# *Group 5: Anupreet, Ashish, Sirisha, Zara and Zena*

# *Project Proposal: Sentiment Analysis and Recommendation System*

## Introduction

This project aims to leverage machine learning (ML) techniques to analyze customer reviews and recommend products using the provided dataset, 'Final Customer Dataset'. Which is a merge of two big data sources from the Kaggle: <https://www.kaggle.com/datasets/shriyashjagtap/e-commerce-customer-for-behavior-analysis>

The dataset is sufficiently large to ensure reliability, consisting of a CSV file with 500,000 rows of data.

The project focuses on two main objectives:

* Sentiment analysis of reviews to understand customer satisfaction and
* Building a recommendation system to enhance user experience and increase sales. These objectives align with real-world applications in the e-commerce domain.

## Problem Statement

In the competitive e-commerce landscape, understanding customer satisfaction and providing personalized recommendations are critical for improving user experience and driving sales. The dataset provides an opportunity to address these challenges using machine learning models that can derive actionable insights from customer behavior and reviews.

## Objectives

1. Sentiment Analysis: Analyze customer reviews to understand satisfaction levels.

2. Recommendation System: Suggest products to customers based on their browsing or purchasing history.

## Proposed Methodology

### *Sentiment Analysis*

* Objective: Analyze customer reviews to understand satisfaction levels.
* Machine Learning Technique: Natural Language Processing (NLP) using a sentiment classifier.

**Steps:**  
1. Preprocess text data: Perform tokenization, stopword removal, and vectorization.  
2. Train a sentiment classifier: Use models such as Naive Bayes or BERT to classify customer reviews as positive, neutral, or negative.  
3. Evaluate the model: Assess performance using metrics like accuracy, precision, recall, and F1-score.

### *Recommendation System*

Objective: Suggest products to customers based on their browsing or purchasing history.

Machine Learning Technique: Collaborative Filtering or Content-Based Filtering.

**Steps:**  
1. Prepare data: Create a user-item interaction matrix using purchase history data.  
2. Apply algorithms: Use collaborative filtering techniques like Matrix Factorization or content-based filtering methods such as K-Nearest Neighbors to recommend products.  
3. Evaluate the system: Use metrics like precision, recall, and mean reciprocal rank (MRR) to measure recommendation quality.

## Dataset Description

The dataset, 'Final Customer Dataset', contains comprehensive customer behavior data, including purchase history and review text. The dataset will be used for both sentiment analysis and recommendation system development.

## Tools and Technologies

- Programming Language: Python  
- Libraries: Scikit-learn, TensorFlow (for advanced NLP models)  
- Data Preprocessing: Pandas, NumPy  
- Visualization: Matplotlib, Seaborn, Leaflet  
- Deployment: Streamlit or Flask for showcasing results

## Expected Outcomes

1. A trained sentiment analysis model capable of categorizing customer reviews into sentiment classes.  
2. A recommendation system that provides personalized product suggestions, improving user experience.  
3. Actionable insights into customer satisfaction and purchasing behavior.

## Conclusion

This project will demonstrate the application of machine learning in addressing real-world e-commerce challenges. By analyzing customer reviews and providing personalized recommendations, the project aims to contribute to improving customer satisfaction and business performance.