

AI TRANSPORT MODE PREDICTOR – DAR ES SALAAM

Introduction

This project demonstrates the application of Machine Learning in predicting transport modes in Dar es Salaam. Using a Decision Tree Classifier, the app provides predictions for common transport options (daladala, bodaboda, bajaj, walking) based on contextual factors such as distance, time of day, weather, traffic, and budget. The aim is to showcase how interpretable models can support everyday decision-making in urban mobility.

Dataset Description

A synthetic dataset of 500 records was created to simulate transport choices. **Features included:**

- Distance: short, medium, long
- Time of Day: morning, afternoon, evening
- Weather: sunny, rainy, cloudy
- Traffic: low, medium, high
- Budget: low, medium, high
- Target variable: Transport Mode (daladala, bodaboda, bajaj, walking)

Errors such as missing values, typos, and duplicates were introduced to demonstrate preprocessing.

Methodology

1. **Data Preprocessing**
 - Removed duplicates and missing values.
 - Corrected typos and standardized categories.
 - Applied one-hot encoding for categorical variables.
2. **Model Training**
 - Split dataset into training (80%) and testing (20%).
 - Trained a Decision Tree Classifier with controlled depth to avoid overfitting.
3. **Evaluation**
 - Measured accuracy, precision, recall, and F1-score.
 - Visualized the decision tree to interpret feature importance.
4. **Deployment**
 - Saved trained model (`decision_tree_transport.pkl`).
 - Built Streamlit app (`app.py`) for user interaction.
 - Deployed on Streamlit Cloud for public access.

Models

Decision Tree

- Trained with `max_depth=5` for interpretability.

- Captures non-linear relationships.
- Provides **feature importance** for interpretability.

Logistic Regression

- Trained with `max_iter=1000`.
- Provides **coefficients per class**.
- Good for linear separability and probability outputs.

Model Comparison

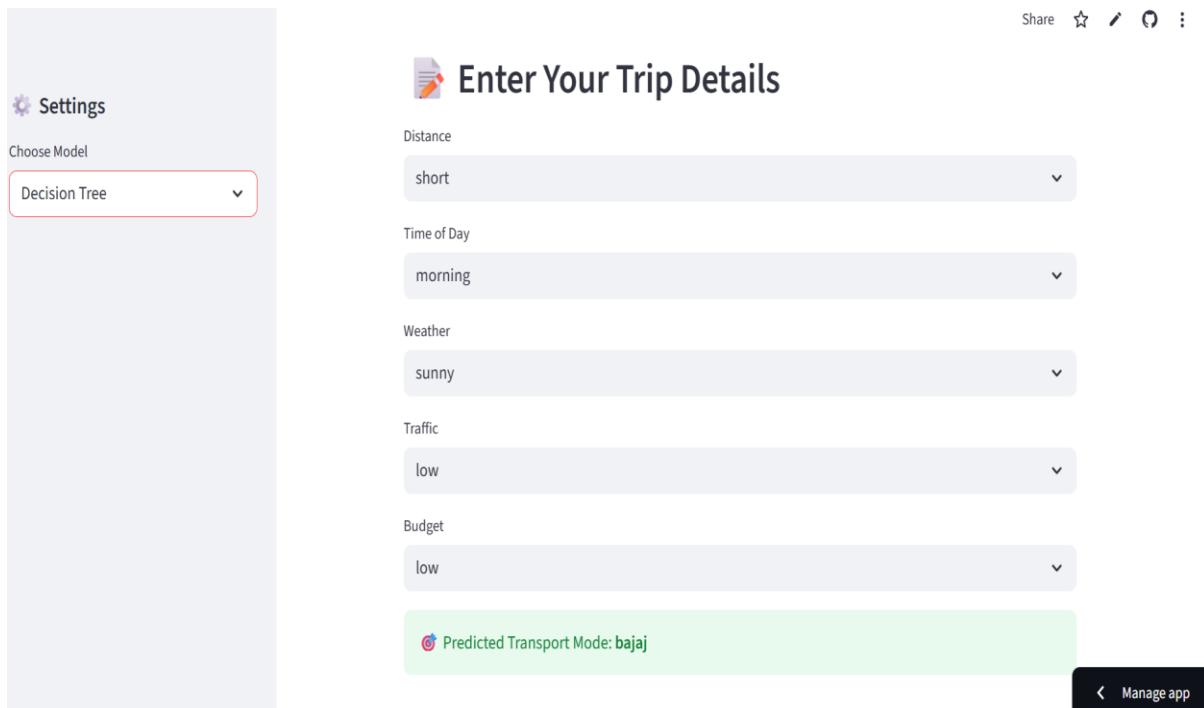
Model	Accuracy	Strengths
Decision Tree	~0.70–0.75	Easy to interpret, feature importance
Logistic Regression	~0.65–0.72	Stable, probability-based predictions

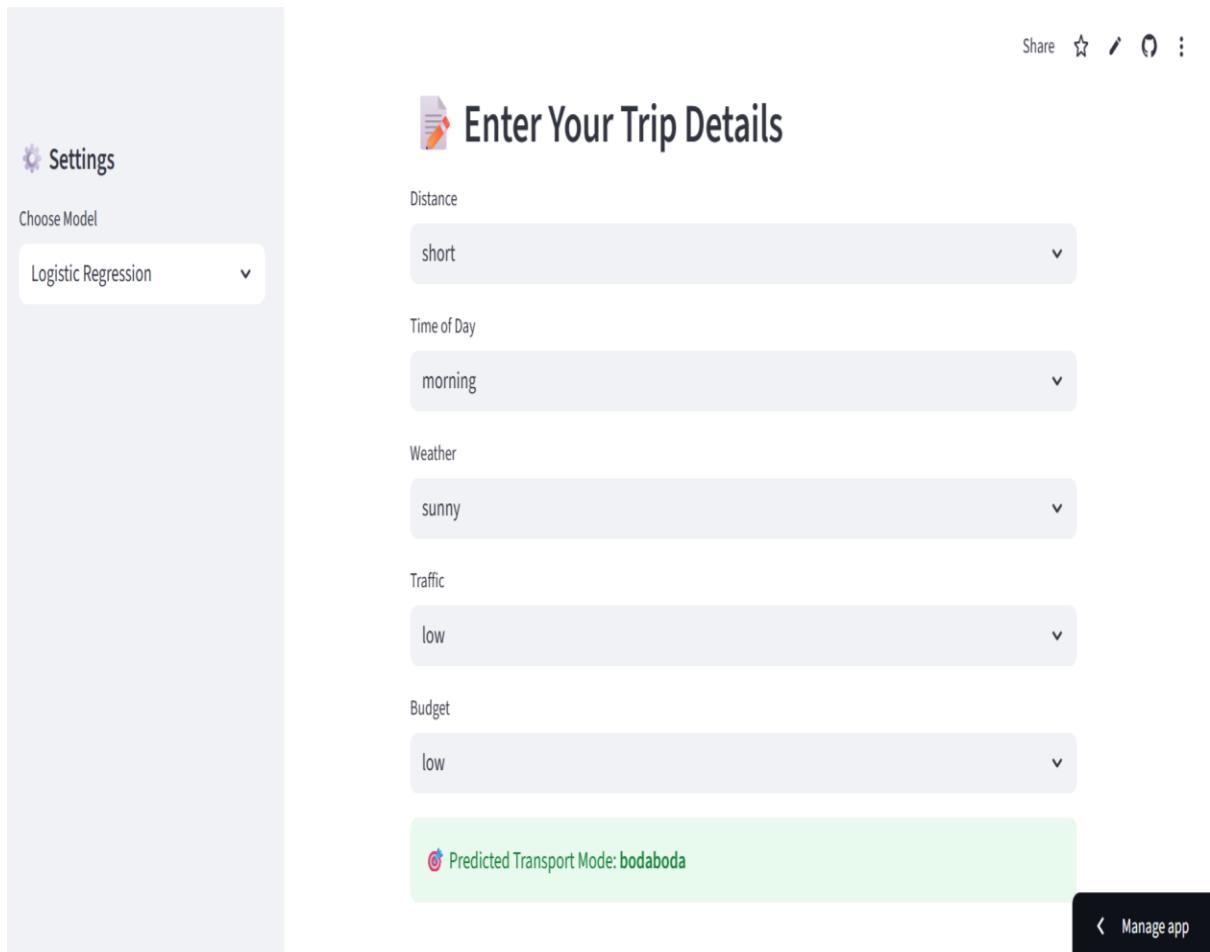
Streamlit App

Features:

- Sidebar for **model selection** (Decision Tree or Logistic Regression).
- User-friendly input fields for trip details.
- Animated interface using **Lottie** for professional look.
- Prediction displayed with styled success box.

Consider the screenshots of app





Conclusion

This project highlights the potential of Machine Learning in addressing urban transport challenges.

- Both models perform reasonably well, with **Decision Tree** slightly stronger in interpretability.
- Logistic Regression provides **probability-based outputs** useful for uncertainty analysis.
- The Streamlit app makes predictions accessible to non-technical users in a professional interface.

Application URL: ↗ <https://ml-transport-laprojectore-yf6wakwb4uepz7efxx7blb.streamlit.app/> (ml-transport-project.streamlit.app in Bing)