NYC Crime & Route Prediction Application

Introduction

The NYC Risk & Route Prediction Application is a web-based tool built using Streamlit to enhance public safety by predicting crime risks and analyzing route safety in New York City. This application leverages machine learning models, geospatial data, and real-time APIs to provide users with actionable insights about crime risks for specific locations or routes. By integrating interactive maps and predictive algorithms, the platform serves both residents and visitors who need to make informed decisions about their travel and safety in NYC.

Features

1. Crime Prediction

The Crime Prediction feature of the application predicts the likelihood of crime occurring at a specific location in New York City, based on historical crime data and machine learning models. This feature empowers users to assess potential risks before heading to a particular area. Here's a breakdown of how it works:

a. Predicts Crime Risk Percentages

The system predicts the probability of a crime occurring at a given location based on the data.

Location Coordinates: The latitude and longitude of the user-selected or inputted address.

b. Predicts Likely Crime Types

In addition to the overall crime risk percentage, the system predicts the types of crime that are most likely to occur in that area. The crime types are drawn from a list of predefined categories such as: Burglary, Assault, Theft, Vandalism, Robbery

c. User Input Methods

The application supports two methods for users to input the location for crime risk prediction:

Text Field (Address Input): Users can type an address into a text field (e.g., "123 Main St, NYC") to get the corresponding crime risk information. The application then uses geocoding (via APIs like GoMaps or geopy) to convert the address into geographic coordinates (latitude and longitude), which are then processed by the crime prediction models.

Interactive Map: Users can click on an interactive map to select a location. This method allows users to visually select areas they are interested in. Once the user clicks on a location on the map, the application retrieves the coordinates of the selected point and uses them to predict the crime risk and likely crime types for that area.

Example Flow:

A user enters an address ("Downtown Manhattan") or clicks on the map to select a point.

The system converts the address or location into coordinates.

The crime risk prediction model processes these coordinates, and the system outputs:

Crime risk percentage (e.g.,40%) and Top three likely crime types (e.g., Burglary 35%, Robbery 25%, Assault 15%).

2. Route Risk Prediction

The Route Risk Prediction feature is designed to evaluate the safety of different routes between a source and a destination within New York City. It helps users assess which route is safest and provides key details to support safer decision-making during travel. The feature works as follows:

a. Evaluates the Safety of Different Routes

When users input both a source and destination, the application uses geocoding and routing algorithms to calculate multiple possible routes between the two points. Each route is evaluated for its safety based on risk scores for different geographic zones the route traverses.

b. Provides Detailed Breakdown of Route Risk Scores, Total Distance, and Travel Duration

For each route, the application displays the following details:

Risk Score: An overall safety score for the route based on the crime risk in the zones the route traverses.

Total Distance: The total length of the route, measured in miles or kilometers.

Travel Duration: The estimated time required to travel the route, based on current or average traffic conditions.

c. Interactive Selection of Source and Destination

The route risk prediction system allows users to choose the source and destination locations using two methods:

Text Field (Address Input): Users can type in the source and destination addresses (e.g., "Central Park, NYC" and "Times Square, NYC"). The system will convert these addresses into geographic coordinates and find the safest routes between them.

Interactive Map: Users can click on the map to select both the starting point and the destination. This method allows users to directly visualize the route and the locations they want to analyze. Upon clicking the source and destination, the map dynamically displays the risk scores for the various route options, highlighting the safest routes.

Example Flow:

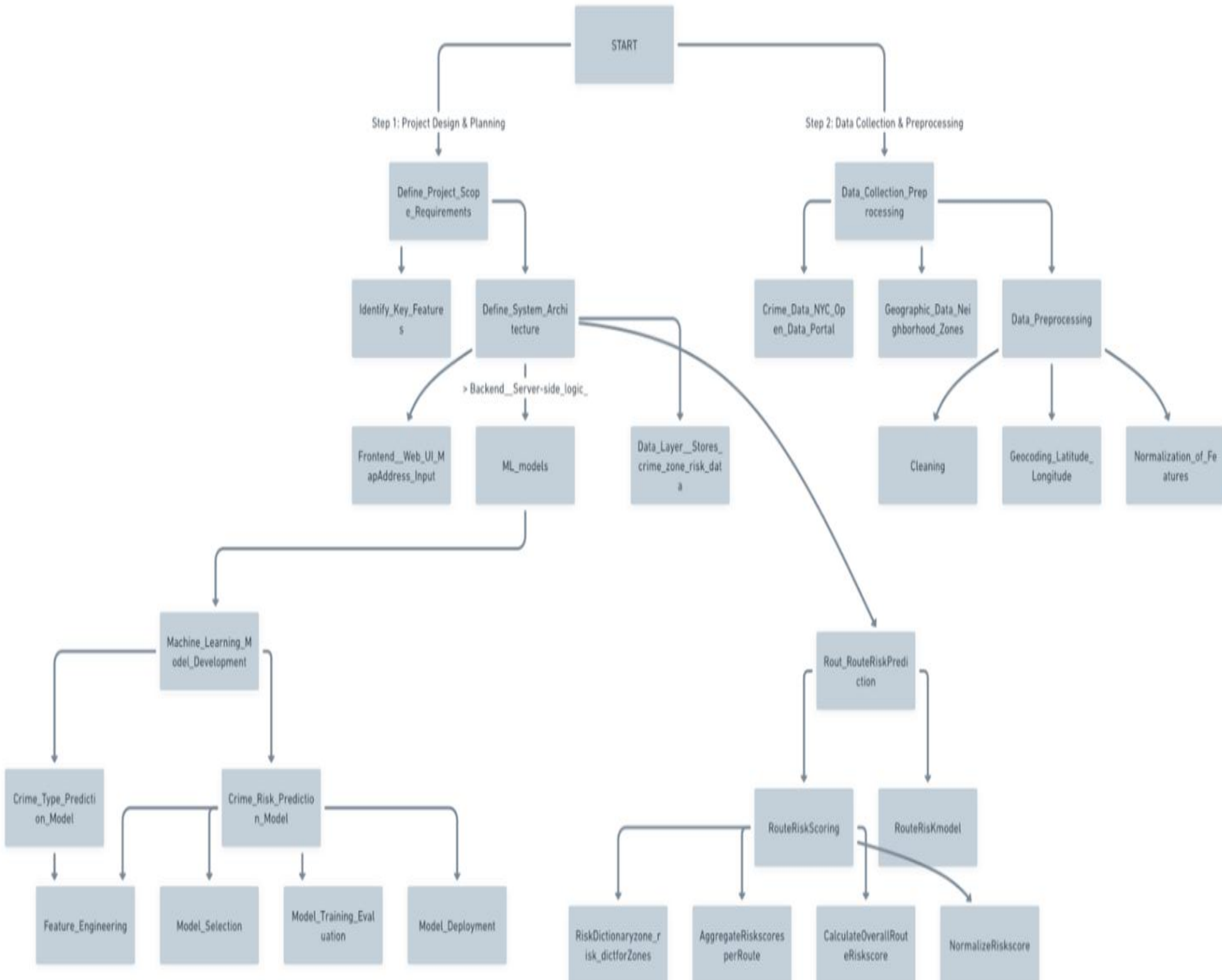
A user enters a source address and a destination address or clicks on the map to select a point.

The system converts the source address and destination address into coordinates.

The route prediction model processes these coordinates, and the system outputs:

The risk score associated with the path, duration of the path along with the distance and walking directions.

Methodology



1. Project Design and Planning

The first step in developing the NYC Crime Detection System was to define the overall project scope and design. This included understanding the requirements of the system, identifying the key features, and laying out the architecture of the application.

The architecture of the system is divided into several key components:

Frontend: A web-based user interface that allows users to input addresses or interact with an interactive map to request crime predictions and route safety assessments.

Backend: The server-side logic that processes user requests, interacts with external APIs (such as geocoding and routing APIs), and runs machine learning models for crime risk predictions.

Data Layer: The data repository that stores historical crime data, zone risk dictionaries, and other relevant datasets used for predictions.

External APIs:

Geocoding API: Used to convert addresses into geographic coordinates.

Routing API: Used to calculate travel routes between source and destination.

Machine Learning Models: Trained models for predicting crime risk percentages and crime types.

2. Data Collection and Preprocessing

Historical Crime Data: Crime Data from NYC Open Data Portal: A publicly available dataset that includes information about reported crimes in New York City, such as crime type, date, time, and location.

Data Preprocessing:

Cleaning: Raw crime data was cleaned by removing duplicate entries, correcting erroneous values, and standardizing format (e.g., converting date-time fields to a consistent format).

Geocoding: The crime data was enriched with geographic coordinates (latitude and longitude) to ensure compatibility with the location-based prediction model.

Normalization: Features such as crime rates in different zones were normalized to ensure consistent scale for machine learning models.

3. Machine Learning Model Development

Crime Risk Prediction Model: A regression model was developed to predict the probability of a crime occurring at a specific location. The process involved:

Feature Engineering:

Crime Type Frequency: The frequency of different crime types in the area.

Time of Day: Hourly crime data to capture temporal patterns.

Neighborhood Socioeconomic Data: Demographic and socioeconomic data to provide context to crime predictions.

Model Selection:

Various machine learning algorithms were considered for the crime risk prediction model, including:

Logistic Regression: For binary classification (crime vs. no crime).

Random Forest/Gradient Boosting: For multi-class classification and handling complex relationships in the data.

After testing different models, Random Forest Classifier was chosen for its high performance in classification tasks and ability to handle feature interactions.

Model Training and Evaluation:

The data was split into training and testing sets.

Hyperparameter tuning was done to optimize the model's performance.

Cross-validation was used to assess model robustness.

The model was evaluated using metrics such as accuracy, precision, recall, and F1-score.

Crime Type Prediction Model: A multi-class classification model was developed to predict the type of crime most likely to occur at a given location. This model was trained using the historical crime data with labels indicating the specific crime types (e.g., robbery, burglary, assault).

Feature Selection:

The features used for this model were similar to those used in the crime risk model, but with a focus on the type of crime in the location.

Model Selection:

A multinomial logistic regression or Naive Bayes classifier was chosen due to its effectiveness in multi-class problems.

Training and Evaluation:

Similar to the crime risk model, the crime type model was trained, tuned, and evaluated using appropriate classification metrics.

Label Encoding:

Since the crime types were categorical, label encoding was applied to convert categorical labels (e.g., "robbery," "burglary") into numerical values for easier processing by machine learning algorithms.

4. Route Risk Prediction

The Google Maps Directions API was integrated to retrieve alternative routes between the source and destination points. The API returns routes along with detailed information such as distance, travel time, and step-by-step directions.

Route Risk Scoring:

Risk Dictionary (zone_risk_dict): A predefined risk dictionary was created for different geographic zones in NYC, based on historical crime data. Each zone (neighborhood or district) was assigned a risk score based on factors such as crime rates and demographics.

Risk Score Calculation:

For each route, the geographic path was divided into smaller segments (zones).

The risk score for each and aggregated to calculate the total risk score for the route.

The final risk score was normalized to ensure consistency across different routes.

Route Risk Model:

A scoring algorithm was developed to evaluate the risk of each route, considering the individual risk scores of the zones the route passes through.

The algorithm computes the overall route risk by aggregating the zone scores, using a weighted average statistical method.

System Architecture

1. Frontend:

The frontend of the system is built using Streamlit, which is a powerful framework for creating interactive web applications with Python. This enables users to interact with the system, providing inputs and viewing the results.

Interactive Maps:

Streamlit integrates with mapping libraries like folium to display dynamic maps on the webpage.

Users can view the geographical context of their selected location or route with interactive zooming and panning functionality.

Crime Predictions and Route Details:

After entering a location, the system provides two types of predictions:

Crime Risk Prediction: A prediction of the likelihood of a crime occurring at the selected location.

Route Risk Prediction: Details of a travel route's safety, including a risk score, estimated distance, and travel time.

2. Backend:

The backend handles the server-side logic, such as processing user input, making API requests, and running machine learning models to generate predictions. It interacts with various services, such as geocoding APIs and machine learning models, and stores relevant data.

Machine Learning Models:

The system uses pre-trained machine learning models stored in joblib files to make predictions:

Crime Risk Prediction Model: Predicts the probability of crime occurring at a given location.

Crime Type Classification Model: Predicts the types of crimes most likely to occur at a specific location.

Google Maps API: Provides services for geocoding (converting addresses to coordinates) and route calculations (determining routes between two points).

3. Visualization:

Streamlit Integration:

Streamlit makes it easy to embed these interactive folium maps directly into the web application, giving users a seamless experience while interacting with both data and visual elements.

Tableau:

By using Tableau using the NYC Crime data, we used visually appealing analytics. Exploring complex crime and route data through dynamic heatmaps, charts, and time-series graphs.

Key Algorithms and Functions

1. **Crime Prediction**:

    get_lat_lng: Geocodes an address into latitude and longitude using GoMaps API.

    label_encoder_crime.inverse_transform(): Converts predicted numerical labels to crime type names.

2. **Route Risk Prediction**:

    get_all_routes_with_coordinates: Fetches alternative routes and details (distance, duration, steps).

    find_zone_id: Maps coordinates to geographic zone IDs.

    identify_routes_risk_score: Calculates the overall risk score for each route.

3. **Mapping and Visualization**:

    folium.Map(): Creates an interactive map.

    st_folium: Integrates Folium maps into the Streamlit application.

    folium.Marker: Adds markers to the map for selected locations.

Challenges and Solutions

1. **API Limitations**:

    Challenge: Dependency on the GoMaps API for geocoding and routing, with potential rate limits or downtime.

    Solution: Implement caching for geocoding results to minimize redundant API calls.

2. **Real-Time Processing**:

    Challenge: Latency in fetching routes and making predictions.

    Solution: Use Streamlit caching to optimize model and data loading.

3. **Data Granularity**:

    Challenge: Sparse data for certain zones affects prediction accuracy.

    Solution: Use interpolation or clustering techniques to improve predictions in sparse areas.

Results

1. **Crime Prediction**:

    Successfully predicts crime risk percentages for selected locations.

    Provides accurate predictions for top three likely crime types.

2. **Route Risk Prediction**:

    Calculates risk scores for multiple routes.

    Displays step-by-step directions and route-specific details for user evaluation.

Future Improvements

1. **Real-Time Data Integration with Google Maps**:

    Incorporate real-time crime reports to improve prediction accuracy and relevance within google maps when travelling.

2. **Mobile Application**:

    Develop a mobile version of the platform for easier access on-the-go.

3. **Dynamic Risk Updates**:

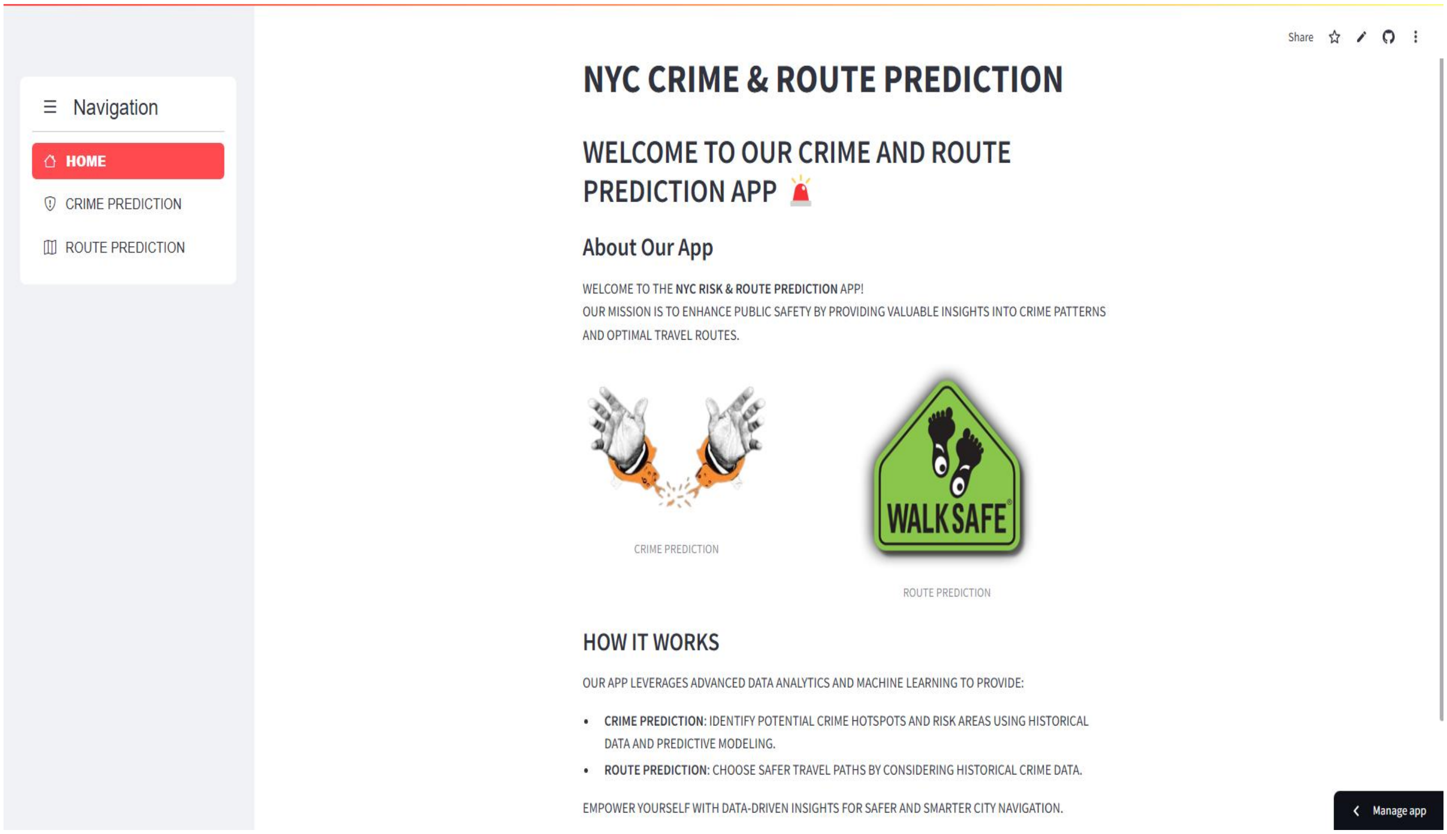
    Integrate dynamic risk updates based on time of day, season, and events.

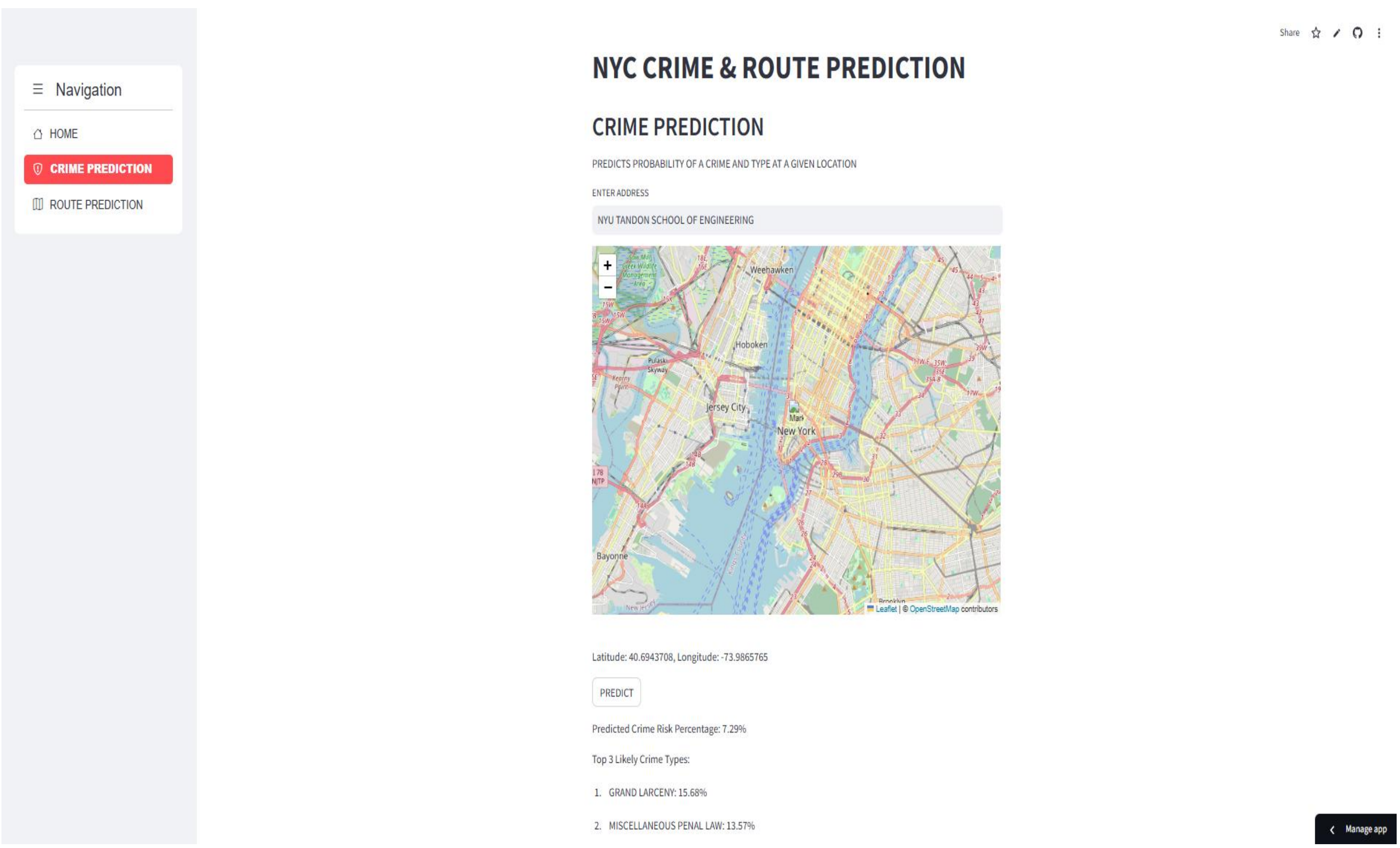
Conclusion

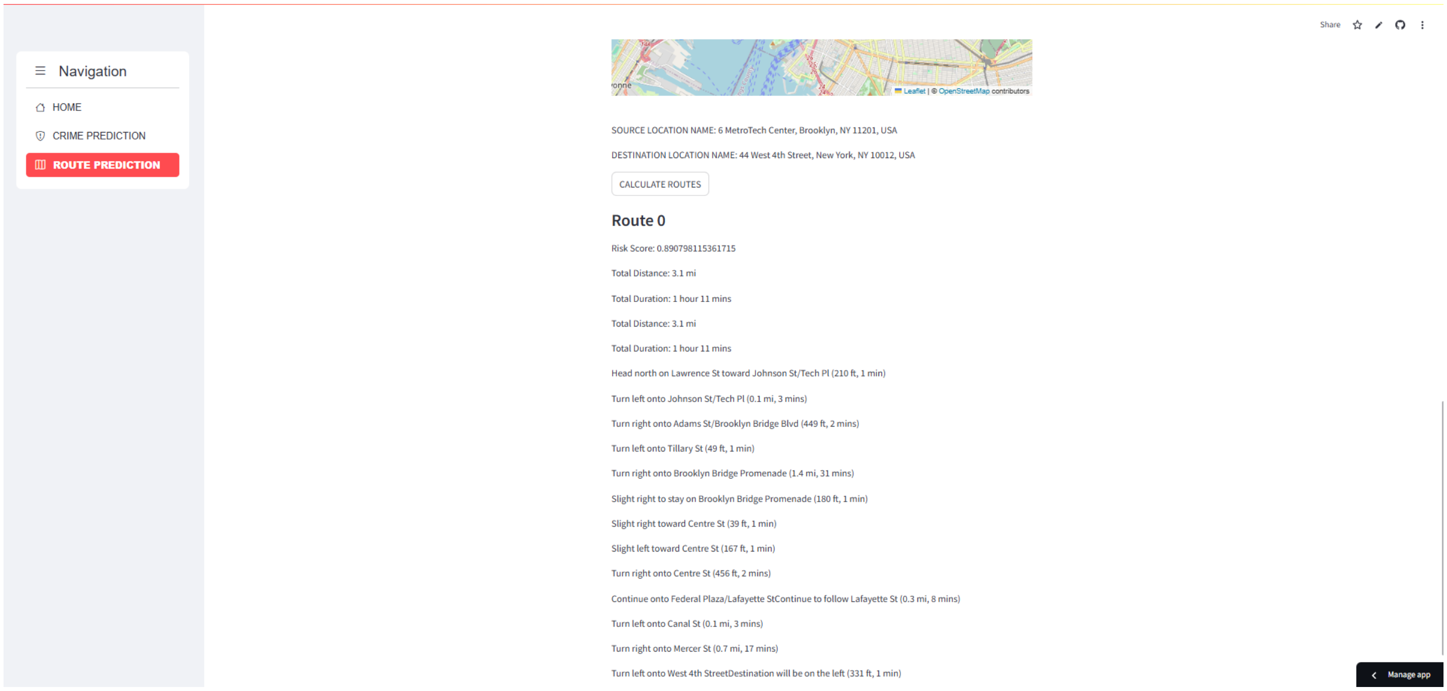
The NYC Risk & Route Prediction Application is a robust tool for enhancing urban safety. By combining predictive analytics with intuitive visualization, the platform empowers users to make safer decisions in navigating New York City. With future enhancements, it has the potential to serve as a vital resource for residents, tourists, and law enforcement agencies.

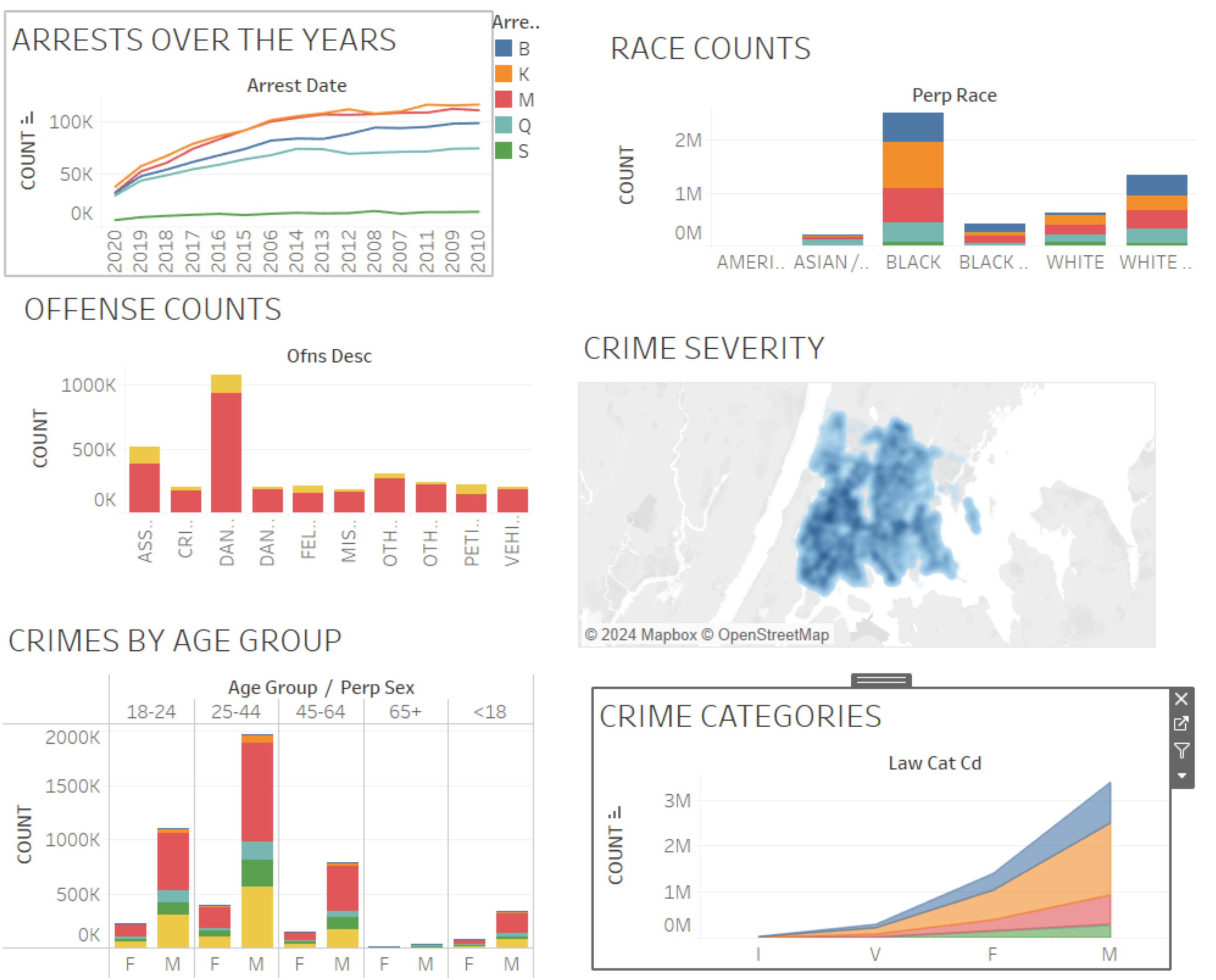
References

1. **NYC Open Data Portal (Crime Data)**: https://opendata.cityofnewyork.us/
2. **NYC Geographic Information Systems (GIS) Data**: https://www1.nyc.gov/site/planning/data-maps/open-data.page
3. **Scikit-learn Documentation**: https://scikit-learn.org/
4. **Random Forest Classification (Breiman, 2001)**: https://doi.org/10.1023/A:1010933404324
5. **Logistic Regression for Crime Type Prediction**:https://www.wiley.com/en-us/Applied+Logistic+Regression%2C+3rd+Edition-p-9780470582473
6. **GoMaps API**: https://gomaps.com/api
7. **Geopy Documentation**: https://geopy.readthedocs.io/


Share
Navigation
HOME
CRIME PREDICTION
ROUTE PREDICTION
NYC CRIME & ROUTE PREDICTION
WELCOME TO OUR CRIME AND ROUTE PREDICTION APP
About Our App
WELCOME TO THE NYC RISK & ROUTE PREDICTION APP!
OUR MISSION IS TO ENHANCE PUBLIC SAFETY BY PROVIDING VALUABLE INSIGHTS INTO CRIME PATTERNS AND OPTIMAL TRAVEL ROUTES.
CRIME PREDICTION
WALK SAFE
ROUTE PREDICTION
HOW IT WORKS
OUR APP LEVERAGES ADVANCED DATA ANALYTICS AND MACHINE LEARNING TO PROVIDE:
• CRIME PREDICTION: IDENTIFY POTENTIAL CRIME HOTSPOTS AND RISK AREAS USING HISTORICAL DATA AND PREDICTIVE MODELING.
• ROUTE PREDICTION: CHOOSE SAFER TRAVEL PATHS BY CONSIDERING HISTORICAL CRIME DATA.
EMPOWER YOURSELF WITH DATA-DRIVEN INSIGHTS FOR SAFER AND SMARTER CITY NAVIGATION.
Manage app


Share
Navigation
HOME
CRIME PREDICTION
ROUTE PREDICTION
NYC CRIME & ROUTE PREDICTION
CRIME PREDICTION
PREDICTS PROBABILITY OF A CRIME AND TYPE AT A GIVEN LOCATION
ENTER ADDRESS
NYU TANDON SCHOOL OF ENGINEERING
Weehawken
Hoboken
Jersey City
New York
Bayonne
Leaflet | © OpenStreetMap contributors
Latitude: 40.6943708, Longitude: -73.9865765
PREDICT
Predicted Crime Risk Percentage: 7.29%
Top 3 Likely Crime Types:
1. GRAND LARCENY: 15.68%
2. MISCELLANEOUS PENAL LAW: 13.57%
Manage app


Navigation
HOME
CRIME PREDICTION
ROUTE PREDICTION
Share
Leaflet | © OpenStreetMap contributors
SOURCE LOCATION NAME: 6 MetroTech Center, Brooklyn, NY 11201, USA
DESTINATION LOCATION NAME: 44 West 4th Street, New York, NY 10012, USA
CALCULATE ROUTES
Route 0
Risk Score: 0.890798115361715
Total Distance: 3.1 mi
Total Duration: 1 hour 11 mins
Total Distance: 3.1 mi
Total Duration: 1 hour 11 mins
Head north on Lawrence St toward Johnson St/Tech Pl (210 ft, 1 min)
Turn left onto Johnson St/Tech Pl (0.1 mi, 3 mins)
Turn right onto Adams St/Brooklyn Bridge Blvd (449 ft, 2 mins)
Turn left onto Tillary St (49 ft, 1 min)
Turn right onto Brooklyn Bridge Promenade (1.4 mi, 31 mins)
Slight right to stay on Brooklyn Bridge Promenade (180 ft, 1 min)
Slight right toward Centre St (39 ft, 1 min)
Slight left toward Centre St (167 ft, 1 min)
Turn right onto Centre St (456 ft, 2 mins)
Continue onto Federal Plaza/Lafayette StContinue to follow Lafayette St (0.3 mi, 8 mins)
Turn left onto Canal St (0.1 mi, 3 mins)
Turn right onto Mercer St (0.7 mi, 17 mins)
Turn left onto West 4th StreetDestination will be on the left (331 ft, 1 min)
Manage app


ARRESTS OVER THE YEARS
Arrest Date
COUNT
100K
50K
0K
2020 2019 2018 2017 2016 2015 2006 2014 2013 2012 2008 2007 2011 2009 2010
Arre..
B
K
M
Q
S
RACE COUNTS
Perp Race
COUNT
2M
1M
0M
AMERI.. ASIAN /.. BLACK BLACK .. WHITE WHITE ..
OFFENSE COUNTS
Ofns Desc
COUNT
1000K
500K
0K
ASS.. CRI.. DAN.. DAN.. FEL.. MIS.. OTH.. OTH.. PETI.. VEHI..
CRIME SEVERITY
© 2024 Mapbox © OpenStreetMap
CRIMES BY AGE GROUP
Age Group / Perp Sex
18-24 25-44 45-64 65+ <18
COUNT
2000K
1500K
1000K
500K
0K
F M F M F M F M F M
CRIME CATEGORIES
Law Cat Cd
COUNT
3M
2M
1M
0M
I V F M