

2 General Description

2.1 Product Perspective

The goal of this project is to build a model that can predict the traffic volume depending on the conditions or factors that causes traffic problems. For, this We have used ensemble learning algorithms to develop the prediction model.

2.2 Problem Statement

To create traffic prediction model for using machine learning algorithms which will help us to predict and analyze the traffic volume based on different conditions. This project aims to implement the following use cases:

- To predict traffic volume at different times of the day
- To predict traffic volume at different weather conditions.

2.3 Proposed Solution

The solution proposed here is a prediction model which will help us to find out the traffic volume based on different factors such as time, weather conditions and holidays. This entire project is build using DVC (Data Version Control) and MLFlow for retraining, experimenting and model monitoring purpose.

2.4 Further Improvements

With continuous retraining and model configurations, model can be improved continuously. In future, other features can be added to improvise our prediction model.

2.5 Data Information

The dataset used in this project is a UCI machine learning dataset. The data at every transformation is uploaded and extracted from mongoDB database. The dataset contains hourly interstate 94 westbound traffic volume for MN DoT ATR station 301. The region of data lies between regions of Minneapolis. It includes features such as holiday, time, weather, etc. which impacts the traffic volume traffic volume directly.

Information of attributes of dataset as follows:

- holiday: Indicates if the date is a holiday and if it specifies the holiday, if not None.
- temp: Indicates the temperature in Kelvin.
- rain_1h: Amount in mm of rain that occurred in the hour.
- snow_1h: Amount in mm of snow that occurred in the hour.
- clouds_all: Percentage of cloud cover.
- weather_main: Short textual description of the current weather.
- weather_description: Longer textual description of the current weather.
- date_time: Hour of the data collected in local CST time.
- traffic_volume: Hourly I-94 ATR 301 reported westbound traffic volume.

2.6 Tools Used

The whole project is built upon Python Programming language. Visual Studio code and jupyter notebooks are used as a prime development interface. Python libraries such as NumPy, Pandas, Matplotlib, DVC, MLflow, etc. are majorly used in the project.



- Visual Studio Code is used as IDE.
- Python is used as the main programming language.
- HTML is used for website design.
- Git is used for continuous integration and version control.
- DVC is used for Data Version Controls
- MLflow is a platform to streamline machine learning development, including tracking experiments, packaging code into reproducible runs, and sharing and deploying models.
- Jupyter is used for analysis notebooks and for testing codes.
- Sklearn is used for performing machine learning operations.
- Python libraries such as numpy, pandas, matplotlib, seaborn, etc. are also used for required functions.
- Railways and Heroku are the deployment platforms used for this project.

2.7 Constraints

The machine learning based prediction system must be user friendly, as automated as possible and users should not be required to know any of the workings.

2.8 Assumptions

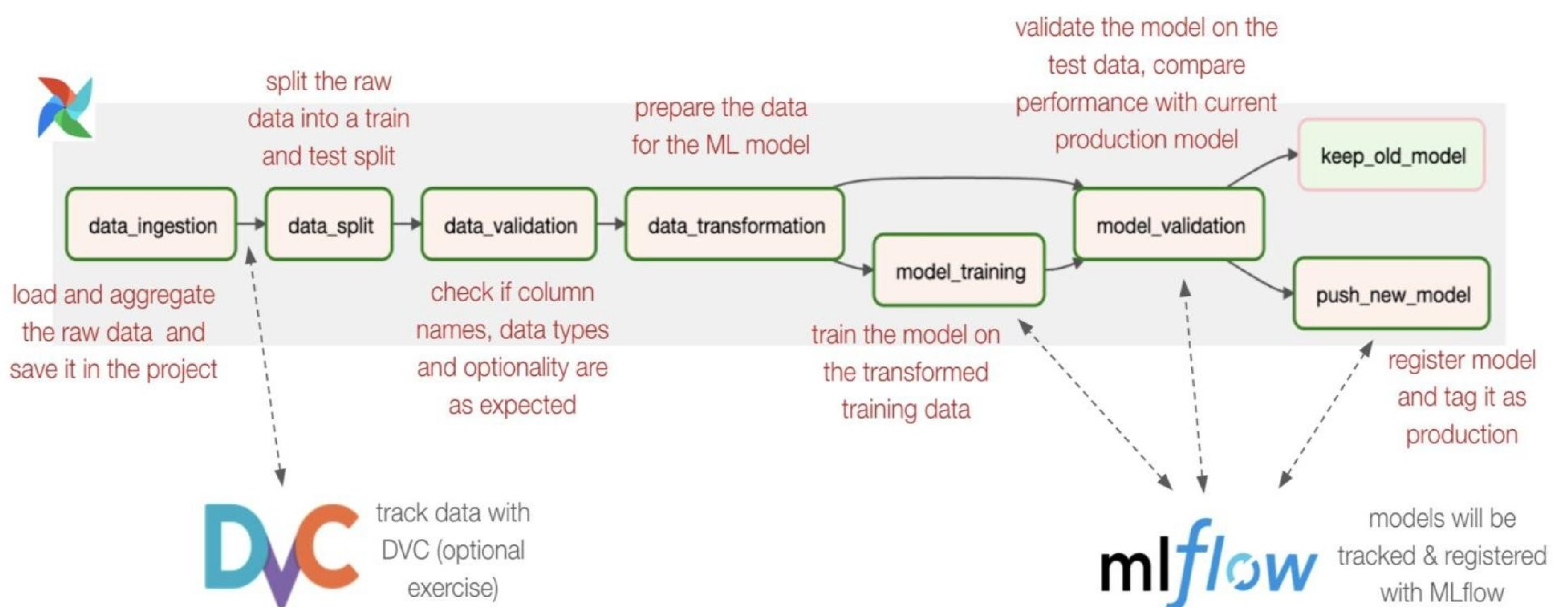
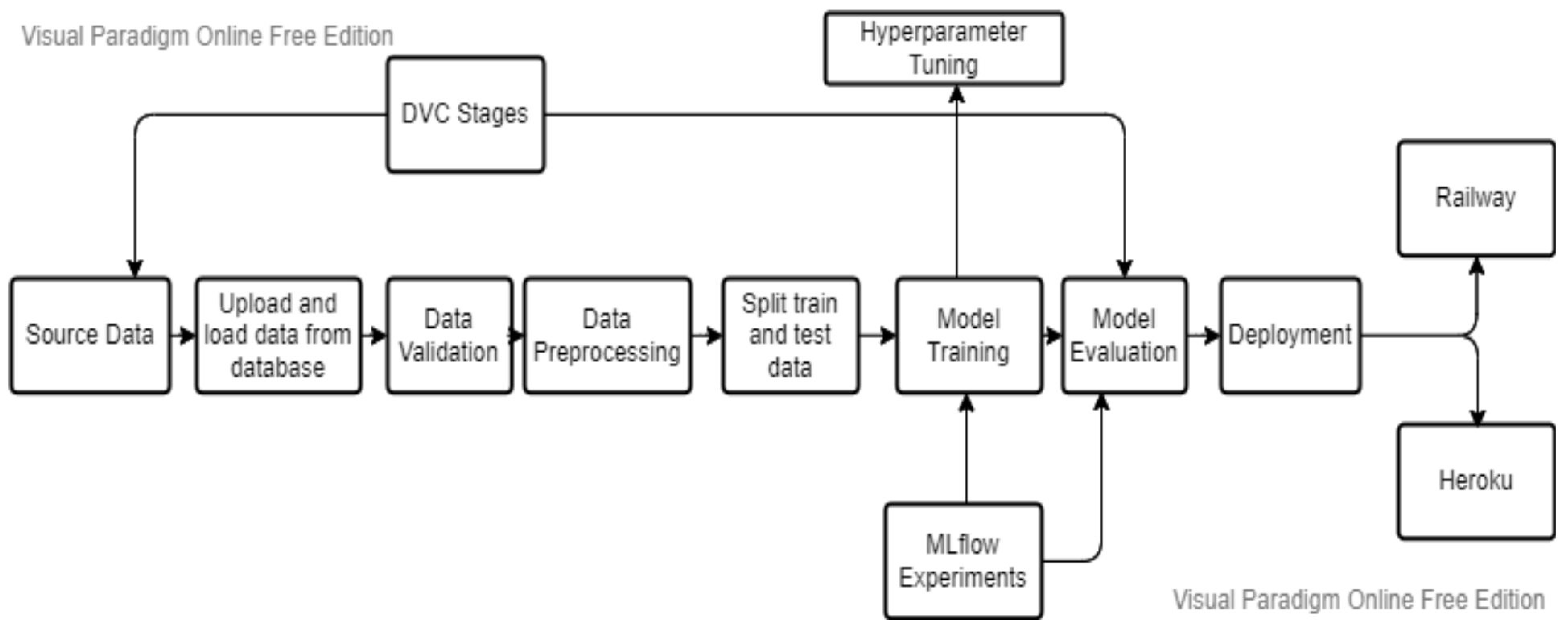
The main objective of this project is to build a predictive model which predicts the traffic Volume on untrained data. XGBoost Regression – an ensemble learning approach is used for model training. It is assumed that the project has the ability to predict accurate results when appropriate requirements are satisfied. Model retraining can be performed with the help of mlflow.

3 Design Details

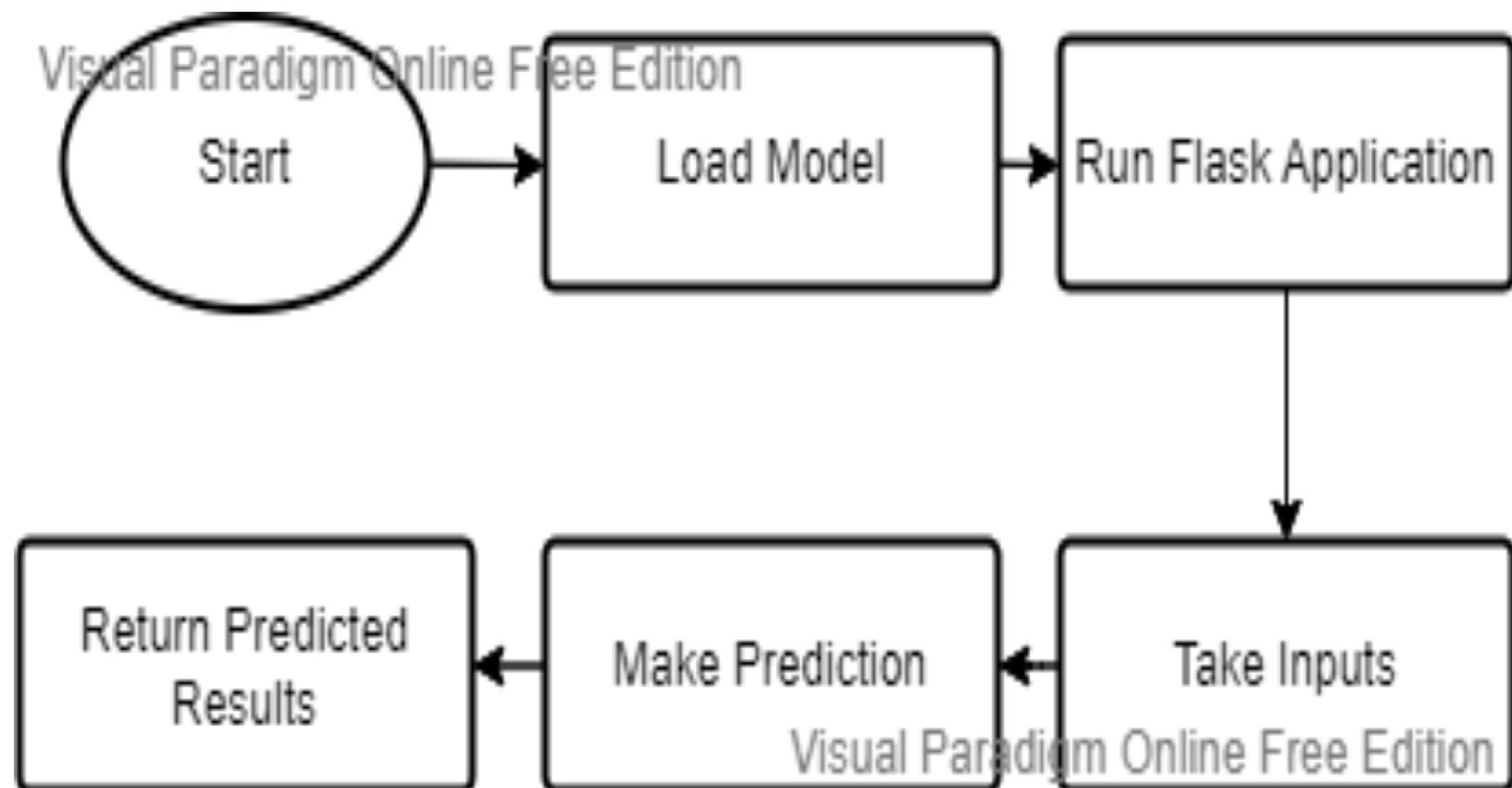
3.1 Process Flow

For prediction model, we will be using machine learning models. Below is the process flow diagram of the project.

3.1.1 Proposed Methodology



3.1.2 Deployment



3.2 Event log

The system should log every event so that the user will know what process is running internally.

Initial Step-By-Step Description:

The System identifies at what step logging required

The System should be able to log each and every system flow.

Developer can choose logging method. You can choose database logging/ File logging as well.

System should not hang even after using so many loggings.

Logging just because we can easily debug issues so logging is mandatory to do.

3.3 Error Handling

Should errors be encountered, an explanation will be displayed as to what went wrong?

An error will be defined as anything that falls outside the normal and indented usage.

4 Performance

The Metro Interstate Traffic Prediction is a prediction model which predicts the traffic Volume based on the weather condition, time and occasions. The model therefore, is expected to be accurate as overfitting or an underfitting model can lead to unexpected results.

4.1 Error Handling

The code written and the components used should have the ability to be reused with no problems. Custom Exception class is used to handle errors.

6 References

- <https://archive.ics.uci.edu/ml/datasets/Metro+Interstate+Traffic+Volume>
- <https://www.youtube.com/playlist?list=PLZoTAELRMXVOk1pRcOCaG5tXxgMalple>