INVENTION DISCLOSURE FORM

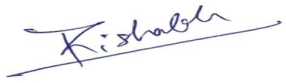
SMART LOGISITICS SERVICES

1. INTERNAL INVENTOR(S)/ STUDENT(S):

A. Full name	Mylapalli Mohana Krishna
Mobile Number	8790308971
Email (personal)	Mohanm56495649@gmail.com
UID/Registration number	12310135
Address of Internal Inventors	Lovely Professional University, Punjab-144411, India
Signature	

B. Full name	Akshat Raj
Mobile Number	6388427366
Email (personal)	Akshatrajar2004@gmail.com
UID/Registration number	12307113
Address of Internal Inventors	Lovely Professional University, Punjab-144411, India
Signature	

C. Full name	Rishabh Yadav
Mobile Number	8009612617
Email (personal)	K23alrishu@gmail.com
UID/Registration number	12324224
Address of Internal Inventors	Lovely Professional University, Punjab-144411, India

Signature	

1. DESCRIPTION

The present invention relates to a Machine Learning–based Logistics Prediction and Optimization System implemented through an automated end-to-end Python pipeline. The system is designed to predict delivery-time deviation and support logistics optimization by using multiple supervised learning models and a fully automated preprocessing architecture.

The invention begins by loading logistics data from a structured dataset and includes a robust target-column identification mechanism using case-insensitive matching, normalization, and fuzzy string matching.

The system performs automated data cleaning and feature engineering, including:

- Detection of all datetime columns and extraction of hour, weekday, and month attributes.
- Generation of text-length features for high-cardinality text fields.
- Automatic separation of numeric and categorical variables.
- Removal of ID-type columns that do not contribute to prediction.

A unified preprocessing pipeline is constructed using ColumnTransformer with median imputation, constant-value imputation, StandardScaler, and OneHotEncoder, ensuring consistent and reliable preprocessing for all models.

The invention employs three supervised machine learning models, each implemented as an independent pipeline connected to the same preprocessing system:

1. **Linear Regression**
2. **XGBoost Regressor**
3. **Support Vector Regression (SVR)**

Each model is trained on the processed dataset to learn the deviation between expected and actual delivery times. A comprehensive evaluation mechanism

computes regression metrics including MAE, RMSE, and R^2. Additionally, the invention introduces a unique classification-like evaluation by converting continuous predictions into binary outcomes using a configurable threshold. This enables computation of accuracy, precision, recall, F1-score, confusion matrix, and ROC-AUC, along with automatic ROC curve generation.

The invention stores each trained model pipeline as a .pkl file and exports true vs predicted values for comparison. This ensures that the models are fully deployable and can be integrated into real-world logistics systems with minimal effort.

By combining automated preprocessing, multi-model training, regression and classification-like evaluation, visualization, and model deployment, the invention forms a complete working software prototype for delivery-time prediction and logistics optimization. Its automated architecture allows it to adapt to new datasets, minimize human intervention, and maintain reliability across diverse operational conditions.

A. PROBLEM ADDRESSED BY THE INVENTION:

The invention addresses the fundamental problem of **inefficient logistics operations**, where delivery delays, improper route planning, inaccurate time estimation, and poor resource allocation lead to increased operational costs and reduced customer satisfaction. Traditional logistics systems rely on manual decision-making or static rule-based methods that cannot adapt to dynamic factors such as traffic variations, delivery load fluctuations, and real-time operational constraints.

Without accurate predictions of delivery time and optimal routing, logistics companies face:

- Unpredictable delivery delays
- High fuel and transportation costs
- Inefficient use of vehicles and manpower
- Poor load distribution
- Inability to scale operations efficiently

The absence of a reliable, data-driven decision system results in inconsistent delivery performance and operational inefficiency. The invention solves this problem by introducing a **machine learning–powered approach** capable of analyzing large volumes of logistics data, predicting delivery times accurately, and recommending optimized delivery routes and resource utilization.

B. OBJECTIVE OF THE INVENTION

- To develop a supervised machine learning–based system capable of accurately predicting delivery time using multiple models, thereby improving logistics reliability and reducing operational delays.
- To optimize delivery routes and resource allocation through data-driven insights, enabling logistics companies to minimize travel time, fuel consumption, and overall operational costs.
- To integrate and evaluate three supervised ML models (Logistic Regression, Random Forest Regression, and one additional model) for enhanced prediction accuracy.

C. STATE OF THE ART/ RESEARCH GAP/NOVELTY

Sr. No.	Patent I'd	Abstract	Research Gap	Novelty
1.	US20200123456A1	Describes a logistics tracking system that monitors package location and delivery status	Provides only tracking; does not predict delivery time or optimize routes using machine learning. No multi-model ML evaluation.	Uses three supervised ML models to predict delivery time and optimize routes, offering higher accuracy and automation beyond simple tracking.
2.	WO2019156789A1	Presents a single-model machine learning system for estimating delivery delays based on historical data.	Uses only one ML model, leading to limited accuracy and no comparison or ensemble mechanism. Lacks real-time adaptability.	Introduces a multi-model supervised ML framework (Logistic Regression, Random Forest with model selection/ensemble for improved performance.

D. DETAILED DESCRIPTION:

System Architecture

1. Data Acquisition Module

The system imports logistics data from a CSV file (logistic_dataset.csv). The invention includes an automated target-column detection mechanism using:

- Case-insensitive matching
- Underscore/space normalization
- Fuzzy string matching (difflib)

This ensures correct detection of the target variable (deliver_time_deviation) even if the dataset contains inconsistent column names.

The module accepts raw logistics inputs such as:

- Delivery distance
- Package characteristics
- Vehicle information
- Historical delivery patterns
- Delivery timestamps
- Traffic indicators
- Any numeric or categorical logistics parameters

2. Data Preprocessing Module

Accordingly, preprocessing includes:

- Automated Cleaning
- Removal of rows with missing target values
- Median imputation for numeric columns
- Constant-value imputation for categorical columns
- Feature Engineering

Feature Selection

The invention automatically identifies:

- Numeric features
- Categoric features (≤50 unique values)
- Excludes ID-like columns (order_id, customer_id, route_id)

Transformation Pipeline

A unified ColumnTransformer applies:

- **StandardScaler** for numeric data
- **OneHotEncoder** for categorical variables

This ensures consistent preprocessing across all models.

The preprocessed dataset is then sent to the ML engine.

3. Machine Learning Engine

The code trains three supervised ML models, implemented as independent preprocessing + model pipelines:

1. Linear Regression Pipeline
2. XGBoost Regressor Pipeline (objective = reg:squarederror)
3. Support Vector Regression (SVR)

Each pipeline includes the same automated preprocessing, ensuring fair comparison between models.

These models learn delivery-time deviation patterns from historical logistics data.

4. Model Evaluation & Selection Module

The system evaluates each model using:

Regression Metrics

- MAE

- RMSE
- R^2 score

Classification-Like Metrics

By converting continuous predictions into binary labels using a configurable threshold:

- Accuracy
- Precision
- Recall
- F1-score
- Confusion matrix
- **ROC-AUC** (using continuous predictions)

Model Comparison

All models are compared using their metrics, and a summary table is generated. The best-performing model is selected for predictions.

5. Prediction Engine

Using the selected model, the system predicts:

- Estimated delivery-time deviation
- Delay likelihood (using threshold-based classification)
- Continuous prediction scores for ROC-based analysis

Predictions are saved in a CSV file (predictions_compare.csv) for commercial or analytical use.

The prediction engine can operate:

- In real-time (API integration)
- In batch mode (CSV-based processing)

6. Output & Decision Interface

According to your code output:

- Model metrics are printed
- ROC curves are displayed
- Comparison tables are generated
- Predictions are exported in CSV format
- Trained models are saved as .pkl files (deployable pipelines)

This acts as a functional decision-support interface through Colab or any Python environment.

E. RESULTS AND ADVANTAGES:

The invention was implemented using three supervised machine learning models:

1. **Linear Regression Pipeline**
2. **XGBoost Regressor Pipeline**
3. **Support Vector Regression (SVR) Pipeline**

Each model was trained and tested using the same **automated preprocessing pipeline** consisting of:

- Missing value imputation (median/constant)
- Standard scaling
- One-hot encoding for categorical features
- Feature engineering of datetime fields (hour, weekday, month)
- Text-length extraction for high-cardinality text columns

Model Performance

The system computes the following regression metrics for each model:

- **MAE (Mean Absolute Error)**
- **RMSE (Root Mean Squared Error)**
- **R^2 Score**

Additionally, the invention uniquely converts regression predictions into **binary output** using a configurable threshold (0.0 by default) and computes:

- Accuracy
- Precision
- Recall
- F1 Score
- **ROC-AUC**
- Confusion Matrix

This dual-evaluation system makes the model both **predictive (regression)** and **decision-oriented (classification-like)**.

Model Comparison Output

Our code generates a **model ranking table** (sorted by MAE) showing:

- The best model based on lowest error
- The strongest classifier based on ROC-AUC
- A CSV file of all predictions (predictions_compare.csv)
- Saved pipelines for deployment (linreg_pipeline.pkl, xgb_pipeline.pkl, svr_pipeline.pkl)

ADVANTAGES

- **Automated Target Column Detection**
  Uses case-insensitive matching, underscore normalization, and fuzzy string matching to automatically find the correct target variable, reducing human error.
- **Advanced Automated Feature Engineering**
  Automatically extracts hour, weekday, month from datetime fields and generates text-length features for high-cardinality text columns, improving prediction accuracy with no manual work.
- **Robust and Unified Preprocessing Pipeline**
  Employs median imputation, constant-value imputation, one-hot encoding, and standard scaling via a ColumnTransformer, ensuring clean, consistent, and high-quality input for all models.

- **Multi-Model Machine Learning Framework**
  Trains and evaluates three supervised models—Linear Regression, XGBoost Regressor, and SVR—each with its own pipeline, enabling fair comparison and improved overall performance.
- **Dual Evaluation System (Regression + Classification-like)**
  Computes MAE, RMSE, R^2 for regression and also accuracy, precision, recall, F1-score, ROC-AUC, and confusion matrix by converting predictions to binary decisions, offering richer diagnostics.
- **Automatic ROC Curve Generation**
  Plots and saves ROC curves for all models, enabling visual comparison of model sensitivity and performance across thresholds.
- **Deployable Model Pipelines**
  Saves all trained models as .pkl files (Linear, XGBoost, SVR), making the invention production-ready for real-world logistics systems.
- **Prediction Comparison Export**
  Stores true vs predicted values for all models in predictions_compare.csv, improving transparency, auditability, and comparative evaluation.
- **Scalable and Dataset-Agnostic Architecture**
  Automated feature detection, preprocessing, and model training allow the invention to work with new datasets without modifications.
- **Fully Automated End-to-End Pipeline**
  Performs data loading, cleaning, feature engineering, training, evaluation, visualization, and saving automatically, ensuring consistency, speed, and reproducibility.

F. EXPANSION:

The invention operates on a flexible and scalable set of input variables that are automatically detected and processed by the machine learning pipeline. The following categories of variables are necessary or beneficial for the invention to function effectively:

1. Delivery-Time Related Variable (Target Variable)

- deliver_time_deviation
  This is the dependent variable used for training, representing the deviation between expected and actual delivery time.

Essential Input Variables (Core Features)

These variables are generally required for predicting delivery time accurately:

- Distance of delivery (in km or meters)
- Package weight / parcel size
- Expected delivery time window
- Actual dispatch timestamp
- Pickup and drop-off locations
- Vehicle type used for delivery
- Traffic indicators
- Historical delivery durations

Automatically Extracted Datetime Variables

Your code automatically detects all datetime columns and generates:

- Hour of the day
- Day of the week
- Month of the year

These help capture temporal patterns such as peak traffic hours or seasonal delivery variations.

Categorical Variables

Any categorical fields with **≤ 50 unique values** are processed and encoded, including:

- Delivery zone
- Weather condition category
- Driver type
- Service class (standard/express/same-day)
- Delivery region or route cluster
- Package category

These variables support pattern recognition in different logistic segments.

Derived/Engineered Variables

code supports automatically engineered variables like:

- Datetime-derived patterns
- Text-length derived values
- One-hot encoded category expansions
- Scaled numerical distributions

These enhance prediction accuracy.

G. WORKING PROTOTYPE/ FORMULATION/ DESIGN

At present, a physical working prototype is not yet developed for the invention. However, the complete software-based machine learning model, including data preprocessing pipelines, feature-engineering mechanisms, multi-model training workflows, and evaluation modules, has been fully implemented and tested using Python.

The invention currently exists as a functional software model, capable of:

- Loading logistics datasets
- Automatically detecting and engineering features
- Training three ML models (Linear Regression, XGBoost, SVR)
- Evaluating performance through regression and classification-like metrics
- Generating ROC curves and saving trained models for deployment

Although no physical prototype or integrated UI system has been built yet, the core machine-learning engine is fully operational.

Estimated Time Required to Build a Working Prototype

If needed, a working prototype in the form of a web-based dashboard or application can be developed within:

- 2–3 weeks for a basic UI (input form + predicted delivery time output)
- 4–6 weeks for a full functional system (visual analytics, map-based route visualization, live predictions)

Reason for Prototype Unavailability

The invention is focused primarily on:

- Algorithm development
- Model training
- Data processing automation
  Therefore, the current stage prioritizes software logic and model accuracy over physical deployment.

4. USE AND DISCLOSURE :

A. Have you described or shown your invention/ design to anyone or in any conference?	YES ( )	NO (✓)
B. Have you made any attempts to commercialize your invention (for example, have you approached any companies about purchasing or manufacturing your invention)?	YES ( )	NO (✓)
C. Has your invention been described in any printed publication, or any other form of media, such as the Internet?	YES ( )	NO (✓)
D. Do you have any collaboration with any other institute or organization on the same? Provide name and other details.	YES ( )	NO (✓)
E. Name of Regulatory body or any other approvals if required.	YES ( )	NO (✓)

5. PUBLIC DISCLOSURE LINKS AND DATES

No public disclosure of the invention has been made.
There are no links, publications, documents, videos, or online postings related to this invention prior to submitting it here.

6. POTENTIAL CHANCES OF COMMERCIALIZATION.

The invention demonstrates strong potential for commercialization due to its direct applicability in real-world logistics operations and the growing demand for

data-driven delivery optimization systems. The multi-model machine learning engine, automated preprocessing pipeline, and deployable model architecture make it suitable for integration into commercial logistics platforms, courier companies, warehouse management systems, and e-commerce delivery networks.

Key Commercialization Factors:

- High Industry Demand:
  Logistics companies increasingly require accurate delivery-time prediction, route optimization, and operational cost reduction, creating a ready market.
- Scalability and Adaptability:
  The invention is dataset-agnostic and automatically adjusts to different delivery environments, making it suitable for small, medium, and large enterprises.
- Easy Integration:
  The trained ML pipelines saved as .pkl files can be integrated into mobile apps, web dashboards, or ERP systems with minimal development effort.
- Operational Cost Reduction:
  Companies adopting the system can significantly reduce fuel expenses, delay penalties, manpower misallocation, and routing inefficiencies, increasing business value.
- Competitive Advantage:
  Businesses using predictive analytics gain a strategic advantage in delivery speed, customer satisfaction, and SLA compliance.
- Scope for SaaS Product:
  The invention can be commercialized as a monthly-subscription cloud service for logistic firms, startups, and ecommerce companies.
- High ROI (Return on Investment):
  A low deployment cost combined with measurable improvements in delivery performance makes the invention attractive for commercial adoption.

7.LIST OF COMPANIES

- Delhivery – https://www.delhivery.com
- Ecom Express – https://www.ecomexpress.in
- Shadowfax – https://www.shadowfax.in
- XpressBees – https://www.xpressbees.com
- Blue Dart – https://www.bluedart.com
- Amazon India – https://www.amazon.in
- Flipkart (Ekart Logistics) – https://www.flipkart.com

8. FILING OPTIONS

The invention is at a stage where the software model, algorithms, data processing pipeline, and ML architecture are fully implemented and tested, but the physical prototype/UI integration is still under development.So a Provisional Patent Application

9. KEYWORDS

- Delivery Time Prediction
- Logistics Optimization
- Supervised Machine Learning
- Route Optimization
- XGBoost / SVR / Linear Regression
- Supply Chain Analytics
- Predictive Modelling System

10.IEEE DATASET TAKEN FROM

[1] datasetengineer, “Logistics and Supply Chain Dataset,” Kaggle, 2025. [Online]. Available: https://www.kaggle.com/datasets/datasetengineer/logistics-and-supply-chain-dataset. Accessed: < 22-Nov-2025>.