

FLIPKART CUSTOMER SUPPORT ANALYSIS & PREDICTION REPORT

1. Introduction

In the competitive e-commerce industry, customer satisfaction plays a crucial role in customer retention and brand reputation. Flipkart, being one of India's leading e-commerce platforms, handles a large volume of customer support interactions every day. Analyzing these interactions and predicting customer satisfaction can help improve service quality, optimize operational efficiency, and enhance the overall customer experience.

This project focuses on performing **Exploratory Data Analysis (EDA)** and **Machine Learning (ML)** on Flipkart's customer support dataset to identify key insights and build a predictive model for Customer Satisfaction (CSAT) scores.

2. Dataset Overview

The dataset used in this project contains detailed records of Flipkart customer support interactions. Each row represents a single customer interaction, and the dataset includes information such as:

- Interaction channel (Inbound/Outbound)
- Issue category and sub-category
- Order details
- Agent information (agent name, supervisor, manager, tenure, shift)
- Handling time
- Item price
- Customer Satisfaction Score (CSAT)

The **CSAT Score**, ranging from 1 to 5, is used as the target variable to measure customer satisfaction.

3. Exploratory Data Analysis (EDA)

3.1 Data Understanding and Cleaning

The initial analysis involved checking the dataset structure, identifying missing values, and verifying duplicate records. Several columns contained missing values, especially customer remarks, item price, and handling time.

- Numerical missing values were handled using **median imputation** to reduce the impact of outliers.
- Columns with excessive missing values and low analytical importance, such as customer remarks, were removed.
- Column names were cleaned to avoid inconsistencies during analysis.

These steps ensured the dataset was clean and ready for analysis.

3.2 Data Visualization and Insights

A total of **15 meaningful visualizations** were created following the **UBM rule** (Univariate, Bivariate, and Multivariate analysis). Key insights include:

- **Inbound interactions** are significantly higher than outbound interactions, indicating that customers primarily reach out for support when facing issues.
- **Order-related and product-related issues** are the most common categories of customer complaints.
- **Experienced agents** tend to receive higher CSAT scores compared to newly onboarded agents.
- The **morning shift** handles a higher volume of customer interactions.
- **Longer handling times** show a slight negative impact on customer satisfaction.
- Supervisor and manager-level analysis indicates that leadership and team management influence customer satisfaction.
- Correlation analysis shows that CSAT scores are influenced by multiple factors rather than a single numerical variable.

These insights help Flipkart identify operational bottlenecks and improvement areas.

4. Hypothesis Testing

Three hypotheses were statistically tested to validate EDA findings:

1. **Agent experience impacts CSAT score**
2. **Handling time affects CSAT score**
3. **Interaction channel influences CSAT score**

Statistical tests such as **Independent T-Test** and **One-Way ANOVA** were used. The results confirmed that agent experience, handling time, and interaction channel have a statistically significant impact on customer satisfaction.

5. Feature Engineering & Preprocessing

To prepare the dataset for machine learning:

- Categorical variables were converted into numerical format using **Label Encoding**.
- Numerical features were scaled using **StandardScaler**.
- Irrelevant identifier columns were removed.
- Outliers in handling time were treated using the **IQR method**.
- The dataset did not require NLP preprocessing, as the textual column was removed due to high missing values.

An **80:20 train-test split** was used to evaluate model performance.

6. Machine Learning Model Implementation

The problem was framed as a **classification task**, where the goal was to predict the CSAT score.

6.1 Models Implemented

Three machine learning models were developed and evaluated:

1. **Logistic Regression** – Used as a baseline model
2. **Decision Tree Classifier** – Captured non-linear relationships
3. **Random Forest Classifier** – An ensemble model to improve accuracy and reduce overfitting

6.2 Model Evaluation and Tuning

- Model performance was evaluated using **accuracy**, confusion matrix, and classification report.
- **Hyperparameter tuning** was performed using **GridSearchCV** for all models.
- Random Forest achieved the **highest accuracy** after tuning and showed better generalization compared to other models.

7. Model Explainability

Feature importance analysis using Random Forest revealed that the most influential features for predicting customer satisfaction are:

- Agent tenure
- Connected handling time
- Issue category
- Interaction channel
- Agent shift

These insights help explain why certain customer interactions result in higher or lower satisfaction scores.

8. Final Model Selection

The **Random Forest Classifier** was selected as the final prediction model because:

- It achieved the highest accuracy
 - It handles feature interactions effectively
 - It reduces overfitting through ensemble learning
 - It provides feature importance for explainability
-

9. Deployment Readiness

The final model was saved using **joblib** and reloaded to perform predictions on unseen data as a sanity check. This confirms that the model is **deployment-ready** and can be integrated into a live system for real-time CSAT prediction.

10. Business Impact

The results of this project can help Flipkart:

- Improve agent training programs
 - Optimize staffing across shifts
 - Reduce handling time
 - Identify recurring customer issues
 - Enhance overall customer satisfaction and retention
-

11. Conclusion

This project successfully combined Exploratory Data Analysis and Machine Learning to analyze Flipkart's customer support data and predict customer satisfaction scores. The EDA provided valuable business insights, while the machine learning models enabled accurate prediction of CSAT scores.

The Random Forest model emerged as the best-performing model and offers a scalable, data-driven solution for improving customer support operations. This project demonstrates how analytics and machine learning can be effectively applied to enhance customer experience and operational efficiency in the e-commerce domain.

Project Completion

Hurrah! You have successfully completed the Flipkart EDA & Machine Learning Capstone Project.