

Flipkart Customer Service Satisfaction Classification - Final Project Report

Table of Contents

1. Introduction
2. Problem Statement
3. Objective
4. Dataset Description
5. Tools & Technologies Used
6. Data Preprocessing
7. Model Development Using Vertex AI
8. Evaluation Metrics
9. Predictions and Visualization
10. Streamlit Dashboard Deployment
11. GitHub Repository Structure
12. Key Insights & Observations
13. Challenges Faced
14. Future Scope
15. Conclusion

1. Introduction

Customer satisfaction is a critical business metric for evaluating the effectiveness of customer service teams. This project focuses on automating the classification of customer satisfaction (CSAT) scores for Flipkart's customer support data using machine learning. Leveraging Google Cloud Platform's Vertex AI AutoML, we created an end-to-end pipeline, visualized outcomes, and deployed an interactive dashboard for real-time analytics.

2. Problem Statement

Manual CSAT evaluation is time-consuming and lacks consistency. Flipkart seeks a scalable ML-driven approach to classify customer satisfaction scores based on support interaction data, thereby enabling real-time performance tracking and feedback loops.

3. Objective

- Build a classification model using Google Cloud Vertex AI AutoML.
 - Predict CSAT score (1 to 5) using historical support interaction data.
 - Evaluate and interpret model performance.
 - Create dashboards and visualizations.
 - Deploy an interactive dashboard for management review.
-

4. Dataset Description

The dataset used is named `Customer_support_data.csv`, containing ~85,000 rows with the following key features:

Input Features:

- `unique_id`: Unique record identifier
- `channel_name`, `category`, `sub_category`
- `customer_remarks`, `order_id`, `order_date_time`
- `issue_reported_at`, `issue_responded`, `survey_response_date`
- `customer_city`, `product_category`, `item_price`
- `connected_handling_time`, `agent_name`, `supervisor`, `manager`
- `tenure_bucket`, `agent_shift`

Target Label:

- `CSAT Score` (Categorical: 1 to 5)

After prediction, the model generates probabilities:

- `csat_score_5_scores`, `csat_score_4_scores`, ..., `csat_score_1_scores`
-

5. Tools & Technologies Used

- **Google Cloud Platform (GCP)**
 - Vertex AI AutoML
 - Cloud Storage
 - **Python Libraries**
 - `pandas`, `matplotlib`, `seaborn`, `google-cloud-storage`, `google-cloud-aiplatform`
 - **Development Environments**
 - Google Colab
 - Jupyter Notebook
 - **Deployment**
 - Streamlit (for dashboard)
 - **Version Control**
 - GitHub
-

6. Data Preprocessing

- Uploaded raw CSV to GCP Cloud Storage.
 - Explored and cleaned missing values in order times and remarks.
 - Transformed `CSAT Score` into a categorical label.
 - AutoML handled further preprocessing internally.
 - Created `predict_input.csv` (features only) for batch prediction.
-

7. Model Development Using Vertex AI

Key Steps:

1. Created a Vertex AI Tabular Dataset.
2. Imported data from Cloud Storage.
3. Selected `CSAT Score` as the target column.
4. Set a training budget of 1 node-hour.
5. AutoML handled feature engineering & model selection.
6. Model trained and evaluated via pipeline.

Resulting Model:

- Name: `flipkart_csat_model_final`
 - Output: Batch predictions with confidence scores
-

8. Evaluation Metrics

- **ROC AUC:** 0.901
- **PR AUC:** 0.727
- **F1 Score:** 0.718
- **Feature Importance:** Handled internally by Vertex AI
- **Confusion Matrix:** Exported from Vertex AI UI

The metrics show strong performance for a categorical classification model.

9. Predictions and Visualization

Generated predictions were saved in `flipkart_csat_predictions.csv`. From the prediction scores, the final class (`predicted_csat`) was derived by identifying the class with the maximum probability.

Charts Generated (Stored in `charts/`):

1. Predicted CSAT distribution
 2. Confidence score distribution
 3. Handling time vs CSAT
 4. Item price vs CSAT
 5. Agent-wise average CSAT
 6. Shift-wise CSAT distribution
-

10. Streamlit Dashboard Deployment

Built an interactive one-page dashboard hosted via Streamlit Cloud.

Dashboard Features:

- Summary Metrics:
- Total Records
- Average Predicted CSAT
- Top Agent by count
- Chart Visualizations
- Download Button for predictions CSV
- Error Handling for missing `predicted_csat`

Live Dashboard auto-loads latest CSV and chart data.

11. GitHub Repository Structure

```
Flipkart-csat/  
├─ flipkart_csat_predictions.csv  
├─ predict_input.csv  
├─ cleaned_customer_support_data.csv  
├─ charts/  
│   ├─ predicted_csat_distribution.png  
│   ├─ prediction_confidence_distribution.png  
│   ├─ handling_time_vs_csat.png  
│   ├─ item_price_vs_csat.png  
│   ├─ agentwise_avg_csat.png  
│   └─ shift_vs_csat.png  
├─ app.py # Streamlit dashboard code  
├─ Flipkart_CSAT_Report.pdf  
├─ flipkart_csat_notebook.ipynb  
└─ README.md
```

12. Key Insights & Observations

- Most predictions skewed toward CSAT 5 and 4 (positive sentiment).
 - Longer handling time may correlate with lower satisfaction.
 - Morning shifts showed better satisfaction scores than evening shifts.
 - Few agents and supervisors contributed to majority CSAT interactions.
-

13. Challenges Faced

- GCP service account permissions (needed IAM roles manually set)
 - Vertex AI quota errors (custom training limits)
 - Streamlit deployment errors due to missing fallback logic
 - GitHub push conflicts due to parallel updates
-

14. Future Scope

- Add real-time prediction endpoint using online prediction
 - Enable CSV upload in dashboard for new predictions
 - Introduce NLP sentiment analysis on customer remarks
 - Compare AutoML with manually tuned models
 - Integrate BigQuery for large-scale data operations
-

15. Conclusion

This project successfully built and deployed a complete ML pipeline on GCP to predict and analyze Flipkart's customer service satisfaction. From data cleaning to dashboard hosting, the workflow was automated and production-ready. The interactive dashboard and model metrics can empower decision-makers to monitor and improve service quality proactively.

Prepared by: Ishan Mishra\ **Contact:** ihrm.aiml@gmail.com\ **GitHub:** [Flipkart CSAT Repo](#)\ **Dashboard:-** <https://flipkart-csat-ishan.streamlit.app/>
