



**INTRODUCTION TO DATA SCIENCE
SEMESTER PROJECT REPORT**

SUBMITTED BY

NOOR-UL-HUDA 23I-2099

MALAIKA ZAINAB 23I-2605

SUBMITTED TO

MISS SAIRA QAMAR

Comparison of the predictive performance of the Regression, Decision Tree, and K-Means Clustering Models:

1. Regression Model

Performance

- Metrics:
 - MAE: 0.0715 (The error is low, which shows good accuracy).
 - MSE: 0.0093 (indicates better handling of large errors).
 - R-squared: 0.8888 (88.8% of the variance in the target variable is explained).
- Strengths:
 - Effective for continuous variable prediction.
 - Provides explainability; coefficients indicate how features impact predictions.
 - Linear regression is computationally inexpensive.
- Limitations:
 - Assumes linear relationships, which may not hold in all customer behavior datasets.
 - Sensitive to outliers and multicollinearity.
- Applicability:
 - Prediction of purchase amounts based on features like age or frequency.
 - Useful for understanding relationships between variables (e.g., Age and Purchase Amount).

2. Decision Tree

Performance

- Metrics:
 - Accuracy: 0.92
 - Precision: 0.92
 - Recall: 1.00
 - F1 Score: 0.96 (balanced performance for imbalanced datasets).
- Strengths:
 - Handles both categorical and continuous features.
 - Non-linear relationships and interactions between features are captured.
 - Easy to interpret visually.

- Limitations:
 - Can be sensitive to small changes in the data.
- Applicability:
 - To Classify customers into behavioral segments based on purchase frequency and amount.
 - To Detect patterns like high-frequency, high-value customers.

3. K-Means Clustering

Performance

- Cluster Centers:
 - Cluster 0: Young, low purchase frequency (0.24).
 - Cluster 1: Older, medium purchase frequency (0.43).
 - Cluster 2: Middle-aged, high purchase frequency (0.84).
- Strengths:
 - Unsupervised; does not need labeled data(removed predicted label)
 - Groups customers with similar behaviors, useful for segmentation.
 - Computationally efficient with small to medium datasets.
- Limitations:
 - Sensitive to scaling; requires normalized data.
 - Requires specifying the number of clusters.
- Applicability:
 - To Identify customer segments for targeted marketing (e.g., low-frequency vs. high-frequency shoppers).
 - To Detect outliers, such as rare purchasing behaviors.

Comparative Summary Table:

Model	Best For	Strengths	Limitations
Regression	Predicting continuous variables	Simple, interpretable, low error rates	Linear assumptions, sensitive to outliers
Decision Tree	Classifying customers into behavior groups	Captures non-linear relationships, interpretable	Overfitting, sensitive to data changes
K-Means	Unsupervised segmentation, behavioral patterns	Effective for grouping, scalable	Requires pre-defined clusters

Real-World Applicability in Customer Behavior Analysis

1. Regression:

- Predicts purchase amount based on individual characteristics (e.g., age or income level).
- Used for forecasting and strategic planning.

2. Decision Tree:

- Identifies purchase patterns for targeted promotions (e.g., frequent small purchasers vs. infrequent big spenders).
- Helps in churn prediction or upselling opportunities

3. K-Means Clustering:

- Segments customers for personalized marketing campaigns (e.g., loyalty rewards for Cluster 2: high-frequency shoppers).
- Insights into age-group purchasing tendencies.

Summary of real-world applicability:

Each model serves a distinct purpose in customer behavior analysis. Regression is ideal for predicting outcomes, decision trees excel in classification tasks, and K-Means is a powerful tool for segmentation. A combination of these models can maximize insights and drive data-driven decision-making.

Actionable Recommendations for the Electronics Section:

1. Tailored Promotions Based on Customer Segments (K-Means Clustering)

- **Cluster 0 (Younger, Low-Frequency Shoppers):**
 - Provide targeted discounts on reasonably priced equipment (such as headphones and gadgets).
 - Conduct advertising efforts on sites that are well-liked by younger audiences (e.g., social media, influencer promotions).
 - Introduce flexible payment plans or subscription-based offers to encourage purchases.
- **Cluster 1 (Older, Medium-Frequency Shoppers):**

- Emphasize promoting dependable and long-lasting goods (such as household appliances).
- Provide senior-friendly support services, like setup assistance.
- Provide loyalty benefits, like rebates or longer warranties, for recurring purchases.
- **Cluster 2 (Middle-Aged, High-Frequency Shoppers):**
 - Target with premium electronics, like high-end smartphones or smart home devices.
 - Provide personalized offers based on purchase history (e.g., bundle deals for complementary products).
 - Highlight the latest trends and innovations to maintain engagement with this high-value segment.

2. Optimize Inventory Based on Predictive Analysis (Regression)

- Use the regression model to forecast demand for popular products during different seasons.
- Stock up on products appealing to middle-aged customers, as they demonstrate high purchase frequency.
- Allocate resources toward age groups most likely to purchase specific categories, e.g., younger shoppers for portable electronics and middle-aged shoppers for home entertainment systems.

3. Using decision tree analysis, enhance Shopping Experience

- **Improve convenience:** Provide features like "Shop by Age" or "Shop by Usage Frequency" sections in online and offline stores.
- **Targeted email campaigns:**
 - Segment users into categories based on their behavior (low, medium, high frequency) and promote relevant products.
 - Use customer data to send reminders for upgrades (e.g., new versions of previously purchased products).

4. Customer Retention and Acquisition

- **For low-frequency shoppers:**
 - Offer entry-level electronics to encourage initial purchases.

- Use referral discounts to attract peers.
- **For medium and high-frequency shoppers:**
 - Incentivize repeat purchases through points systems or tiered memberships.
 - Highlight exclusive access to new launches or pre-order opportunities.

5. Seasonal and Holiday Campaigns

- Align with the decision tree's recall of customer patterns to predict high-demand periods.
- Use festive offers and discounts targeted at high-value segments (e.g., Cluster 2) to maximize revenue.

Conclusion

A data-driven approach that makes use of insights into consumer behavior may help the electronics division enhance customer loyalty, optimize inventory, and customize experiences. Both revenues and customer happiness will increase as a result of these activities.