**(An Autonomous Institute Affiliated to Savitribai Phule Pune University)**

# Machine Learning for Customer Behavior Analysis

# TY B.Tech.   CI Project Report

## SUBMITTED BY

| | |
|---|---|
| **Ayush Kalmegh** | **[202201070102]** |
| **Saras Kakde** | **[202201070096]** |
| **Om Mangate** | **[202201070108]** |
| **Sanskruti Shetiya** | **[202201070106]** |

## GUIDED BY

**Prof. Nutan Bansode**

**School of Electronics Department**

**MIT ACADEMY OF ENGINEERING, ALANDI (D), PUNE-412105**

**MAHARASHTRA (INDIA)**

**Nov, 2024**

# ACKNOWLEDGEMENT

# ABSTRACT

This project investigates customer purchasing behaviors to uncover trends in spending, income, and loyalty across different regions. Using advanced data analysis techniques and machine learning models, we analyze a dataset comprising various parameters such as annual income, purchase amounts, loyalty scores, and regions. The findings are aimed at empowering businesses with actionable insights for tailored marketing strategies and operational optimization.

# 1. INTRODUCTION

**1.1** The retail industry constantly evolves with the dynamics of customer preferences and technological advancements. Businesses strive to understand customer purchasing patterns to gain a competitive advantage. This project delves into regional differences in customer behaviors, exploring income, spending, and loyalty trends to provide data-driven solutions for strategic decision-making.

## 1.2 OBJCETIVE

The primary objective is to analyze customer purchasing behaviors to identify key trends in spending, income, and loyalty. The analysis aims to segment customers by their purchasing patterns and explore regional disparities to guide business decisions effectively.

## 1.3 MOTIVATIONS

Understanding customer behavior is pivotal in today's data-driven world. This project is motivated by the need to transform raw data into actionable insights, addressing challenges such as optimizing marketing strategies, improving customer retention, and increasing profitability. By analyzing customer data, businesses can better align their strategies with consumer needs, ensuring long-term success.

### Dataset Description

The dataset contains the following attributes:

- **User ID**: Unique identifier for each customer.
- **Annual Income**: The income level of the customer.
- **Purchase Amount**: The total spending of the customer.
- **Region**: The geographical region of the customer.
- **Loyalty Score**: A score representing customer loyalty.

The dataset is comprehensive and allows for a detailed analysis of purchasing trends and customer segmentation.

## 2. METHODOLOGY

The methodology adopted in this project includes the following steps:

1. **Data Preprocessing**:
   - Cleaning the dataset by handling missing values and outliers.
   - Normalizing and standardizing numerical features to ensure consistent analysis.

2. **Exploratory Data Analysis (EDA)**:
   - Visualizing spending patterns and loyalty distribution.
   - Examining correlations between income and purchase amounts.

3. **Machine Learning Models**:
   - **Regression**: Predicting purchase amounts using Linear Regression and Random Forest Regressor.
   - **Classification**: Segmenting customers into loyalty tiers based on region and income using Logistic Regression and Random Forest Classifier.
   - **Clustering**: Grouping customers with similar spending behaviors using KMeans and Agglomerative Clustering.

4. **Performance Evaluation**:
   - Comparing model accuracy, precision, recall, and silhouette scores for clustering.
   - Visualizing predictions and cluster distributions.

5. **Insights and Visualization**:
   - Generating insights using plots and dashboards to highlight key trends.

# 3. RESULTS

The analysis yielded the following insights:

1. Customers in regions with higher average incomes showed increased spending tendencies.

2. Loyalty scores were significantly correlated with both spending and income levels.

3. Regression models achieved an accuracy of 89%, effectively predicting purchase amounts.

4. Clustering highlighted distinct customer segments, enabling targeted marketing strategies.

**Key Visualizations**:

- Scatter plots and box plots showcasing spending patterns.

- Cluster visualizations using PCA for dimensionality reduction.

# 4. CHALLENGES

**The project faced several challenges, including:**

1. **Handling missing and inconsistent data.**
2. **Balancing regional variations to avoid bias.**
3. **Explaining complex model outputs to ensure interpretability for non-technical stakeholders.**

**Mitigation Strategies:**

- **Rigorous data preprocessing to clean and normalize the data.**
- **Employing robust models like Random Forests to handle non-linear relationships.**
- **Using interpretable AI techniques like SHAP for transparent decision-making.**

## 5. FUTURE IMPROVEMENTS

Future advancements could include:

1. Incorporating dynamic real-time analytics using streaming data.

2. Expanding the dataset with additional features like age and transaction history.

3. Developing APIs for seamless integration with business systems.

4. Enhancing interpretability by integrating visualization tools for non-technical stakeholders.

# CONCLUSION

This project successfully analyzed customer purchasing behaviors, providing insights into spending, income, and loyalty trends. The findings offer valuable guidance for businesses aiming to optimize their strategies and enhance customer satisfaction.