Data Requirements

Afbeelding met tekst, logo, Lettertype, Graphics

Automatisch gegenereerde beschrijving

18-09-2023

Evvision

Cédric, Matthijs, Rik and Arthur

# Introduction

We have for the group project an assignment from Jugo. In this project, we aim to develop a predictive model for traffic congestion during rush hours along various highway routes. Different highway routes have been assigned to respective groups, each susceptible to varying levels of traffic jams. Understanding and predicting traffic conditions will enable our consultants to plan their workdays effectively.

The primary objective is to provide insights on the least congested days for commuting to the client's office and to our Hertogenbosch office for the upcoming week.

Furthermore, if feasible, we will propose alternative routes or travel times to mitigate potential delays. The project's success will contribute to enhanced time management and improved efficiency in our consultants' schedules.

Let's delve into the analysis and prediction of traffic congestion to optimize our workweek and streamline daily operations.

Table of contents

[Introduction 2](#_Toc145929678)

[Requirements 4](#_Toc145929679)

[Determine Data Volume 5](#_Toc145929680)

[Define Data Quality Standards 6](#_Toc145929681)

[Consider Ethical and Legal Aspects 7](#_Toc145929682)

# Requirements

|  |  |  |  |
| --- | --- | --- | --- |
| Requirement | Data Type | Units | Range |
| Weather Temperature | Numerical | Celsius | -25°C to 50°C |
| Rain forecast | Numerical | Percent | 0 - 100 |
| Day | Categorical |  | Monday - Sunday |
| Date | Numerical | Day – month – year – hour – minute - seconds | 01 – 31 | 01 – 12 | Years provided from dataset | 0 – 24 | 0 - 60 | 0 - 60 |
| Vehicle | Categorical |  | Car – truck - motor |
| Amount of vehicles | Time Series | Count per hour/minutes | 0 – max amount of vehicles | 0 - 60 |
| Average speed | Numerical | Kilometers per hour | 0 - 130 |
| Cause of traffic jam | Categorical |  | Accident – weather – road work – High traffic volume - etc |
| Time spent | Numerical | Hour/minute |  |

# Determine Data Volume

At this point, it is difficult to establish the precise data volume required, and relying on older data may not be advisable because the values could be inaccurate. Given the uncertainty surrounding data needs and the potential for outdated information, it is essential to exercise caution when considering older data. Too much data can also lead to loss of detail. For example, if you take the last 10 years of traffic jams and the last year there were a lot of traffic jams, the average may not match the recent values.

# Define Data Quality Standards

We may want to consider using data from 2020 onwards since there have been changes in speed limits. This more recent data will provide a more accurate representation of the current situation.

# Consider Ethical and Legal Aspects

We do not incorporate personal data, such as license plate information, and we also do not make use of extensive data about the vehicle itself. We also do not look when a car gets on the highway and were it leaves the highway.