Edge Assisted Crime Prediction and Evaluation Framework for Machine Learning Algorithms

MALINENI PERUMALLU EDUCATIONAL SOCIETY'S EDUCATIONAL INSTITUITIONS

Bachelor of Technology in

COMPUTER SCIENCE AND ENGINEERING

Under The Guidance Of

Mr. K N B CHARI

M.TECH

ASSISTANT PROFESSOR

TEAM MEMBERS:

- M.Anand Sai
- M.Vinod
- CH.Sathish
- B.Manikanta
- Gopi Krishna

ABSTRACT

The growing populations in major cities have led to new challenges, particularly in public safety and regulation. This paper presents a strategy for predicting crime occurrences using historical data and demographic observations. It proposes a crime prediction framework utilizing machine learning algorithms at the network edge. The study evaluates four types of crimes—murder, trial speed, women and children repression, and narcotics—and validates the framework's effectiveness. Machine learning models, including Decision Trees, Neural Networks, K-Nearest Neighbors, and Impact Learning, are used for crime prediction, with the Decision Tree model achieving a maximum accuracy of 81%. The results highlight how machine learning can enhance public safety by predicting criminal events.

INTRODUCTION :

Public safety is a primary concern for the global population, particularly with the rise of urbanization. As people increasingly migrate to cities, it's predicted that over 70% of the world’s population will reside in urban areas by 2050 [1]. Additionally, the number of terrorist attacks, defined by the Global Terrorism Database as acts of violence aimed at civilian populations to achieve political objectives, has reached the highest levels ever recorded in the past decade. Machine learning (ML) techniques play a crucial role in addressing these challenges, particularly in smart city applications, by leveraging data to improve public safety [2]. This paper presents a graphical representation of crime across various regions, including Bangladesh, using data from 2012 to 2019. The crime prediction model reveals that region 1 experiences higher crime rates compared to region 2, which has a lower crime rate. Based on this, a predictive model for crime in 2021 was developed. The study compares crime incidents between 2019 and 2021, providing valuable insights for residents and municipal officials. This information aids in identifying the most dangerous areas, helping to improve patrol route planning and offering tourists essential guidance on safer areas to visit, ultimately enhancing community safety.

EXISTING SYSTEM :

- Thus, a complete analysis of four distinct sorts of crimes, such as murder, rapid trial, repression of women and children, and narcotics, validates the efficiency of the proposed framework.
- The complete study and implementation process have shown a visual representation of crime in various areas of country.
- The total work is completed by the selection, assessment, and implementation of the Machine Learning (ML) model, and finally, proposed the crime prediction.
- Criminal risk is predicted using classification models for a particular time interval and place.

DISADVANTAGES OF EXISTING SYSTEM

- Correlations do not necessarily imply causality or predictive power. Unknown factors could be responsible.
- Predictions will be based on past data which reflects historical biases in enforcement, surveillance, reporting.
- Hard to determine how far into the future predictions remain accurate or useful. Model uncertainty compounds over time.
- Nearly impossible to capture interactions between complex social and psychological variables that drive criminal acts.

PROPOSED SYSTEM :

ML methods such as Decision Trees, Neural Networks, K-Nearest Neighbors, and Impact Learning are being utilized, and their performance is compared based on the data processing and modification used. A maximum accuracy of 81% is obtained for Decision Tree algorithm during the prediction of crime. Crime data for training models is limited - we don't know the full scope of crimes committed, just those reported/recorded. And recorded data has biases based on enforcement patterns. There's an inherent randomness/unpredictability in human behavior and events. Tiny differences can cascade in unexpected ways. Detailed prediction far into the future is a exponentially harder problem versus short-term forecasting.

ADVANTAGES OF PROPOSED SYSTEM

- Machine learning algorithms to analyze crime patterns and make predictive models
- Edge computing framework to enable localized, low-latency predictions
- Visualizations and comparisons of predicted crime rates across regions
- Use of real crime data to train and evaluate the models
- Goal of enabling data-driven crime prevention and improving public safety_.

ALGORITHMS

RANDOM FOREST:

Random Forest is a machine learning algorithm that combines multiple decision trees to make predictions, improving accuracy and reducing overfitting compared to a single decision tree.

Advantages:

High Accuracy

Handles Missing Data

Robust to Outliers

Ease of Use

Reduces Overfitting

MODULES

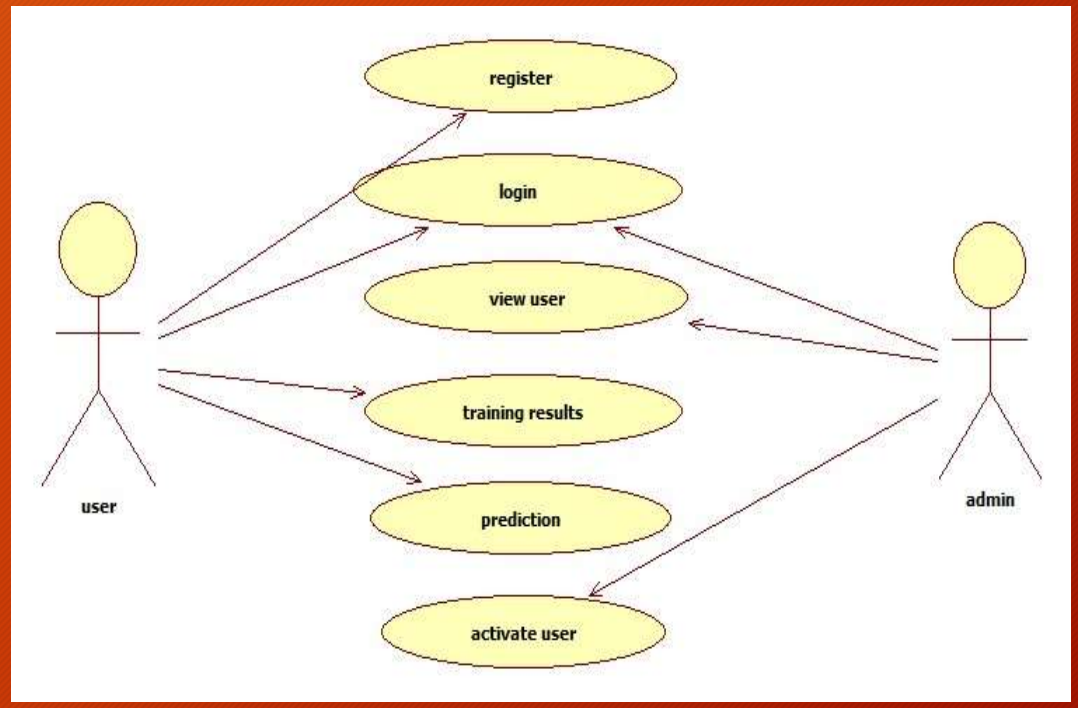
- **User Module**: Users first register with their email and mobile number. After the admin activates the user account, the user can log in, upload a dataset (like the Employment Scam Aegean Dataset), and preprocess the data in float format. The user can then apply machine learning algorithms for classification, which will calculate accuracy, macro average, and weighted average. They can view the results and predictions on the web interface.
- **Admin Module**: Admins can log in, activate user accounts, and view overall data. Admins can also access the classification results, including accuracy and averages for all algorithms, and see the final classification results after all algorithms have been executed.

- **Data Preprocessing**: The dataset undergoes preprocessing to remove noise and HTML tags. Feature selection is applied to reduce the number of attributes. Support Vector Machine (SVM) is used for feature selection, while an ensemble classifier based on Random Forest detects fraudulent job posts with an impressive 97.4% accuracy.
- **Machine Learning**: Various classifiers (like KNN, Decision Tree, SVM, Naive Bayes, Random Forest, Multilayer Perceptron, and Deep Neural Network) are used to classify job posts as real or fraudulent. The system calculates and displays accuracy, macro average, and weighted average for each classifier, helping to determine the best one based on performance.

UML DIAGRAMS

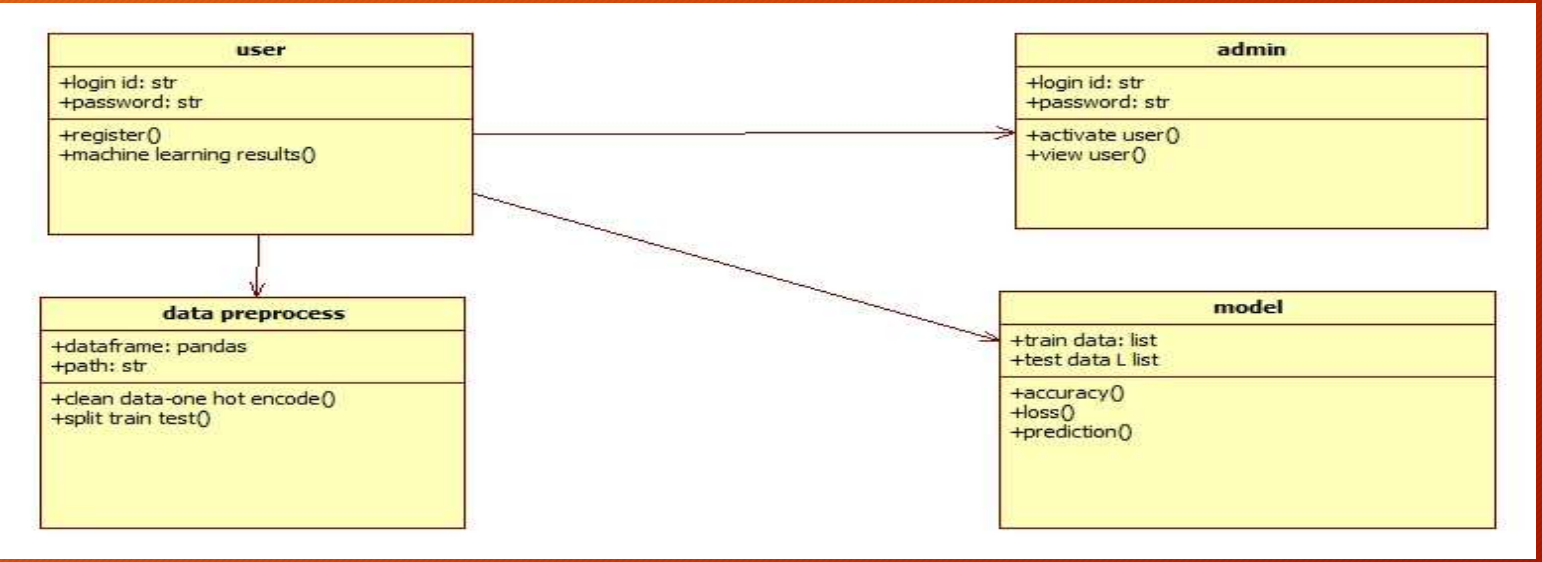
Use Case Diagrams :

A use case diagram in the Unified Modeling Language (UML) is a type of behavioral diagram defined by and created from a Use-case analysis. Its purpose is to present a graphical overview of the functionality provided by a system in terms of actors, their goals (represented as use cases), and any dependencies between those use cases. The main purpose of a use case diagram is to show what system functions are performed for which actor. Roles of the actors in the system can be depicted.



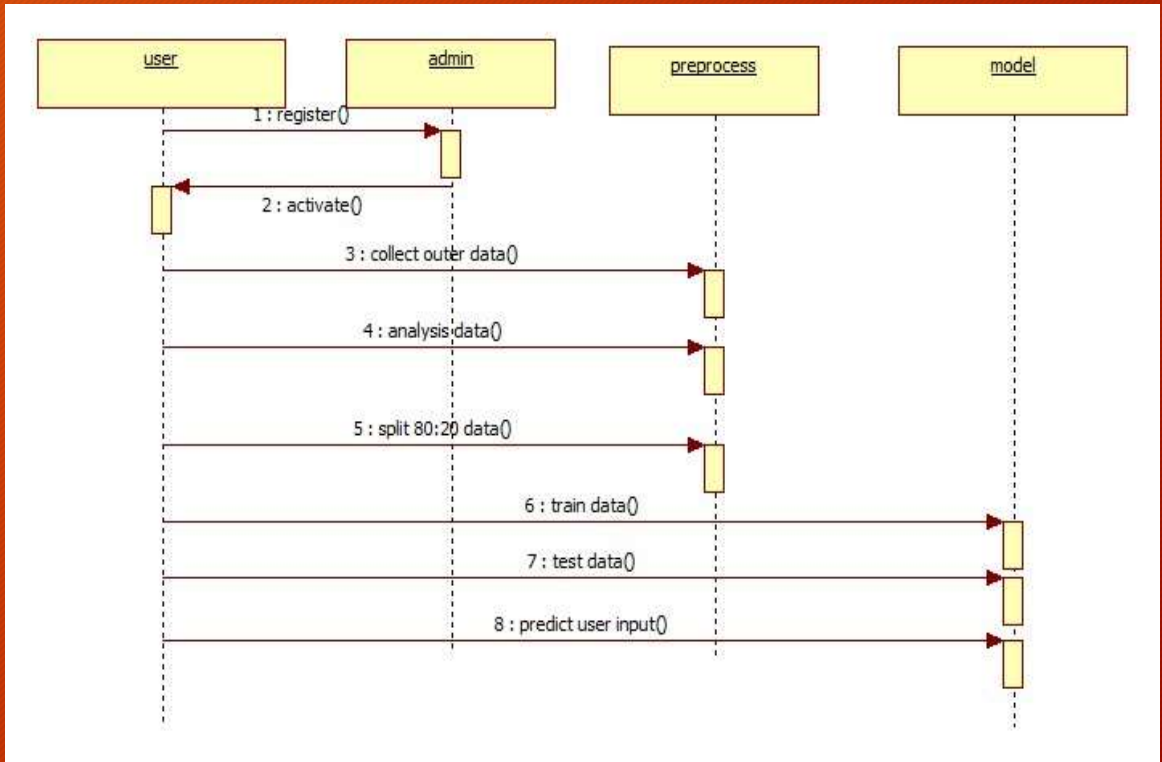
Class Diagram

In software engineering, a class diagram in the Unified Modeling Language (UML) is a type of static structure diagram that describes the structure of a system by showing the system's classes, their attributes, operations (or methods), and the relationships among the classes. It explains which class contains information.



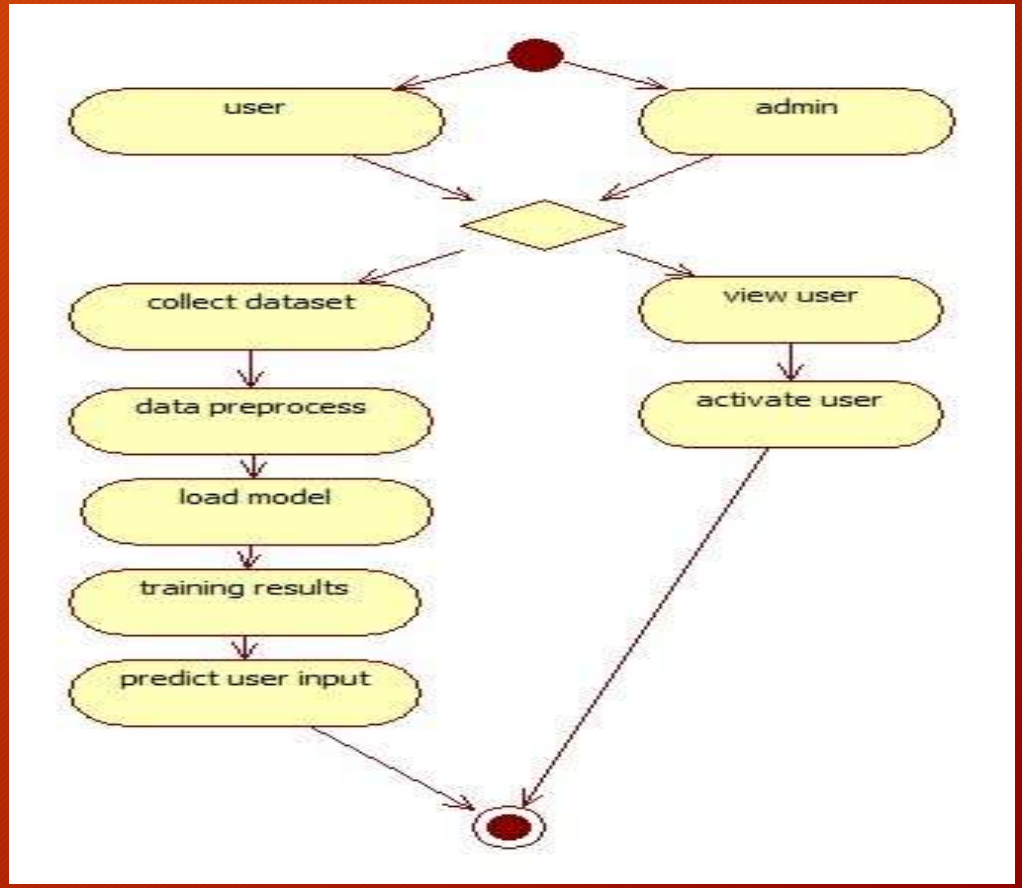
Sequence Diagram

A sequence diagram in Unified Modeling Language (UML) is a kind of interaction diagram that shows how processes operate with one another and in what order. It is a construct of a Message Sequence Chart. Sequence diagrams are sometimes called event diagrams, event scenarios, and timing diagrams.

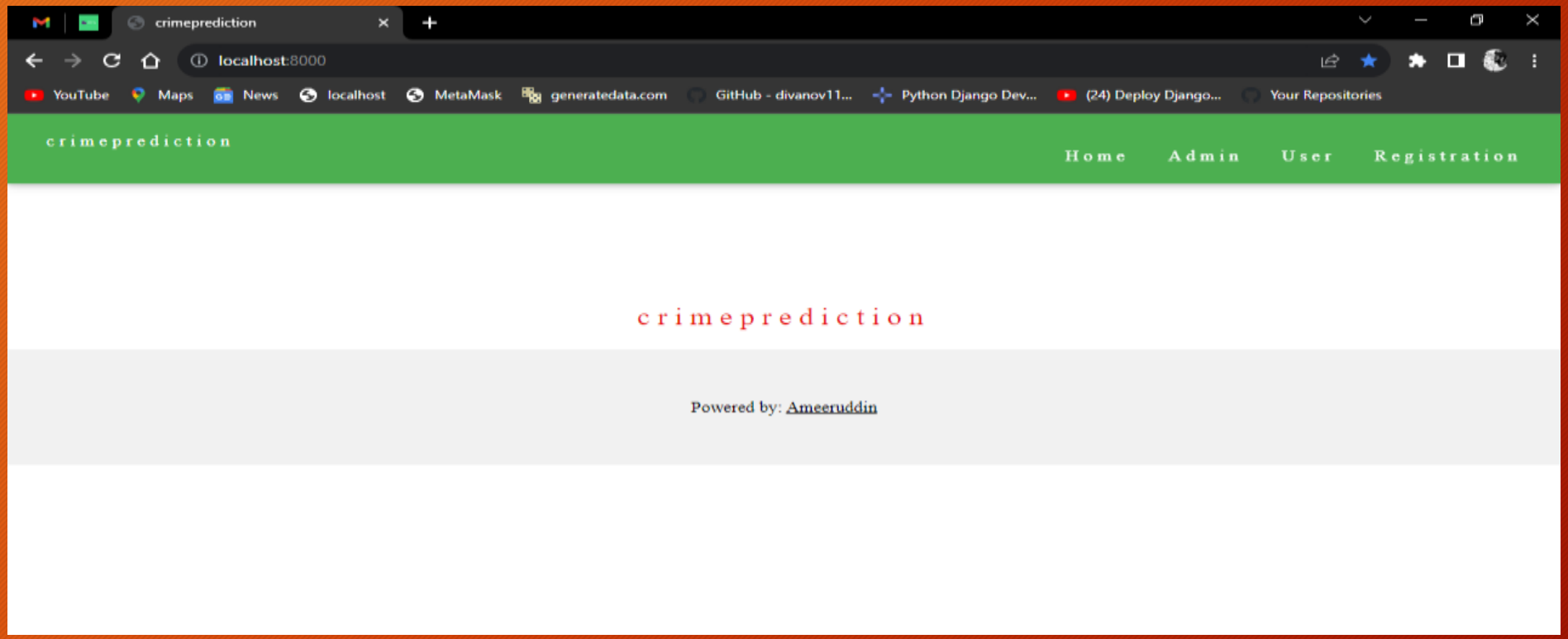


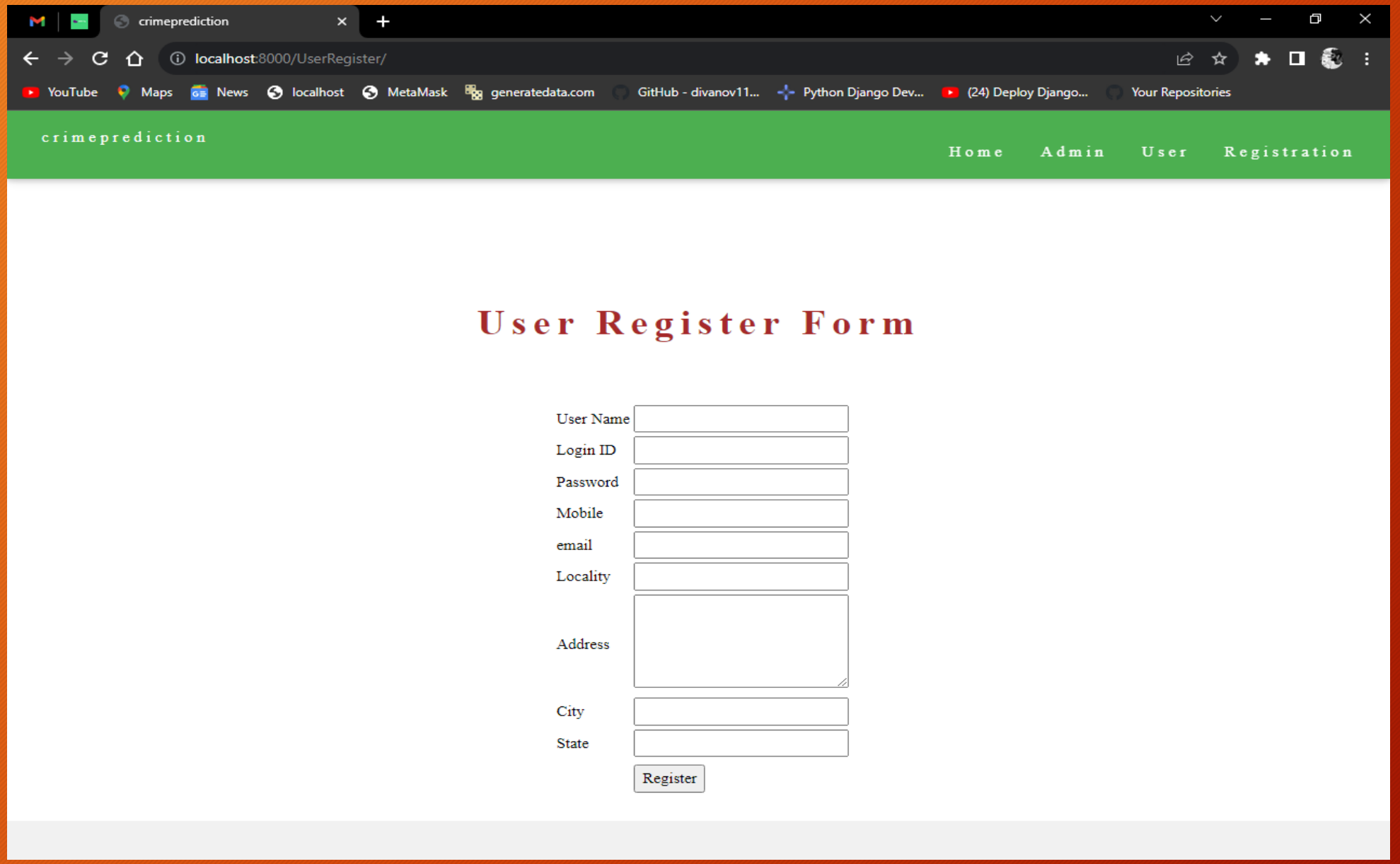
Activity Diagram

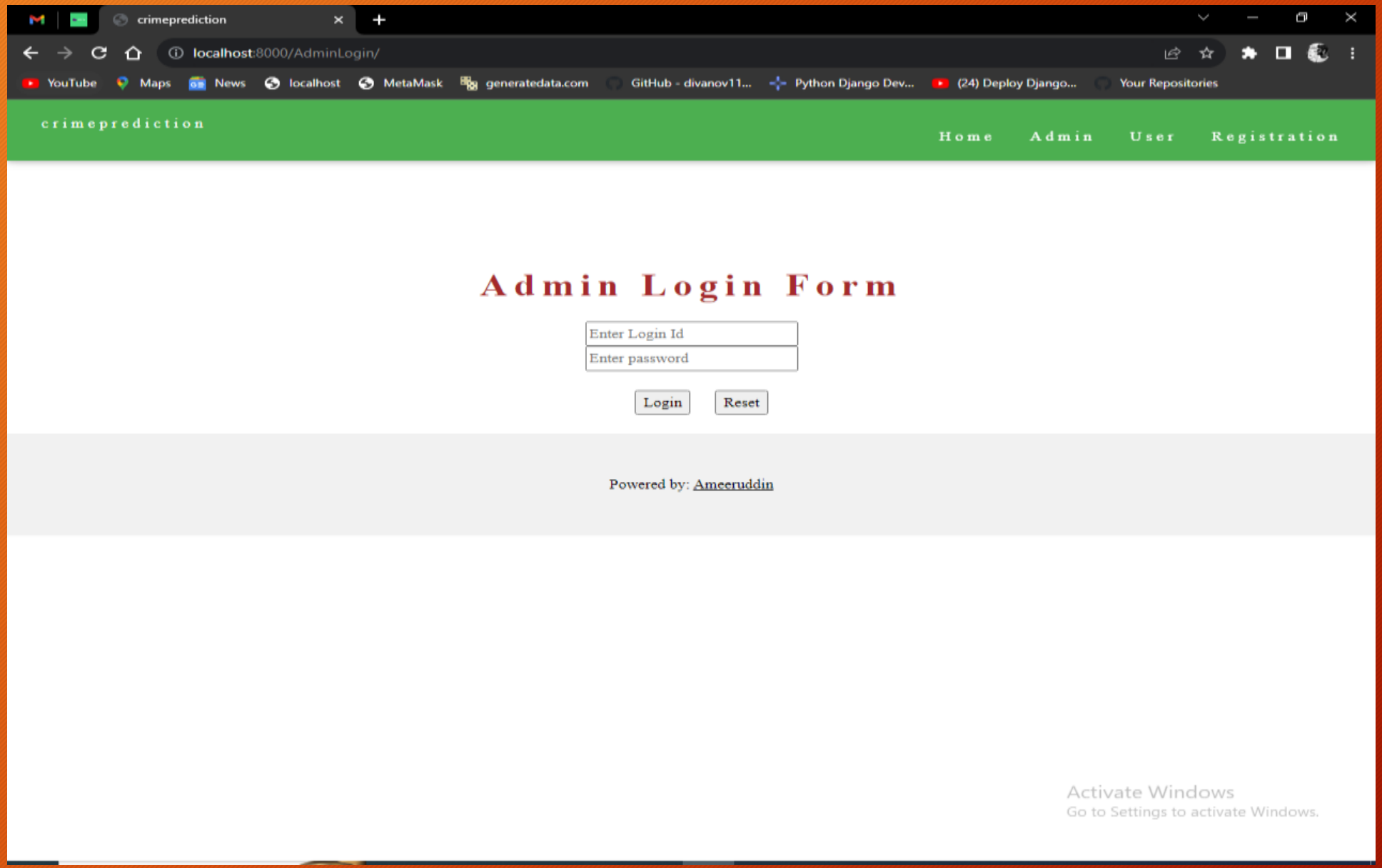
Activity diagrams are graphical representations of workflows of stepwise activities and actions with support for choice, iteration and concurrency. In the Unified Modeling Language, activity diagrams can be used to describe the business and operational step-by-step workflows of components in a system. An activity diagram shows the overall flow of control.

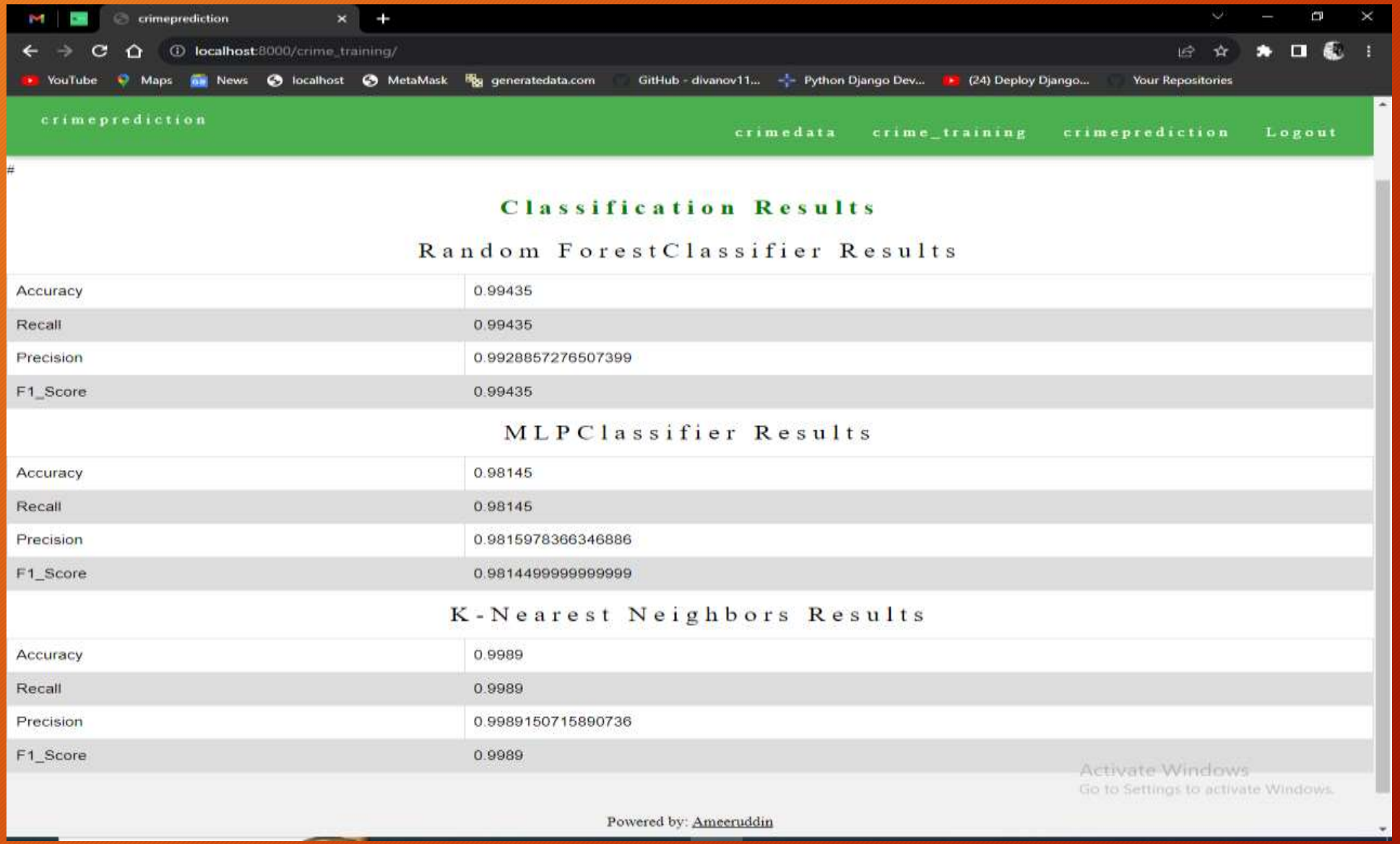


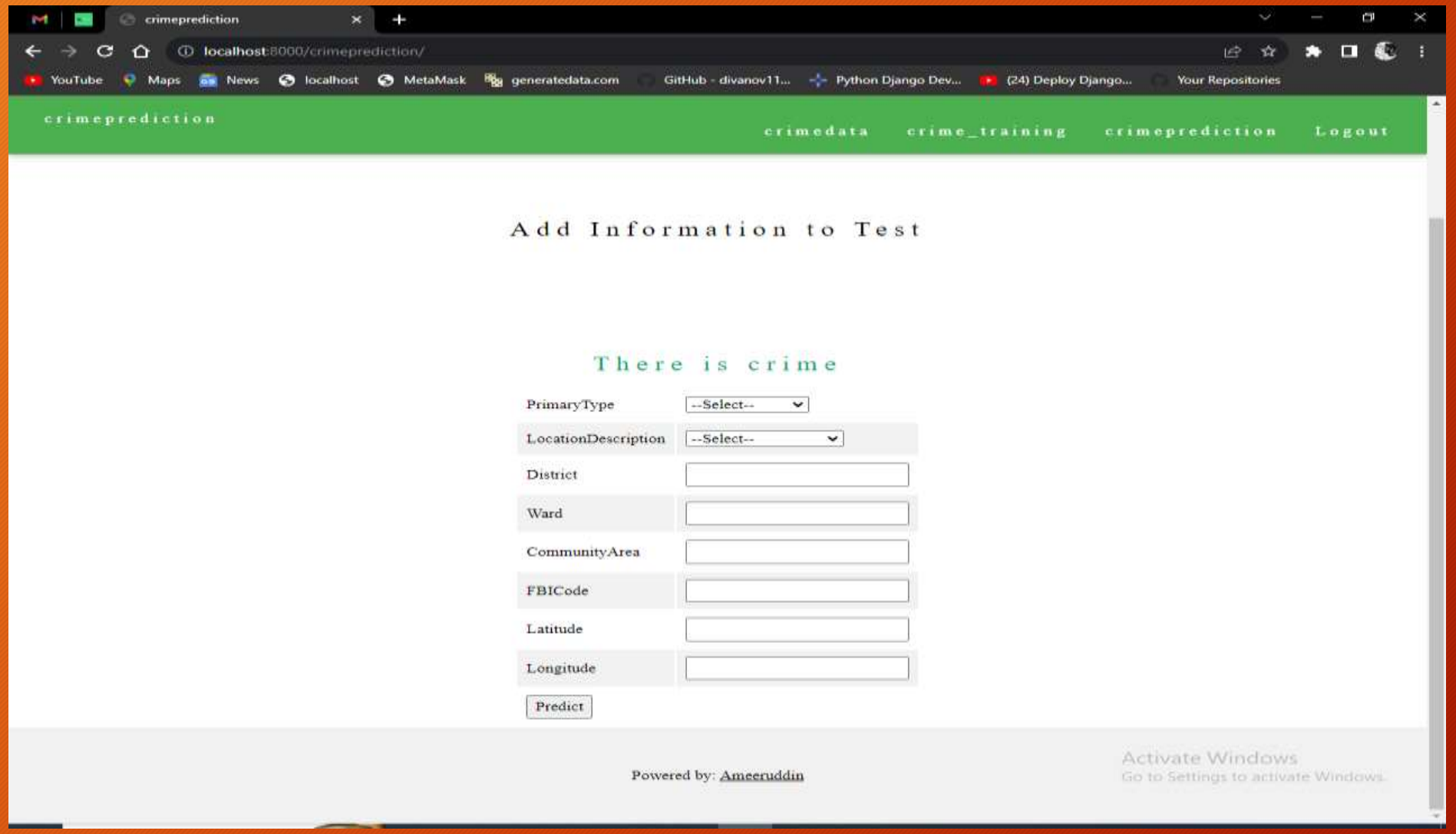
Screen shots




crimeprediction
localhost:8000/UserRegister/
YouTube
Maps
News
localhost
MetaMask
generatedata.com
GitHub - divanov11...
Python Django Dev...
(24) Deploy Django...
Your Repositories
crimeprediction
Home
Admin
User
Registration
User Register Form
User Name
Login ID
Password
Mobile
email
Locality
Address
City
State
Register


crimeprediction
localhost:8000/AdminLogin/
YouTube
Maps
News
localhost
MetaMask
generatedata.com
GitHub - divanov11...
Python Django Dev...
(24) Deploy Django...
Your Repositories
crimeprediction
Home
Admin
User
Registration
Admin Login Form
Enter Login Id
Enter password
Login
Reset
Powered by: Ameeruddin
Activate Windows
Go to Settings to activate Windows.


crimeprediction
localhost:8000/crime_training/
YouTube
Maps
News
localhost
MetaMask
generatedata.com
GitHub - divanov11...
Python Django Dev...
(24) Deploy Django...
Your Repositories
crimeprediction
crimedata
crime_training
crimeprediction
Logout
#
Classification Results
Random ForestClassifier Results
Accuracy 0.99435
Recall 0.99435
Precision 0.9928857276507399
F1_Score 0.99435
MLPClassifier Results
Accuracy 0.98145
Recall 0.98145
Precision 0.9815978366346886
F1_Score 0.9814499999999999
K-Nearest Neighbors Results
Accuracy 0.9989
Recall 0.9989
Precision 0.9989150715890736
F1_Score 0.9989
Activate Windows
Go to Settings to activate Windows.
Powered by: Ameeruddin


crimeprediction
localhost:8000/crimeprediction/
YouTube
Maps
News
localhost
MetaMask
generatedata.com
GitHub - divanov11...
Python Django Dev...
(24) Deploy Django...
Your Repositories
crimeprediction
crimedata
crime_training
crimeprediction
Logout
Add Information to Test
There is crime
PrimaryType
--Select--
LocationDescription
--Select--
District
Ward
CommunityArea
FBICode
Latitude
Longitude
Predict
Powered by: Ameeruddin
Activate Windows
Go to Settings to activate Windows.

SYSTEM REQUIREMENTS

SOFTWARE REQUIREMENTS:

- **Operating system** : Windows 10.
- **Coding Language** : Python.
- **Front-End** : Html. CSS
- **Designing** : Html,css,javascript.
- **Data Base** : SQLite.

HARDWARE REQUIREMENTS:

- System : Intel i3
- Hard Disk : 1 TB.
- Monitor : 14’ Colour Monitor.
- Mouse : Optical Mouse.
- Ram : 4GB.

Conclusion

In this work, we have introduced a crime prediction and evaluation framework for machine learning algorithms of network edge. We collected data from 2012 to 2019 to analyze and evaluate our forecast. We used machine learning approaches to anticipate crime events, which can make a significant contribution to improving city public safety, which is a big problem in many cities across the world. It was fascinating to see how pre-processing, and transformation may affect the model's output, particularly when breaking the day into many time periods. Due to the provenance of the data, this solution was created for a specific city in the country. However, if equivalent data is made accessible, the technique may be applied to other cities. Based on the training set input for the four algorithms, we find the Decision tree method to be extremely successful and accurate in predicting crime data. The Decision Stump algorithm's poor performance could be attributed to a certain amount of randomness in the various crimes and associated features (shows a low correlation coefficient among the four algorithms); the KNN's branches are more rigid and only give accurate results if the test set follows the pattern modelled.

References

1. “68% of the world population projected to live in urban areas by 2050, says UN — UN DESA — United Nations Department of Economic and Social Affairs.” https://www.un.org/development/desa/en/news/population/2018- revision-of-world urbanization-prospects.html (accessed Oct. 14, 2021).
2. Y. Wu, W. Zhang, J. Shen, Z. Mo, and Y. Peng, “Smart city with Chinese characteristics against the background of big data: Idea, action and risk,” Journal of Cleaner Production, vol. 173, pp. 60–66, Feb. 2018.
3. M. Kowsher, A. Tahabilder, and S. A. Murad, “Impact-learning: A robust machine learning algorithm,” in ACM International Conference Proceeding Series, pp. 9–13, Jul. 2020.
4. Priyanka and D. Kumar, “Decision tree classifier: A detailed survey,” International Journal of Information and Decision Sciences, vol. 12, no. 3, pp. 246–269, 2020.

RATE OF MURDERS AND NARCOTICS IN VARIOUS REGIONS

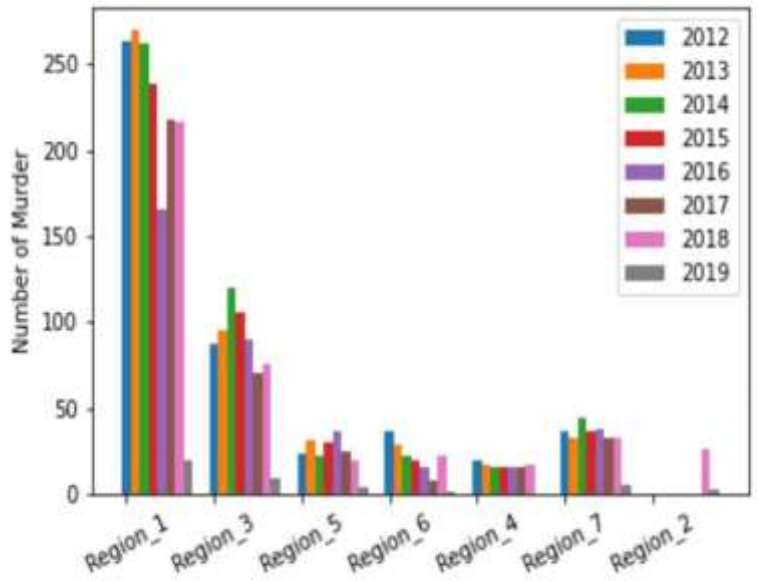


Fig. 4. The rate of Murder in various region.

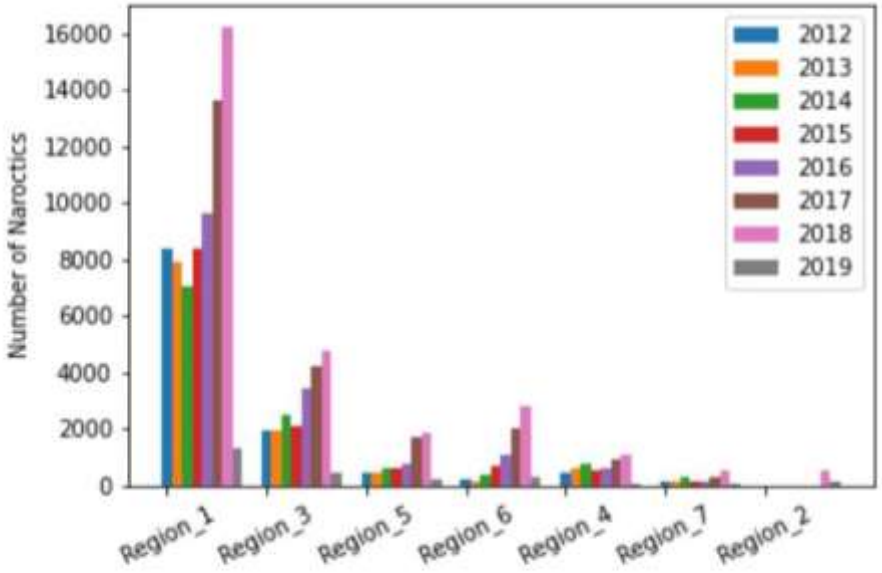


Fig. 7. The rate of Narcotics in various region.

THANK YOU