Metro Interstate Traffic Volume Prediction System

High-Level Design Documentation

- ***Fazal Rehman***

Contents

Abstract

The Metro Interstate Traffic Volume Prediction System aims to build a machine learning model that can accurately predict traffic volume on interstate highways based on various factors such as weather conditions, time, and holidays. By forecasting traffic volume, city planners and individuals can make informed decisions to reduce congestion and improve commuting experiences. The dataset used for this project has been sourced from the UCI Machine Learning Repository and contains traffic volume recorded hourly, along with corresponding weather and time features.

Introduction

What is High-Level Design?

The purpose of this High-Level Design (HLD) Document is to add the necessary detail to the current project description to represent a suitable model for coding. This document is also intended to help detect contradictions prior to coding, and can be used as a reference manual for how the modules interact at a high level.

The HLD will:

- Present all of the design aspects and define them in detail.
- Describe the user interface being implemented.
- Describe the hardware and software interfaces.
- Describe the performance requirements.
- Include design features and the architecture of the project.
- List and describe the non-functional attributes like:
    - Security
    - Reliability
    - Maintainability
    - Portability
    - Reusability
    - Application compatibility
    - Resource utilization
    - Serviceability

Scope

The HLD documentation presents the structure of the system, such as the database architecture, application architecture (layers), application flow (Navigation), and technology architecture. The HLD uses non-technical to mildly technical terms, which should be understandable to the administrators of the system.

Definitions

Term	Description
MITVPS	Metro Interstate Traffic Volume Prediction System
UI	User Interface
Flask	Python-based web framework
CSV	Comma-Separated Values file format

General Description

Product Description

MITVPS is a machine learning-driven system that predicts hourly traffic volume based on real-time user inputs regarding weather, time, and holidays. The system provides insights and facilitates better planning to minimize congestion.

Problem Statement

Nowadays, traffic is a major issue for everyone, and it is a source of stress for anyone who has to deal with it on a daily basis. The growth of the population delays traffic and makes it worse day by day. The settlement of modern civilization looks at it, but it is unable to act in such a way as to protect people. We can watch traffic, collect data, and anticipate the next and subsequent observations using a variety of approaches and patterns. The observation agency then makes observations, which are then analysed, and predictions are made. Being stuck in a cosmopolitan city's traffic is the most common occurrence in one's life.

The goal of this project is to build a prediction model using multiple machine learning techniques and to use a template to document the end-to-end stages. We're trying to forecast the value of a continuous variable with the Metro Interstate Traffic Volume dataset, which is a regression issue.

Proposed Solution

The solution for the Metro Interstate Traffic Volume Prediction System involves data cleaning, feature engineering, and training various machine learning models. The best-performing model will be integrated into a Flask web application for real-time predictions. The system will include logging, error handling, and efficient resource management for reliability and performance.

Further Improvements

Real-time traffic and weather data integration can improve prediction accuracy. The system can be extended to special event traffic handling. AI-based anomaly detection can identify sudden traffic changes. Mobile app support can offer live traffic updates and alternate routes.

Technical Requirements

As technical requirements, we don't need any specialised hardware for virtualisation of the application. The users should have a device that has access to the web and a fundamental understanding of providing input.

Data Requirements

Data requirements completely depend on our problem statement. We need the dataset from the MN Department of Transportation to improve the accuracy of the model. The Required Dataset should contain the following features:

- Date-Time – Weekday, Month, Hour can be extracted from it.
- We should know that, do we have a holiday on that particular day or not.
- Temperature on the specified date and time.
- Weather conditions like Clouds, Clear, Rain, Snow, Fog, etc., on a specified date and time.
- The Numeric Percentage of cloud cover on the specified datetime.

Tools Used

- Python Framework such as NumPy, Pandas, Scikit-learn, Matplotlib, Seaborn, etc.
- Integrated Development Environment – Jupyter Notebook
- Backend – Flask
- Frontend – HTML/CSS/JavaScript
- Visualisation – PowerBI
- Version Control/Open Source – GitHub
- Source Code Editor – Visual Studio Code

Hardware Requirements

- Windows Server, Linux, or any other operating System that can run as a web server, capable of delivering HTML5.
- 2 GB of free storage.
- 512 MB of RAM.
- 3 GB of hard disk space.

Constraints

The MITVPS must be user-friendly and should not require anyone to have any prior knowledge in order to use it.

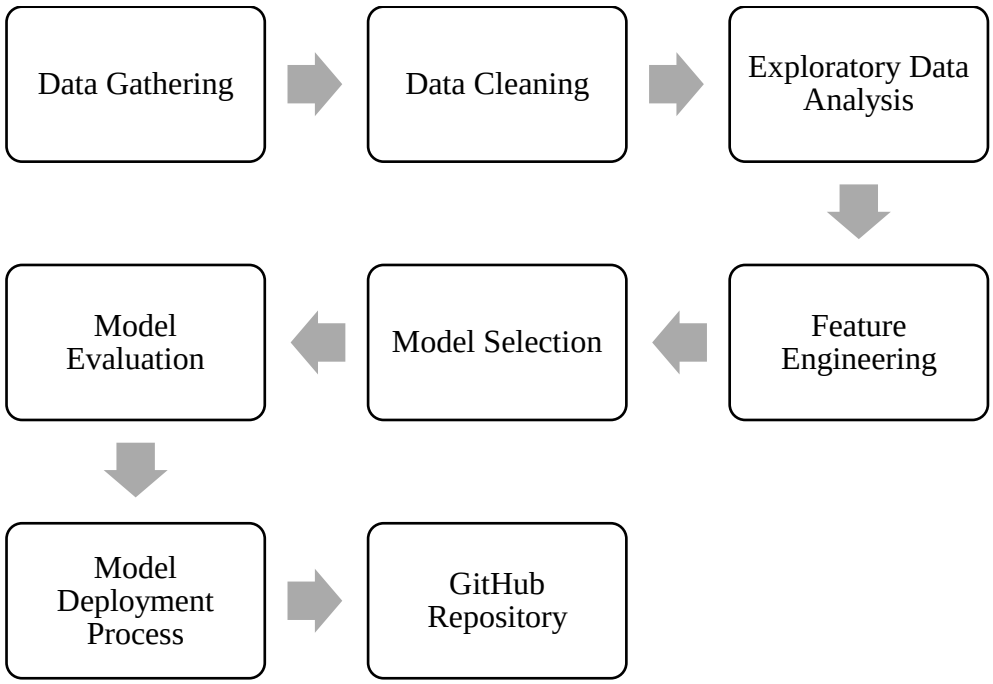
Assumptions

The main objective of the project is to implement the use cases for the new dataset that comes through the UI. A Machine Learning model is employed to process the user input for prediction. It is assumed that each aspect of this project has the flexibility to figure along the approach the designer is expecting.

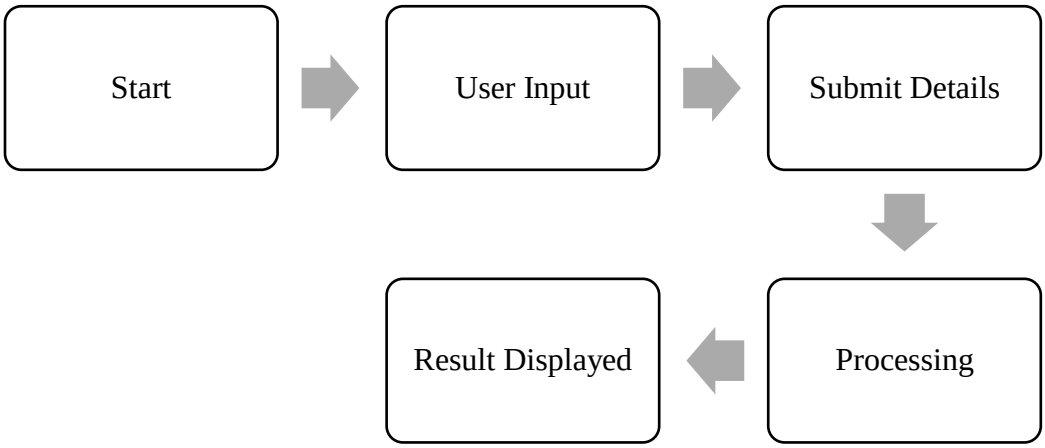
Design Details

Process Flow

The Process flow diagram is shown below:



Deployment Process



Logging

In logging, at each step, if an error or exception occurs, the event is logged into the system log file with the reason and timestamp. These help the developer to debug the system bugs and rectify the errors.

Error Handling

Errors should be encountered; an explanation will be displayed as to what went wrong. An error will be defined as anything that fails outside the normal and intended usage.

Performance

The MITVPS is used to predict the traffic volume. It can be used by concerned authorities; therefore, it is supposed to be as accurate as possible. So that it doesn't mislead the authorities. Also, Model retraining is very important to improve the performance

Reusability

The code written and the components used should have the ability to be reused with no problems.

Application Compatibility

The different components of the system are communicating or using Python as an interface between them. All the components have their own tasks to perform, and it is a job of Python to ensure proper transfer of Information.

Resource Utilisation

When any task is performed, it will likely use all the processing power available until that function is finished.

Deployment

The Model can be deployed using any cloud services such as Amazon Web Services, Microsoft Azure, Google Cloud, etc.

Dashboards

The dashboards will display charts over time with progress on various indicators and factors.

Conclusion

The MITVPS will give the traffic volume predictions instantly and has the potential to help the concerned agencies.