**Project Report**

**Introduction to Data Science**

**Section A**

**Roll no:**

**I22-2011**

**23k-2062**

**1. Comparison of Models:**

**a) Regression Model:**

**Strengths:**

Predicts a continuous outcome (e.g., predicted purchase amount or customer lifetime value).

Simple to understand and implement.

Gives insights into how different features (like age, income, etc.) impact the predicted value.

Limitations:

Assumes a linear relationship between features and the target variable (which may not always be true).

Sensitive to outliers.

Not ideal for complex, non-linear relationships.

Real-World Applicability:

Useful when predicting specific values, like the amount a customer will spend or the number of purchases they will make.

Works well when there is a clear linear relationship between features and the target variable.

b) Decision Tree Model:

Strengths:

Handles both numerical and categorical data.

Can capture non-linear relationships between features.

Easy to interpret, as decisions are based on a flowchart-like structure.

Limitations:

Prone to overfitting (i.e., fitting the model too closely to training data and not generalizing well to new data).

Can be unstable, meaning small changes in data can lead to different trees.

Decision boundaries are often axis-aligned, which may not always be ideal.

Real-World Applicability:

Useful for classification tasks (e.g., determining whether a customer will purchase a product or not based on their behavior).

Effective when the data contains complex interactions or thresholds.

c) K-Means Clustering:

Strengths:

Can identify distinct customer segments based on behavior.

Unsupervised, so no need for labeled data.

Helps identify patterns and groups (e.g., high spenders vs. low spenders).

Limitations:

Assumes that the number of clusters (k) is known in advance.

Sensitive to the initial placement of centroids, which may affect the results.

Not suitable for data with non-spherical or irregularly shaped clusters.

Real-World Applicability:

Great for customer segmentation and market analysis.

Helps businesses target specific customer groups with tailored strategies.

**2. Actionable Recommendations for the Electronics Section:**

Targeting High-Spending Customers: Based on the clusters, identify high-spending customer groups. Focus on these clusters for premium marketing strategies, offering them exclusive discounts or early access to new products.

Personalized Promotions: Using insights from clustering, tailor promotions to each customer group. For instance, customers in one cluster might prefer discounts on gadgets, while another group may be interested in accessories or upgrades.

Improve Customer Retention: The regression model might predict the likelihood of repeat purchases. Focus on customers with high predicted purchase amounts and high purchase frequency to improve retention, offering loyalty programs or targeted communication.

**Conclusion:**

In this analysis, we compared three different models to understand and predict customer behavior:

Regression Model is great for predicting continuous outcomes like how much a customer will spend, but it works best when relationships between features are linear.

Decision Tree Model provides clear, easy-to-interpret rules for decision-making, and is helpful for understanding complex customer behavior, though it can overfit and be unstable at times.

K-Means Clustering is ideal for identifying distinct customer groups based on similar behaviors, helping businesses understand their customer segments better. However, it requires knowing the number of clusters beforehand and may struggle with irregularly shaped groups.

By combining these models, we can not only predict customer behavior but also group customers with similar behaviors, helping the electronics section create targeted strategies to boost sales, improve customer loyalty, and optimize marketing efforts.