

Smart Traffic Prediction System: Enhancing Urban Mobility Through Intelligent Traffic Prediction and Dynamic Route Optimization

VEDANT MAVALE

`vedant.mavale_24uds@sanjivani.edu.in`

sanjivani university <https://orcid.org/0009-0004-8583-537X>

Research Article

Keywords: Traffic Prediction, Urban Mobility, Machine Learning, Traffic Classification, Route Optimization, Feature Importance, Intelligent Transportation Systems, Smart Cities, Decision Support Dashboard

Posted Date: October 13th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11335507/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: The authors declare no competing interests.

Smart Traffic Prediction System: Enhancing Urban Mobility Through Intelligent Traffic Prediction and Dynamic Route Optimization

Vedant Mavale
Sanjivani University
Kopargaon, Maharashtra, India

ABSTRACT

This study presents an implemented Smart Traffic Prediction System that classifies urban traffic conditions as Low, Medium or High from five measurable inputs - vehicle count, average speed, hour of day, weather and road type - and uses the predicted condition to recommend the fastest of several candidate routes. The system is delivered as an interactive dashboard that shows the traffic prediction and a route-comparison table in a single view. On the displayed test results the classifier correctly labelled 1,993 of 2,000 observations (99.65% accuracy; macro-averaged F1 of about 0.995), and feature-importance analysis showed that vehicle count (about 0.67) and average speed (about 0.31) account for almost all of the model's predictive signal, with hour, road type and weather contributing very little. In the demonstrated example (80 vehicles, 35 km/h, hour 18, clear weather, main road) the system predicts Low traffic and recommends Route A with an estimated travel time of 20 minutes over Route B (24 min) and Route C (27.5 min). The paper reports the measurement framework, sampling considerations, feature identification, ethical aspects, dataset bias and limitations of the implementation.

KEYWORDS: Traffic Prediction, Urban Mobility, Machine Learning, Traffic Classification, Route Optimization, Feature Importance, Intelligent Transportation Systems, Smart Cities, Decision Support Dashboard

CCS CONCEPTS: Applied computing → Transportation; Computing methodologies → Machine learning; Information systems → Decision support systems.

1 INTRODUCTION

Urban road networks face growing congestion, which increases travel time, fuel consumption and emissions. Intelligent transportation systems address this problem by converting traffic observations into predictions that help drivers and planners act before congestion builds up. However, a prediction is only useful if it is presented in a form that supports a decision. A raw vehicle count or speed reading does not tell a commuter which route to take.

Route-selection tools that rely only on the base (free-flow) travel time ignore the current traffic state, and may therefore recommend a nominally short route that is actually congested. The central problem addressed in this work is the lack of a simple, interpretable pipeline that links a predicted traffic category to route-level travel-time estimates and a final recommendation.

This study implements a Smart Traffic Prediction System in which a machine-learning classifier predicts the traffic category from traffic and environmental inputs, and a route-recommendation component adjusts the base time of each candidate route according to its traffic condition. The purpose is to show that combining traffic prediction with route evaluation produces a more informed recommendation than comparing base travel times alone.

2 RESEARCH OBJECTIVES

The study has the following objectives:

1. To identify and operationalise measurable variables (vehicle count, average speed, hour, weather and road type) that describe urban traffic conditions.
2. To develop and evaluate a machine-learning model that classifies traffic as Low, Medium or High, and to determine which input features contribute most to the prediction.
3. To implement a dynamic route-recommendation component and an interactive dashboard that presents the prediction and the recommended route together.

3 RELATED WORK

Research on traffic forecasting spans statistical time-series models, classical machine learning and deep learning. The studies below are foundational references for the techniques used or discussed in this work. A common gap is that forecasting studies rarely connect the predicted traffic state to an end-user route decision in a single interpretable interface.

1. Short-Term Traffic Forecasting: Where We Are and Where We're Going [1]

Authors and year: Vlahogianni, Karlaftis and Golias, 2014. *Methodology / approach:* A comprehensive review of short-term traffic forecasting research covering statistical and computational-intelligence methods. *Research gap:* As a review, it surveys prediction methods but does not propose an end-to-end system that turns predicted traffic categories into route recommendations.

2. Traffic Flow Prediction with Big Data: A Deep Learning Approach [2]

Authors and year: Lv, Duan, Kang, Li and Wang, 2015. *Methodology / approach:* A deep-learning model (stacked autoencoders) for traffic-flow prediction using large-scale data. *Research gap:* The emphasis is on prediction accuracy of flow values with complex models, rather than a lightweight, interpretable classification model with feature-importance analysis and a user-facing route recommendation.

3. Random Forests [3]

Authors and year: Breiman, 2001. *Methodology / approach:* Introduced the random forest ensemble, which combines many decision trees and provides feature-importance measures. *Relevance:* Ensemble tree methods that expose feature importance are a natural fit for tabular traffic data containing both numeric and categorical inputs. *Research gap:* A general-purpose method paper; it does not address traffic prediction or route recommendation.

Note: this list should be extended with additional recent traffic-prediction and route-optimization studies chosen by the author.

4 RESEARCH METHODOLOGY

The methodology follows a controlled pipeline in which traffic and contextual observations are collected and prepared, a classifier is trained and evaluated, and the predicted traffic condition is used to estimate the travel time of candidate routes.

4.1 Traffic Data Collection

A traffic-prediction project depends on historical or real-time observations of vehicle counts, speeds, timestamps, weather and road classification. The implementation developed in a Google Colab notebook uses these feature categories; the exact external source of the training data is not visible from the dashboard and should be stated by the author in the final version.

4.2 Data Cleaning and Feature Preparation

Missing values, incorrect records and inconsistent formats are handled during cleaning. Numerical variables (vehicle count, average speed, hour) and categorical variables (weather, road type) are then prepared for

modelling, with categorical variables encoded numerically.

4.3 Traffic Classification Model

A supervised classification model is trained on historical observations to predict one of three classes: Low, Medium or High traffic. The model exposes feature-importance scores, which are used to interpret its behaviour. Training, validation and testing data are kept separate: the training set learns model parameters, the validation set tunes the model, and the test set is reserved for final evaluation.

4.4 Route Recommendation Logic

Each candidate route has a base travel time. The predicted traffic condition of the route scales this base time to give an estimated time. In the displayed example the estimates are consistent with multipliers of 1.00 for Low, 1.25 for Medium and 1.60 for High traffic ($20 \times 1.00 = 20$, $22 \times 1.25 = 27.5$, $15 \times 1.60 = 24$). The route with the lowest estimated time is recommended.

4.5 Evaluation

The classifier is evaluated with a confusion matrix and the derived accuracy, precision, recall and F1-score. For estimated travel time, regression measures such as MAE or RMSE can additionally be used when ground-truth travel times are available.

Workflow summary.

1. Data Collection - collect traffic, time, weather, road and travel-time observations.
2. Data Cleaning - handle missing values, incorrect records and inconsistent formats.
3. Feature Preparation - prepare numerical and categorical variables for modelling.
4. Model Training - train the traffic classifier on historical observations.
5. Prediction - predict the traffic condition for the selected input values.
6. Route Evaluation - estimate travel time for each route under its traffic condition.
7. Recommendation - select and display the route with the lowest estimated time.

5 MEASUREMENT CONCEPTS

Measurement refers to converting traffic conditions and environmental information into quantifiable variables. Because 'traffic condition' cannot be read from a single physical sensor, it is operationalised through measurable indicators: vehicle count represents traffic volume, average speed represents movement efficiency, hour represents time-of-day effects, and weather and road type provide context.

5.1 Types of Variables

1. **Independent variables:** Vehicle Count, Average Speed, Hour, Weather and Road Type.
2. **Dependent variables:** Predicted traffic class (Low / Medium / High) and estimated route travel time.
3. **Control considerations:** the same preprocessing, feature encoding and model are applied each time so that identical inputs reproduce identical outputs.

Table 1: Levels of Measurement Used in the Study

Variable	Scale	Description
Vehicle Count	Ratio	Number of vehicles observed; has a meaningful zero.
Average Speed	Ratio	Average vehicle speed in km/h; zero means no movement.
Hour	Interval	Time-of-day value (0-23) capturing traffic variation.
Weather	Nominal	Categorical condition such as Clear or Rain.
Road Type	Nominal	Categorical road classification such as Main Road.
Traffic Prediction	Nominal (ordered categories)	Predicted class: Low, Medium or High.
Estimated Time	Ratio	Estimated travel time in minutes.

5.2 Reliability and Validity of Measurement

Reliability. The dashboard is a repeatable interface: entering the same inputs reproduces the same prediction, provided the underlying model and data are unchanged. Reliability of the data itself depends on the consistency of the traffic observations and on applying the same preprocessing every time.

Validity. The selected variables are directly related to traffic conditions and route travel time. Vehicle count and speed are direct numerical traffic indicators, and estimated travel time provides an interpretable outcome for route selection.

6 SAMPLING TECHNIQUES

The conceptual population is all traffic observations on urban roads, described by combinations of traffic volume, speed, time of day, weather and road type. The sampling unit is one traffic observation (a time-based traffic record), and the prediction target is the traffic category. The exact dataset size and sampling procedure used in the notebook are not visible from the dashboard and are therefore not assumed here.

Table 2: Sampling Techniques Considered

Technique	Description	Suitability in this study
Simple Random Sampling	Every record has an equal chance of selection.	Suitable for a general baseline evaluation set.
Stratified Random Sampling	Population divided into strata (e.g., traffic class, hour, weather), then sampled within each stratum.	Recommended: keeps rare classes (High traffic) and rare conditions (rain) represented.
Systematic Sampling	Every k-th record from an ordered list.	Useful for quick development checks, but may align with daily cycles.
Time-based Split	Earlier records for training, later records for testing.	Appropriate for time-series data to avoid leakage from the future.

The confusion matrix reported for the model covers 2,000 test observations, distributed unevenly across the classes (Table 3). The imbalance shows why accuracy alone is insufficient and why per-class precision and recall are also reported.

Table 3: Class Distribution of the Evaluated Test Set (derived from the confusion matrix)

Actual class	Test observations	Share
Low	1,215	60.8%
Medium	621	31.1%
High	164	8.2%
Total	2,000	100%

7 DATASET ANALYSIS AND FEATURE IDENTIFICATION

The dashboard separates input features from prediction and recommendation outputs. Table 4 summarises the features and the example values used in the demonstration.

Table 4: Feature Identification for the Traffic Prediction System

Feature	Example value	Role in the system
Vehicle Count	80	Traffic volume indicator (input)
Average Speed	35 km/h	Movement / speed indicator (input)
Hour	18	Time-of-day factor (input)
Weather	Clear	Environmental factor (input)
Road Type	Main Road	Road-context factor (input)
Traffic Prediction	LOW TRAFFIC	Model output (classification)
Base Time	15 / 20 / 22 min	Free-flow time of Routes B / A / C
Estimated Time	24 / 20 / 27.5 min	Traffic-adjusted time of Routes B / A / C

7.1 Exploratory Data Analysis (EDA)

- Distribution of observations across Low, Medium and High traffic classes to assess class balance.
- Average vehicle count per hour of the day to examine temporal variation (Section 10.4).
- Distribution of weather and road-type categories to detect under-represented conditions.
- Feature-importance analysis to identify the most influential predictors (Section 10.2).
- Missing-value and outlier checks on vehicle count and speed.

8 DATA COLLECTION, ETHICS AND DATASET BIAS

8.1 Ethical Considerations

1. **Purpose limitation:** traffic data should be used for transportation analysis and system improvement, not for tracking individuals.
2. **Privacy:** if vehicle identifiers, location traces or other personal information are collected, they must be anonymised and protected.
3. **Transparency:** the system should state that predictions are model-based and may contain uncertainty.
4. **Safety:** route recommendations should be reviewed for safety and must not encourage unsafe driving behaviour.
5. **Data minimisation:** only data necessary for prediction and route optimisation should be retained.

8.2 Dataset Bias

Table 5: Dataset Bias and Potential Impact

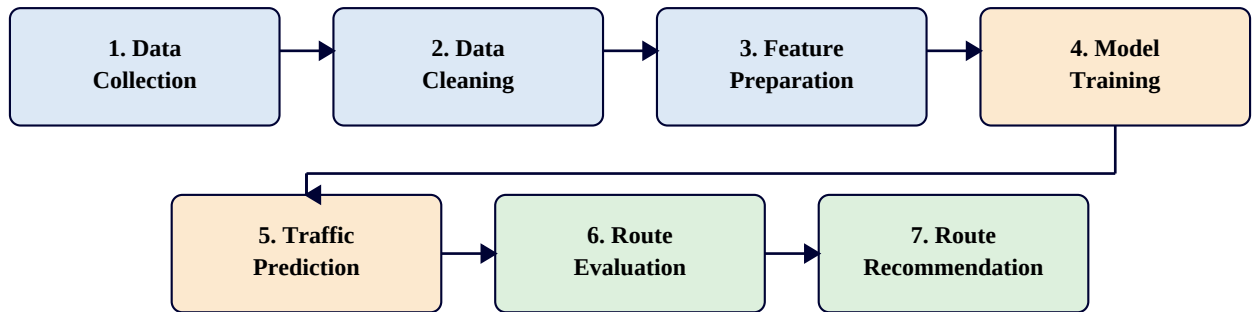
Type of Bias	Description	Potential impact
Geographic Bias	Data from a limited set of roads or cities.	Poor generalisation to other road networks.
Time-of-day Bias	Mostly daytime or mostly off-peak observations.	Weaker predictions during late-night or peak hours.
Weather Imbalance	Rain or fog observations are rare.	Inaccurate predictions under adverse weather.
Road-type Imbalance	Some road categories dominate.	Weaker performance on under-represented road types.
Class Imbalance	Low traffic is far more frequent than High traffic (Table 3).	Higher error risk on the minority High class.

8.3 Bias Mitigation Steps

1. Use stratified sampling so that rare classes and conditions appear in the evaluation set.
2. Report per-class precision, recall and F1 rather than accuracy alone.
3. Collect additional observations for rainy weather, night hours and under-represented road types.
4. Document dataset coverage limitations as part of the evaluation.

9 IMPLEMENTATION AND SYSTEM WORKFLOW

The system was implemented as an end-to-end pipeline, from data collection to route recommendation, and is exposed through an interactive dashboard (Figure 2). The user enters the vehicle count, average speed, hour, weather and road type, and presses the predict button. The dashboard then displays the predicted traffic category and a route-recommendation table comparing base time, traffic condition and estimated time, followed by the recommended route.



Inputs: vehicle count, average speed, hour, weather, road type | Output: traffic class + recommended route

Figure 1: Implemented Smart Traffic Prediction workflow.

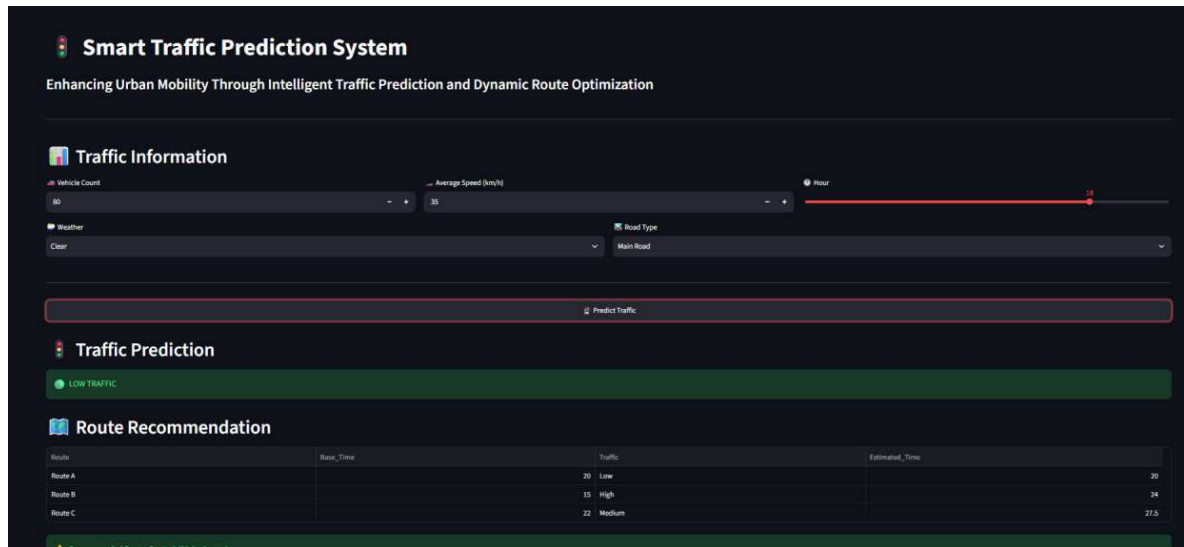


Figure 2: Implemented Smart Traffic Prediction System dashboard showing inputs, traffic prediction and route recommendation.

10 RESULTS AND EVALUATION

10.1 Displayed Prediction and Route Recommendation

For the dashboard example (80 vehicles, 35 km/h, hour 18, Clear weather, Main Road), the predicted traffic is LOW TRAFFIC. Table 6 shows the route comparison. Route B has the lowest base time (15 min), but its High traffic condition raises the estimate to 24 min. Route A, with Low traffic, keeps its 20 min estimate and is recommended. Route C has the highest estimate at 27.5 min.

Table 6: Route Recommendation Result from the Dashboard

Route	Base Time (min)	Traffic	Estimated Time (min)	Recommended
Route A	20	Low	20	Yes
Route B	15	High	24	No
Route C	22	Medium	27.5	No

10.2 Feature Importance

Figure 3 shows that vehicle count is the dominant predictor (importance about 0.67), followed by average speed (about 0.31). Hour contributes only slightly, while road type and weather have very low importance in this model. Traffic volume and vehicle speed are therefore the main factors driving the prediction.

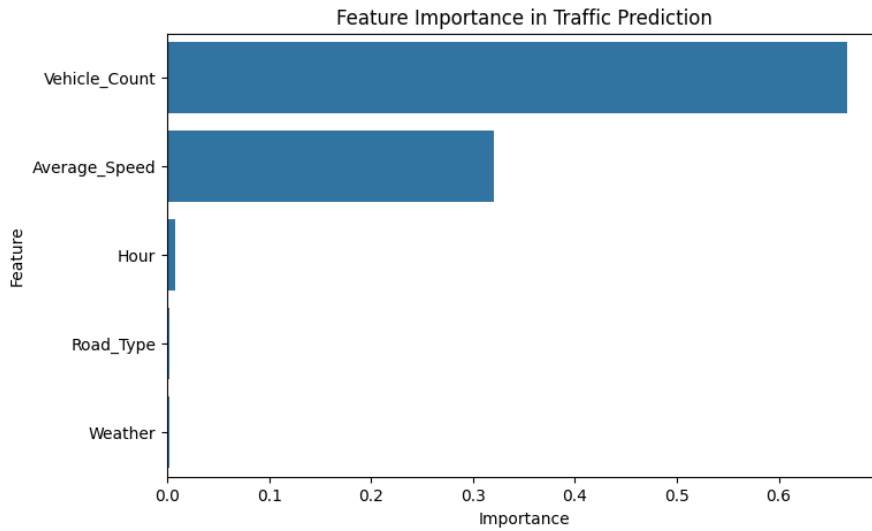


Figure 3: Feature importance in traffic prediction.

10.3 Classification Performance

The confusion matrix in Figure 4 shows that the model correctly predicted 163 High, 1,214 Low and 616 Medium observations. Only 7 of 2,000 observations were misclassified, giving an overall accuracy of 99.65%. Table 7 gives the per-class metrics derived from the matrix.

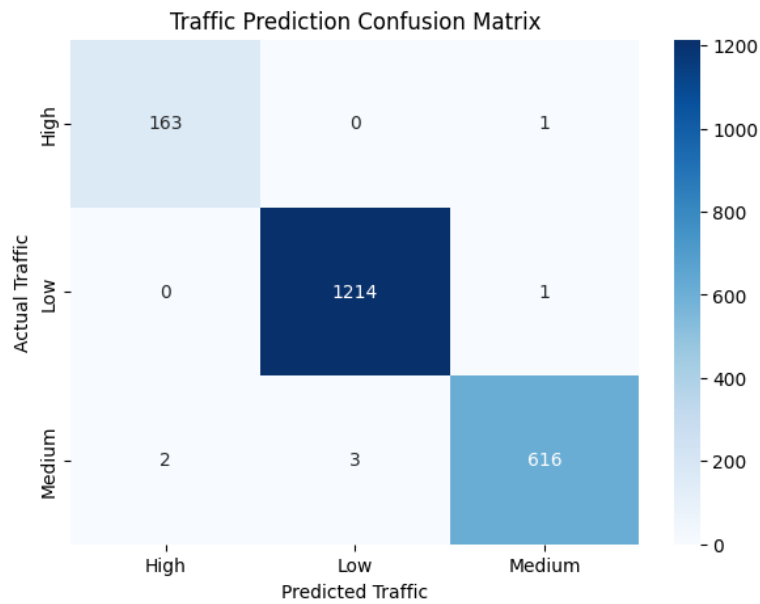


Figure 4: Traffic prediction confusion matrix.

Table 7: Per-Class Metrics Derived from the Confusion Matrix

Class	Precision	Recall	F1-score	Support
High	0.988	0.994	0.991	164
Low	0.998	0.999	0.998	1,215
Medium	0.997	0.992	0.994	621
Macro average	0.994	0.995	0.995	2,000
Accuracy			0.9965	2,000

The very high scores should be read with care. Without information on the split strategy it is not possible to rule out that closely related observations appear in both training and test sets, and the near-perfect result

suggests the classes are largely separable from vehicle count and speed alone.

10.4 Hourly Variation in Traffic Volume

Figure 5 shows the average vehicle count for each hour. The lowest value occurs around hour 3 (about 80.5 vehicles) and the highest around hour 10 (about 89.7 vehicles), with further peaks near hours 15, 19 and 23. The variation is modest in absolute terms, which is consistent with Hour having low importance in the model, but it confirms that demand changes through the day.

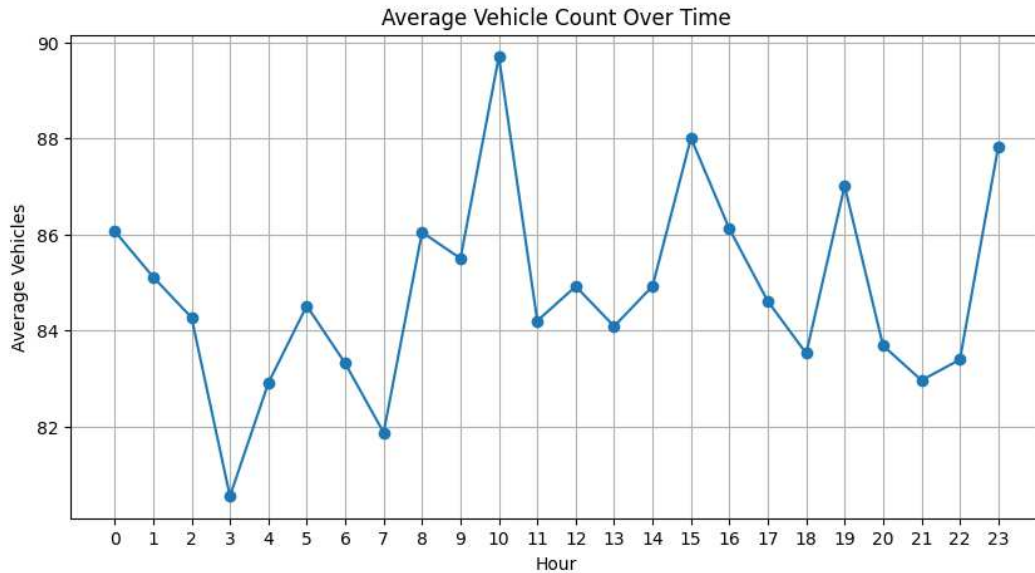


Figure 5: Average vehicle count over time.

10.5 Route Travel-Time Comparison

Figure 6 compares travel time across the three routes: Route B is lowest (about 18.8 min), Route A is about 20 min and Route C is highest (about 22 min). These values differ from the dashboard example in Table 6, which depends on the specific inputs and traffic condition of each route, so the chart is best read as an independent comparison. It illustrates the main design point of the system: the route with the shortest nominal time is not necessarily the best recommendation once traffic conditions are taken into account.

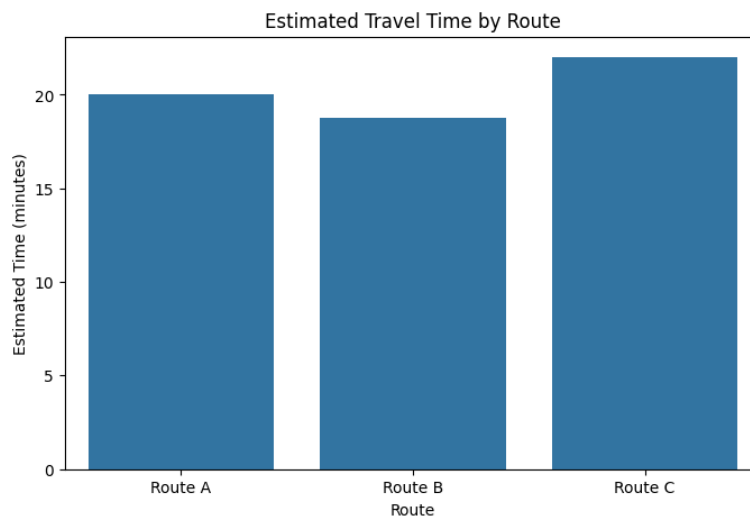


Figure 6: Estimated travel time by route.

11 LIMITATIONS

1. The dataset size, sampling procedure, train/validation/test split and model algorithm are not documented in the dashboard and must be stated for full reproducibility.
2. The reported performance is tied to the current dataset and may not generalise to other cities, road networks or weather conditions.
3. The route-time multipliers appear to be fixed rules rather than learned values, and routes are compared on traffic condition and base time only, without distance, incidents or road closures.
4. The classes are imbalanced (High traffic is only 8.2% of the test set), and weather and road type contribute almost nothing to the model, which may reflect limited variation in the data.
5. Predictions are not based on live traffic feeds, so they cannot react to sudden events.

Future improvements. Integrating real-time traffic feeds, adding distance, incident and road-closure data to the route score, collecting more diverse peak/non-peak and adverse-weather observations, and monitoring model performance over time as traffic patterns change.

12 CONCLUSION

This study implemented a Smart Traffic Prediction System that combines a machine-learning traffic classifier with a traffic-aware route-recommendation component in a single interactive dashboard. The classifier reached 99.65% accuracy on the 2,000 displayed test observations, with vehicle count and average speed as by far the most influential features. In the demonstrated case the system predicted Low traffic and recommended Route A (20 min) over Route B (24 min) and Route C (27.5 min), showing that the shortest base time does not always give the best route once traffic is considered. The work supports the use of prediction-driven route recommendation for urban mobility, while highlighting the importance of documented data sources, balanced coverage, real-time data and responsible use.

REFERENCES

- [1] Eleni I. Vlahogianni, Matthew G. Karlaftis, and John C. Golias. 2014. Short-term traffic forecasting: Where we are and where we're going. *Transportation Research Part C: Emerging Technologies* 43 (2014), 3-19.
- [2] Yisheng Lv, Yanjie Duan, Wenwen Kang, Zhengxi Li, and Fei-Yue Wang. 2015. Traffic flow prediction with big data: A deep learning approach. *IEEE Transactions on Intelligent Transportation Systems* 16, 2 (2015), 865-873.
- [3] Leo Breiman. 2001. Random forests. *Machine Learning* 45, 1 (2001), 5-32.