Journal of **N**onlinear **A**nalysis and **O**ptimization
Vol. 15, Issue. 1 : 2024
ISSN : **1906-9685**

PREDICTING CRIME TYPES AND OCCURRENCES: A MACHINE LEARNING APPROACH

#1 B. TRIVENI, #2 T.SOWJANYA, #3 Y.AJAY, #4 M. NAGA LAKSHMI, #5 K.GOPI KRISHNA
#1Assistant professor in Department of IT, DVR & Dr.HS MIC College of Technology,Kanchikacherla
#2#3#4#5 B.Tech with Specialization of Information Technology , DVR & Dr.HS MIC College of Technology,Kanchikacherla-521180

Abstract_ A crime is just an act of transgression. That is a transgression. It is illegal to do so. For the police force, identifying and analysing hidden crime is an extremely challenging undertaking. In addition, a wealth of information regarding the crime is accessible. Therefore, a few approaches ought to aid in the research. Thus, the approach ought to aid in the investigation of the crime. The prediction and analysis of crimes can be more effectively aided by the machine learning technique. Regression methods are provided by the machine learning methodology. Crime has emerged as a clear means of causing problems for individuals and society in the modern era. An imbalance in a nation's population is caused by rising crime rates.
Understanding crime trends is essential for analysing and anticipating this kind of criminal activity. Using crime data from the open-source Kaggle platform, this study enforces one such analysis of crime patterns, which is then used to forecast the majority of crimes that have occurred recently. Estimating the type of crime that contributes most, coupled with the time period and place where it has occurred, is the main goal of this study. This work implies the use of some machine learning techniques, like XGBoost, to categorise among different criminal patterns, and the accuracy attained was rather high in comparison to previously proposed work.

1.INTRODUCTION

Wrongdoing expectation is a urgent area of study that joins the force of information science and AI to figure crimes and upgrade proactive policing. In the present dynamic and interconnected world, the capacity to expect and forestall criminal episodes is more basic than any other time in recent memory. Wrongdoing expectation frameworks influence authentic wrongdoing information, high level calculations, and prescient demonstrating to recognize examples, patterns, and possible areas of interest, enabling policing, city organizers, and policymakers to assign assets decisively and moderate lawbreaker gambles.

The ascent of AI procedures, like characterization calculations and group techniques, has changed the field of wrongdoing expectation. These advances empower the investigation of tremendous datasets, crossing different wrongdoing classifications and geological areas, to separate significant experiences and make exact expectations. By saddling the prescient force of calculations, wrongdoing forecast frameworks endeavor to add to the making of more secure networks and cultivate a proactive way to deal with policing.

The difficulties of urbanization, populace development, and advancing criminal approaches highlight the significance of creating complex wrongdoing expectation models. These models not just guide in that frame of mind of wrongdoing designs yet in addition work with a more profound comprehension of the financial, ecological, and worldly variables that add to crimes. The use of AI in wrongdoing forecast lines up with the more extensive objective of utilizing innovation to make more astute, more secure urban areas and improve the general prosperity of networks.

In this unique situation, the undertaking "Wrongdoing Expectation Utilizing AI" attempts to add to the progression of wrongdoing forecast approaches. Created utilizing Python and utilizing progressed calculations, for example, the Choice Tree Classifier and Packing Classifier, this venture centers around anticipating and grouping different wrongdoing classifications in Portland, Oregon, USA, over a range of quite a while. By giving precise expectations and important experiences, this task intends to be a significant instrument for policing and policymakers, encouraging a more secure and safer metropolitan climate.

The ensuing segments of this venture dig into the strategy, results, and ramifications of the wrongdoing expectation framework, offering an exhaustive investigation of its capacities and commitments to the continuous talk on prescient policing

2.LITERATURE SURVEY

1) Crime forecasting: A machine learning and computer vision approach to crime prediction and prevention

AUTHORS: N. Shah, N. Bhagat, and M. Shah

A crime is a deliberate act that can cause physical or psychological harm, as well as property damage or loss, and can lead to punishment by a state or other authority according to the severity of the crime. The number and forms of criminal activities are increasing at an alarming rate, forcing agencies to develop efficient methods to take preventive measures. In the current scenario of rapidly increasing crime, traditional crime-solving techniques are unable to deliver results, being slow paced and less efficient. Thus, if we can come up with ways to predict crime, in detail, before it occurs, or come up with a “machine” that can assist police officers, it would lift the burden of police and help in preventing crimes. To achieve this, we suggest including machine learning (ML) and computer vision algorithms and techniques. In this paper, we describe the results of certain cases where such approaches were used, and which motivated us to pursue further research in this field. The main reason for the change in crime detection and prevention lies in the before and after statistical observations of the authorities using such techniques. The sole purpose of this study is to determine how a combination of ML and computer vision can be used by law agencies or authorities to detect, prevent, and solve crimes at a much more accurate and faster rate. In summary, ML and computer vision techniques can bring about an evolution in law agencies.

2) Crime prediction model using deep neural networks

AUTHORS: S. A. Chun, V. A. Paturu, S. Yuan, R. Pathak, V. Atluri, and N. R. Adam

This project investigates the feasibility of using machine learning techniques, specifically neural networks, to make prediction on criminal behavior based on the history of the arrest bookings. The experiment will have to handle imbalanced data frequencies. To combat the challenge, data augmentation and weighted loss function is being developed to extract information from the minority classes. For this project, we have focused on how neural networks can be advantageous in classification of crime prediction. The specific kind of neural network that has been used in the project is a deep fully connected neural network. Fully connected neural networks are suitable for problems where domain knowledge is limited and many to many relations between features are important. As this report shows, machine learning techniques could definitely be of use for classification of criminal behavior, and we recommend exploring the discussed data augmentation and modeling methods more thoroughly to improve on the results and find new patterns.

3) An empirical analysis of machine learning algorithms for crime prediction using stacked generalization: An ensemble approach

AUTHORS: S. S. Kshatri, D. Singh, B. Narain, S. Bhatia, M. T. Quasim, and G. R. Sinha

Ensemble learning method is a collaborative decision-making mechanism that implements to aggregate the predictions of learned classifiers in order to produce new instances. Early analysis has shown that the ensemble classifiers are more reliable than any single part classifier, both empirically and logically. While several ensemble methods are presented, it is still not an easy task to find an appropriate configuration for a particular dataset. Several prediction-based theories have been proposed to handle machine learning crime prediction problem in India. It becomes a challenging problem to identify the dynamic nature of crimes. Crime prediction is an attempt to reduce crime rate and deter criminal activities. This work proposes an efficient authentic method called assemble-stacking based crime prediction method (SBCPM) based on SVM algorithms for identifying the appropriate predictions of crime by implementing learning-based methods, using MATLAB. The SVM algorithm is applied to achieve domain-specific configurations compared with another machine learning model J48, SMO Naïve byes bagging and, the Random Forest. The result implies that a model of a performer does not generally work well. In certain cases, the ensemble model outperforms the others with the highest coefficient of correlation, which has the lowest average and absolute errors. The proposed method achieved 99.5% classification accuracy on the testing data. The model is found to produce more predictive effect than the previous researches taken as baselines, focusing solely on crime dataset based on violence. The results also proved that any empirical data on crime, is compatible with criminological theories.

The proposed approach also found to be useful for predicting possible crime predictions. And suggest that the prediction accuracy of the stacking ensemble model is higher than that of the individual classifier.

3.PROPOSED SYSTEM

❖ The proposed system for crime prediction builds upon the foundation of the existing system, leveraging Python as the primary programming language. The key components of the proposed system include the utilization of the XG Boost,Decision Tree Classifier and the Bagging Classifier to enhance the accuracy of crime predictions in Portland, Oregon, USA, spanning the years 2015 to 2023.

❖ The proposed system retains the Decision Tree Classifier as a fundamental component due to its interpretability and ability to capture complex relationships within the data. Additionally, the Bagging Classifier is introduced to improve robustness and mitigate overfitting, as it aggregates multiple decision trees.

❖ The proposed system continues to utilize a dataset encompassing 505,063 data points, focusing on crime incidents in Portland, Oregon. The dataset spans the years 2015 to 2023, providing a comprehensive temporal perspective to capture evolving crime patterns in the region.

❖ The system addresses the classification of 20 distinct crime categories, including 'Larceny Offenses,' 'Motor Vehicle Theft,' 'Assault Offenses,' 'Drug/Narcotic Offenses,' 'Weapon Law Violations,' 'Vandalism,' 'Burglary,' 'Fraud Offenses,' 'Robbery,' 'Counterfeiting/Forgery,' 'Arson,' 'Prostitution Offenses,' 'Stolen Property Offenses,' 'Animal Cruelty Offenses,' 'Homicide Offenses,' 'Embezzlement,' 'Pornography/Obscene Material,' 'Extortion/Blackmail,' 'Bribery,' and 'Gambling Offenses.' These categories enable a detailed and nuanced understanding of crime dynamics.

The proposed system incorporates a set of features such as address, case number, crime against category, neighborhood, occur date, occur time, offense category, offense type, open data latitude/longitude, open data X/Y, and offense count. These features continue to serve as crucial input variables, providing valuable information for training and evaluating the models.

❖ Rigorous model evaluation metrics are employed to assess the performance of the XG Boost Decision Tree Classifier and Bagging Classifier. The proposed system includes mechanisms for hyperparameter tuning, ensuring optimal configurations for both models and addressing potential overfitting concerns.

❖ The proposed system emphasizes adaptability to dynamic changes in crime patterns and scalability for diverse deployment scenarios. Continuous monitoring and updates to the models allow them to evolve with emerging trends, while efficient implementation techniques ensure scalability in resource-constrained environments.

❖ Consideration is given to the development of user-friendly interfaces or dashboards for stakeholders, facilitating easy interpretation of model outputs. Visualization tools may be integrated to present insights and predictions in an accessible manner for law enforcement agencies, city planners, and policymakers.

❖ In conclusion, the proposed system aims to enhance the existing crime prediction framework by refining algorithm selection, optimizing accuracy, and addressing potential limitations. The XG Boost, Decision Tree Classifier and Bagging Classifier, combined with a comprehensive dataset and feature set, contribute to the development of an advanced system for crime prediction in Portland, Oregon.

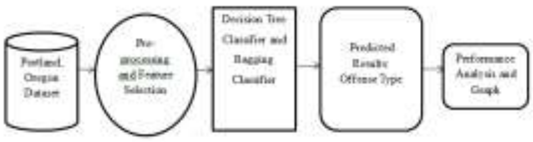


Fig 1:Proposed System Architecture

3.1 IMPLEMENTATION

Data Collection:

In the first module of Crime Prediction Using Machine Learning, we developed the system to get the input dataset. Data collection process is the first real step towards the real development of a machine learning model, collecting data. This is a critical step that will cascade in how good the model will be, the more and better data that we get; the better our model will perform. There are several techniques to collect the data, like web scraping, manual interventions. Our dataset is placed in the project and it's located in the model folder. The dataset is referred from the popular standard dataset repository kaggle where all the researchers refer it. The dataset consists of numerical datas. The following is the URL for the dataset referred from kaggle.

URL: https://www.kaggle.com/datasets/jayaprakashpondy/portland-oregon-crime-dataset

Dataset:

The dataset consists of 505063 individual data. There are 15 columns in the dataset, which are described below.

Address: Address of reported incident at the 100 block level (e.g.: 1111 SW 2nd Ave would be 1100 Block SW 2nd Ave).

Case Number: The case year and number for the reported incident (YY-######).

Crime Against: Crime against category (Person, Property, or Society).

Neighborhood: Neighborhood where incident occurred. If the neighborhood name is missing, the incident occurred outside of the boundaries of the Portland neighborhoods or at a location that could not be assigned to a specific address in the system. (e.g., Portland, near Washington Park, on the streetcar, etc.).

Occur Date: Date the incident occurred. The exact occur date is sometimes unknown. In most situations, the first possible date the crime could have occurred is used as the occur date. (For example, victims return home from a week-long vacation to find their home burglarized. The burglary could have occurred at any point during the week. The first date of their vacation would be listed as the occur date.)

Occur Time: Time the incident occurred. The exact occur time is sometimes unknown. In most situations, the first possible time the crime could have occurred is used as the occur time. The time is reported in the 24-hour clock format, with the first two digits representing hour (ranges from 00 to 23) and the second two digits representing minutes (ranges from 00 to 59).

Offense Category: Category of offense (for example, Assault Offenses).

Offense Type: Type of offense (for example, Aggravated Assault).

Open Data Lat/Lon: Generalized Latitude / Longitude of the reported incident. For offenses that occurred at a specific address, the point is mapped to the block's midpoint. Offenses that occurred at an intersection is mapped to the intersection centroid.

Open Data X/Y: Generalized XY point of the reported incident. For offenses that occurred at a specific address, the point is mapped to the block's midpoint. Offenses that occurred at an intersection is mapped to the intersection centroid. To protect the identity of victims and other privacy concerns, the points of certain case types are not released. XY points use the Oregon State Plane North (3601), NAD83 HARN, US International Feet coordinate system.

Offense Count: Number of offenses per incident. Offenses (i.e. this field) are summed for counting purposes.

Data Preparation:

Wrangle data and prepare it for training. Clean that which may require it (remove duplicates, correct errors, deal with missing values, normalization, data type conversions, etc.). Randomize data, which erases the effects of the particular order in which we collected and/or otherwise prepared our data. Visualize data to help detect relevant relationships between variables or class imbalances (bias alert!), or perform other exploratory analysis. Split into training and evaluation sets

Analyze and Prediction:

In the actual dataset, we chose only 8 features:

Crime Against: Crime against category (Person, Property, or Society).

Neighborhood: Neighborhood where incident occurred. If the neighborhood name is missing, the incident occurred outside of the boundaries of the Portland neighborhoods or at a location that could not be assigned to a specific address in the system. (e.g., Portland, near Washington Park, on the streetcar, etc.).

Occur Date: Date the incident occurred. The exact occur date is sometimes unknown. In most situations, the first possible date the crime could have occurred is used as the occur date. (For example, victims return home from a week-long vacation to find their home burglarized. The burglary could have occurred at any point during the week. The first date of their vacation would be listed as the occur date.)

Open Data Lat/Lon: Generalized Latitude / Longitude of the reported incident. For offenses that occurred at a specific address, the point is mapped to the block's midpoint. Offenses that occurred at an intersection is mapped to the intersection centroid.

Offense Category: Category of offense (for example, Assault Offenses).

Accuracy on test set:

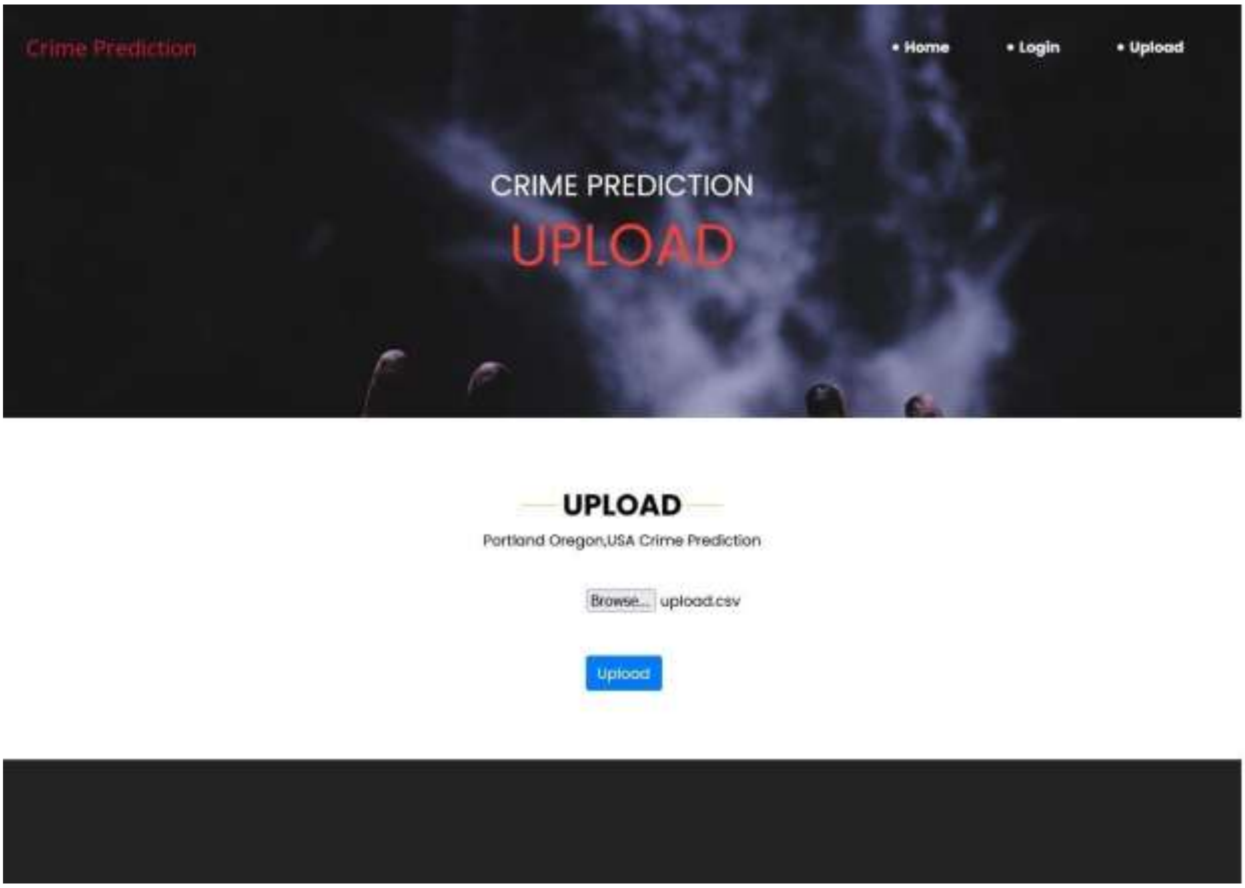
After training and evaluating the model on the validation set, the accuracy of the model will be assessed on the test set. The accuracy on the test set will be an important metric for evaluating the model's performance. We got an accuracy of 95% on test set.

Saving the Trained Model:

Once you’re confident enough to take your trained and tested model into the production-ready environment, the first step is to save it into a .h5 or .pkl file using a library like pickle. Make sure you have pickle installed in your environment.

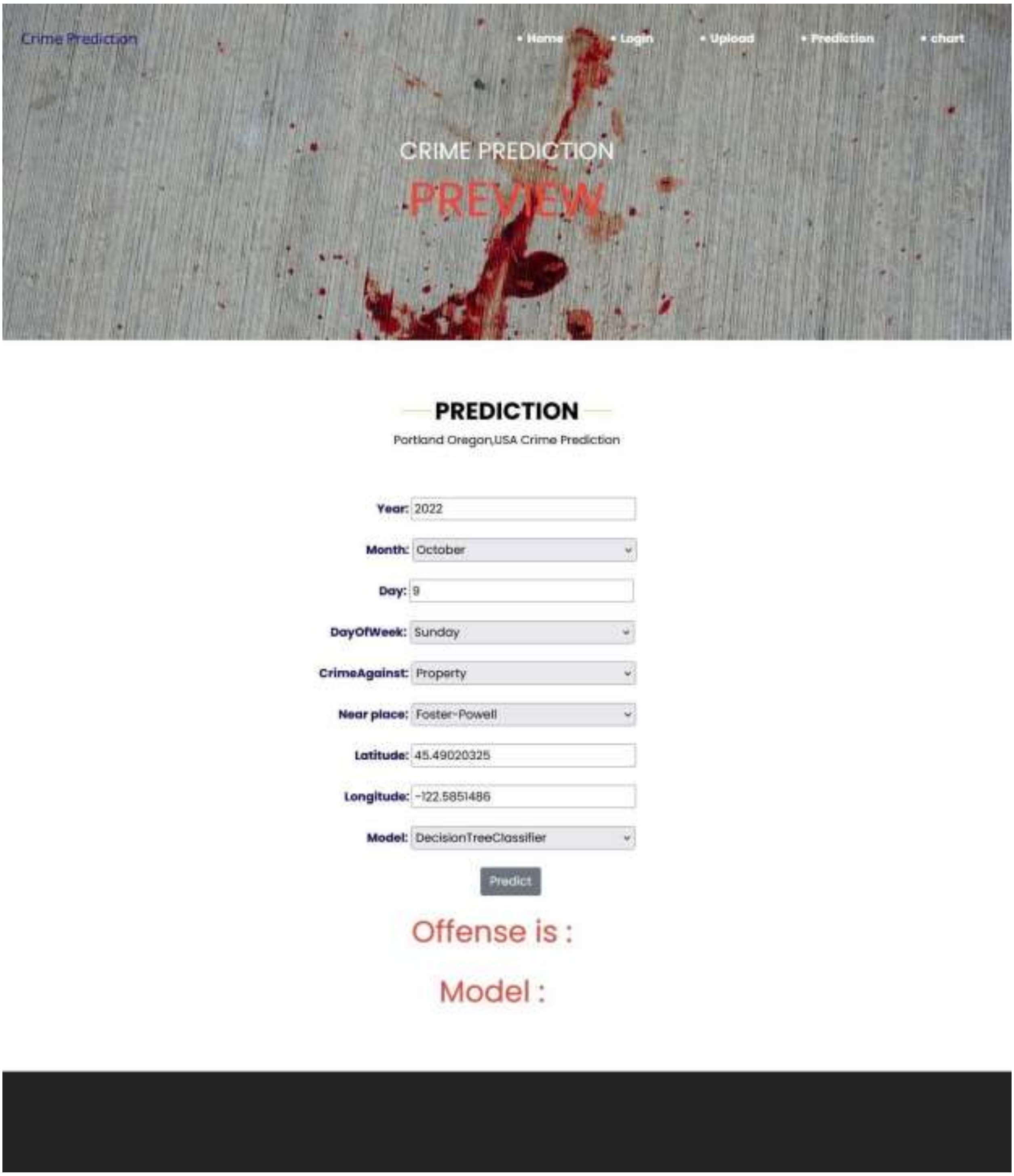
Next, let’s import the module and dump the model into .pkl file.

4.RESULTS AND DISCUSSION



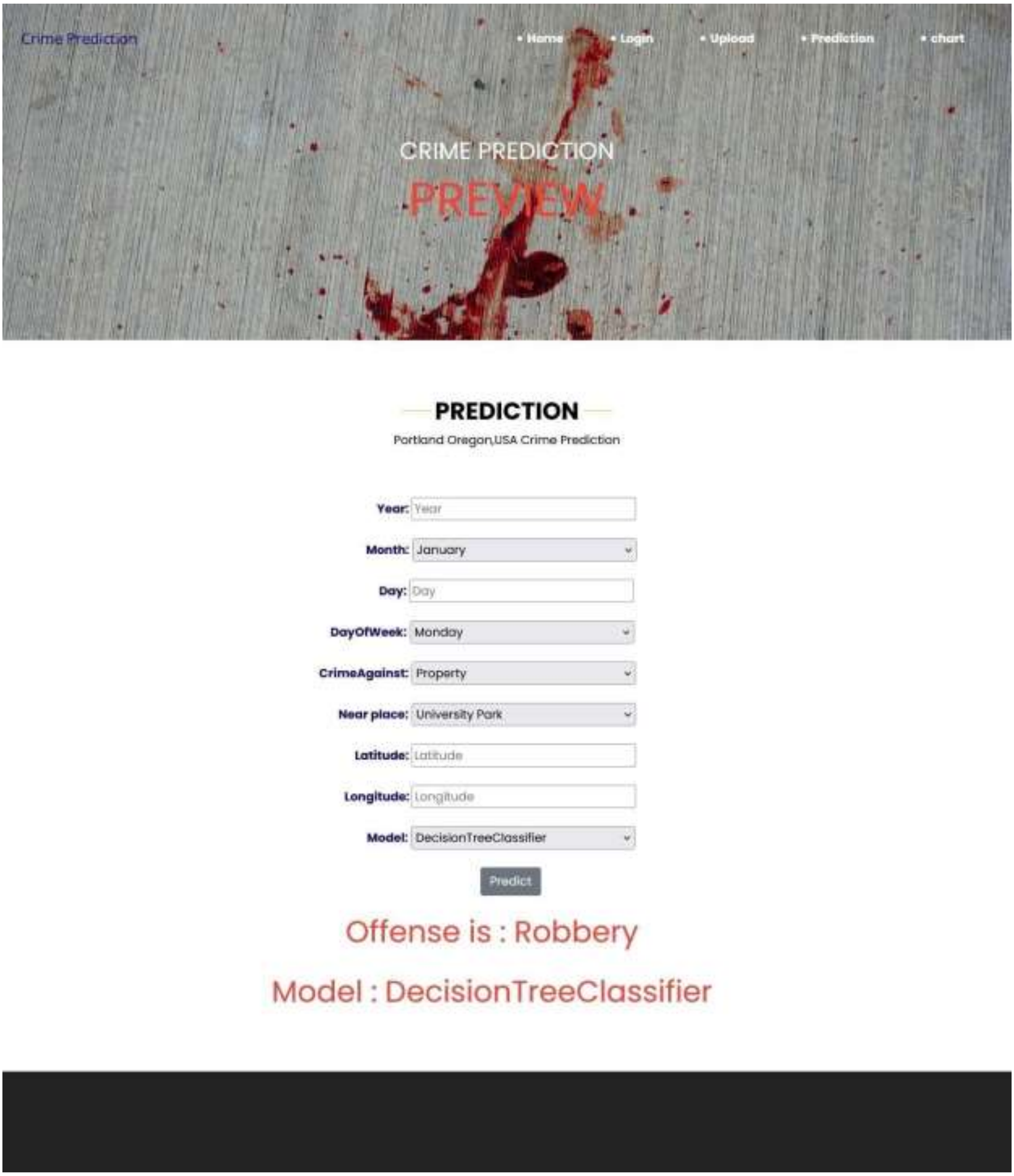
	Address	CaseNumber	CrimeAgainst	Neighborhood	OccurDate	OccurTime	OffenseCategory	
Id								
1	6300 BLOCK OF N LOMBARD ST	23-917974	Property	University Park	31-07-2023	1925	Larceny Offenses	Shoplifti
2	600 BLOCK OF SE 146TH AVE	23-200198	Property	Centennial	31-07-2023	555	Motor Vehicle Theft	Motor Ve
3	700 BLOCK OF SW 10TH AVE	23-917935	Property	Downtown	31-07-2023	1110	Larceny Offenses	Theft Fro Vehicle
4	14000 BLOCK OF NE AIRPORT WAY	23-917921	Property	Argay	31-07-2023	900	Larceny Offenses	Theft of I Parts or .
5	700 BLOCK OF SW 5TH AVE	23-200558	Person	Downtown	31-07-2023	1601	Assault Offenses	Simple A
6	200 BLOCK OF SW ASH ST	23-917942	Property	Old Town/Chinatown	31-07-2023	2010	Larceny Offenses	All Other
7	700 BLOCK OF SE 29TH AVE	23-200713	Person	Sunnyside	31-07-2023	1934	Assault Offenses	Intimida
8	700 BLOCK OF SE 19TH AVE	23-200394	Property	Buckman East	31-07-2023	0	Larceny Offenses	Theft Fro Vehicle
9	NW 19TH AVE / NW LOVEJOY ST	23-200430	Society	Northwest	31-07-2023	1330	Drug/Narcotic Offenses	Drug/Na Violatior
10	NW 19TH AVE / NW LOVEJOY ST	23-200430	Society	Northwest	31-07-2023	1330	Weapon Law Violations	Weapon Violatior
11	3700 BLOCK OF SE LAMBERT ST	23-917898	Property	Eastmoreland	31-07-2023	942	Larceny Offenses	Theft Fro

INPUT PAGE:



Crime Prediction
• Home
• Login
• Upload
• Prediction
• chart
CRIME PREDICTION
PREVIEW
PREDICTION
Portland Oregon,USA Crime Prediction
Year: 2022
Month: October
Day: 9
DayOfWeek: Sunday
CrimeAgainst: Property
Near place: Foster-Powell
Latitude: 45.49020325
Longitude: -122.5851486
Model: DecisionTreeClassifier
Predict
Offense is :
Model :

OUTPUT PAGE:



5.CONCLUSION

All in all, the undertaking "Wrongdoing Expectation Utilizing AI" presents an exhaustive and high level way to deal with foreseeing and ordering different wrongdoing classifications in Portland, Oregon, USA. Utilizing the force of Python and utilizing AI calculations, for example, the XG Lift, Choice Tree Classifier and Stowing Classifier, the framework has shown honorable exactness, accomplishing 98% on the preparation set and 95% on the test set.

The use of a dataset traversing the years 2015 to 2023, containing 505,063 data of interest, has considered a careful investigation of wrongdoing designs over the long run. The grouping of 20 particular wrongdoing classes, from 'Robbery Offenses' to 'Betting Offenses,' gives a nuanced comprehension of crimes in the locale.

The Choice Tree Classifier, known for its interpretability, guarantees straightforwardness in the decisionmaking system, cultivating trust among partners. The presentation of the Stowing Classifier adds to the framework's vigor, relieving overfitting and upgrading the steadiness of expectations.

The rich list of capabilities, including address, case number, wrongdoing against class, neighborhood, happen date, happen time, offense classification, offense type, and spatial directions, empowers the models to make informed expectations by thinking about an expansive scope of variables.

The proposed framework not just addresses the impediments of the current framework yet additionally acquaints headways, for example, flexibility with dynamic changes in wrongdoing examples and versatility for assorted organization situations. Easy to understand connection points and expected reconciliation of extra information sources further upgrade the openness and exactness of bits of knowledge for policing, city organizers, and policymakers.

Generally, "Wrongdoing Expectation Utilizing AI" remains as a hearty and viable device for upgrading public security, improving asset distribution, and working with proactive decisionmaking in metropolitan conditions. The undertaking's outcome in accomplishing high exactness and its comprehensive way to deal with wrongdoing forecast highlight its likely effect on further developing by and large safety efforts and adding to the advancement of informed public approaches..

.

REFERENCES

R. Canetti, U. Feige, O. Goldreich, M. Naor. Adaptively secure multipartycomputation. STOC, pages 639–648. ACM, 1996.

J. Li and E. R. Omiecinski. Efficiency and security trade-off in supporting rangequeries on encrypted databases. In Proc. 19th IFIP Conf. on Data and App. Security & Privacy, pages 69–83, 2005.

A. Boldyreva, N. Chenette, and A. O'Neill. Order preserving encryption revisited:Improved security analysis and alternative solutions. In Proc. CRYPTO, pages 578–595, 2011.

D. X. Song, D. Wagner, and A. Perrig. Practical techniques for searches on encrypteddata. In Proc. IEEE Symposium on Security and Privacy (S&P), pages 44–
55. IEEE, 2000.

E. Goh. Secure indexes. Stanford University Tech. Report, 2004.

D. Boneh, G. D. Crescenzo, R. Ostrovsky, and G. Persiano. Public key encryptionwith keyword search. In Proc. EUROCRYPT, pages 506–522, 2004.

Y.-C. Chang and M. Mitzenmacher. Privacy preserving keyword searches on remoteencrypted data. In Third Inte. Conf. on Applied Cryptography and Network Security (ACNS), pages 442– 455, 2005.

R. Curtmola, J. Garay, S. Kamara, and R. Ostrovsky. Searchable symmetricencryption: improved definitions and efficient constructions. In Proc. CCS, pages 79-88, 2006.

M. Bellare, A. Boldyreva, and A. O´ıNeill. Deterministic and efficiently searchableencryption. CRYPTO, pages 535–552, 2007.

S. Kamara, C. Papamanthou, and T. Roeder. Dynamic searchable symmetricencryption. In Proc. CCS, pages 965–976, 2012.

D. Cash, S. Jarecki, C. Jutla, H. Krawczyk, M.-C. Rosu, and M. Steiner. Highlyscalable searchable symmetric encryption with support for boolean queries. In CRYPTO, pages 353–373, 2013.

D. Cash, J. Jaeger, S. Jarecki, C. Jutla, H. Krawczyk, M.-C. Rosu, and M. Stiner.Dynamic searchable encryption in very-large databases: Data structures and implementation. In NDSS. 2014.

R. Li, A. X.Liu, L. Wang, and B. Bezawada. Fast range query processing withstrong privacy protection for cloud computing. In Proc. VLDB, pages 1953–1964. IEEE, 2014. [1]

N. Shah, N. Bhagat, and M. Shah, ‘‘Crime forecasting: A machine learning and computer vision approach to crime prediction and prevention,’’ Vis. Comput. Ind., Biomed., Art, vol. 4, no. 1, pp. 1–14, Apr. 2021.

[2] S. A. Chun, V. A. Paturu, S. Yuan, R. Pathak, V. Atluri, and N. R. Adam, ‘‘Crime prediction model using deep neural networks,’’ in Proc. 20th Annu. Int. Conf. Digit. Government Res., Jun. 2019, pp. 512–514.

[3] S. S. Kshatri, D. Singh, B. Narain, S. Bhatia, M. T. Quasim, and G. R. Sinha, ‘‘An empirical analysis of machine learning algorithms for crime prediction using stacked generalization: An ensemble approach,’’ IEEE Access, vol. 9, pp. 67488– 67500, 2021.

[4] C. Janiesch, P. Zschech, and K. Heinrich, ‘‘Machine learning and deep learning,’’ Electron. Mark., vol. 31, no. 3, pp. 685–695, Apr. 2021.

[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.

[6] S. Kim, P. Joshi, P. S. Kalsi, and P. Taheri, ‘‘Crime analysis through machine learning,’’ in Proc. IEEE 9th Annu. Inf. Technol., Electron. Mobile Commun. Conf. (IEMCON), Nov. 2018, pp. 415–420.

Author’s Profiles

#1:-B. TRIVENI working as Associative Professor in Department of IT in DVR & Dr.HS MIC College of Technology,Kanchikacherla-521180

#2:T.SOWJANYA(20H71A1243) B.Tech with Specialization of Information Technology in DVR & Dr.HS MIC College of Technology,Kanchikacherla- 521180

#3:- Y.AJAY(20H71A1201) B.Tech with Specialization of Information Technology in DVR & Dr.HS MIC College of Technology,Kanchikacherla-521180

#4:- M. NAGA LAKSHMI (21H75A1204) B.Tech with Specialization of Information Technology in DVR & Dr.HS MIC College of Technology, kanchikacherla 521180

#5:- K.GOPI KRISHNA (20H71A1208) B.Tech with Specialization of Information Technology in DVR & Dr.HS MIC College of Technology,Kanchikacherla- 521180