Project Proposal

COMP 8118
Data Mining
Fall 2021

September 2021

Navid Mohammad Imran
U00761100
nimran@memphis.edu

Hosneara Ahmed
U00744746
hahmed1@memphis.edu

Should be submitted and approved by September 10, 2021

What is the problem/application?

Traffic volume prediction

What methods will be tested or implemented?

- Trend analysis (trend in volume from history will affect the prediction)
- Regression (how much traffic flows in a certain location at a given time)

What data sets will be used?

- Minnesota Interstate Dataset: It is collected in Minneapolis-St.Paul in Minnesota. This dataset is geared towards predicting the traffic volume depending on weather and other environmental factors. To be more specific, it is a multivariate time series dataset which contains 48204 instances of hourly interstate traffic volume along with weather and holiday features from 2012-2018.

What are the potential challenges for implementation?

We do not have much experience in working with big datasets so there may be a learning curve for us to finish everything within the deadline. Also, our hardware may not have enough memory to train the model efficiently.

What are the expected deliverables?

Initial Observation

- Evaluation of the dataset
- Related works and methods

Description of implementation

- Data mining techniques used
- Discussion of algorithms used to train models

Result and Analysis

- Knowledge discovery
- Time and space complexity analysis
- Prediction accuracy

Conclusion

- Possibility of further study

What are the responsibilities of each team member?

- Navid will focus on data cleaning, transformation, and mining
- Hosneara will concentrate on training models for prediction and evaluation