RouteSavvy

CS-GY 6513 Big Data Project Proposal

Team Members:

- Krapa Karthik [**NET ID**: kk5754 **NYU ID**: N12039854]
- Shreyansh Bhardwaj [NET ID: sb10261 NYU ID: N17664537]
- Sourik Dutta [NET ID: sd5913 NYU ID: N19304628]
- Dev Thakkar [NET ID : djt8795 NYU ID : N19070379

Introduction

Urban areas, especially metropolises like New York City, face persistent challenges related to traffic congestion and public transportation inefficiencies. Commuters waste countless hours annually due to unpredictable delays, overcrowded transit options, and suboptimal route choices. The increasing availability of real-time data from various sources presents an unprecedented opportunity to enhance commuter experiences and optimize urban mobility.

The Urban Mobility Optimizer aims to revolutionize how people navigate through dense urban environments by leveraging big data technologies to process and analyze real-time transportation data. By integrating traffic patterns, public transit schedules, crowd density metrics, and historical trends, our system will provide personalized route recommendations that adapt dynamically to changing conditions, ultimately saving commuters time and reducing frustration.

Unlike existing navigation applications that rely primarily on static maps and limited real-time updates, our solution will incorporate comprehensive data streams from multiple sources to create a holistic view of the urban transportation landscape, enabling truly intelligent routing decisions.

Objective

The primary objective of this project is to build a scalable and effective Urban Mobility Optimizer that generates personalized route recommendations for commuters based on real-time traffic conditions, public transit schedules, and crowd densities. By using big data processing and real-time analytics, this system will help commuters reduce travel times, avoid congested areas, and make more informed transportation choices.

Methodology

1. Data Collection and Ingestion

- Data Sources: MTA subway data, NYC Department of Transportation traffic information, Citibike availability, weather APIs, and social media reports

- Data Streaming: Apache Kafka will be used to handle continuous streams of data from various transportation sources

2. Data Processing

- Apache Spark Structured Streaming will aggregate and analyze real-time data from Kafka, transforming it into actionable insights for route optimization.

3. Machine Learning / Analytics

- Predictive models will be developed to forecast traffic congestion, transit delays, and crowd density.

- Route optimization algorithms will consider multiple factors beyond distance to recommend optimal paths.

4. User Interface

- A mobile application will provide users with real-time route recommendations, updates on delays, and alternative suggestions.

- Interactive maps will visualize routes, congestion areas, and transit options

5. Evaluation

- The system's performance will be evaluated based on response time, prediction accuracy, and user satisfaction to ensure optimal commuter experience.

Expected Results

- A scalable and efficient system capable of processing and analyzing real-time urban mobility data.

- Accurate and timely route recommendations that adapt to current traffic and transit conditions.
- Enhanced commuter experience through reduced travel times and improved reliability.

Technologies and Tools

- Python
- Apache Kafka
- PySpark
- [cloud platforms, e.g., Azure, AWS]
- [Additional libraries or frameworks]

Conclusion

By harnessing the power of big data technologies, this project endeavors to revolutionize urban commuting through real-time, data-driven route optimization. The integration of diverse data sources, advanced processing frameworks, and predictive analytics will culminate in a solution that enhances the efficiency and reliability of urban transportation systems..