

Calgary Crime Data Analysis and Neural Network Prediction

Introduction

The aim of this project is to analyze and predict crime trends in Calgary using the Crime and Disorder Data provided by the City of Calgary's data website. The dataset spans from 2018 to 2024, detailing the number of crimes occurring each month in various communities. By thoroughly analyzing this data, we seek to understand the underlying patterns and build a neural network model to forecast future crime occurrences.

Strategy

The project is divided into several key steps:

- 1. Loading and Understanding the Data:** Importing the dataset and examining its structure to get an initial sense of its content.
- 2. Data Preprocessing:** Cleaning and preparing the data for analysis, including handling missing values and ensuring correct data types.
- 3. Exploratory Data Analysis (EDA):** Conducting detailed analysis to uncover trends and patterns in the data.
- 4. Building a Neural Network Model:** Developing a model, specifically an LSTM (Long Short-Term Memory) neural network, to predict future crime occurrences.
- 5. Optimizing and Training the Model:** Fine-tuning the model to enhance its predictive performance.
- 6. Predicting Future Crimes:** Using the trained model to forecast the number of crimes in future months.

Insights from Exploratory Data Analysis

The exploratory data analysis revealed several key insights about crime patterns in Calgary:

1. Community Distribution:

- Beltline emerged as the most dangerous community, accounting for 11.4% of the top crimes, followed by Forest Lawn (10.7%) and Downtown Commercial Core (10.2%).
- The safest community was 13M, representing 22.7% of the least crimes, followed by 02K and 02B, each with 13.6%.

2. Crime Category Distribution:

- Theft from Vehicle was the most prevalent crime category, constituting 21.7% of all crimes, followed by Theft of Vehicle (16.7%) and Break and Enter - Commercial (13.8%).

3. Crime Trends Over the Years:

- The year 2019 saw the highest number of crime reportings, followed by 2022 and 2018. The data for 2024 was incomplete, with only a few months available.

4. Crime Patterns by Month:

- Monthly crime distribution showed variations, indicating potential seasonal trends.

5. Community and Category Analysis:

- Certain crime categories were more prevalent in specific communities. For example, Forest Lawn had a high incidence of Break & Enter - Other Premises, while Marlborough had low incidences of Commercial Robbery.

6. Year and Category Analysis:

- Detailed analysis showed the distribution of different crime categories over the years, helping to identify any shifts or trends in crime types.

Neural Network Model

To predict future crime occurrences, an LSTM (Long Short-Term Memory) neural network was employed. LSTM networks are particularly suited for time series prediction tasks due to their ability to capture long-term dependencies.

Model Building and Training

1. Sequence Preparation: Creating sequences from the data to be used as inputs for the LSTM model.

2. Train-Test Split: Dividing the data into training, validation, and test sets to evaluate the model's performance.

- 3. Model Architecture:** Constructing an LSTM model with 50 units, followed by a dropout layer to prevent overfitting, and a dense layer for output.
- 4. Model Compilation:** Using the Adam optimizer and Mean Squared Error (MSE) as the loss function.
- 5. Training:** Training the model over 100 epochs with validation to monitor and improve performance.

Performance

- The model's loss decreased significantly over the training epochs, indicating improved predictive accuracy. The validation loss provided an estimate of the model's performance on unseen data.

Conclusion

This project successfully utilized a comprehensive strategy involving data preprocessing, exploratory analysis, and neural network modeling to predict crime trends in Calgary. The insights gained from the analysis can inform public safety strategies and resource allocation, while the predictive model offers a tool for anticipating future crime occurrences, aiding in proactive crime prevention measures.