Traffictelligence: Advanced Traffic Volume Estimation With Machine

Milestone 1: Introduction

Activity 1: Project Overview

Traffic management and planning are crucial for the efficient functioning of urban environments. The usual approaches for estimating traffic volume include manual counting, use of sensors, and simple models which could be expensive or difficult to sustain due to labour. Inaccuracy may also be an issue at times. A new innovation dubbed "Traffictelligence" is on a mission, it aims to use machine learning in developing a reliable, effective, and easy-to-implement solution for traffic volume estimation.

Activity 2: Objectives

- Develop a machine learning model: Not a simple one but a sophisticated model which is able to predict the amount of traffic based on both historical data and the data fed in real time.
- Blend different streams of data into one coherent whole: Let these be not only historical traffic information, but also weather conditions, date, time and temperature.
- Real time processing: It must deliver current estimates of the volume that flows through particular roads instantaneously as soon as new information comes in.
- Scalability and Efficiency: Guarantee that the solution can scale to accommodate significant urban regions and is implemented with ease without major adjustments in infrastructure.

Milestone 2: Project Initialization and Planning Phase

The "Project Initialization and Planning Phase" marks the project's outset, defining goals, scope, and stakeholders. This crucial phase establishes project parameters, identifies key team members, allocates resources, and outlines a realistic timeline. It also involves risk assessment and mitigation planning. Successful initiation sets the foundation for a well-organized and efficiently executed machine learning project, ensuring clarity, alignment, and proactive measures for potential challenges.

Activity 1: Define Problem Statement

As the city's traffic manager, real-time traffic congestion prediction would be one of the primary factors towards efficient management. Now, analysis of historical data and sensor inputs can feed machine learning models in order to predict patterns of congestion. At the same time, proactive deployment of control measures such as signal timing adjustment or rerouting vehicles would alleviate the problem of traffic jams and make the commuting experience far smoother.

Activity 2: Project Proposal (Proposed Solution)

The project aims to predict traffic volume on city road networks using past and current data. It will cover model creation learning from past traffic info, and checking for accuracy with the right performance measures. We want to make traffic flow better, cut down on jams, and boost overall transport effectiveness. We'll use machine learning methods to study past traffic data and live sensor inputs. By using advanced prediction models, we hope to forecast traffic volume with precision. This will allow for hands-on traffic control flexible signal management, and smarter resource use for public transit and parking. The project will explain the importance of these forecasts, our planned approach, and how it might change urban travel and people's daily commutes.

Activity 3: Initial Project Planning

Initial Project Planning involves outlining key objectives, defining scope. It encompasses setting timelines, allocating resources, and determining the overall project strategy. During this phase, the team establishes a clear understanding of the dataset, formulates goals for analysis, and plans the workflow for data processing. Effective initial planning lays the foundation for a systematic and well-executed project, ensuring successful outcomes.

SmartLender Project Planning Report: Click Here

Milestone 3: Data Collection and Preprocessing Phase

The Data Collection and Preprocessing Phase involves executing a plan to gather relevant traffic volume prediction data from Kaggle, ensuring data quality through verification and addressing missing values.

Preprocessing tasks include cleaning, encoding, and organizing the dataset for subsequent exploratory analysis and machine learning model development.

Activity 1: Data Collection Plan, Raw Data Sources Identified, Data Quality Report

The dataset for "TrafficTelligence: Advanced Traffic Volume Estimation with Machine Learning" is sourced from Skillwallet. It includes temperature, weather, holidays, and timestamps. Data quality is ensured through thorough verification, addressing missing values, and maintaining adherence to ethical guidelines, establishing a reliable foundation for predictive modeling.

SmartLender Data Collection Report: Click Here

Activity 2: Data Quality Report

The dataset for "TrafficTelligence: Advanced Traffic Volume Estimation with Machine Learning" is sourced from Kaggle. It includes applicant details and financial metrics. Data quality is ensured through thorough verification, addressing missing values, and maintaining adherence to ethical guidelines, establishing a reliable foundation for predictive modeling.

SmartLender Data Quality Report: Click Here **Activity**

3: Data Exploration and Preprocessing

Data Exploration involves analyzing the traffic volume dataset to understand patterns, distributions, and outliers. Preprocessing includes handling missing values, scaling, and encoding categorical variables. These crucial steps enhance data quality, ensuring the reliability and effectiveness of subsequent analyses in the traffic volume prediction project.

SmartLender Data Exploration and Preprocessing Report: Click Here

Milestone 4: Model Development Phase

The Model Development Phase entails crafting a predictive model for traffic volume prediction. It encompasses strategic feature selection, evaluating and selecting models (Random Forest, Polynomial Regression, SVR, and XGB), initiating training with code, and rigorously validating and assessing model performance for informed decision-making in the traffic volume prediction process.

Activity 1: Feature Selection Report

The Feature Selection Report outlines the rationale behind choosing specific features (e.g., temp, rain, day) for the traffic volume prediction model. It evaluates relevance, importance, and impact on predictive accuracy, ensuring the inclusion of key factors influencing the model's ability to predict the traffic volume.

SmartLender Feature Selection Report: [Click Here]

Activity 2: Model Selection Report

The Model Selection Report details the rationale behind choosing Random Forest, Polynomial Regression, SVR, and XGB models for traffic volume prediction. It considers each model's strengths in handling complex relationships, interpretability, adaptability, and overall predictive performance, ensuring an informed choice aligned with project objectives.

SmartLender Model Selection Report: [Click Here]

Activity 3: Initial Model Training Code, Model Validation and Evaluation Report

The Initial Model Training Code employs selected algorithms on the traffic volume dataset, setting the foundation for predictive modeling. The subsequent Model Validation and Evaluation Report rigorously assesses model performance, employing metrics like accuracy and precision to ensure reliability and effectiveness in predicting traffic volume outcomes.

SmartLender Model Development Phase Template: [Click Here]

Milestone 5: Model Optimization and Tuning Phase

The Model Optimization and Tuning Phase involves refining machine learning models for peak performance. It includes optimized model code, fine-tuning hyperparameters, comparing performance metrics, and justifying the final model selection for enhanced predictive accuracy and efficiency.

Activity 1: Hyperparameter Tuning Documentation

The Gradient Boosting model was selected for its superior performance, exhibiting high accuracy during hyperparameter tuning. Its ability to handle complex relationships, minimize overfitting, and optimize predictive accuracy aligns with project objectives, justifying its selection as the final model.

Activity 2: Performance Metrics Comparison Report

The Performance Metrics Comparison Report contrasts the baseline and optimized metrics for various models, specifically highlighting the enhanced performance of the Gradient Boosting model. This assessment provides a clear understanding of the refined predictive capabilities achieved through hyperparameter tuning.

Activity 3: Final Model Selection Justification

The Final Model Selection Justification articulates the rationale for choosing Gradient Boosting as the ultimate model. Its exceptional accuracy, ability to handle complexity, and successful hyperparameter tuning align with project objectives, ensuring optimal traffic volume predictions.

SmartLender Model Optimization and Tuning Phase Report: Click Here

Milestone 6: Project Files and Documentation

For project file in GitHub, kindly click the link and refer to the flow. **Click Here**

For the documentation, Kindly refer to the link. **Click Here**

Milestone 7: Project Demonstration

In the upcoming module called Project Demonstration, individuals will be required to record a video by sharing their screens. They will need to explain their project and demonstrate its execution during the presentation.

Results:

Model Training & Testing

```
[70]: from sklearn.model_selection import train_test_split
      x_train, x_test, y_train, y_test = train_test_split(x_scaled, y_value, test_size=0.2)
```

```
[71]: import xgboost as xgb
      model=xgb.XGBRegressor()
      model.fit(x_train, y_train)
```

```
[71]:                         XGBRegressor
XGBRegressor(base_score=None, booster=None, callbacks=None,
             colsample_bylevel=None, colsample_bynode=None,
             colsample_bytree=None, device=None, early_stopping_rounds=None,
             enable_categorical=False, eval_metric=None, feature_types=None,
             gamma=None, grow_policy=None, importance_type=None,
             interaction_constraints=None, learning_rate=None, max_bin=None,
             max_cat_threshold=None, max_cat_to_onehot=None,
             max_delta_step=None, max_depth=None, max_leaves=None,
             min_child_weight=None, missing=nan, monotone_constraints=None,
             multi_strategy=None, n_estimators=None, n_jobs=None,
             num_parallel_tree=None, random_state=None, ...)
```

```
[72]: y_pred=model.predict(x_test)
      from sklearn.metrics import mean_squared_log_error, r2_score
      r2_score(y_test, y_pred)
```

```
[72]: 0.9564729093269886
```

Fig: Model training and testing output

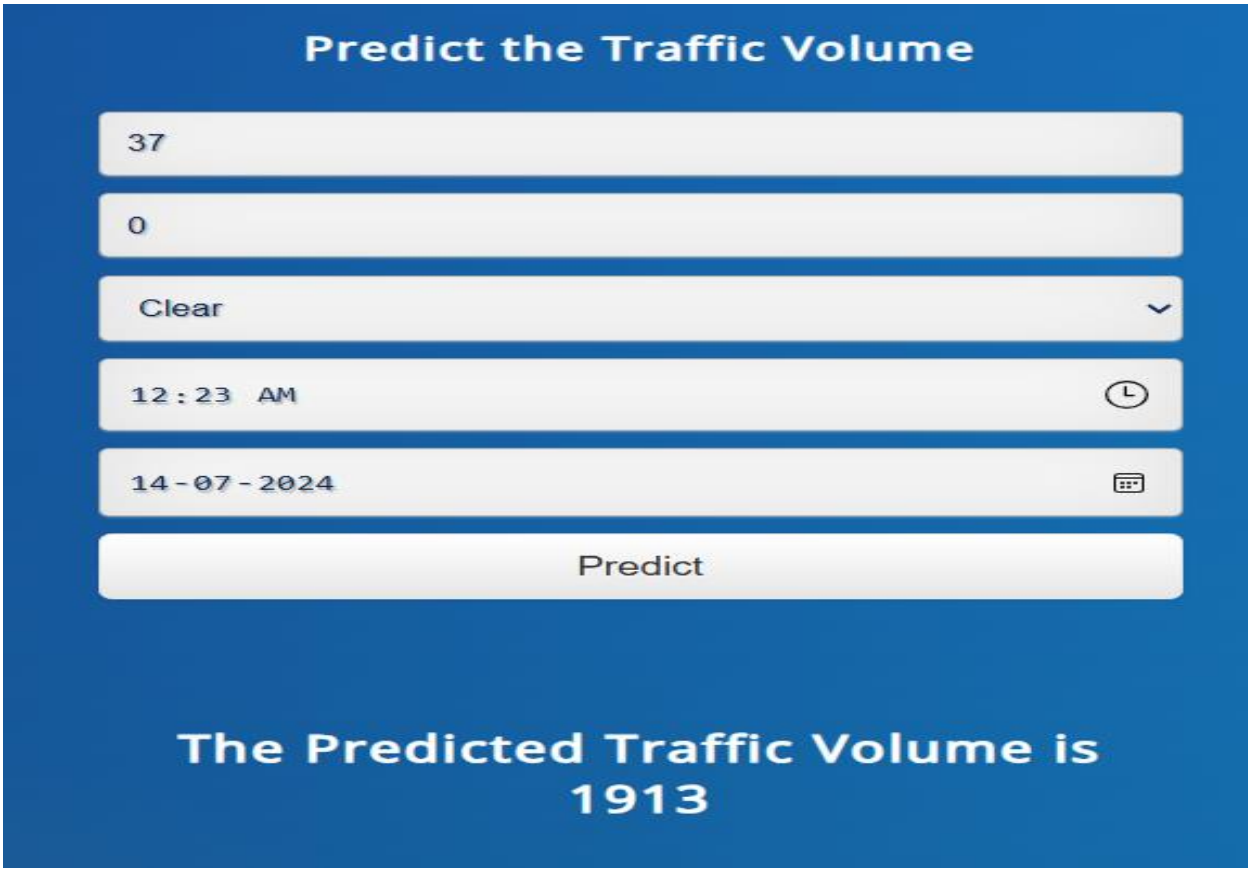

Fig: Deployed model

Advantages and Disadvantages:

Advantages:

It helps in unentrapped and smooth flow of traffic.
Travellers can now able to plan their route for trip, offices in rush hour.
This model helps the administration to rerouting the traffic and adjusting traffic signals
Helps any navigation to suggest most efficient route.
It can help emergency services plan routes more effectively, ensuring quicker response times.

Disadvantages:

Model is heavily dependent on data, so its accuracy depends on quality and quantity of data.
It needs to continuously update its data for accurate prediction.
It is computationally very costly

Future Scope:

- This will help manage traffic lights, tolls, and public transportation in smart cities.
- It models will improve navigation and safety for autonomous vehicles.
- These models can reduce traffic congestion, leading to cleaner air and better urban planning.
- Public transport systems can use traffic predictions to optimize schedules and improve passenger experience.

Conclusion:

Machine learning's traffic prediction abilities hold immense promise for revolutionizing how cities move. These models can optimize traffic flow, plan routes more effectively, allocate resources better, and even benefit the environment. While data dependency, complexity, and infrastructure need pose challenges, the future looks bright. Applications span smart cities, self-driving cars, public transit, and more. As machine learning evolves, it will play a key role in creating efficient, safe, and sustainable urban environments, ultimately leading to smarter, more connected cities with a higher quality of life.