FOOD DELIVERY TIME ESTIMATION
Machine Learning Based Regression Model

Harshana Yadav
Ayushi Rathore

Business Problem

Food delivery platforms struggle with:

- Inaccurate delivery time estimates
- Poor customer satisfaction
- Order cancellations
- Delivery partner inefficiencies

01

02

03

Why This Matters?

Inaccurate delivery predictions lead to:

- Reduced customer trust
- Increased operational cost
- Lower retention rates

Business Impact

If solved effectively, the system can:

- 📈 Increase customer satisfaction
- 🚚 Optimize delivery partner allocation
- 📉 Reduce late deliveries
- 💰 Improve operational efficiency

Business Value:

Accurate ETA → Higher Retention → Higher Revenue

Dataset Overview

Dataset contains order, delivery partner, and location-based features.

Feature	Description
Delivery Partner Age	Age of delivery executive
Delivery Partner Rating	Rating score (1–5)
Restaurant Latitude/Longitude	Pickup location
Delivery Latitude/Longitude	Drop location
Type of Order	Food category
Type of Vehicle	Bike, Scooter

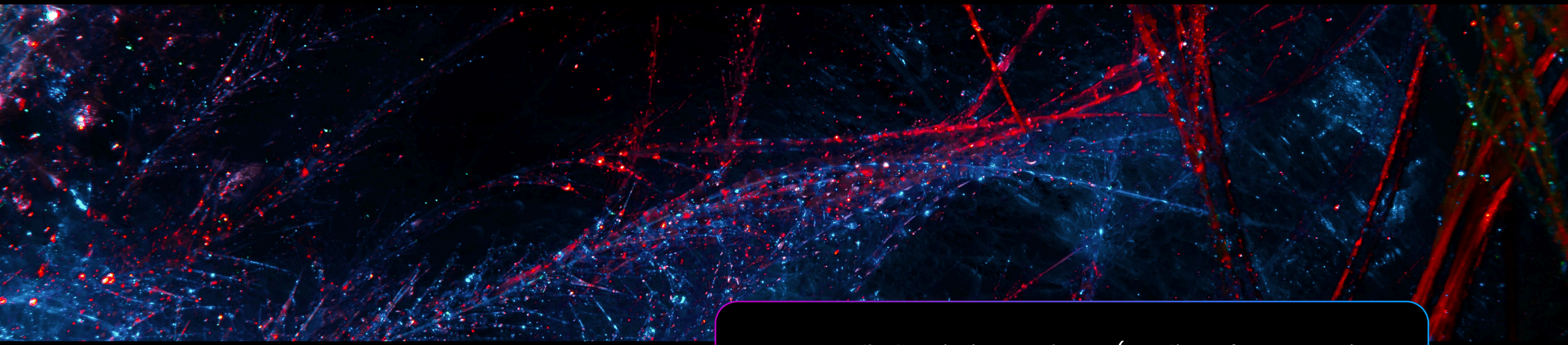
Exploratory Data Analysis (EDA)

Observed patterns:

- Delivery time increases with distance
- Lower-rated delivery partners slightly increase delays
- Vehicle type impacts delivery speed
-
- Age of partner has minimal effect

Visualizations used:

- Correlation heatmap
- Distance vs Delivery Time scatter plot
- Boxplots for vehicle types
- Distribution of delivery times



Data Preprocessing

- Handled missing values (median for numeric, mode for categorical).
- Removed duplicates.
- Outliers treated using IQR method.
- Encoded categorical variables using One-Hot Encoding.
- Scaled numerical features using StandardScaler.
- Split dataset into train/test (80-20).

Feature Engineering

- Distance Calculation: Haversine formula between restaurant & delivery locations.
- Interaction Features:
- Distance × Vehicle Type
- Rating × Distance
- Encoded categorical variables for vehicle and order type.
- Features added to sklearn Pipeline for reproducibility.

Model Selection & Architecture

- Models trained and compared:
- Linear Regression
- Random Forest Regressor
- Gradient Boosting Regressor
- XGBoost
- LightGBM
- 5-fold cross-validation used for robust evaluation.
- Hyperparameter tuning performed using RandomizedSearchCV.
- Pipeline architecture ensures reproducibility and easy deployment.

Evaluation Metrics

- Metrics used to evaluate models:
- MAE (Mean Absolute Error) – average prediction error in minutes
- RMSE (Root Mean Squared Error) – penalizes large errors
- R^2 Score – model's explanatory power

```
Mean Absolute Error (MAE): 3.15
Mean Squared Error (MSE): 15.65
Root Mean Squared Error (RMSE): 3.96
R-squared (R2) Score: 0.82
```

- Feature importance analysis shows the top drivers of delivery time:

1. Distance between restaurant and delivery location
2. Vehicle type
3. Delivery partner rating
4. Interaction: Distance × Vehicle
5. Type of Order

- SHAP analysis used for tree-based models for additional interpretability.

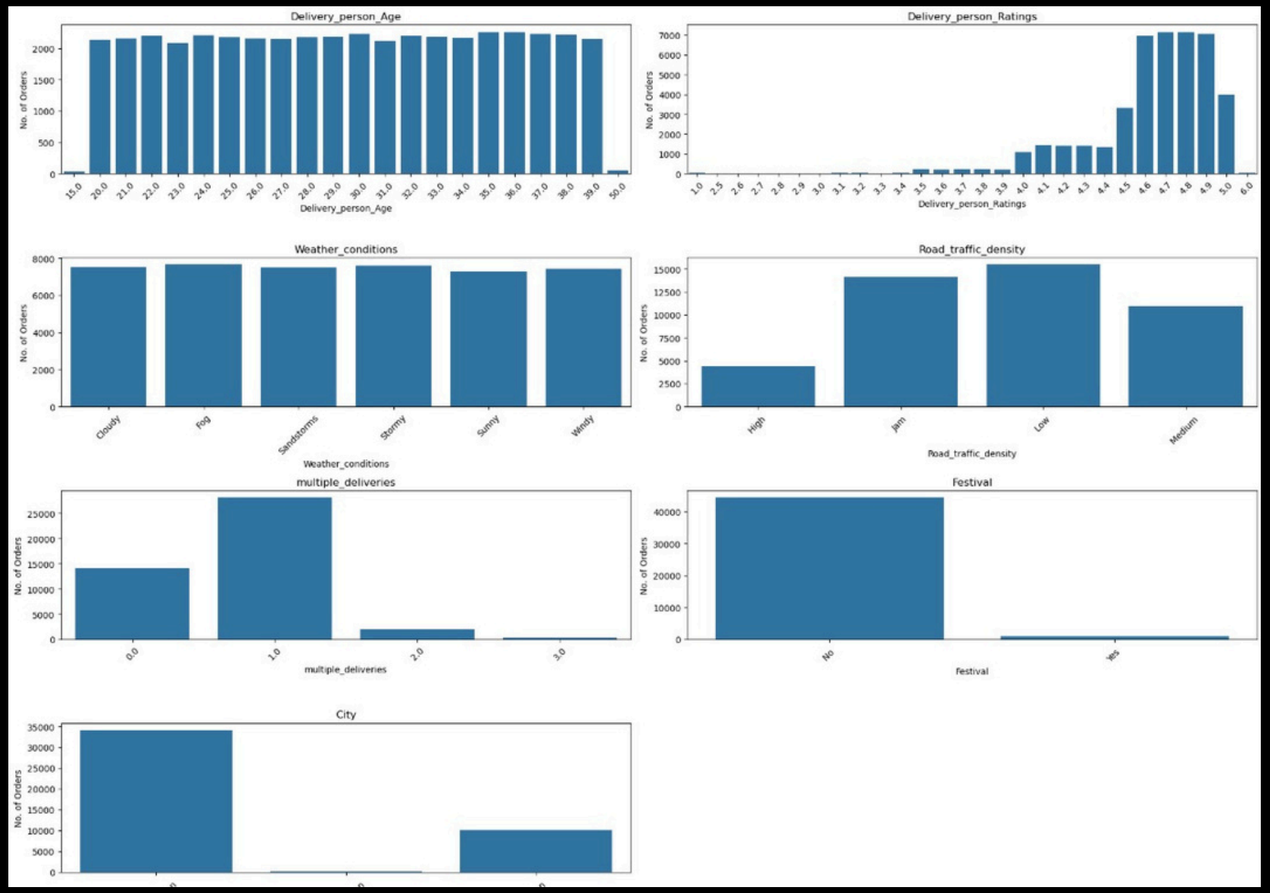
Model Explainability

```
LinearRegression
Best parameters:  {}
Best R2 score:  0.42195334486780717

DecisionTreeRegressor
Best parameters:  {'max_depth': 7}
Best R2 score:  0.7170948555927096

RandomForestRegressor
Best parameters:  {'n_estimators': 300}
Best R2 score:  0.811648817798079

XGBRegressor
Best parameters:  {'max_depth': 7, 'n_estimators': 20}
Best R2 score:  0.8212592396880698
```

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=9, max_leaves=None,
             min_child_weight=None, missing=nan, monotone_constraints=None,
             multi_strategy=None, n_estimators=20, n_jobs=None,
             num_parallel_tree=None, random_state=None, ...)
```

	count	mean	std	min	25%	50%	75%	max
Restaurant_latitude	45593.0	17.017729	8.185109	-30.905562	12.933284	18.546947	22.728163	30.914057
Restaurant_longitude	45593.0	70.231332	22.883647	-88.366217	73.170000	75.898497	78.044095	88.433452
Delivery_location_latitude	45593.0	17.465186	7.335122	0.010000	12.988453	18.633934	22.785049	31.054057
Delivery_location_longitude	45593.0	70.845702	21.118812	0.010000	73.280000	76.002574	78.107044	88.563452


Delivery_person_Age
No. of Orders
Delivery_person_Age
Delivery_person_Ratings
No. of Orders
Delivery_person_Ratings
Weather_conditions
No. of Orders
Cloudy
Fog
Sandstorms
Stormy
Sunny
Windy
Weather_conditions
Road_traffic_density
No. of Orders
High
Jam
Low
Medium
Road_traffic_density
multiple_deliveries
No. of Orders
multiple_deliveries
Festival
No. of Orders
No
Yes
Festival
City
No. of Orders

	count	mean	std	min	25%	50%	75%	max
Restaurant_latitude	45593.0	17.017729	8.185109	-30.905562	12.933284	18.546947	22.728163	30.914057
Restaurant_longitude	45593.0	70.231332	22.883647	-88.366217	73.170000	75.898497	78.044095	88.433452
Delivery_location_latitude	45593.0	17.465186	7.335122	0.010000	12.988453	18.633934	22.785049	31.054057
Delivery_location_longitude	45593.0	70.845702	21.118812	0.010000	73.280000	76.002574	78.107044	88.563452

ons	Road_traffic_density	Vehicle_condition	Type_of_order	Type_of_vehicle	multiple_deliveries	Festival	City	Time_taken(min)
nny	High	2	Snack	motorcycle	0	No	Urban	(min) 24
rmy	Jam	2	Snack	scooter	1	No	Metropolitian	(min) 33
ons rms	Low	0	Drinks	motorcycle	1	No	Urban	(min) 26
nny	Medium	0	Buffet	motorcycle	1	No	Metropolitian	(min) 21
udy	High	1	Snack	scooter	1	No	Metropolitian	(min) 30

	Delivery_person_Age	Delivery_person_Ratings	Restaurant_latitude	Restaurant_longitude	Delivery_location_latitude	Delivery_locatio
0	37.0	4.9	22.745049	75.892471	22.765049	
1	34.0	4.5	12.913041	77.683237	13.043041	
2	23.0	4.4	12.914264	77.678400	12.924264	
3	38.0	4.7	11.003669	76.976494	11.053669	
4	32.0	4.6	12.972793	80.249982	13.012793	

Future Improvements

- Include real-time traffic and weather data for dynamic predictions
- Peak hour prediction and alerts
- Restaurant preparation time feature
- Deep learning models for further performance gains
- Dashboard for delivery analytics

Conclusion

- *Built an end-to-end ML pipeline for delivery time prediction*
- *Engineered key features like distance and interaction terms*
- *Trained multiple regression models and selected XGBoost as best*
- *Deployed a Streamlit app for real-time predictions*
- *Solution is scalable, interpretable, and reliable*
- *Ready for hackathon demo and business impact*