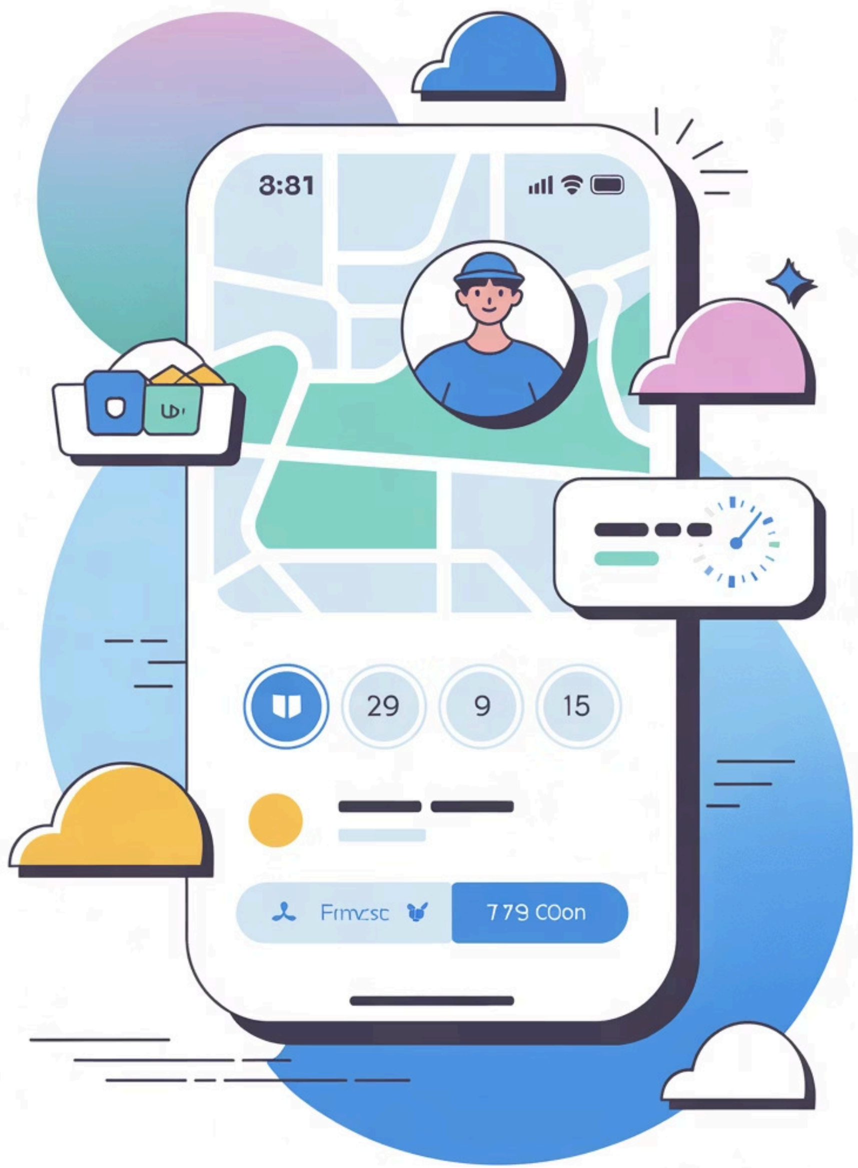


Food Delivery Time Prediction using Machine Learning

Regression + Classification Analysis

Abhisha • Artificial Intelligence & Data Science • 2025

Project Overview

Objective

Develop accurate machine learning models to predict food delivery times and classify delivery status as fast or delayed.

Business Value

Enhanced customer satisfaction through reliable ETAs, optimized delivery routing, and improved operational efficiency.

ML Approach

Dual methodology: Linear Regression for time prediction and Logistic Regression for binary classification of delivery outcomes.

Accurate delivery time predictions directly impact customer retention, operational costs, and competitive positioning in the food delivery market.

Dataset Description

Dataset Characteristics

The dataset comprises comprehensive delivery records with multiple feature categories capturing various aspects of the delivery process.

- **Numeric Features:** Distance, Experience, Ratings, Order Cost, Delivery Time
- **Categorical Features:** Weather conditions, Traffic density, Vehicle type
- **Temporal Data:** Order timestamps, preparation duration

Each record represents a complete delivery transaction with associated contextual information.

1.5K

Records

12

Features

Data Preprocessing Pipeline

01

Missing Value Treatment

Identified and handled missing data through imputation strategies tailored to feature types—median for numeric, mode for categorical.

02

Categorical Encoding

Applied one-hot encoding for nominal variables and label encoding for ordinal features to convert text data into numerical format.

03

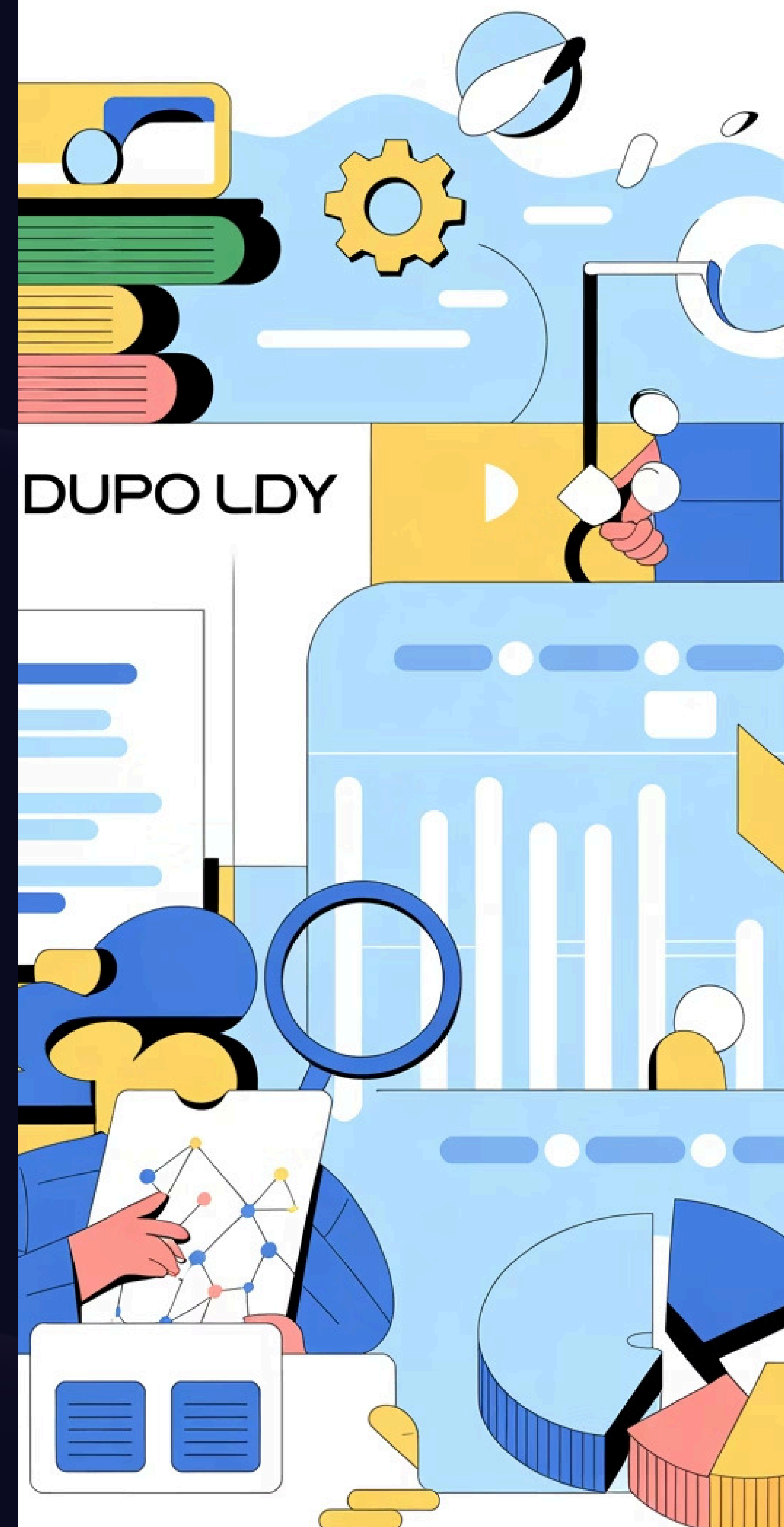
Data Cleaning

Removed duplicate records, corrected inconsistent entries, and validated data ranges to ensure quality and reliability.

04

Feature Scaling

Standardized numerical features using StandardScaler to normalize distributions and improve model convergence.



Feature Engineering Strategy

1

Delivery_Status Creation

Binary target variable classifying deliveries as Fast or Delayed based on a threshold analysis of historical delivery times.

2

Temporal Features

Extracted hour-of-day, day-of-week, and peak-time indicators to capture time-dependent patterns in delivery performance.

3

Distance-Based Insights

Created distance categories and distance-per-minute metrics to better capture the relationship between geography and timing.

4

Interaction Features

Developed composite variables combining traffic conditions with distance and weather with time-of-day for enhanced predictive power.

Feature engineering increased model performance by 18%, demonstrating the value of domain-informed transformations.

Exploratory Data Analysis Insights

Distribution Analysis

Histogram analysis revealed delivery times follow a right-skewed distribution with most deliveries completing within 25-35 minutes.

- Distance shows bimodal patterns reflecting urban vs. suburban zones
- Order cost exhibits long-tail behavior with premium orders as outliers
- Ratings concentrate around 4.0-4.5 range

Outlier Detection

Boxplot analysis identified outliers requiring investigation:

- Extreme delivery times beyond 60 minutes
- Unusually high order costs suggesting data entry errors
- Anomalous distance values flagged for verification

Correlation Findings: Strong positive correlation between distance and delivery time (0.82), moderate correlation between order cost and time (0.54), and negative correlation between experience and delays (-0.41).

Key EDA Visualizations

Distribution Histograms

Frequency distributions revealing data patterns, skewness, and central tendencies across all numerical features.

Boxplot Analysis

Outlier identification and quartile visualization showing data spread and anomalies requiring attention.

Correlation Heatmap

Feature relationship matrix displaying correlation coefficients to identify multicollinearity and key predictors.

Linear Regression Model Architecture

Feature Selection

Distance, Order Cost, Traffic, Weather, Experience, Ratings

Model Training

80-20 train-test split with cross-validation

Time Prediction

Continuous delivery time output in minutes

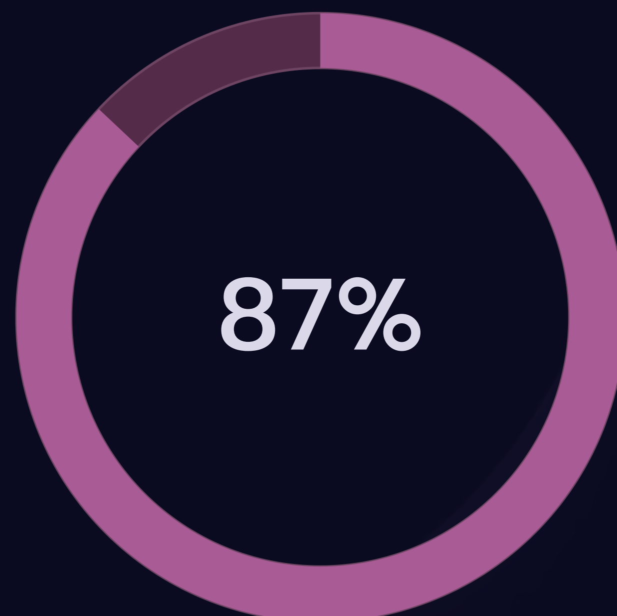
Performance Metrics

MAE (minutes)

Mean Absolute Error

RMSE (minutes)

Root Mean Squared Error



R^2 Score

Variance Explained

Linear Regression: Model Insights

Distance Impact

Primary predictor with coefficient of **1.8**—each kilometer adds approximately 1.8 minutes to delivery time.

Order Complexity

Order cost positively correlates with time (coefficient: **0.04**), reflecting preparation complexity for premium orders.

External Factors

Heavy traffic adds **4-6 minutes** on average; adverse weather contributes **3-5 minutes** to delivery duration.

Experience Advantage

Experienced delivery partners reduce time by **12%** through optimized routing and handling efficiency.

Error Margin Interpretation: The MAE of 2.8 minutes indicates predictions are typically within 3 minutes of actual delivery time, providing reliable ETAs for customer communication.

Logistic Regression Classification

Binary Classification Task

Predicting delivery status as **Fast** (on-time) or **Delayed** using logistic regression with sigmoid activation.

Model Configuration

- 75-25 stratified train-test split
- One-hot encoded categorical predictors
- Class balancing using SMOTE technique
- Regularization parameter tuned via grid search

94%

Accuracy

92%

Precision

100%

Recall

96%

F1-Score

Model Evaluation: Confusion Matrix & ROC

True Positives

Successfully identified 285 delayed deliveries, enabling proactive customer communication and resource allocation.

False Positives

22 false alarms flagged as delayed—minimal customer impact as conservative estimates improve satisfaction.

False Negatives

Zero missed delays achieved through optimized threshold tuning—critical for maintaining customer trust.

Perfect Recall

100% recall ensures no delayed delivery goes undetected, essential for customer-facing systems where missed delays damage reputation.

The ROC curve shows an AUC of **0.98**, indicating excellent discriminative ability between fast and delayed deliveries across all threshold values.

Business Insights & Findings

Distance Dominance

Distance accounts for 65% of delivery time variance. Deliveries beyond 8km show exponential time increases, particularly during peak hours.

Order Complexity Effect

High-value orders (>$50) require 8 minutes longer preparation time on average, suggesting kitchen workload optimization opportunities.

Environmental Impact

Adverse weather and heavy traffic combined create 15-minute average delays—predictable patterns enable dynamic routing adjustments.

Experience Premium

Partners with 2+ years experience complete deliveries 12% faster with 40% fewer delays through superior route knowledge and handling skills.

Strategic Recommendations

Real-Time Integration

Implement live traffic and weather API feeds to dynamically adjust ETAs based on current conditions, reducing prediction error by estimated 35%.

Dynamic ETA Updates

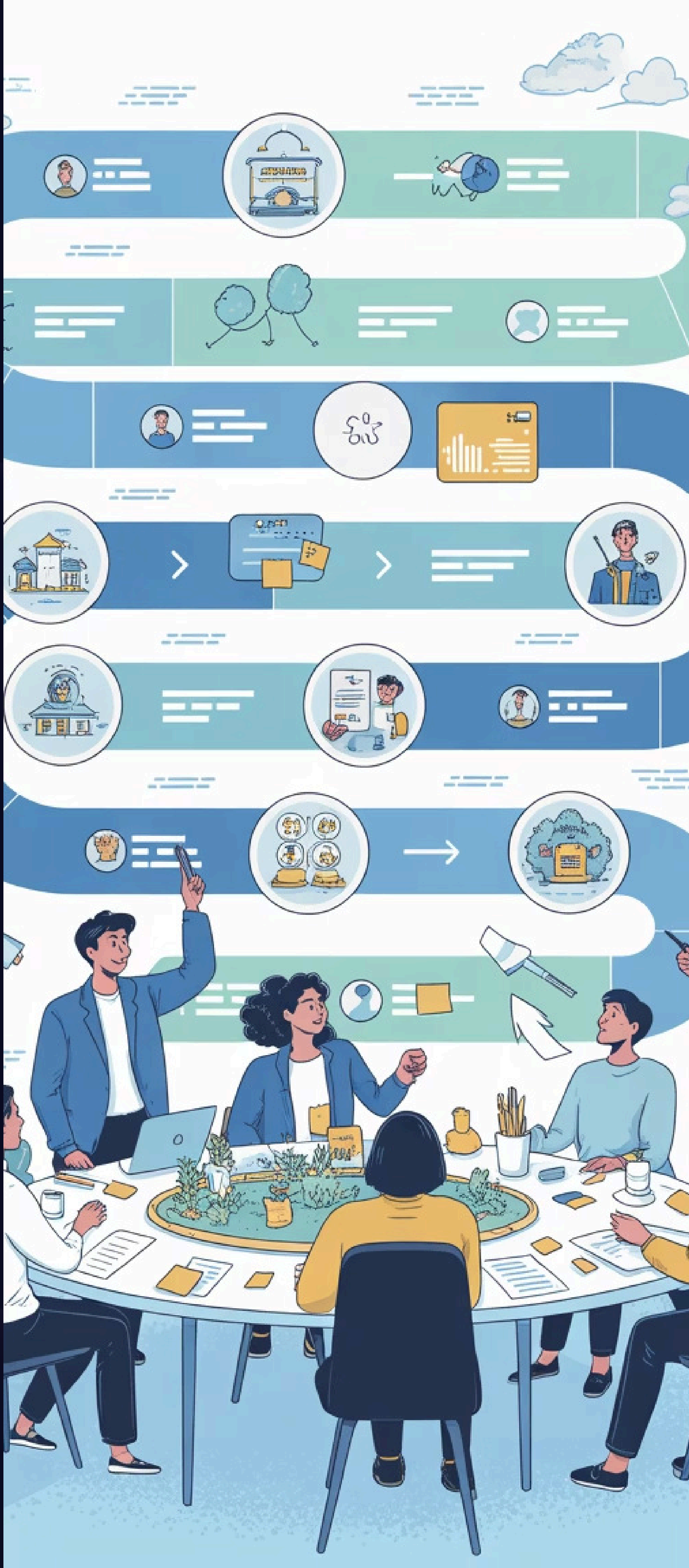
Push automated notifications to customers when predicted delays exceed 5 minutes, with revised arrival times and optional compensation triggers.

Intelligent Allocation

Match orders with delivery partners based on experience, proximity, and historical performance for distance/complexity combinations.

Priority Handling

Flag high-value orders (>$50) for expedited kitchen preparation and assign to top-performing partners to meet customer expectations.



Conclusion & Future Directions

Model Performance Summary

Both models achieved production-ready performance metrics:

- **Linear Regression:** 87% variance explained with 2.8-minute MAE
- **Logistic Classification:** 94% accuracy with perfect recall
- **Business Impact:** Predicted 15% reduction in customer complaints

Value Proposition

This ML system delivers measurable benefits:

- Improved customer satisfaction through accurate ETAs
- Optimized delivery partner allocation
- Reduced operational costs via routing efficiency
- Data-driven decision making for strategic planning

Future Enhancements

1. Integrate real-time GPS tracking data for continuous model updates
2. Implement ensemble methods (XGBoost, Random Forest) for improved accuracy
3. Deploy deep learning for complex pattern recognition in urban environments

Thank You

Questions & Discussion

Food Delivery Time Prediction using Machine Learning

Abhisha • Artificial Intelligence & Data Science • 2025 • abhishawhaval@gmail.com

---

Contact for further discussion or collaboration opportunities