# Customer Sentiment Analytics in Rail Transportation (IRCTC)

Capstone Project Report
Submitted by: Teesa Sharma
Roll Number: 24MBMB25
Course: MBA - Business Analytics
Institution: University of Hyderabad
Academic Year: 2024-2025
Submission Date: November 2025

---

ABSTRACT

---

This project analyzes customer feedback for IRCTC using Big Data Analytics. Data was collected from the Google Play Store,
processed using Apache Spark on Databricks, and analyzed with NLP and machine learning techniques. The unified pipeline
(Tokenization -> Stopword Removal -> TF-IDF -> Classification) ensured consistent and scalable processing. The study identified
sentiment patterns, complaint trends, and customer loyalty insights for IRCTC.

---

INTRODUCTION

---

IRCTC handles millions of transactions daily, with its app being a crucial interface. Understanding customer sentiment helps
optimize service quality. Big Data Analytics enables the transformation of unstructured text data into meaningful patterns
to guide operational decisions.

---

METHODOLOGY & PIPELINE

---

1. Data Cleaning and Preprocessing
2. NLP Pipeline: Tokenization -> Stopword Removal -> TF-IDF
3. Model Training: Logistic Regression, Random Forest, Naive Bayes
4. Evaluation using MLlib
5. Visualization using Matplotlib

---

USE CASE SUMMARY

---

Use Case 1: Sentiment Classification - Classified reviews with 89% accuracy using Random Forest.
Use Case 2: Service Aspect Analysis - Identified most discussed service areas (food, booking, app).
Use Case 3: Time-Based Trends - Observed monthly sentiment changes.
Use Case 4: Complaint Prediction - Detected complaint reviews with 88% accuracy.
Use Case 5: Loyalty Insights - Clustered users into loyal, neutral, disengaged using K-Means.

---

RESULTS AND FINDINGS

---

- Majority of users (70%) expressed positive sentiments.
- Booking and app issues were the leading causes of negative reviews.
- Loyal customers represented 45% of users.
- Complaint trends showed predictable spikes during peak seasons.

---

CONCLUSION & FUTURE SCOPE

---

This project demonstrates how Big Data Analytics and NLP can extract actionable insights from railway feedback.
Future enhancements include real-time streaming dashboards, multi-language review handling, and Power BI integration.

---

ACKNOWLEDGMENT

---

Gratitude to the University of Hyderabad's Department of Business Analytics and all mentors who guided this project.