# E-commerce Cart Abandonment Analysis

**Table of Contents**

## 1. Overview

This project addresses the challenge of cart abandonment in e-commerce using machine learning. By analyzing customer session behavior, we build and deploy a predictive model that estimates whether a user is likely to abandon their cart or complete the purchase. The model is integrated into a Streamlit web application for real-time decision-making, enhancing user engagement and boosting conversion rates.

### 2. Introduction

In today’s digital economy, online retailers face a major hurdle in converting potential buyers into customers. A significant percentage of users abandon their shopping carts before completing a purchase. Predictive analytics powered by machine learning can provide insights into user behavior and allow for preemptive engagement strategies. This project aims to build a complete AI/ML pipeline to detect and respond to potential cart abandonment.

### 3. Problem Statement

**High cart abandonment rates on the e-commerce platform result in lost sales and reduced customer lifetime value.**

The objective is to leverage behavioral session data to build a predictive model capable of identifying abandonment-prone sessions.

### 4. Business Value

- Improve overall conversion rate and revenue
- Trigger timely interventions like discounts or reminders
- Automate marketing and retention workflows
- Understand user behavior patterns to optimize UX

### 5. Dataset Description

The dataset used consists of anonymized e-commerce session logs with the following fields:

| Feature | Description |
|---|---|
| session_id | Unique identifier for each session |
| pages_visited | Number of pages visited |
| time_on_site | Time spent on site (in seconds) |
| cart_value | Total value of items in the cart (USD) |
| abandoned | Target variable (1 = abandoned, 0 = purchase) |

There are no missing values. Class imbalance was handled during model training using class_weight='balanced'.

## 6. Methodology

---

### 6.1 Data Preprocessing

- Verified data types, ranges, and nulls
- Scaled continuous features for clustering
- Used class weighting for model imbalance

---

### 6.2 Exploratory Data Analysis (EDA)

- Visualized relationships between time, cart value, and abandonment
- Used histograms, boxplots, and pairplots
- Identified lower engagement as a leading cause of abandonment

---

### 6.3 Model Building

- Models used:
  - Logistic Regression
  - Random Forest Classifier
- Split: 80% train / 20% test
- Used sklearn pipeline for reproducibility

---

### 6.4 Model Evaluation

- Evaluation Metrics:
  - Accuracy, F1-score, Precision, Recall
  - ROC Curve and AUC
- Feature Importance:
  - time_on_site > pages_visited > cart_value
- Random Forest outperformed Logistic Regression with ~69% accuracy

---

### 6.5 Unsupervised Learning

- Applied KMeans to identify user clusters
- Used PCA to reduce dimensions to 2D
- Helped visualize session behavior groups for segmentation

### 7. Deployment

The trained Random Forest model was saved as final_model.pkl and used in a **Streamlit app** for live predictions.

**App Features:**

- Inputs: pages visited, time on site, cart value
- Outputs:
  - **"Likely to Abandon Cart"** or **"Likely to Complete Purchase"**
  - **Probability Score** of prediction

### 8. Results and Observations

- Random Forest performed best on evaluation metrics
- Sessions with low engagement had significantly higher abandonment risk
- Clustering revealed distinct behavior segments
- Model predictions aligned with expected patterns from EDA

### 9. Conclusion

This project demonstrates an end-to-end ML workflow to solve a real-world business problem. The deployed app allows stakeholders to identify and act on at-risk sessions in real-time. The model can be further improved by incorporating features like user device, traffic source, or browsing history.

### 10. Tools and Technologies Used

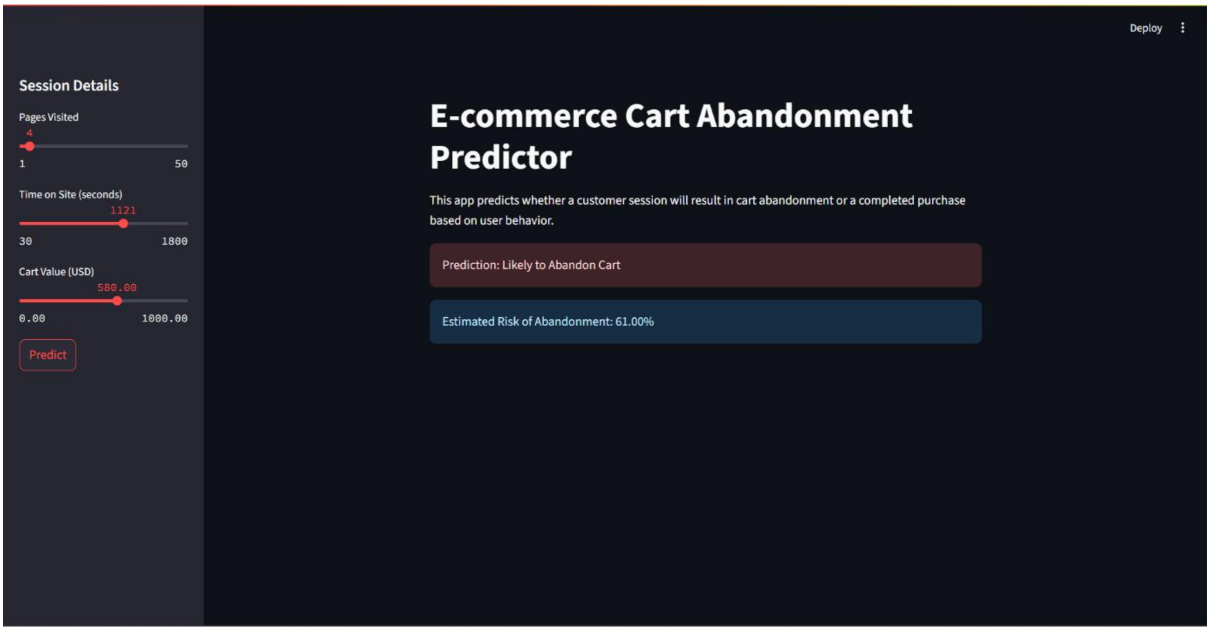
- **Language:** Python
- **Libraries:** pandas, numpy, matplotlib, seaborn, scikit-learn, joblib
- **Deployment:** Streamlit
- **Environment:** Jupyter Notebook, VS Code

### 11. References

- scikit-learn Documentation: https://scikit-learn.org/
- Streamlit Docs: https://docs.streamlit.io/
- Kaggle e-commerce behavior datasets
- Research blogs on cart abandonment prediction

## 12. Appendix

- *Streamlit App Interface*



**Session Details**

Pages Visited: 4 (1 – 50)

Time on Site (seconds): 1121 (30 – 1800)

Cart Value (USD): 580.00 (0.00 – 1000.00)

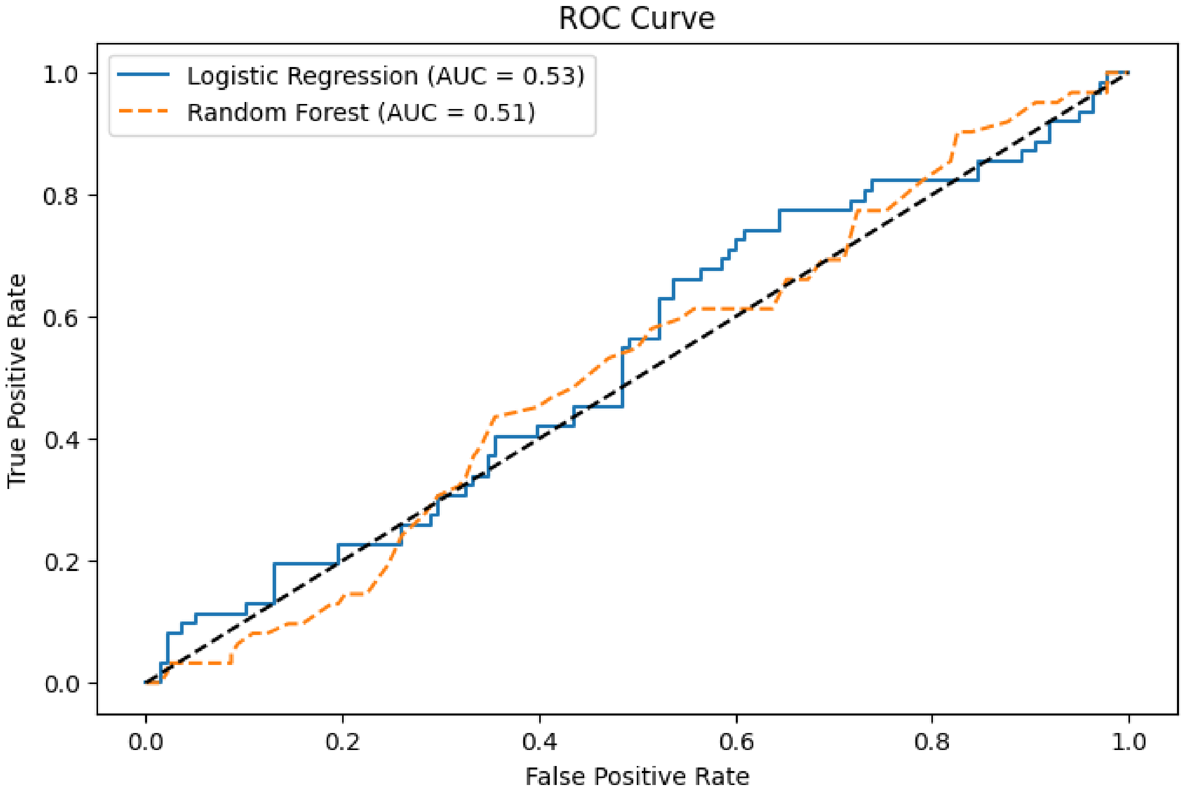
# E-commerce Cart Abandonment Predictor

This app predicts whether a customer session will result in cart abandonment or a completed purchase based on user behavior.

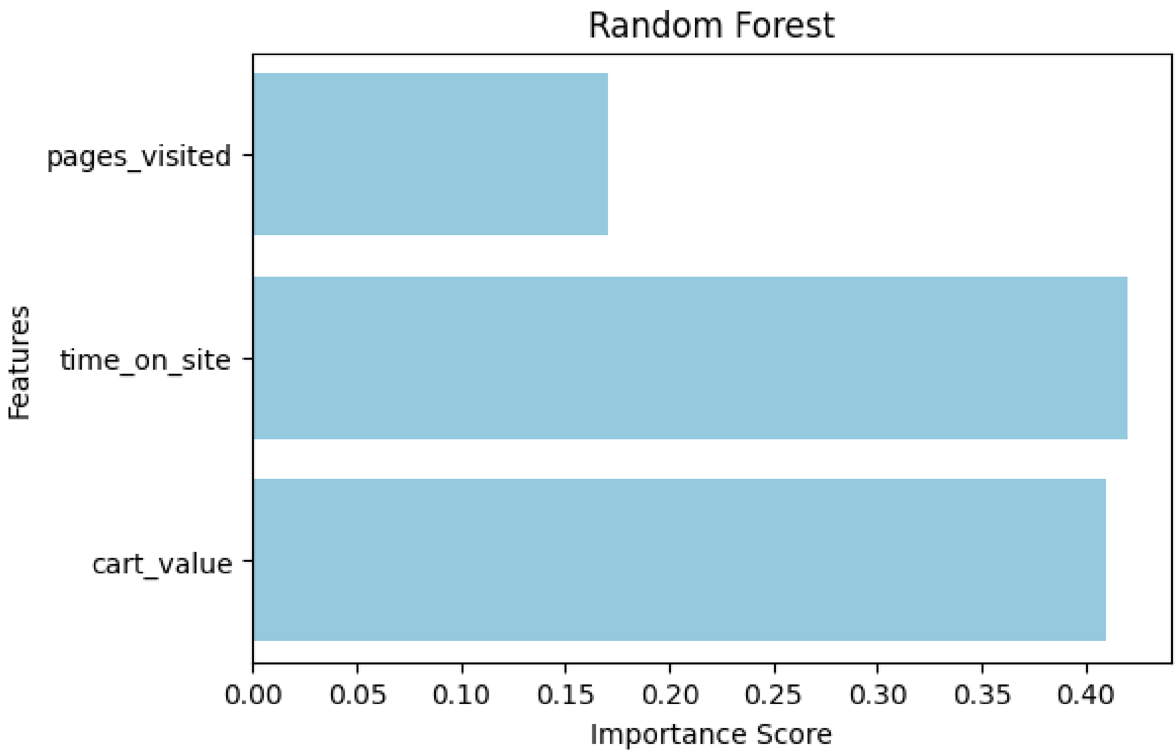
Prediction: Likely to Abandon Cart

Estimated Risk of Abandonment: 61.00%

- *ROC Curve*



- *Feature Importance Plot*



---

**13. Submitted By**

**Name:** Koushik Palakurthi
**Email:** koushik_palakurthi@srmap.edu.in
**University:** SRM University, AP
**Project Title:** E-commerce Cart Abandonment Analysis
**Submission Date:** 5th July 2025