



Presented by Group 10

# CPSC 5310 Machine Learning Group Project

Enhancing Customer Retention in  
E-Commerce Through Predictive Analytics



# *Team Members*



Mehar Sukhi  
Buruguru



Nandan Varma  
Pericharla



Shyamalan Kannan



Rupeshwar Rao



# Introduction

---



- This dataset provides a comprehensive view of customer interactions within an e-commerce platform. It includes demographic details, behavioral patterns, and transaction records of customers, enabling insights into their preferences, satisfaction levels, and likelihood of continued engagement. The analysis focuses on solving key business problems such as:
    - Predicting customer churn (classification)
    - Estimating customer satisfaction scores (regression)
    - Forecasting order trends and cashback amounts (prediction)
  - Through this dataset, we aim to empower decision-making by identifying patterns that drive customer retention and optimizing operational strategies.
- 



# About the Dataset

- The dataset belongs to a leading online E-Commerce company. An online retail (E-commerce) company wants to know the customers who are going to churn, so accordingly they can approach customer to offer some promos.
- The dataset contains 20 columns with a mix of categorical (e.g., **PreferredLoginDevice**, **Gender**) and numerical (e.g., **Tenure**, **CashbackAmount**) data.
- The target variable for predicting churn is Churn, which is binary (1 for churned, 0 otherwise).
- [Ecommerce Customer Churn Analysis and Prediction - Kaggle](#)



# Goal

The Goal is to accomplish:

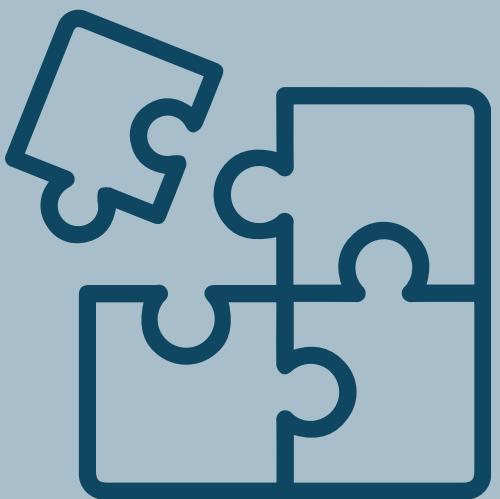
- Predict Churn (whether a customer will return or leave).
- Identify and classify customers who are likely to return to the shop.



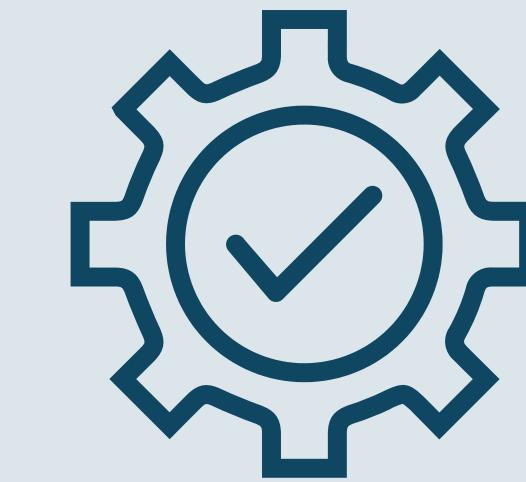
# *Plan*



**Data Cleaning &  
Preprocessing**



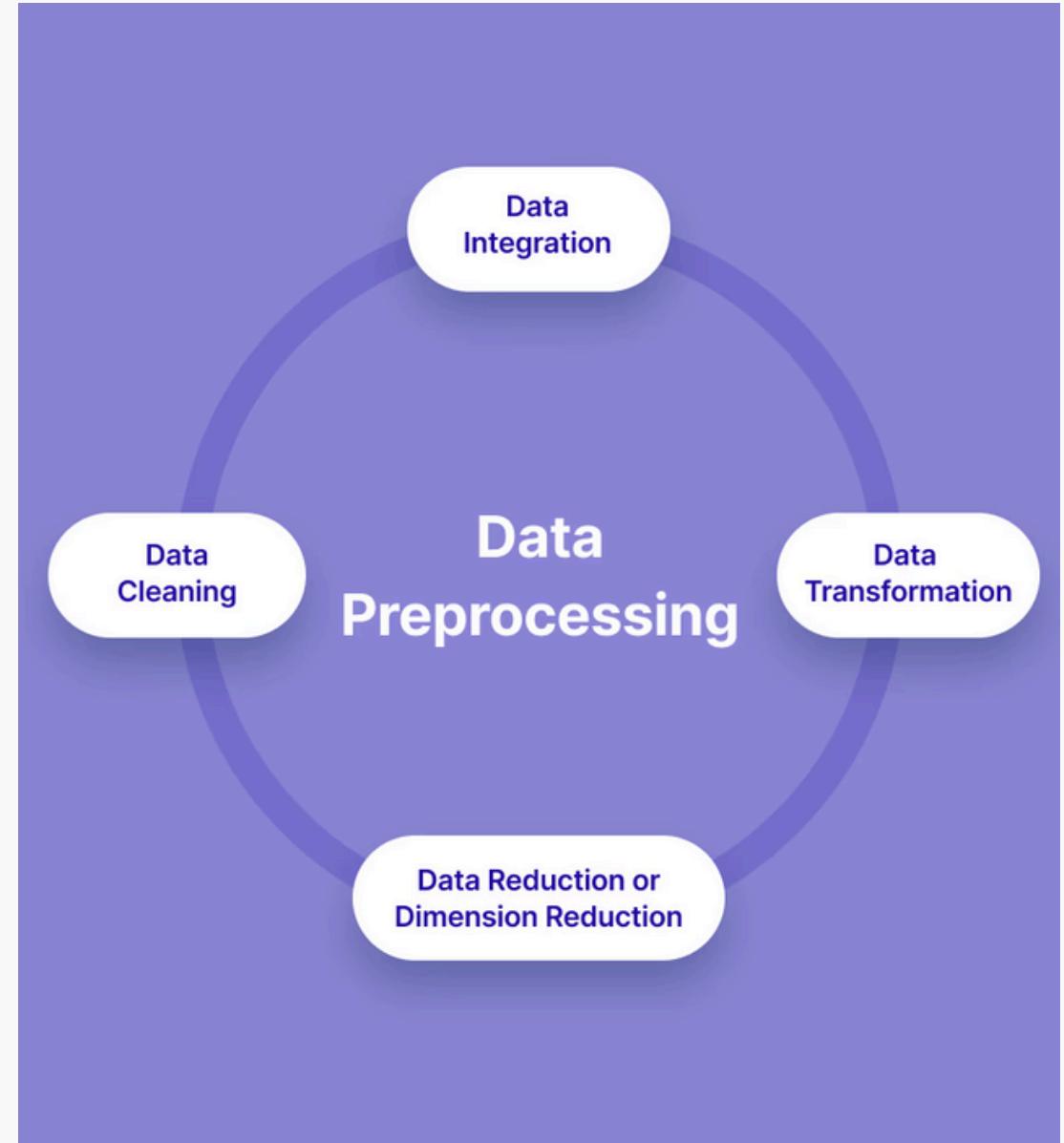
**Model Selection & Training**



**Testing & Performance  
Evaluation**

# *Data Preprocessing*

- Handle missing values (e.g., **Tenure**, **HourSpendOnApp**, etc.).
- Encode categorical columns such as **PreferredLoginDevice**, **PreferredPaymentMode**, and **Gender**.
- Normalize numerical columns like **WarehouseToHome**, **CashbackAmount**, etc.
- Remove or impute outliers in numeric data where necessary.

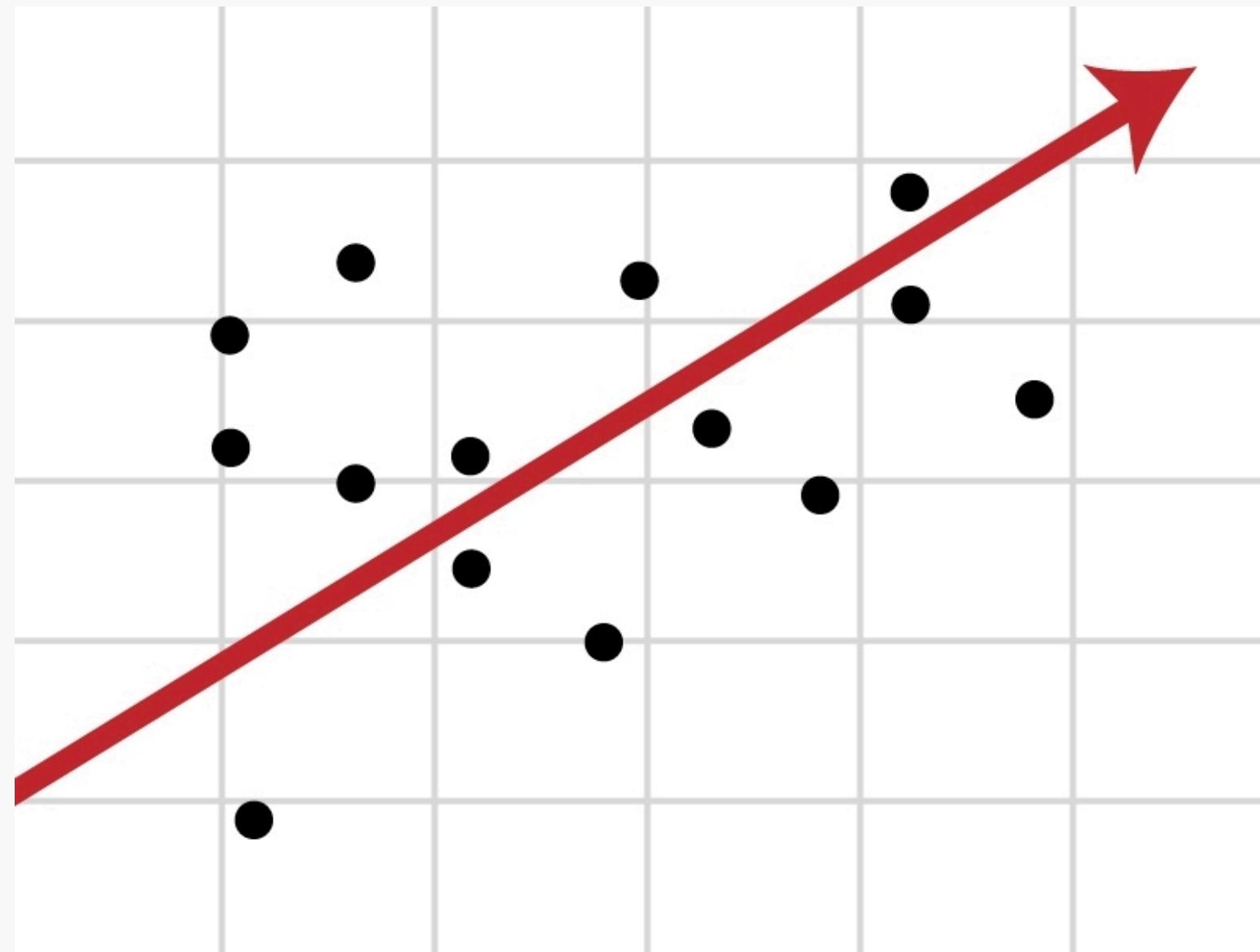


# *Data Prediction Using Regression*

Regression techniques can be used for predicting numeric features, like:

- Tenure
- DaySinceLastOrder
- CashbackAmount

• • • •



# *Classes That Can Be Classified*

- The primary classification is Churn
  - Binary Format:
    - 1 = Yes
    - 0 = No.
- Other potential classification tasks could include:
  - Predicting PreferredPaymentMode based on other customer features.
  - Predicting customer satisfaction (SatisfactionScore) as a multi-class classification.



# Conclusion



---

The analysis of this e-commerce dataset reveals valuable insights into customer behavior and engagement patterns. By leveraging machine learning techniques, we can predict churn, forecast purchasing trends, and optimize customer satisfaction strategies. These insights empower businesses to:

- Develop targeted retention programs.
- Enhance customer experience through personalized services.
- Drive revenue growth by anticipating and meeting customer needs.





*Thank you*

