Conestoga College

School of Applied Computer Science & Information Technology

SENG8080 - Case Studies Big Data

Traffic Prediction Using Machine Learning

Aparna Santhosh - 8899699

Astle Russel Fernandez - 8894387

Sreelekshmi Pulickichiraveli - 8901347

October 21, 2023

**Abstract**

Efficient traffic management is essential to improving daily mobility and overall quality of life in cities like Kitchener. This case study focuses on the application of machine learning techniques to predict and manage traffic conditions in Kitchener, using data from Bing Maps Traffic Incident Data API with the aim of providing a more data-driven approach to traffic forecasting and event planning has been achieved.

The main objective of this case study is to develop a machine learning algorithm that can predict traffic conditions and events in Kitchener. This predictive model will help city officials, commuters, and transit services to anticipate and proactively manage traffic issues.

Methodology:

1. Data Collection: The analysis begins by collecting traffic data from the Bing Maps Traffic Incident Data API. This data set includes real-time information about accidents, road closures and construction incidents.

2. Data Preprocessing: The raw data will be processed first for cleaning, processing, and optimization. Attributes such as event type, point, and historical traffic data will be extracted.

3. Machine Learning Models: Various machine learning algorithms, including LSTM (long short-term memory) model and ARIMA model will be used to predict traffic conditions. Historical data will be used to train and validate these models.

4. Real-time Data Integration: For real-time forecasting, the model will be integrated with existing traffic incident data API from Bing map. This allows them to continually update and see events in real time.

5. Visualization: Predicted traffic conditions will be visualized on maps. Dashboards will provide insights into traffic patterns.

6. Evaluation: The model's accuracy, precision, recall, and F1-score will be evaluated, and its performance will be assessed in real-world scenarios.

Our model will enable proactive traffic management by reducing congestion and response times. Also, the City authorities can efficiently allocate resources and responders to incidents as they occur. The data and predictive analytics approach contributes to more efficient and informed traffic management in the city.

Contents

[Introduction 4](#_Toc147960212)

[Data Research and Integration 4](#_Toc147960213)

[Data Collection 4](#_Toc147960214)

[Data Storage and Maintenance 4](#_Toc147960215)

[Data Quality 4](#_Toc147960216)

[Data Analysis and Visualization 5](#_Toc147960217)

[Extrension 5](#_Toc147960218)

[References 6](#_Toc147960219)

# Introduction

In cities like Kitchener, improving everyday mobility and overall quality of life is based on efficient traffic management. To provide a more data-driven approach to traffic forecasting and event planning, this case study focuses on the application of machine learning techniques to predict and manage traffic conditions in Kitchener. Data from Bing Maps Traffic Incident Data API is used.

The primary goal of this case study is to create a machine learning algorithm that can forecast Kitchener traffic patterns and events. To anticipate and proactively manage traffic concerns, our predictive model will be helpful to commuters, city officials, and transportation networks.

# Data Research and Integration

Our primary data source we are developing to provide advanced traffic forecasting for Kitchener is the Bing Maps Traffic Incident Data API, which provides real-time traffic incident data such as accidents, roadblocks, and construction and the secondary data source is the traffic volume data from the Kitchener GeoHub. We capture, pre-process, and incorporate this data into our machine learning model. We will combine both data sources based on the location. Real-time information ensures that the updated forecasts and event management are current.

Real-time data updates are important to our approach, as we seek to provide instant incident detection and continuously updated traffic predictions. Data is visualized on maps, enabling easy comprehension of traffic conditions and incident locations.

# Data Collection

Data collection is an important ongoing process in our traffic prediction project for Kitchener, mainly driven by the real-time data source, the Bing Maps Traffic Incident Data API. This API is our resource for the most recent information on traffic incidents, covering a wide range of incident patterns such as accidents, road closures, and construction work. We ensure that we always have the most recent information on traffic incidents in Kitchener by retrieving the data via frequent HTTP requests.

We have used many methods and tools to effectively utilize this data source. The Requests library to enable seamless API interactions, the Pandas library for data preprocessing, and Python for scripting HTTP requests and data retrieval are the primary tools in our data collection process.

# Data Storage and Maintenance

We have chosen MongoDB as our primary data storage solution. It is a NoSQL database that is particularly good at managing massive amounts of real-time data and it is a perfect fit for the constant stream of traffic incident data that comes in from the Bing Maps Traffic Incident Data API.

Data durability, scalability, and high availability are guaranteed using a dedicated MongoDB server(localhost) for data storage. With this, we can effectively store data for the project's present requirements while also preparing for future expansion and growing data needs. MongoDB also has the capability to seamlessly manage semi-structured and unstructured data, such as traffic incident records, with ease. It offers excellent performance for read and write operations, making it well-suited for real-time data integration and retrieval.

# Data Quality

In our traffic prediction project for Kitchener, it is critical to ensure high-quality data, especially when working with real-time data via the Bing Maps Traffic Incident Data API and storing it in MongoDB.

Data quality begins with the preprocessing phase. We thoroughly clean the raw data to handle anomalies, missing values, and inconsistencies. This ensures that the data entered to our MongoDB database is accurate and reliable. Any irregularities are addressed in this phase to enhance data quality.

# Data Analysis and Visualization

We will use dashboards to provide detailed information on traffic patterns and will use map-based visualization to present expected traffic conditions. This approach will improve traffic management abilities by providing stakeholders with data-driven decision-making tools.

# Extension

A careful method is needed for scaling out Kitchener's traffic prediction project, which is driven by data from the Bing Maps Traffic Incident Data API and MongoDB storage. It is essential to evaluate the volume and complexity of the data to fulfil current storage requirements. This stage makes sure that past datasets are kept, and that the storage solution of choice can manage the data streams from the API without any issues.

A scalable storage solution that can support expansion should be incorporated into the project as it moves forward to respond to the growing volumes of data and changing requirements. Being adaptable is essential since this method enables the increase of storage capacity as the project grows. Additionally, by keeping access to more recent, less often visited data while preserving older, less frequently accessed data, data archiving techniques help reduce storage expenses.

# Proposed Allocation Project Team Roles

Astle Russel Fernandez – Astle's primary responsibility is to lead the design and visualization aspects of the project. This includes creating user-friendly visualizations of traffic predictions, ensuring data is presented effectively, and the design of dashboards for insights. Additionally, Astle will be involved in data analyzation, assessing the results of predictive models.

Sreelekshmi Pulickichiraveli – Sreelekshmi's primary role is to manage the project's written documentation, including reports, project plans, and progress updates. She will also take charge of data preprocessing, preparing data for analysis, and ensuring data quality. Sreelekshmi will oversee data storage and maintenance, managing databases and data retention policies.

Aparna Santhosh – Aparna will be responsible for project write-ups, ensuring that the project's documentation is comprehensive and well-structured. She will also focus on data research and integration, identifying valuable data sources and incorporating them into the project. Aparna will collaborate on data preprocessing to prepare the data for analysis.

We are holding regular team meetings to discuss progress, challenges, and next steps. These meetings will take place according to project milestones and can be conducted virtually to accommodate team members' schedules.

# Project Timeline

|  |  |  |
| --- | --- | --- |
| **Date** | **Deliverable** | **Responsible** |
| Oct 24 | Data Collected and planned | Aparna Santhosh |
| Nov 10 | 1st Draft Circulated to Team | Aprana Santhosh |
| Nov 15 | 1st Draft of Presentation Circulated | Aparna Santhosh, Sreelekshmi Pulickichiraveli |
| Nov 23 | User testing by the team and errors/refinements identified. | Astle Russel Fernandez |
| Nov 26 | Final Adjustments made and checked | Astle Russel Fernandez |
| Nov 28 | Process and Report Due at 10pm | Aparna Santhosh, Sreelekshmi Pulickichiraveli |

# 

# References

1. Flavius-Stoichescu. (2023, September 28). Traffic incident data - Bing Maps. Microsoft Learn. https://learn.microsoft.com/en-us/bingmaps/rest-services/traffic/traffic-incident-data

2. Rbrundritt. (2022, June 8). Getting a Bing Maps key - Bing Maps. Microsoft Learn. https://learn.microsoft.com/en-us/bingmaps/getting-started/bing-maps-dev-center-help/getting-a-bing-maps-key

3. Anunaya, S. (2023). Data Preprocessing in Data Mining – A Hands On Guide (Updated 2023). Analytics Vidhya. https://www.analyticsvidhya.com/blog/2021/08/data-preprocessing-in-data-mining-a-hands-on-guide/#:~:text=Data%20preprocessing%20is%20the%20process,learning%20or%20data%20mining%20algorithms.

4. Brush, K., & Burns, E. (2022). data visualization. Business Analytics. https://www.techtarget.com/searchbusinessanalytics/definition/data-visualization