# Sentiment-Aware Churn Prediction in Quick Commerce

## BE Project Report

विश्वजीवनामृतं ज्ञानम्


2023IMT-050 Malladi Nagarjuna

2023IMT-059 Prasanna Mishra

2023IMT-060 Prasun Baranwal

2023IMT-073 Shivam Deolankar

**ABV-Indian Institute of Information Technology and Management Gwalior**

This project implements the methodology from the following Scopus-indexed article:

- **Title:** Data-driven strategic customer segmentation considering cart abandonment behavior: Insights from e-grocery delivery platforms
- **Publication:** SJMSoM, IIT Bombay (2025)
- **Scopus Link:** https://www.scopus.com/pages/publications/105006995581

**Abstract:** The growing e-commerce grocery (e-grocery) sector drives e-grocers to pursue higher customer retention rather than acquisition... Traditional behavioral segmentation metrics such as recency (R), frequency (F), and monetary value (M) encounter accuracy and decision-making challenges... This study enhances the RFM model by incorporating the delivery ratio (D) metric representing customer cart behavior and separates 'R' to present an innovative R + FMD model. Results demonstrate the superiority of the R + FMD model... with the GMM showing the best performance... The insights from this study provide businesses with powerful tools to target customer retention through precise segmentation.

## 2. Dataset Utilized

The analysis was performed on a publicly available e-commerce dataset representing the operations of Blinkit.

- **Dataset Name:** Blinkit Marketing and Customer Feedback Dashboard
- **Source:** Kaggle
- **Link:** https://www.kaggle.com/datasets/yashmotiani/blinkit-marketing-and-customer-powerbi-dashbord

**Dataset Overview:** The dataset contains comprehensive information across six files, including customer details, order history, product information, marketing campaign performance, and direct customer feedback with sentiment labels. This rich collection of data is ideal for replicating and extending the research paper's methodology.

## 3. Project Goal & Methodology

The primary goal of this project is to build a **sentiment-aware churn prediction model** for the Blinkit dataset by implementing the core concepts of the reference research paper.

The methodology followed a four-phase approach:

1. **Feature Engineering (R+FMD Model):** We calculated **Recency (R)**, **Frequency (F)**, **Monetary (M)**, and the innovative **Delivery Ratio (D)** for each customer. The Delivery Ratio (Delivered Orders / Total Orders) serves as a powerful proxy for cart abandonment behavior.
2. **Customer Segmentation (GMM):** Using the **Gaussian Mixture Model (GMM)**, customers were grouped into distinct segments based on their FMD profiles.
3. **Sentiment Analysis:** Customer feedback was analyzed to extract key metrics like average rating and negative_feedback_count, adding a crucial layer of customer sentiment.
4. **Churn Prediction:** A **Random Forest Classifier** was trained on the combined R+FMD and sentiment features to predict the probability of a customer churning.

### 1. Key Analytical Findings

Our analysis confirmed the paper's hypotheses and yielded actionable insights for Blinkit:

- **Finding 1: R+FMD Model is Superior:** The inclusion of the **Delivery Ratio (D)** metric resulted in more distinct and meaningful customer segments compared to a traditional RFM model. It successfully separated reliable customers from those with high cancellation rates.
- **Finding 2: High-Value Customers Have Critical Issues:** An analysis of the High-Value Customers" segment revealed that their primary complaints are not about price, but about core operations. The top issues were **Product Quality**, **Missing Items**, and **Late Deliveries**.
- **Finding 3: Key Drivers of Churn Identified:** The predictive model identified the three most significant factors that lead to customer churn:

1. **Recency:** The number of days since a customer's last order remains the strongest predictor.
2. **Negative Feedback Count:** Customers who leave even one piece of negative feedback are significantly more likely to churn.

3. **Delivery Ratio:** A low order completion rate is a powerful early warning sign of disengagement.

## 2. Interactive Dashboard Implementation

To make these findings accessible and actionable, an interactive dashboard was developed using Python and the Streamlit library. The dashboard serves as a central hub for business intelligence.

**Key Dashboard Features:**

- **Executive Summary:** High-level KPIs on revenue, customers, and order statuses, providing an at-a-glance view of business health.
- **Interactive 3D Segmentation:** A 3D scatter plot that allows stakeholders to visually explore the customer segments based on their Frequency, Monetary value, and Delivery Ratio.
- **High-Value Customer Analysis:** A dedicated section that visualizes the top complaints and most frequently purchased products for the most valuable customer segment.
- **Proactive Churn List:** A sortable table listing the top customers at the highest risk of churning, along with their churn probability score.

## 3. Conclusion & Business Impact

This project successfully translates academic research into a practical business solution. By implementing the sentiment-aware R+FMD model, we have created a powerful tool that allows Blinkit to:

- **Understand Customers More Deeply:** Move beyond simple transaction metrics to understand complex behaviors like cart abandonment.
- **Retain Valuable Customers:** Proactively identify at-risk customers and address the specific issues frustrating the most profitable segments.
- **Optimize Marketing Efforts:** Focus retention budgets on customers who are most likely to churn, maximizing return on investment.