PREDICTIVE CRIME ANALYSIS PROJECT REPORT

Introduction

The project focuses on the application domain of crime analysis, specifically within the city of Portland. The primary objective is to harness historical crime data to develop a predictive model that identifies potential hotspots of criminal activity across different districts. The underlying motivation is to enhance the efficiency of law enforcement agencies in resource allocation, thereby aiding in crime prevention and overall public safety.

The project involves a comprehensive analysis of historical crime data from Portland, encompassing various types of crimes and their locations over a specified time period. Machine learning algorithms are then employed to build a predictive model that can identify locations of criminal activity. Data preprocessing, and model training are integral components of the approach.

The results of the project aim to provide law enforcement agencies with actionable insights into potential crime hotspots. The predictive model's effectiveness is evaluated through testing against historical data and, when applicable, real-time data. The goal is to demonstrate the model's ability to accurately forecast areas where criminal activities are likely to occur, enabling law enforcement to deploy preventive measures and patrols strategically. The success of the project is measured by the model's predictive accuracy.

In summary, this project endeavors to harness the power of data and machine learning to empower law enforcement agencies in Portland to proactively address and prevent criminal activities. The application of predictive modeling provides a forward-looking strategy, allowing for more effective and targeted deployment of resources, with the ultimate goal of creating safer communities.

Related Work

Here-in are the related work:-

1. Predicting Crime in Portland Oregon by Jorie Koster-Hale (2016): This blog post describes the winning entry for the National Institute of Justice Real-Time Crime Forecasting Challenge. The team used a combination of police reports, open-source data (e.g., US census, Foursquare), and time-series modeling to predict crime hotspots in Portland. Their approach achieved promising results, demonstrating the feasibility of using diverse data sources for crime prediction.
2. Open source crime Prediction for the National Institute of Justice (2016): This document details another winning entry for the NIJ challenge. The authors used open-source software and freely available data to develop a crime

prediction model for Portland. Their approach demonstrates the potential of using readily available resources for crime analysis, making it accessible to law enforcement agencies with limited budgets.

3. Crime Hotspots in Taiwan: A Spatiotemporal Analysis by Chih-Chien Lee et al. (2020): This study explores the use of a Poisson regression model for identifying crime hotspots in Taiwan. The authors demonstrate the effectiveness of this simple and transparent model in predicting crime hotpots, suggesting its potential application in other contexts, including Oregon districts.

Methods

We used Python, Pandas, and relevant libraries for data manipulation, GeoPandas for spatial analysis, and machine learning libraries for model development and evaluation. Below are the methodologies used:-

i. Data Collection: We utilized crime forecasting datasets recommended by the National Institute of Justice (NIJ) from https://nij.ojp.gov/funding/real-time-crimeforecasting-challenge-posting. The data included crime types, temporal ranges, and geographic coordinates, sourced from three Excel files. We employed Python and Pandas to import, extract, and concatenate the data into a comprehensive Data Frame.
ii. Data Exploration: Examined the dataset's structure to understand its characteristics and distributions.
iii. Data Preprocessing: Handled missing values and duplicates to ensure data quality.
iv. Data Analysis: Merged point data with a GeoDataFrame of Portland district boundaries. This step involved mapping Portland districts with a grid overlay, visually exploring patterns and relationships.
v. Model Building: Constructed a machine learning model for crime location prediction. We achieved this by building an LSTM model with one LSTM layer and one Dense output layer. We trained the model on the training set for 50 epochs.
vi. Model Testing: Assessed the model's performance on a designated dataset.
vii. Model Performance Evaluation: We evaluated the LSTM model on the test set and printed the Mean Squared Error (MSE), Mean Absolute error and Root Mean Squared Error. We made predictions on the test set using the trained model. We also plotted predicted vs. actual coordinates to visualize differences and assess model performance.

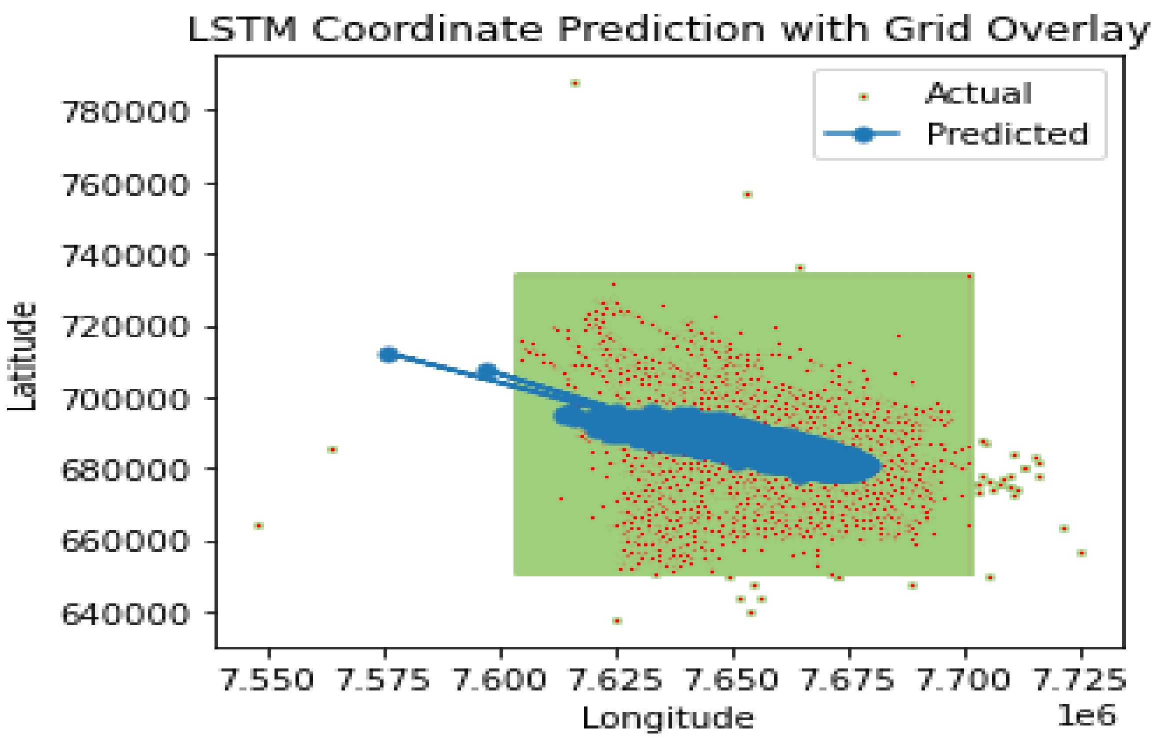
Results and Discussions

The obtained Mean Squared Error (MSE) of 0.0067 serves as a crucial metric for evaluating the overall performance of the model. The small value of MSE is indicative of the model's ability to predict crime occurrences accurately on unseen data. Both the training and test set MSE values being low suggest that the model performs consistently across different datasets. This establishes the reliability of the model in predicting time series coordinates related to crime.

The Mean Absolute Error (MAE) of 0.06 provides an average measure of the absolute differences between predicted and actual values. This metric indicates that, on average, individual predictions can deviate by up to 0.06 units.

The Root Mean Squared Error (RMSE) of 0.082 offers an additional perspective on the error magnitude, considering the original scale of the target variable. This metric is valuable for contextualizing the errors in a meaningful way, especially when comparing the model's performance to other forecasting methods or benchmarks.

In summary, the obtained MSE, MAE, and RMSE values collectively indicate that the model exhibits strong overall performance in predicting crime occurrences. The low error metrics suggest the model's potential for generating reliable forecasts, which can significantly contribute to the development of effective crime prevention strategies.



Conclusion

In conclusion, our machine learning model has demonstrated promising results in predicting forecasted crime coordinates using historical data from Portland, Oregon. The low Mean Squared Error observed in our evaluation indicates the model's ability to provide meaningful forecasts of potential crime hotspots. This predictive capability can be invaluable for law enforcement agencies in optimizing resource allocation and strategically deploying personnel to areas identified as high risk through our visualizations.

However, it is crucial to acknowledge certain limitations and shortcomings in our approach. Firstly, our model's performance may be influenced by changes in socio-economic factors, law enforcement strategies, or other external variables that are not explicitly considered in our dataset. Additionally, the model's generalizability to different cities or regions might be a point of concern, as crime patterns can vary significantly based on local factors.
For future work, it is recommended to explore ways to enhance the model's robustness and adaptability to different contexts. Incorporating real-time data feeds and continuously updating the model could improve its accuracy and relevance over time. In terms of project extensions, the integration of additional data sources, such as weather patterns, social events, or community dynamics, could offer a more comprehensive understanding of crime trends. Collaborations with local communities and stakeholders can provide valuable insights for refining the model and tailoring it to specific needs. Finally, the development of user-friendly interfaces for law enforcement agencies and community members would facilitate effective utilization of the predictions, fostering a collaborative and data-driven approach to crime prevention and intervention.

Data and Software Availability
Here is the link to our GitHub repository: https://github.com/GVSU-CIS635/gvsu-cis635-term-project-predictive-crime-analysis.

In this project, we utilized Python software. The crime forecasting datasets used were obtained from the National Institute of Justice (NIJ) and can be accessed at https://nij.ojp.gov/funding/real-time-crime-forecasting-challenge-posting#data.

Specifically, we utilized the following folders from the NIJ, which contain the three datasets used and the Portland district map:

1. Dataset 1: https://nij.ojp.gov/sites/g/files/xyckuh171/files/media/document/010113_123113_Data.zip

2. Dataset 2: https://nij.ojp.gov/sites/g/files/xyckuh171/files/media/document/010114_123114_Data.zip

3. Dataset 3: https://nij.ojp.gov/sites/g/files/xyckuh171/files/media/document/010115_123115_Data.zip

4. Portland District Map: https://nij.ojp.gov/sites/g/files/xyckuh171/files/media/document/portland-police-districts.zip

Additionally, we have included the three datasets and the Portland map in our GitHub repository.

References

1. Real-Time Crime Forecasting Challenge Posting https://nij.ojp.gov/funding/real-time-crime-forecasting-challenge-posting#data
2. Artificial intelligence & crime prediction: A systematic literature review https://www.sciencedirect.com/science/article/pii/S2590291122000961
3. TensorFlow Package https://www.tensorflow.org/learn