DS 5110: Big Data Systems | Final Project

State of Virginia Traffic Reliability – MAP21

Christian Schroeder (dbn5eu), Timothy Tyree (twt6xy), Colin Warner (ynq9ya)

1. **Abstract**
2. **Introduction**

*Primary Objective:* Use actual Virginia highway traffic data from 2017-2020 to accurately predict the reliability of the state’s traffic projections. If that model is found, we can use the state’s forecasted metrics through 2024 to classify *future* unreliable highway segments.

*Context:* In 2012, President Obama signed into law the Moving Ahead for Progress in the 21st Century Act (MAP-21). Among other initiatives, this act transforms the process used for allocating funds towards the improvement of highway, transit, bike, and pedestrian programs - allowing a programmatic framework to inform whether or not a road is in need of transformation.

As part of an ongoing project at the Virginia Department of Transportation (VDOT), our team has been asked to explore more advanced classification models to predict if a MAP-21 reporting segment is reliable.

1. **Overview of Process**

Below we provide a quick description of each stage in our process. Please see specified sections in parentheses for a more thorough description of each stage.

*Data Import and Preprocessing (section ii.a.):* Import 12 separate csv files from Virginia Department of Transportation, joining files on highway segment and year, and completing extensive transformations on several independent variables.

*Data Splitting (section ii.b.):* Split combined data into ‘actual’ and ‘forecasted’ data sets prior to Exploratory Data Analysis. Split the ‘actual’ data into train (90%) and test (10%) segments. Forecasted data is held out to use for classifying future unreliable segments.

*Exploratory Data Analysis (section ii.c.):* Evaluate distributions of numeric variables to determine necessary transformations. Three numeric variables benefit from log transformations. Also determined it necessary to drop geographic categorical variables due to certain instances of these variables not having examples of unreliable segments.

*Model Construction (section iii.a.):*

*Model Evaluation (section iii.b.):*

1. **Summary of Finding**
2. **Data and Methods**
3. **Data Import and Preprocessing**
4. **Data Splitting**
5. **Exploratory Data Analysis**
6. **Results**
7. **Model Construction**
8. **Model Evaluation**
9. **Conclusions**

The conclusions section can include future work, if there was more time.