



The Implementation of Predictive Modelling to Forecast the Duration of Car Sales





The Implementation of Predictive Modelling to Forecast the Duration of Car Sales

Presented by Datacrats

Overview

Data-driven solution that predicts the estimated time to sell a car on Auto Trader, helping sellers and dealerships optimize sales while reinforcing Auto Trader's position as a trusted marketplace.

Impact on Auto Trader's Ecosystem

For Sellers:

Optimize pricing and listing strategies for faster sales.

For Dealerships:

Optimize inventory management and pricing strategies to maximize profitability.

For Auto Trader:

Enhance market leadership with innovative data-driven tools.

Regressor Used:
**Gradient Boosted Trees
(GBT)**

GBTs Utilized

CatBoost

Handles categorical variables without encoding, reduces overfitting.

Why GBT?

Tree-ensembling based on decision trees

Combines weak learners to create stronger models

Reduced bias and variance

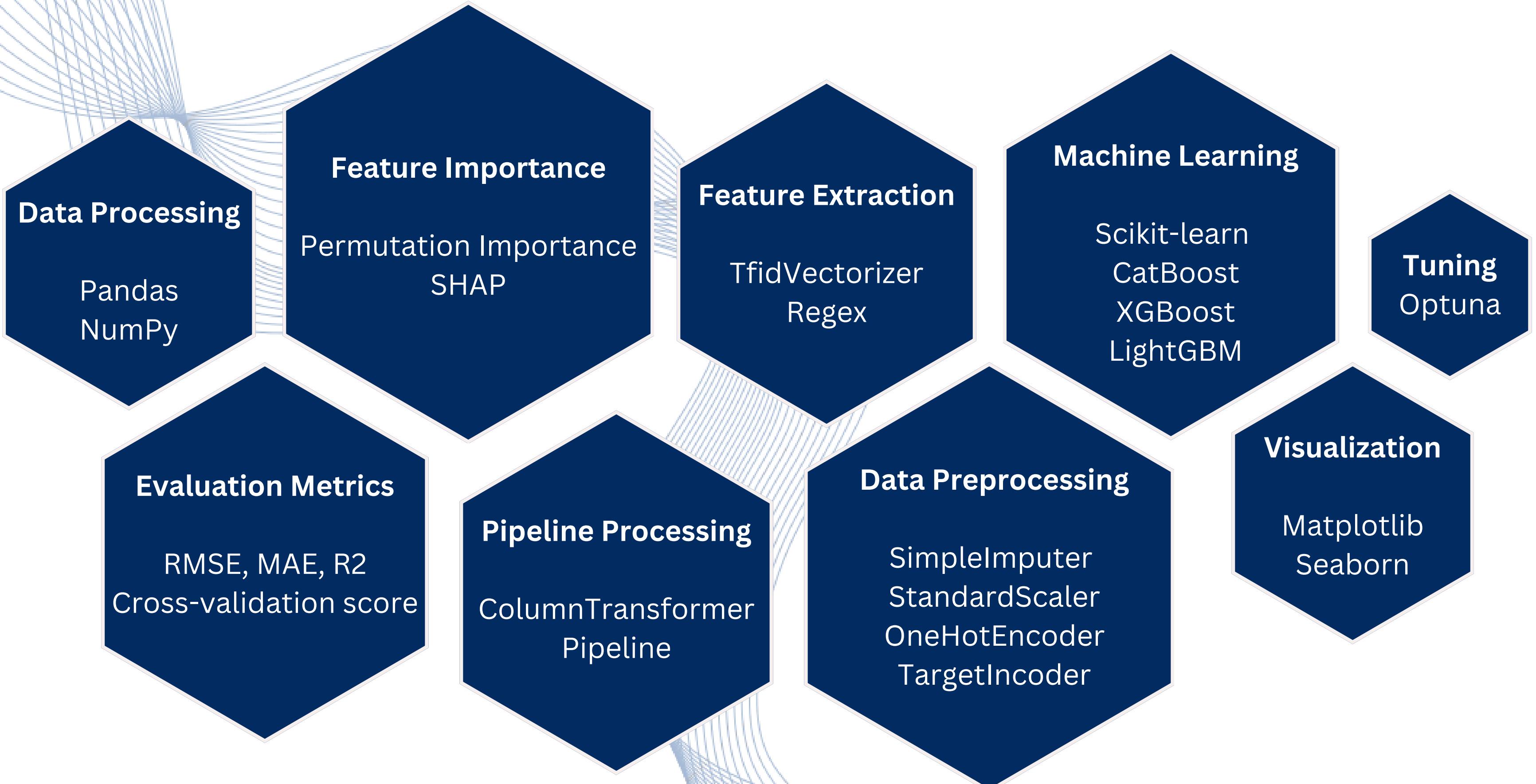
XGBoost

- Optimized gradient boosting with built-in regularization
- Scalable and fast

Light GBM

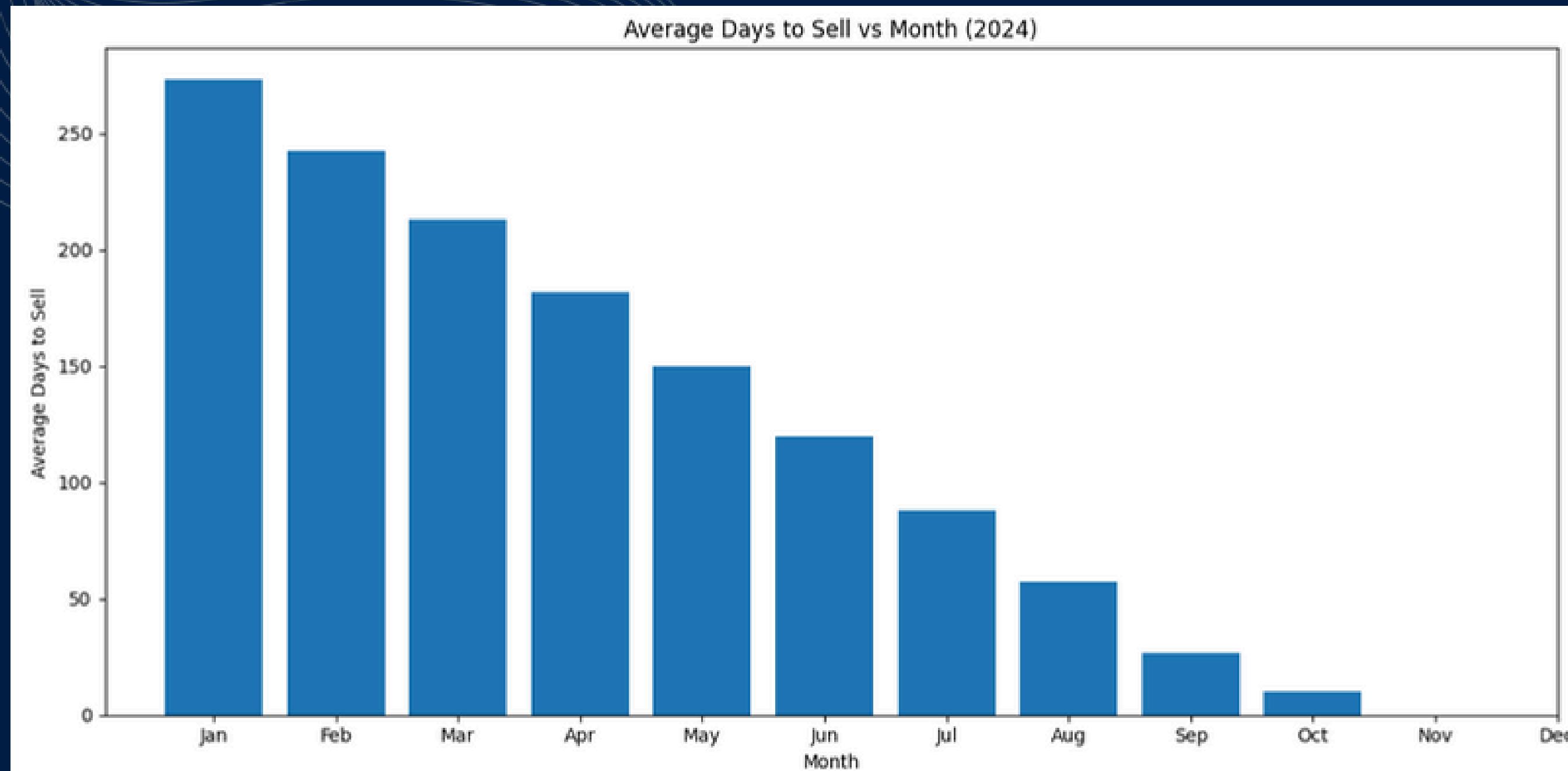
- Smart sampling and smart feature engineering
- Memory efficient and fastest

Tech Stack, Tools and Libraries



Exploratory Data Analysis

First seen date and days to sell show a linear relationship—if a car isn't sold within a certain period, it's likely to stay listed longer. This pattern holds for almost all the years.



Data Preprocessing

1 Data Cleaning

- **Imputation Techniques:**
 - By grouping relevant features
 - Stratified filling
 - Feature extraction
- **Consistency Checks:**
 - Data-type mismatches
 - Duplicates
 - Anomalies

3 Feature Engineering

- Binning
- Feature Creation
- Feature Selection
- Column-type Conversions

2 Outlier Detection & Handling

- **Detection Methods:**
 - **Numerical** - Boxplots, scatterplots
 - **Categorical** - Frequency-based analysis, barplots, histograms
- **Handling Strategies:**
 - **Numerical** - Quantiles, IQR, Z-score
 - **Categorical** - Distribution-based, grouping low-frequency categories

Model Building

1. Algorithms Used:

- ✓ XGBoost (XGB)
- ✓ CatBoost
- ✓ LightGBM

2. Tuning:

Optuna

3. Re-training:

Models trained with
best-found parameters

4. Evaluation:

RMSE, MAE, R2

Cross-validation score

5. Feature Importance:

- Feature Importance
- Permutation Importance
- SHAP

Dimensionality Reduction

Round 1

- Removed features with importance 0-100.
- Retained **27** key features.
- Predictions - **Not Compromised**

Round 3

- Removed features with importance 200-300.
- Retained **11** features.
- Predictions - **Compromised**

Round 2 - (Selected)

- Removed features with importance 100-200
- Retained **17** key features.
- Predictions - **Not Compromised**
- Refined the pipeline for Model Building
- Tested on Unknown Data – Ensured robustness & generalization

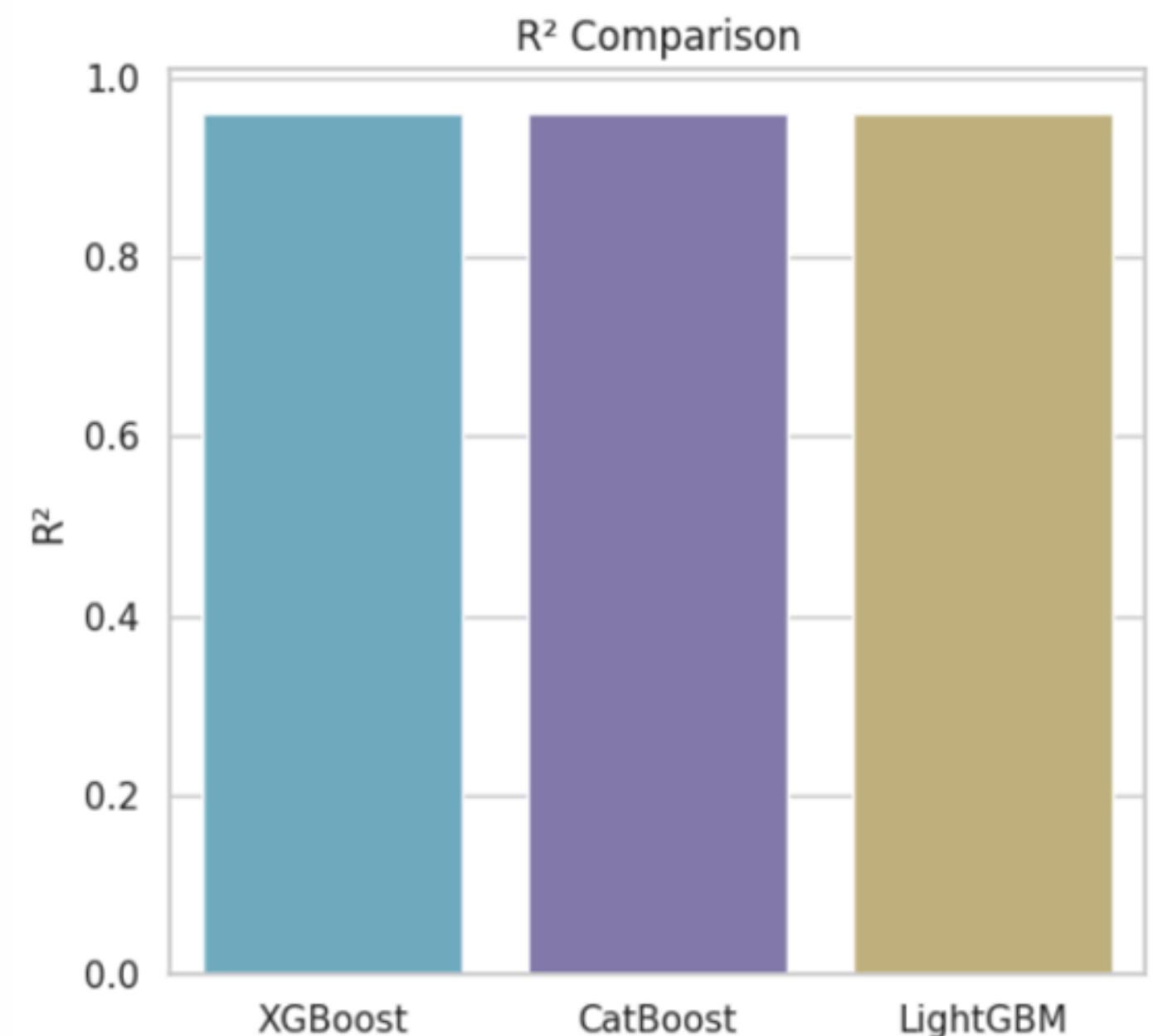
Model Evaluation and Selection

- Models performed similar in terms of accuracy
- Model Selected: CatBoost (best evaluation scores)

Evaluation Scores

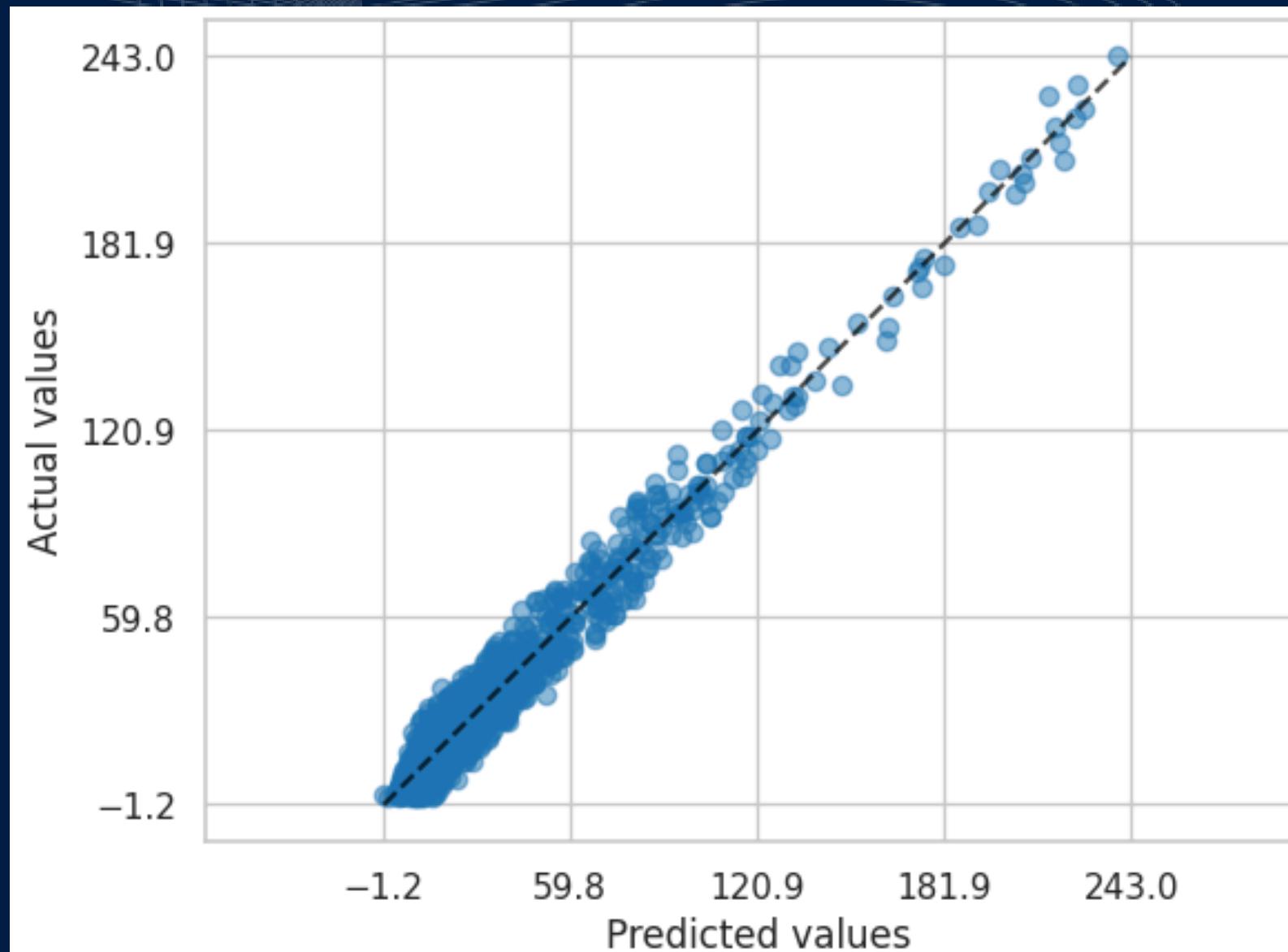
Model	MAE	RMSE	R ²
XGBoost	6.1956	7.4915	0.9594
CatBoost	6.0452	7.3714	0.9607
LightGBM	6.1200	7.4468	0.9599

R² Scores

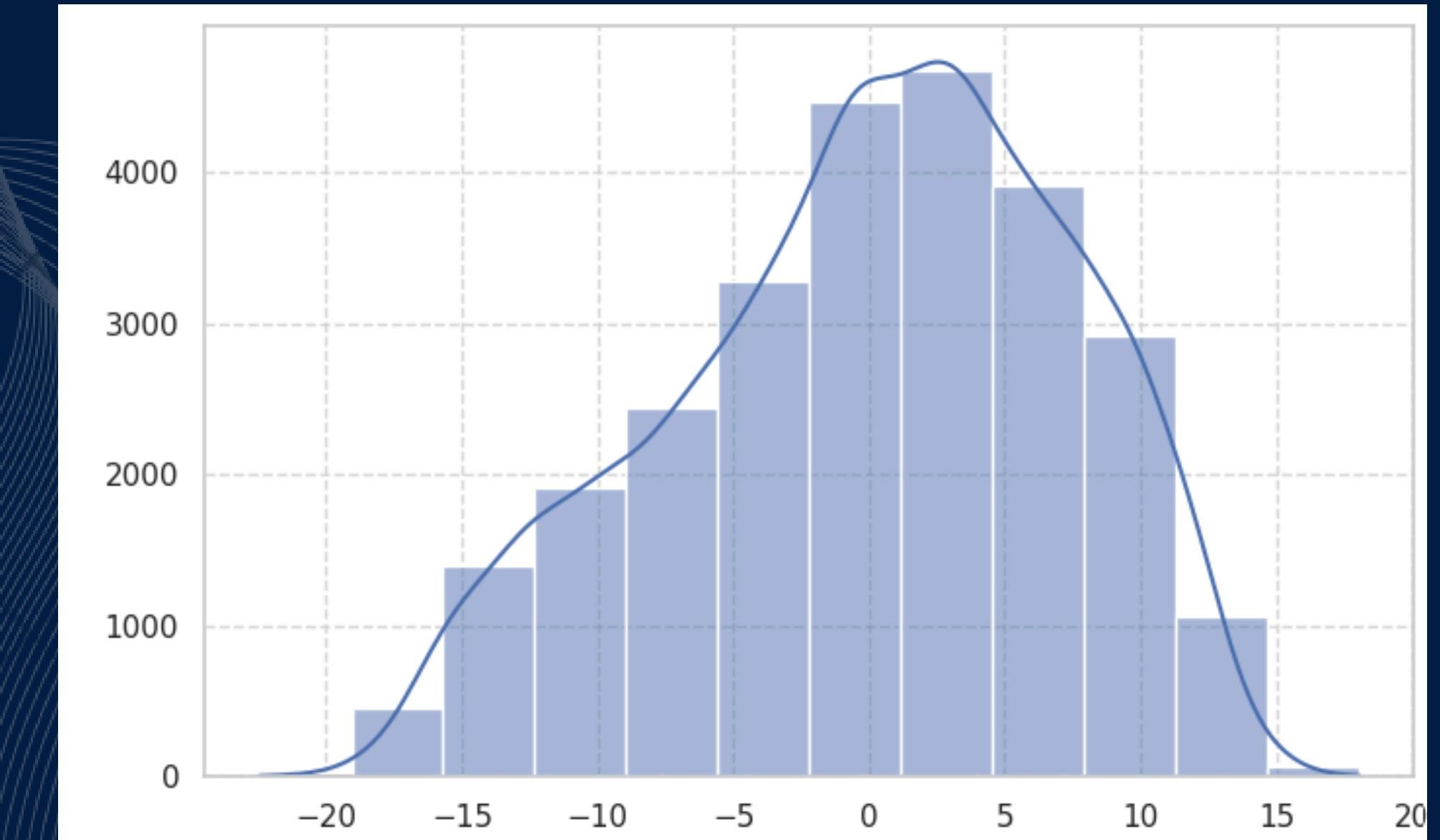


Residuals of CatBoost Model

Actual Vs Predicted Data Points



Residual Distribution



Results

Actual	Predicted	Residual (Actual - Predicted)
92	95.528864	-3.528864
205	204.307743	0.692257
9	8.844917	0.155083
39	45.723326	-6.723326
102	105.397369	-3.397369

Feature Insights

The Following features has more impact toward model decision:

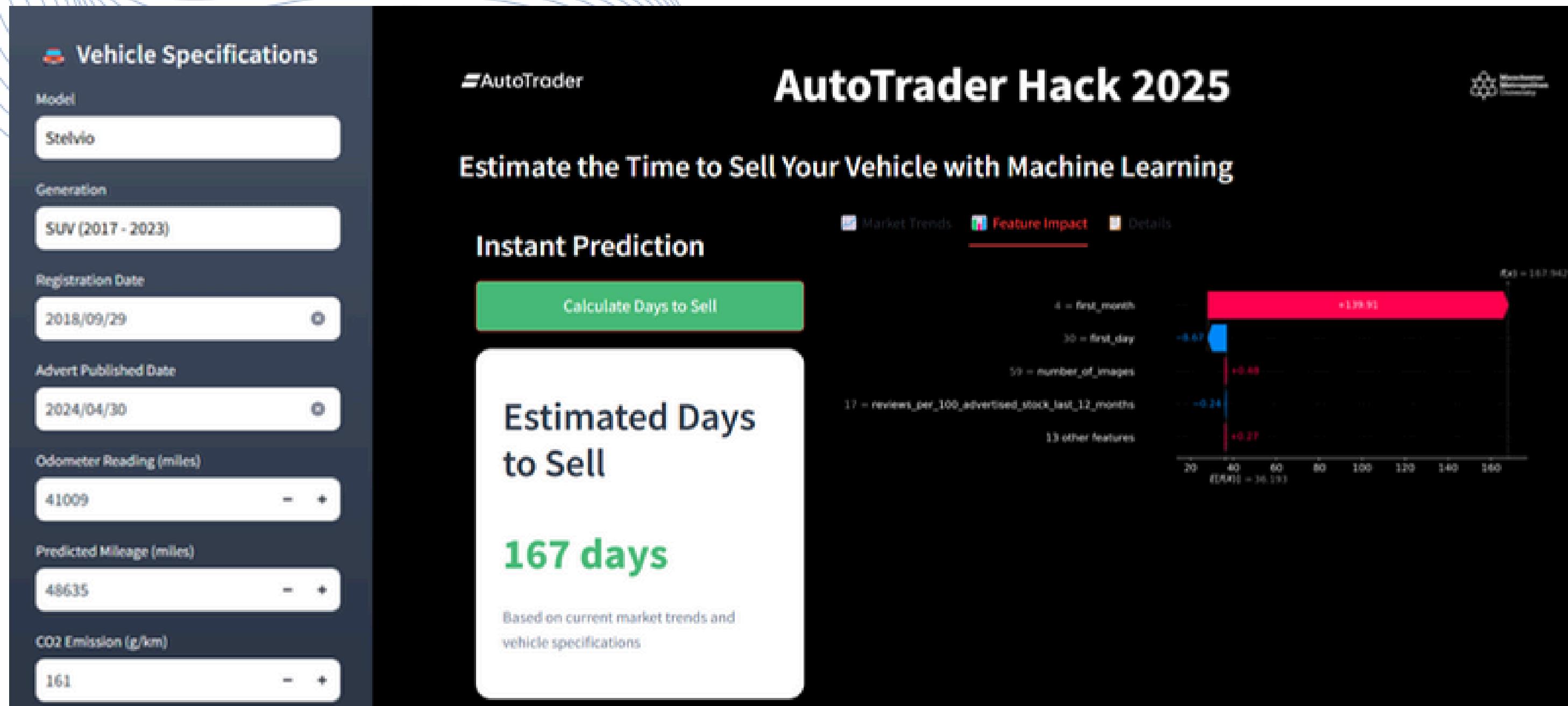
- First seen date
- Reviews
- Postcode Area

Business Impact

- Sales Optimization – Helps sellers set realistic expectations.
- Faster Decision-Making – Enables dealerships to optimize inventory.

Highlights

- StreamLit GUI - To interact with the model
- Recursive Dimensionality Reduction
- NLP Based TFIDF ranking for attention grabber
- Explainable AI



Demo

AutoTrader Hack 2025

Manchester Metropolitan University

Odometer Reading (miles)

20000 - +

Predicted Mileage (miles)

41332 - +

CO2 Emission (g/km)

164 - +

Fuel Economy (WLTP Combined mpg)

32 - +

Length (mm)

4263 - +

Boot Space (Seats Up, litres)

343 - +

First Retailer Asking Price (£)

25399 - +

Price Indicator Rating

5

AutoTrader

Estimate the Time to Sell Your Vehicle with Machine Learning

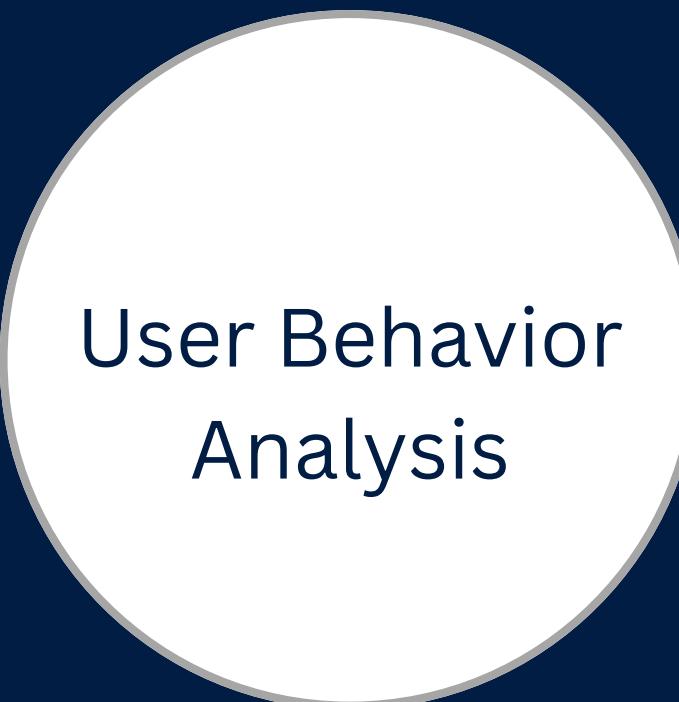
Market Trends Feature Impact Details

Instant Prediction

Calculate Days to Sell

This screenshot shows a web-based machine learning application for estimating the time to sell a vehicle. On the left, there's a sidebar with various vehicle parameters: Odometer Reading (miles), Predicted Mileage (miles), CO2 Emission (g/km), Fuel Economy (WLTP Combined mpg), Length (mm), Boot Space (Seats Up, litres), First Retailer Asking Price (£), and Price Indicator Rating. Each parameter has a numeric input field with minus and plus buttons for adjustment. The main area features the AutoTrader logo and the title 'AutoTrader Hack 2025'. Below the title is the subtitle 'Estimate the Time to Sell Your Vehicle with Machine Learning'. There are three tabs at the top right: 'Market Trends' (selected), 'Feature Impact', and 'Details'. In the center, there's a section titled 'Instant Prediction' with a large green button labeled 'Calculate Days to Sell'.

Future Scope



Learnings

- GBT excels in complex regression tasks.
- Data preprocessing is key to model performance.
- Optuna outperforms Grid Search CV for tuning.
- Permutation Importance suits non-tree models best.
- PCA works best for highly correlated data.

Challenges

- Data Cleaning
- Finding best hyper-params for each model
- Recursive Dimensionality Reduction
- Balancing model complexity and interpretability.

Summary

This project aimed to develop a robust machine learning model to predict days to sell a car. The approach involved several stages: data preprocessing, model selection, hyperparameter tuning, and feature importance analysis. We utilized three powerful models – XGBoost, CatBoost, and LightGBM – to determine the best one for our dataset.

Conclusion

After conducting a thorough evaluation, the best-performing model was selected based on key metrics such as R^2 , RMSE, and MAE. The chosen model showcased exceptional predictive accuracy, which was further validated through feature importance analysis, ensuring the relevance of the selected features. This solution provides Auto Trader with a powerful tool to enhance the platform's value proposition, offering significant benefits to all stakeholders involved.