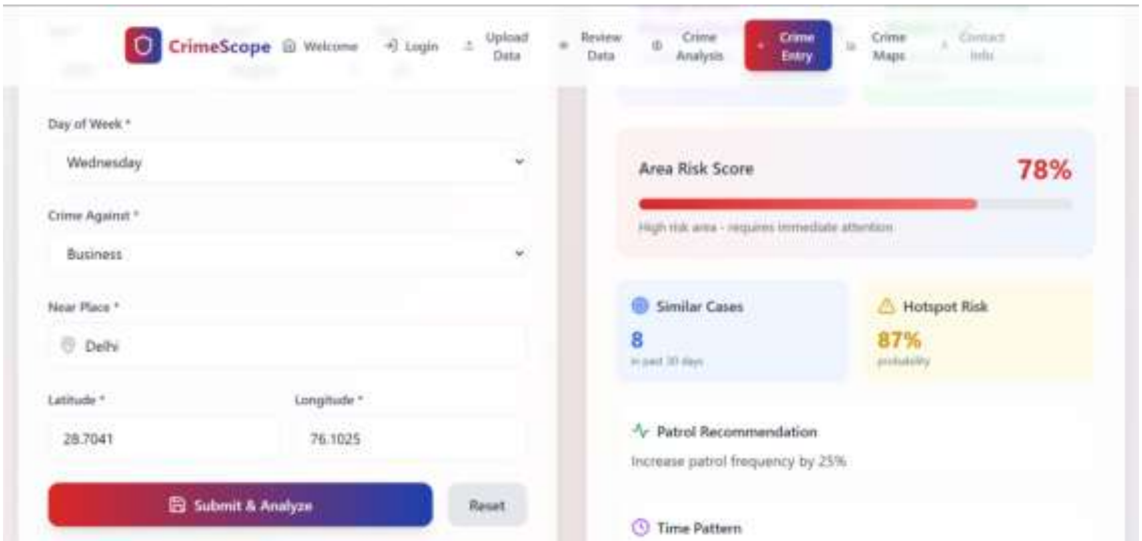

Figure 4: Prediction Page

CRIME TYPE DISTRIBUTION

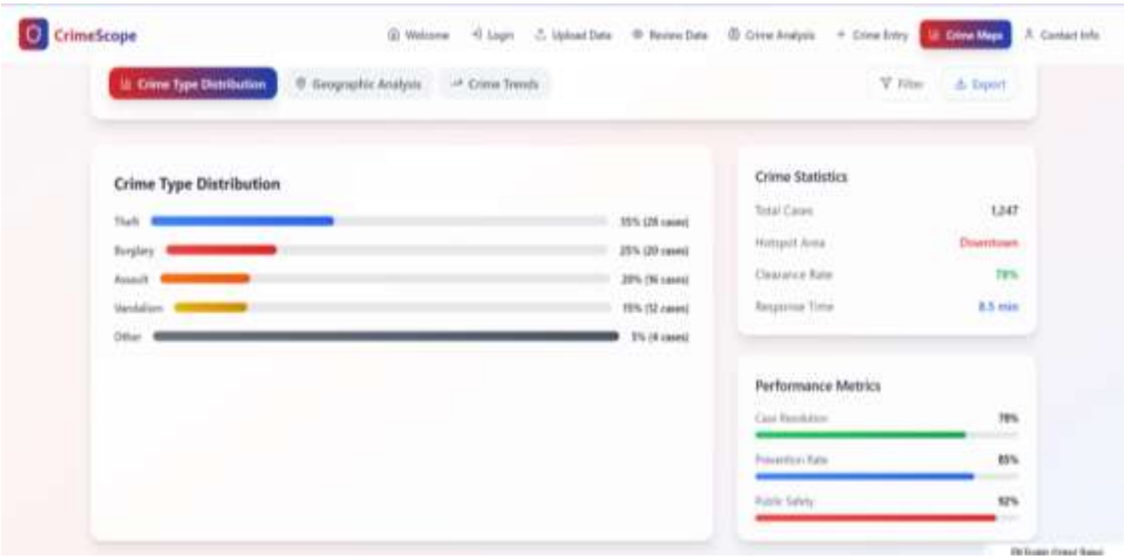

Figure 5: Crime Maps Page

CHARTS PAGE

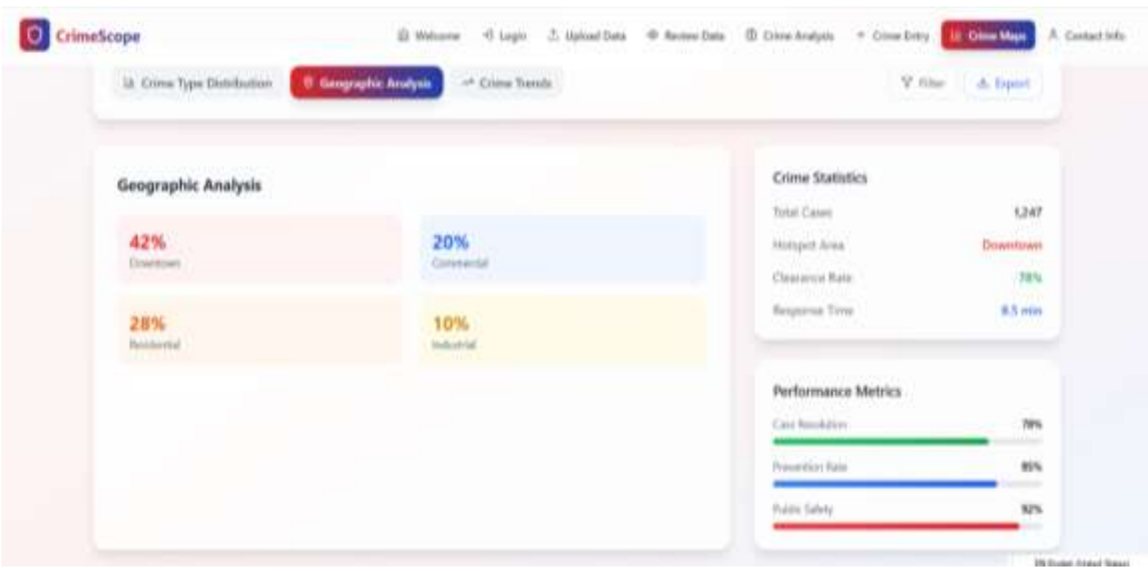

Figure 6: Charts Page

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)
(Int Peer Reviewed Journal)
Vol. 05, Issue 09, September 2025, pp : 194-198
www.ijprems.com
editor@ijprems.com
e-ISSN : 2583-1062
Impact Factor : 7.001

CONTACT PAGE

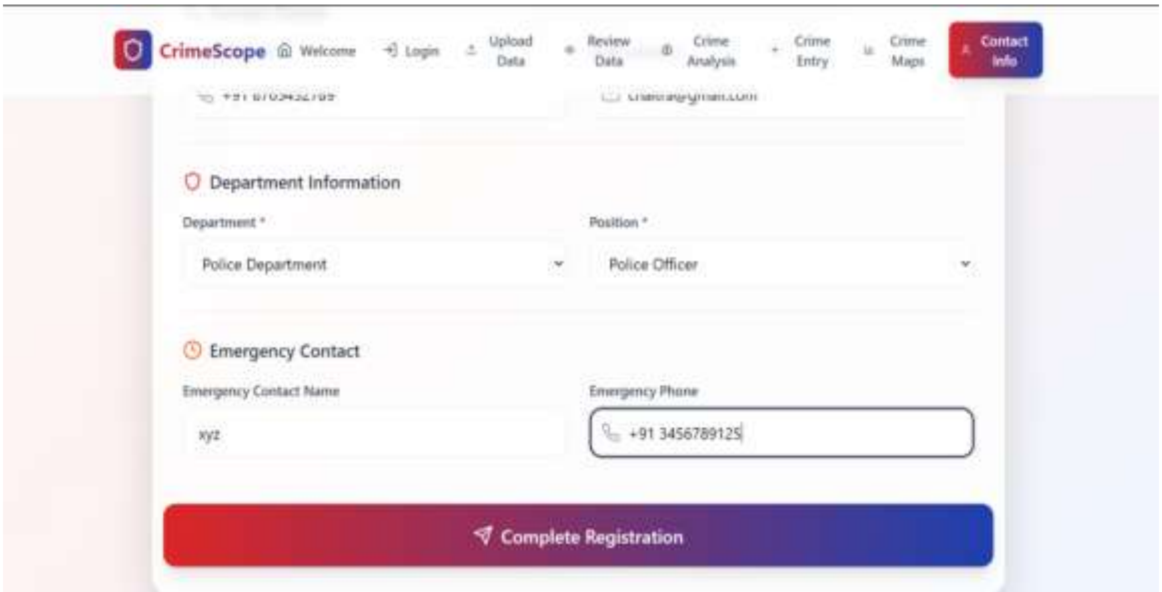

Figure 7: Contact Page

6. CONCLUSION

This project shows how machine learning can be used to predict crime and analyze crime data. The resulting system effectively combines predictive models with an interactive web-based user interface, allowing experts to both make predictions based on data and explore crime data visually. Random Forest obtained the highest accuracy and proved that it can be used for crime classification purposes.

Future work could complement this work with deep learning methods, larger and broader data sets, and deployment in real-time to prevent crimes.

7. REFERENCES

[1] T. Wang, C. Rudin, D. Wagner, and R. Sevieri. Machine learning finds patterns of crime Journal of Computational and Graphical Statistics.

[2] Pedregosa, F., et al. (2011). Scikit-learn: Scikit-learn: Learning To Code Machine Learning in Python. Machine Learning Research Journal.

[3] UCI Machine Learning Repository - Crime Data Available at: archive.ics.uci.edu

[4] Title Page of IJPREMS Paper Format Guidelines - www.ijprems.com