COMP 3610: Big Data Analytics
Project Proposal Guidelines

1   Project Proposal Requirements

Each group must develop a project proposal that clearly outlines their intended research and development plan. The proposal should contain the following sections:

- **Problem Statement:** Clearly define the problem your group intends to solve.

- **Justification:** Explain why solving this problem is important or valuable.

- **Preliminary Approach:** Provide initial thoughts on how you plan to solve the problem, including potential algorithms, tools, or methods.

- **Data Considerations:** Indicate the dataset(s) you plan to use or describe how you intend to collect the necessary data.

- **Group Members:** List all group members along with their ID numbers.

The proposal must be one page in length and submitted in PDF format.

Project Complexity: The project must be appropriately challenging as it accounts for 50% of the course grade. However, it should be feasible to complete within the semester. The lecturer will provide guidance and adjustments to proposals to ensure they are suitable. If a group is unable to come up with a viable project idea, a project will be assigned to them.

2 Sample Proposal

Title: Predicting Traffic Congestion Using Machine Learning

Group Members:
John Doe (123456)
Jane Smith (654321)
Alice Brown (789012)
Bob White (210987)

Problem Statement:
Traffic congestion is a persistent issue in urban areas, leading to wasted time, increased pollution, and economic losses. Our project aims to develop a machine learning model to predict traffic congestion levels based on historical data and real-time inputs.

Justification:
Traffic congestion affects millions of people daily, increasing commute times and fuel consumption. By providing accurate congestion forecasts, city planners can implement traffic control measures and commuters can optimize their travel schedules, reducing overall congestion.

Preliminary Approach:
Our approach involves collecting and preprocessing historical traffic data, real-time GPS tracking, and weather conditions. We will apply feature selection techniques to determine relevant factors influencing congestion. We plan to implement machine learning models such as random forests and deep learning techniques like LSTMs to enhance predictive accuracy. Model evaluation will be performed using metrics such as RMSE and mean absolute error.

Data Considerations:
We will use publicly available traffic datasets from the city's transportation department, incorporating live traffic feeds through APIs like Google Maps or OpenStreetMap. Additional data, such as weather conditions, will be sourced from NOAA. If real-world data gaps exist, we will generate synthetic traffic data for model training and validation.

Expected Outcomes:

- A predictive model capable of forecasting short-term congestion levels.

- An interactive web dashboard displaying congestion predictions and insights.

- A comprehensive report detailing methodology, data processing steps, and findings.

- Recommendations for optimizing urban traffic flow based on predictive insights.