Course Project#1

Title: Predicting Crime Rates in Chicago

Project Description:

Crime in Chicago has long been a compelling subject of discussion, both due to its significant societal impact and the wealth of data available for analysis. Having lived in Chicago for over two years, I've often found crime to be a recurring topic in conversations with friends and family. This personal connection, combined with the accessibility of high-quality crime datasets, makes it an ideal subject for deeper exploration.

This project approaches the analysis of crime in Chicago from the perspective of a resident seeking to better understand the city's dynamics and navigate it more effectively. The objective is to develop an intelligent crime prediction model capable of forecasting key factors such as the type of crime, its timing, and its location, thereby contributing to a more informed and proactive approach to urban safety.

Problem Statement:

- How has crime in Chicago changed over the years? Was 2016 really the bloodiest year in two decades?
- Are some types of crimes more likely to happen in specific locations or specific times of the day or specific day of the week than other types of crimes?

Dataset Details:

- The Chicago Crime dataset contains a summary of the reported crimes occurred in the City of Chicago from 2005 onwards.

- Dataset has been obtained from the Chicago Police Department's CLEAR (Citizen Law Enforcement Analysis and Reporting) system.
- Number of Attributes: 23

Attribute Information:

Here are some of the important fields used in the dataset:

- There are columns like ID and Case Number which help us uniquely identify our crime record.
- Columns like Date and Year tell us when this crime happened.
- X/Y Coordinate, Latitude/Longitude, Location tells where exactly the crime happened.
- Beat, District, Ward, Community Area tells us in which area of Chicago this crime happened.

Data source: https://www.kaggle.com/currie32/crimes-in-chicago

Method:

The dataset offers a wealth of detailed information, including the date, time, location, and type of crime, among other attributes. For this analysis, we will focus specifically on the timing and type of crime. Crime types are categorized using standardized Illinois Uniform Crime Reporting (IUCR) codes, where each code corresponds to a specific crime. A comprehensive list of these codes is accessible through the City of Chicago’s website.

To analyze the data effectively, we will employ **K-means clustering**, a method that groups data points by minimizing differences within the same cluster. This approach will help us partition the dataset into meaningful clusters, revealing patterns and relationships that can inform our understanding of crime dynamics.

Ethical Problem:

Predictive policing systems utilize extensive databases of historical crime data to forecast potential crime hotspots. However, the reliability of these forecasts hinges on the impartiality of the underlying data. If the historical arrest records or crime reports are biased, the predictions generated by these systems will also be skewed, perpetuating existing inequities. Biases may arise from incomplete information, lack of context, or systemic issues within law enforcement practices.

Several cities, including Los Angeles, New York, and Chicago, have discontinued the use of predictive policing programs due to ethical concerns. These programs often failed to adhere to principles of justice and fairness, as biased data reinforced systemic racism and disproportionately targeted marginalized communities. This underscores the critical need for transparent, unbiased methodologies in crime prediction to ensure ethical and equitable outcomes.

Challenges/Issues:

Crime poses significant challenges for any society, requiring innovative approaches to effectively address and mitigate its impact. Traditional analytical techniques often fall short when dealing with the vast and complex nature of crime data, which includes temporal, spatial, and categorical elements. To overcome these limitations, advanced deep learning techniques can be employed.

Recurrent Neural Networks (RNNs), particularly Long Short-Term Memory (LSTM) networks, offer a powerful solution. These models excel at processing sequential data, making them ideal for analyzing temporal patterns in crime trends. By leveraging LSTM's ability to retain long-term dependencies, the model can effectively predict future crime patterns, such as the type, time, and location of crimes, based on historical data. This approach enables more accurate forecasting, paving the way for data-driven strategies to reduce and prevent crime proliferation.

References:

1. Currie32. (n.d.). *Crimes in Chicago*. Retrieved from https://www.kaggle.com/currie32/crimes-in-chicago
2. Ogunbode, F. (n.d.). *EDA of Crime in Chicago from 2012-2016*. Retrieved from https://www.kaggle.com/femiogunbode/eda-of-crime-in-chicago-from-2012-2016/discussion
3. Djonafegnem. (n.d.). *Chicago Crime Data Analysis*. Retrieved from https://www.kaggle.com/djonafegnem/chicago-crime-data-analysis
4. Pandas Community. (n.d.). *Pandas Tutorials*. Retrieved from http://pandas.pydata.org/pandas-docs/stable/tutorials.html