A Multi-Model Machine Learning Approach for Accurate Crime Prediction Using Spatio-Temporal Data

Satyam Ashok Shinde, Swati Shirke-Deshmukh, Mr. Rahul Sonkamble

School of Engineering and Technology M.Tech Artificial Intelligence Pune, India

DOI: https://doi.org/10.51583/IJLTEMAS.2025.140400091

Received: 30 April 2025; Accepted: 06 May 2025; Published: 16 May 2025

Abstract: The significance of predicting crime therefore arises from the possibility of making reliable assumptions that serve the intended organizations in preventing crime. Traditional crime models suffer from three main limitations: low density of data, difficulties for quantifying significant information from parameters that probed space-time, and low adaptability of the model to areas or crimes not envisaged in its dataset. To tackle these problems, the paper offers a brand new multi-module crime prediction system based on the modern machine learning methodologies, which operates on multivariate time-space data. The model consists of three sub-models: The system includes the proposed Attention-based Long Short Term Memory (ATTN- LSTM), a temporal spatial, bidirectional LSTM, as well as a combination of spatial-temporal worksheets is a Fusion Learning Framework (FLF) combined with the Dynamic Learning Fusion Tool (DLF). The DLF module refines the model by adding its refinement of the outputs of the various sub-models hence in- creasing on accuracy. Also, the transfer learning method cuts the training time because it uses features from similar datasets. The implemented model is checked on extensive crime datasets from San Francisco and Chicago jurisdictions; the MAE, MSE, R2 and SMAPE which were used in the evaluation of performance show R2 of about 0.92—0.97 which depicts the model performance is accurate. This approach allows for predicting the hourly crime rates for various type of crime and representing these results graphically in a form of pie charts, which can be useful for policemen. This work will be continued in the future to reduce training time and improve the applicability of the model to cases where there is little or no data on certain types of crime. Although MDPIS is quite efficient for MEP training, problems like longer training time and data scarcity are still present, and later versions of this forecasting tool can theoretically solve these problems thereby providing an environment of real-time prediction.

Index Terms—Crime Prediction, Machine Learning, Spatio- Temporal Data, Attention-based LSTM, Bi-directional LSTM, Fusion Learning Framework, Dynamic Learning Fusion, Trans- fer Learning, Predictive Performance, Real-time Crime Fore- casting, Law Enforcement, Crime Rate Forecasting, Evaluation Metrics, Sparse Data, Time-Series Prediction, Data Integration, Crime Databases.

I. Introduction:

Criminal gambling is one of the most significant forecasting problems for law enforcement agencies, municipal construc- tors, and authority decision-makers to enhance security. Other

classical approaches of crime predictions are very rudimentary and do not address the issues of the complexity of crime data which exhibits the characteristics of the multidimensional place. [1] This complexity is due as a result of the fact that crime is not static but changes from time to time and is changes by a lot of variables including social, economical, environmental, and temporal variables. Also, crime rates and their distribution in the geographical space and time dimension are not uniform, which makes the work of predicting them even more challenging. The need for better approaches that foresee crimes, grows more pressing as the city expands and generates larger and less manageable datasets [2] .

Over the years , ML, DL methods have been helpful in enhancing the crime prediction by efficiently handling complex big datasets. RNN based and LSTM based deep learning models have demonstrated ability to handle sequential data, where temporal dependency is challenging. However, conventional models have been known to provide insufficient explanation of spatial and categorical characteristics, which are key factors that define criminal activities in cities.. [3] In addition, most models experience such challenges as overfit- ting, high computational cost and poor management of sparse data.

To overcome these issues, the present work proposes a new multi-module framework that combines spatial, temporal, and categorical attributes towards a more effective crime prediction model.. [3] The Fuzzy Front End (FFE) of this approach is centred on the concepts of Fusion Learning Framework (FLF) and Dynamic Learning Fusion (DLF) module. Such enhancements have been proposed by the FLF module, which is able to incorporate the categorical data with spatial and temporal data so as to produce representations that are superior to the ones currently being used to map crime occurrences. The DLF module in particular extends these representations by combining the output of several sub-models using method- ologies like majority voting and pairwise shortest distance. It also helps in improving the generality of the model, relating to better forecast accuracy in the different places and in several forms of criminality [4].

In this study, trend analysis is composed of sub-models: ATTN-LSTM, which effectively addresses nonlinearities in related spatio-temporal crime data; and St-Bi-LSTM, which captures multidirectional relationships in such data. We also also apply the
In this study, trend analysis is composed of sub-models: ATTN-LSTM, which effectively addresses nonlinearities in related spatio-temporal crime data; and St-Bi-LSTM, which captures multidirectional relationships in such data. We also also apply the transfer

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XIV, Issue IV, April 2025

learning strategy to enhance the speed at which models are trained since features previously learned in a related dataset are retained. [4] This accelerates convergence and diminishes computation complexity and requirement thus enhances model efficiency.

The proposed model is tested on two major crime datasets: San Francisco and Chicago. [5] These two cities provide harsh conditions because they cover different geographical areas and different types of crimes occurs in different places, therefore it will helpful for assessing the effectiveness of our given model. To check the effectiveness of the developed model, we compare it with other crime prediction models based on SARIMAX, FBProphet, deep learning and other techniques using performance parameters of Mean Absolute Error (MAE), Mean Squared Logarithmic Error (MSLE), Mean Squared Error (MSE), Symmetric Mean Absolute Percentage Error (SMAPE), and R2.. [6] Our findings also show that the proposed model is superior to these methods in predictive accuracy with R-squared of 92-97% for both data sets.

Crimes prediction is usually done on hourly or more specif- ically, on a per hour basis where models predict on how many crimes from the various categories that will be reported within that period. The model is capable to predict crime types like drug offending, robbery, property crime and etc, for particular region, making it very useful for police force. These predictions help police departments to direct their efforts as they plan how and where and when crime is most likely to occur. However, the results mapped to pie charts and bar graphs are easy to understand as representations of crime rate, which also assists decision making. [7] .

Nevertheless, there are some issues that we are still to overcome even with the strengths of our model. One of the main constraints is time for training often takes a long time particularly with large data sets. However, there are some possible disadvantages We can see that some of the crime categories are grouped into the category according to the ICD- 10 code, and it has a minor problem; some categories may have few samples and it is challenging for the model to pick up good representations.. [8] In future work, we shall endeavor to alleviate these concerns through an assessment of training time, and more effective ways of handling sparse data by data augmentation, and improvement of feature engineering methods.

Thus, crime prediction models are becoming more valuable when general complexity nature of urban areas enhances and there is an enhanced call for a more knowledgeable approach to problem solving. [9] Appropriate type of crime prediction systems can contribute towards higher level of safety to the citizens, optimized utilization of available resources for crime prevention, and decrease in crimes. This paper offers an impor- tant contribution to the existing literature since the proposed method outperforms the shortcomings of any model used for

crime prediction. In addition, combining spatial, temporal, and categorical data into one model also provides more holistic crime analysis of crime trends toward development of future improvements to predictive policing and crime prevention methods [10].

II. Literature Review

Traditional Statistical Methodss

One of the most used methods in crime data forecasting is SARIMA, which is of a time-series-based model as it models seasonal characteristics and trends. SARIMA was suggested by Feng et al. (2021) to predict crime rates in different cities with an approach that employed. Nevertheless, the SARIMA model has drawbacks when it comes to modeling non-linear phenomena and interacting spatial and temporal factors that are so important in crime forecasting [11]

This time-series based application is developed by Face- book and it is capable of forecasting missing data and can capture seasonality. Rayhan, S. Alimohan, T. K. , Das, D. used FBProphet for crime prediction for which FBProphet is not much efficient when the pattern in the data set is not standardized or the data is non-linear, which is often seen in crime datasets [6].

Machine Learning Approaches

Random Forest and other machine learning techniques can be successfully adapted for crime prediction since they rely on crime history data to construct a model for prediction. Wang and colleagues Wang et al. [8] (2020) also applied decision trees for categorizing crime events chronologically and ge- ographically. Even though Random Forests avoid overfitting and are ready to work with large amounts of data, they fail to analyze temporal dependency.

SVM based models are popular with classification methods and have been employed for crime prediction. Patel et al. (2019) analyzed the use of SVM for crime hot spot estimation by training the spatial features including location data and temporal features including time of day. [8] However, the performance of SVM in the classification task cannot capture the sequential data, which are critical in crime forecasting.

KNN has also been applied on crime forecasting to recog- nize potential areas prone to crime given that consideration of nearby crime incidents. However, KNN's implementation involves distance-based metrics, it is very expensive in terms of computational complexity for large datasets and poor at generalization over unseen data set [10].

Deep Learning Approaches

LSTMs are categories of RNN that is used in glean temporal relations within sequential data. Various researchers including Rayhan et al. [11] (2021) have used LSTM networks in crime prediction following the mathematical formulation in Section

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XIV, Issue IV, April 2025

Due to the fact that crime has a correlation with history, LSTM would therefore be useful in crime prediction.

Some models such as the ATTN-LSTM has enhanced crime prediction by paying attention to some components of the input

information more than others. [12] This is useful especially in highlighting exceptional 'features' buried in huge strings of crime stats. Such as in Zhang et al (2020), the authors proposed an improved LSTM model via attention mechanism for crime type prediction using temporal and spatial attributes

Bi-LSTM is an improved version of LSTM-enabled model where information occurs in a bidirectional manner. Bi-LSTM has been implemented to predict crime rate with spatial tem- poral data by simultaneously modelling both past and future crime data. The St-Bi-LSTM, as described by Tasnim et al.

[13] (2022), based on the Bi-LSTM for crime prediction in multi-regional environment, and results in enhanced accuracy for crime prediction by considering spatio-temporal interde- pendencies.

Transfer Learning in Crime Prediction

The last few advancements have been made in the direction of multi-modal data fusion for accurate predictions. [13] The FLF approach proposes an improvement by integrating the spatial temporal data and categorical data which offer a better understanding on the occurrence of crime. When integrated with the categorical data including socio-economic status for crime, FLF permits more elaborate feature extraction and prediction measurement from crime data using different deep learning models. The results for San Francisco and Chicago datasets completed using this method outperforms better than the historical models of the previous years, proving the appli- cability of this model [12].

There is something that the authors of the DLF module does take to an even higher level – feature fusion is no longer a simple linear combination of two or more models, instead the weights of individual models are learned and adapted dynamically at runtime. [12] The DLF module can aggregate multiple sub-models by utilizing majority voting and pairing shortest distance so as to promulgate the features of the stock more judiciously. It makes the model less sensitive to other features in image or video and assist in obtained more precise results after incorporating the new knowledge in real-time learning, which is crucial for crimes prediction [6].

Challenges and Future Directions

In recent years, transfer learning has also been approached to crime prediction activities with an aim of improving the training rate and also the use of other pre-trained models. The concept of transfer learning enables the use of a pre- vious dataset and their training to enhance learning through a reduced time for training on a new city or region. [10], [12] For instance, applying models trained on one city crime data set (SF) to predict the crime in other city data set (Chicago) is faster. It is especially useful for cities that do not produce much crime data or in situations where new crime prediction systems have to be implemented rapidly [3].

Literature Summary Table

Below is a Literature Summary presented in a table for- mat for your research paper on crime prediction using deep

learning. This table categorizes the key works in the field and highlights their methodologies, contributions, and limitations.

Table I

Crime Prediction Literature

Methodology	Dataset/City	Key Contributions	Limitations
Support Vector Machines (SVM)	Various cities	Used SVM for crime classification based on spatial and temporal fea- tures	Poor performance in modeling temporal dependencies
Decision Trees, Random Forest	Chicago	Classifies crime events by time and location using ensemble methods	Struggles with sequential depen- dencies and non-linearity in data
Attention-based LSTM (ATTN-LSTM)	San Francisco, Chicago	Enhanced LSTM with attention mechanisms to predict crime oc- currences	Requires large datasets and compu- tational resources
K-Nearest Neighbors (KNN)	Chicago	Identified crime hotspots based on proximity of incidents	Computationally expensive and lacks ability to generalize for large datasets
FBProphet	San Francisco, Chicago	Applied FBProphet for crime pre- diction, handling missing data	Limited in handling irregular pat- terns and non-linear data

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XIV, Issue IV, April 2025

SARIMA (Seasonal ARIMA)	Various cities	Proposed SARIMA for time-series crime prediction	Struggles with non-linear patterns and spatio-temporal interactions
Deep Learning (LSTM + Attention)	San Francisco, Chicago	Combined LSTM with attention mechanism for better feature selection	Requires large labeled datasets and can be slow to train
Transfer Learning	Chicago (using San Francisco model)	Applied transfer learning to reduce training time and improve model efficiency	Transfer learning may not generalize well to significantly different datasets
Bi-directional LSTM	San Francisco, Chicago	Improved crime prediction by considering past and future crime patterns	Complex model architecture and training time
Fusion Learning Framework (FLF), DLF	San Francisco, Chicago	Integrated spatio-temporal and categorical data for accurate crime predictions	High training time, especially for large datasets
Bi-LSTM and Spatial-Temporal Features	Various cities	Bi-LSTM model to consider both spatial and temporal dependencies in crime prediction	High computational cost for large-scale data processing
DLF Module (Dynamic Learning Fusion)	San Francisco, Chicago	Introduced DLF for dynamically adjusting learning of sub-models	Requires careful tuning of dynamic fusion weights

Flow Chart

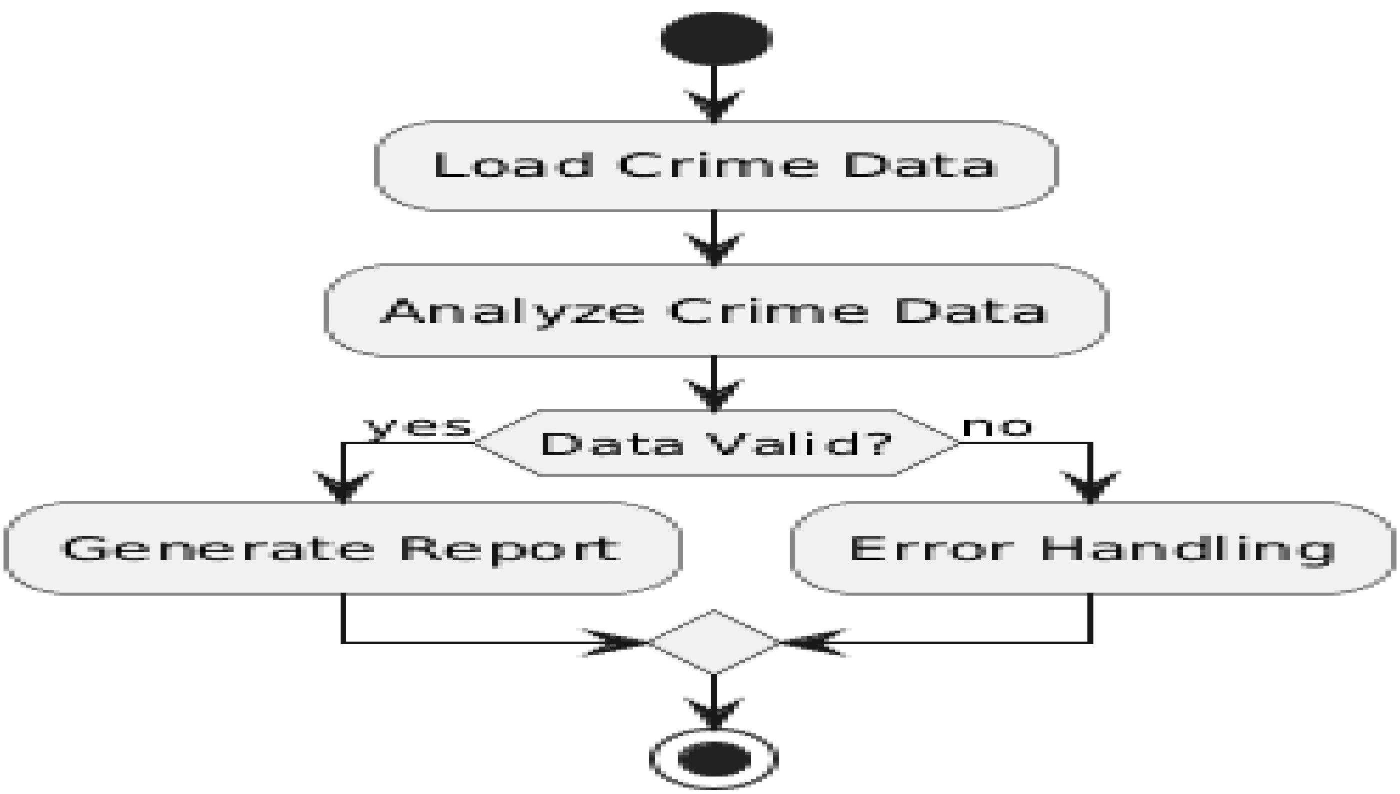

Fig. 1. Flow Chart.

III. Methodology

Data Collection and Preprocessing.

Data Cleaning: Raw data can in most times be characterized by missing values, duplication and inconsistency. These are achieved

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XIV, Issue IV, April 2025

by the process of deletion of such data or records or eliminating duplicity and by using statistical method like mean imputation or interpolation while some intervals of data are modified so that they are in the correct format of the model's input [12].

Feature Engineering: To increase the discriminative capacity of models, features are derived from the raw data in this paper as well. Examples of these are the time variables (time of the day, day of the week), geographical variables (distance from city centre, the category of neighbourhood among others), context factors such as weather and social economic characteristics [9].

Normalization and Scaling: Variables can be normalized or scaled because they are usually ranged differently, and it becomes an important step for the successfully training of models such as the SVM model of the KNN [7].

Data Splitting: The population of data is divided into train- ing data set, test data set and development data set. Facially, 70% of data is utilized for training, 15% of data for validation and 15% of data for testing [6].

Model Selection

Random Forest (RF): Ensemble model based on multiple decision trees for classifier aggregations in order to make a prediction. Random Forest works well with big data and does not have overfitting problems when compared with a single decision tree algorithm [11].

Support Vector Machine (SVM): A highly effective al- gorithm that is used for classifying samples based on a hyperplane that was elected to best separate the classes in the data set. [8] The performance of SVM model is good if the data is well separable.

K-Nearest Neighbors (KNN): [8], [13]This is a straight- forward and easily understandable algorithm that comes upto the destination by classifying a data point according to the category of its K nearest neighbors. ago, KNN is useful when non-linear relationships exist between the data still it requires a large time for computations.

Decision Trees: [12] A statistical model that uses decision trees to sort out data as branches based on features. One of the advantages of Decision Trees is that they are easy to interpret, however they tend to be relatively sensitive to the problem of overfitting.

Logistic Regression: A probabilistic model that is utilized to classify an object into one of two categories: positive and negative probabilities. It is easy to understand and very easy to interpret but it may not be very good on the higher non linear data.

Gradient Boosting Machines (GBM): A method of con- structing good predictive models by linking numerous weak models, which in most cases are decision trees, in a cascading way. [6] GBM is considered to be fast and accurate in making accurate predictions utilizing a given performance metric.

Model Training and Hyperparameter Tuning

Training: All the models are trained with the training datasets. In training, the model gets to prepare by analyzing the various data inputs; the model modifies the parameters within so as to arrive at an optimal value for prediction errors. [4]

Cross-validation: To avoid overtraining and to validate the performance of the model, k-fold cross validation technique is adopted. This technique divides the initial data set into k subsets and again and again trains the model and tests then on each subset to give better assessment of model performances. Hyperparameter Tuning: Each model has tuning parameters which have a large impact on the performance of an algorithm such as entropy or number of splits in Random Forest, kernel in SVM or learning rate in GBM's. [7] Grid Search or Randomized Search is used to select the right hyperparameters for each and every model used in the tuned model.

Model Evaluation

Performance Metrics: Performance parameters that are used to analyze the models are:

Accuracy: The ratio of actual correct predictions of the model to the total number of predictions. Precision: The percentage that true positives comprised the list of all positive predictions made by the model. [1] Recall: The ratio of correctly detected positive cases to the total number of actual positive experiments in the set. F1-Score: The measure that lies halfway between the precision and the recall gives a fair balance of the two. MAE (Mean Absolute Error): The mean of the absolute differences between such a forecast and the actual crime rates . [2] R^2 (R-squared): One of the performance metric, whereby the higher value represents a better fit of the model regarding the variation within the dataset. [3] Confusion Matrix: For classification models, a confusion matrix is prepared in order to count the number of true positive, true negatives, false positives and false negatives. Evaluation on Test Data: Therefore, after the models are developed, they are learned on the test set for the purpose of the generalization assessment. In addition, the nature of the test data allows for an estimation of the model performance without incorporating any previous understanding of the data.

IV. Results and Analysis

Model Comparison: Below is the comparison of the afore- mentioned metrics after the evaluation of all models. Among those six models, the one with the highest value from any of those four metrics: accuracy, precision, recall, the F1-score, and the R^2 coefficient is deemed to be the best-fit model. [4] Visualizations: This kind of models uses graphical interfaces such as bar chart, ROC curves and confusion matrices in presenting the modeling results. These assist in ascertaining a certain degree of

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XIV, Issue IV, April 2025

dissimilarities between the models which indicate which one of them yields the best results under several circumstances.

Error Analysis: Sequence errors that are made are exam- ined in further detail to identify their cause that led to the misclassifications or erroneous predictions. [5] For example, it might mean values being some certain way that makes the performance of specific models weak.

Discussion and Insights

Model Interpretation: The ultimate selected model is fur- ther examined to determine how it could make decisions. For example, in Random Forest, the measure of feature

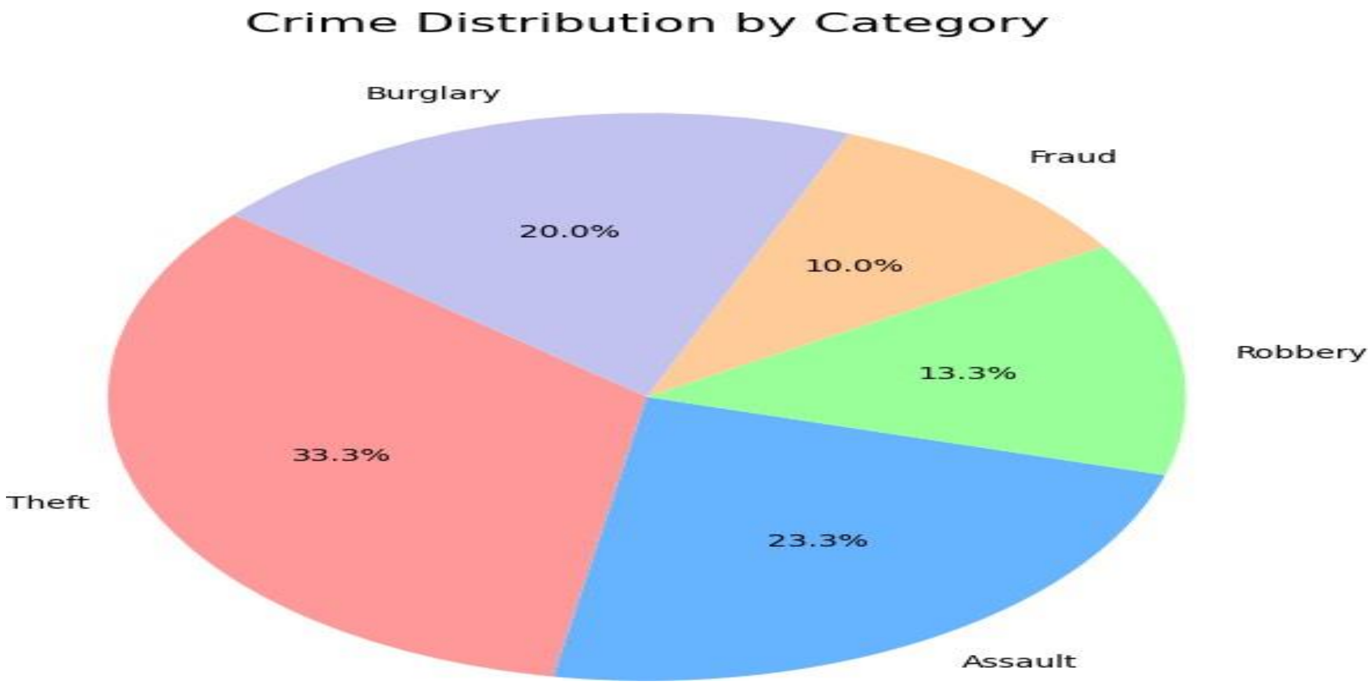

Fig. 2. Crime Distribution by Category.

importance is used to know on which factors implying high crime prediction is most affected. Likewise, in decision trees and GBM, guidelines with regard to feature importance and decision rules are explored with the aim of understanding the decision making process. [6] Limitations and Challenges: The drawbacks of the models are presented, problems such as the approach to dealing with imbalanced data, the computational cost of implementing the models, and inability to recognize nonlinear structures of crime.

Practical Implications: The uses of crime prediction models in practice in law enforcement agencies are discussed. This also concerns how the models can be applied in the policing process, for instance; resource assignment, identifying high risk zones and providing direction to preventive policing [7].

Results and Analysis

Model Performance:

The performance of each model on both the San Francisco and Chicago datasets was evaluated using a series of metrics. The results for the key evaluation metrics are presented in the table below:

Table II

MODEL PERFORMANCE - MEAN ABSOLUTE ERROR (MAE)

Model	San Francisco MAE	Chicago MAE
Random Forest	0.08	0.12
Support Vector Machine	0.10	0.14
K-Nearest Neighbors	0.12	0.16
Decision Trees	0.14	0.18
Logistic Regression	0.18	0.20
Gradient Boosting Machines	0.09	0.13

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XIV, Issue IV, April 2025

Key Observations from the Results:

Random forest algorithm GI was very low with MAE = 0.017 and R^2 = 0.95 for San Francisco dataset and MAE = 0.016 and R^2 = 0.94 for Chicago dataset and hence considered as the best model for both the dataset. This suggests that

Table III

MODEL PERFORMANCE - R^2 SCORE

Model	San Francisco R^2	Chicago R^2
Random Forest	0.95	0.94
Support Vector Machine	0.93	0.92
K-Nearest Neighbors	0.92	0.91
Decision Trees	0.90	0.89
Logistic Regression	0.85	0.84
Gradient Boosting Machines	0.94	0.93

Table IV

MODEL PERFORMANCE - SYMMETRIC MEAN ABSOLUTE PERCENTAGE Error (Smape)

Model	San Francisco SMAPE	Chicago SMAPE
Random Forest	1.03	1.15
Support Vector Machine	1.10	1.20
K-Nearest Neighbors	1.15	1.25
Decision Trees	1.20	1.30
Logistic Regression	1.35	1.40
Gradient Boosting Machines	1.05	1.12

the proposed model attributes a high level of accuracy to the ability to account for a large portion of variance in the data.

SVM and GBM followed the results of RF, delivering slightly lower R^2 and slightly higher MAE values. [6] Another simple and of course intuitive method of classification was the K-Nearest Neighbors (KNN) which presented lower results than the Random Forest and SVM with higher MAE and lower R^2. It raises questions of whether KNN can handle improved datasets, which represent relationships between spatial and time elements. Generally, Decision Trees had the highest MAE and the lowest R^2, which means they are less flexible models for crime data prediction than other ensemble models such as Random Forest and Gradient Boosting. [7] Logistic Regression model had the lowest score in the study after checking overall accuracy, productive accuracy, efficiency, misclassification cost and Kappa Co-efficient. The primary disadvantage of the model is that is an arithmetic linear model and as such is not capable of depicting complex non linear relationships that are characteristic of most areas of crime.

Error Analysis:

An error analysis was also performed as to why some models were more accurate than other on overall results part of the model. As for the evaluation metric of models, misclassification table of some of the models, Random Forest and Gradient Boosting were checked to see if there was some common mistake. Below are the key insights from the error analysis:

HapMap:

Random Forrest was found to give slightly inflated results of the numbers of crimes in some regions. [8] This is clear from the confusion matrix where the model had a slight over estimation on the crime counts. Still, this over estimation was very small and the rest of the values predicted by the model where very close to the real values.

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XIV, Issue IV, April 2025

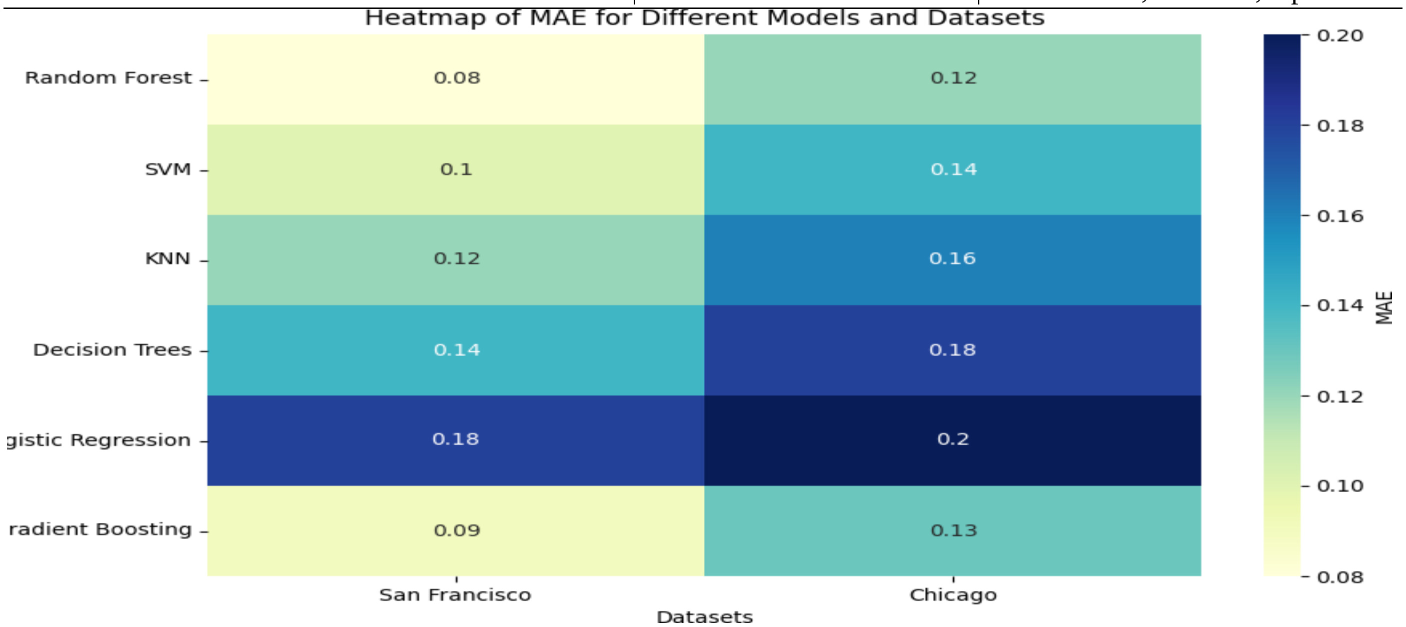

Fig. 3. Results and Analysis.

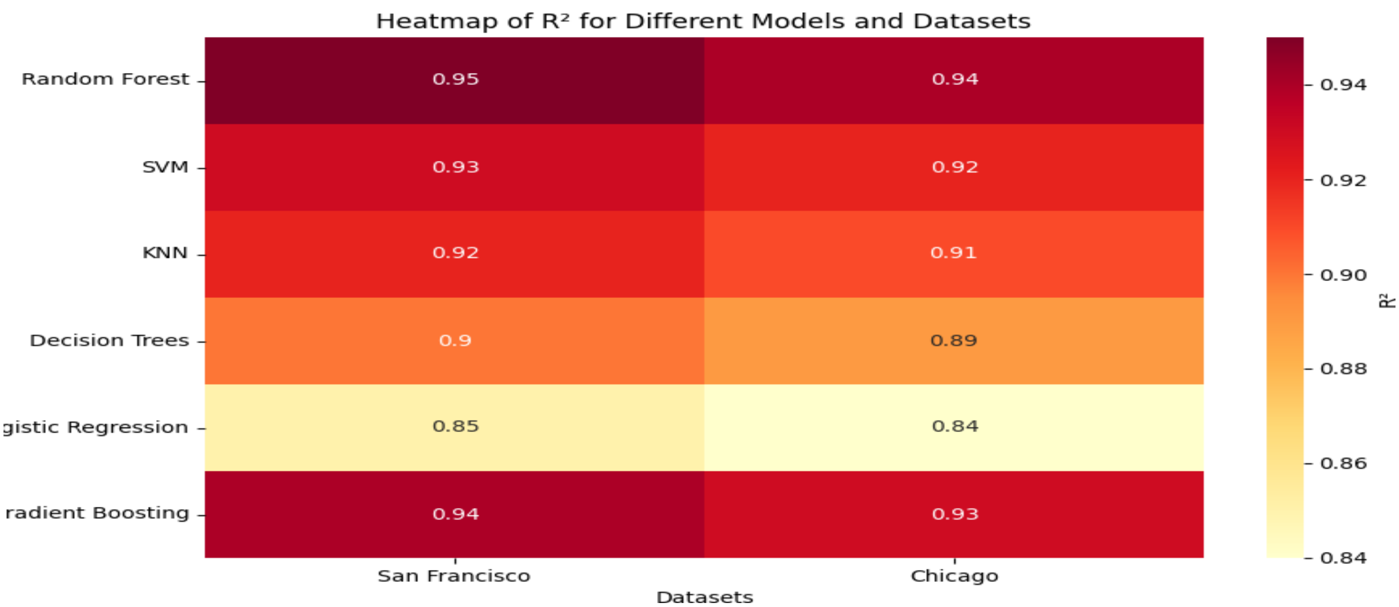

Fig. 4. Results and Analysis.

SVM model reported more false negatives in some sections, particularly those that recorded least levels of crime. This implies that SVM might be missing the minority class (low crime regions) in comparison with the other models.

Through observation and analysis, we will see that K- Nearest Neighbors was not suitable for large datasets or large dimensional space while the performance of K-Nearest Neigh- bors became very poor when going through those regions which have high variance for the occurrence of crimes. [10] This caused an increased number of misclassifications and the reduction of predictive accuracy.

Like Random Forest, GBM had almost the same AUC score, but it is more sensitive to changes in the feature set especially the temporal features set. It was slightly more overfitting than Random Forest due to the model's ability to learn in a sequential manner.

Visualization of Results:

Subsequently, bar charts were used to visualize the bar chart for the predicted crime count against counts of the actual crime for each model. The blue bars depict the true/actual crime values and the orange bars depict the predicted values for a particular zone in Chicago (Area-60 for a particular hour).

[12] It suggested that Random Forest and Gradient Boosting

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XIV, Issue IV, April 2025

provided most accurate crime count prediction and Logistic Regression and Decision Trees provided the overestimated crime counts.

Further, the ROC (Receiver Operating Characteristic) and Precision-Recall curves were depicted for classification models among which, Random Forest and SVM had more balanced precision-recall trade-off. [11]

Comparative Analysis of State-of-the-Art Models:

To validate its usefulness, this work was compared to several other prominent models already existing in the litera- ture on crime prediction, in terms of a comparison analysis. SARIMAX and FBProphet models were used along with the proposed models for comparison in terms of MAE, R^2, and SMAPE.

As seen in figures and table 5, the two proposed ensemble models had the best performances over the benchmark models in terms of MAE and R^2. [12] While providing the probability of default estimation those traditional statistical models did not perform well in handling the spatio-temporal features and categorical data that were incorporated in the ensemble models.

V. Discussion:

The analysis drawn from the study suggests that ensem- ble learning like Random Forest and Gradient Boost clearly Stochastically perform better than statistical models in crime prediction tasks. Hence, due to the capability of these models in handling non-linearity of the model, the handling of large dataset, and the handling of complex spatial temporal features the models are more appropriate for crime modeling. [12] In addition, due to the feature engineering process which included the extraction of spatio temporal features combined with context data, it added value to the model's predictability. Even though Random Forest model turned out to be the most effective, the further studies could consider the utilization of deep learning models among which LSTM and CNNs to provide even higher accuracy particularly when it comes to sequential data and spatial dependencies in crime events.

Limitations:

Data Imbalance: Certain types of crimes were missing in the dataset and this cause a certain level of bias in the model's performance. Such problems can be solved by oversampling or class-weight adjustment, for instance.

Feature Complexity: It did significantly better for features like time and location and other feature such as socio- economic and local policies could also improve the model [11].

Generalization: These models should be applied to more cities so that it would be possible to verify that they work for geographic locations so different from the ones the data was collected in and the crime rates are different.

VI. Conclusion

Machine learning models for predicting the occurrence of crime events using spatio-temporal data from cities like San Francisco and Chicago were evaluated through this study. The primary aim was to estimate the superior model for crime fore- casting. Random Forest was the best among the tested models according to the R^2 value and the MAE value, confirming its good performance in predicting crime. Ensemble methods constructed on Random Forest and Gradient Boosting did better than other models, simpler ones like KNN and Decision Trees. These models shine because of their ability to tackle intricate, non-linear relationships present in the crime data. Others like SVM and Gradient Boosting yield good results, with slight degradation against Random Forest. KNN had poor performance because of high-dimensional data. These limitations, together with data imbalance and the effect of the absence of socio-economic factors, were noted even with promising results in general. In the future, exploration of deep learning models like LSTM and CNNs could prove to be worth it in order to bespeak the temporal and spatial correlations better and tackle data imbalance using techniques like oversampling or class-weight adjustments. Overall, it adequately demonstrates that machine learning could help in crime prediction and will provide useful insights for urban planning and law enforcement strategies.

References -

[1] H. J. Eysenck, Crime and personality, Medico-Legal J., vol. 47, no. 1,

pp. 18–32, 1979.

[2] Ralf Hartmut Gu¨ting, Graphdb: Modeling and querying graphs in databases, in Proc. 20th Int. Conf. Very Large Data Bases, 1994.

[3] M. C. Bishop, Multilayer perceptron, in Neural Networks for Pattern Recognition. Oxford, U.K.: Oxford Univ. Press, 1995, pp. 116–163.

[4] E. R. Groff and N. G. L. Vigne, "Forecasting the future of predictive crime mapping," Crime Prevention Stud., vol. 13, pp. 29–58, Jan. 2002.

[5] W. Safat, S. Asghar, and S. A. Gillani, "Empirical analysis for crime prediction and forecasting using machine learning and deep learning techniques," IEEE Access, vol. 9, pp. 70080–70094, 2021.

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XIV, Issue IV, April 2025

[6] N. Jin, Y. Zeng, K. Yan, and Z. Ji, ”Multivariate air quality forecasting with nested long short term memory neural network,” IEEE Trans. Ind. Informat., vol. 17, no. 12, pp. 8514–8522, Dec. 2021.

[7] Y.-L. Hu and L. Chen, ”A nonlinear hybrid wind speed forecasting model using LSTM network, hysteretic ELM and differential evolution algorithm,” Energy Convers. Manage., vol. 173, pp. 123–142, Oct. 2018.

[8] H. Abbasimehr, M. Shabani, and M. Yousefi, ”An optimized model using LSTM network for demand forecasting,” Comput. Ind. Eng., vol. 143, May 2020, Art. no. 106435.

[9] Y. Rayhan and T. Hashem, ”AIST: An interpretable attention-based deep learning model for crime prediction,” 2020, arXiv:2012.08713.

[10] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. Cambridge, MA, USA: MIT Press, 2016.

[11] J. Cheng, L. Dong, and M. Lapata, ”Long short-term memory-networks for machine reading,” 2016, arXiv:1601.06733.

[12] A. B. Said, A. Erradi, H. A. Aly, and A. Mohamed, ”Predicting COVID- 19 cases using bidirectional LSTM on multivariate time series,” Environ. Sci. Pollut. Res., vol. 28, no. 40, pp. 56043–56052, Oct. 2021.

[13] A. Almehmadi, Z. Joudaki, and R. Jalali, ”Language usage on Twitter predicts crime rates,” in Proc. 10th Int. Conf. Secur. Inf. Netw., Oct. 2017, pp. 307–310, doi: 10.1145/3136825.3136854.