Module 5: Comparison and Conclusion

1. Compare the predictive performance of the regression, decision tree, and K-Means clustering models:

In this section, we compare the predictive performance of the following models in the context of customer behavior analysis:

1. *Linear Regression*
2. *Decision Tree Classifier*
3. *K-Means Clustering*

The models were applied to predict customer behavior, including spending and purchasing patterns. Here's a summary of their performance:

# Linear Regression:

R² (R-squared) was at 0.9 , suggesting that the linear regression model captured the complexities in the data well.

- **Strengths:** Simplicity, ease of interpretation, and efficient for linear relationships.

- **Limitations**: handles non-linear relationships well and can be sensitive to outliers.

- **Real-world Applicability:** Suitable for predicting continuous variables like spending, assuming linear relationships. However, improvements in feature engineering or model choice are needed for better accuracy.

**Decision Tree Classifier:**

- Evaluated using Accuracy, Precision, Recall, and F1 Score . A balanced

Precision and Recall (around 0.75) indicated good classification of customer behavior into purchasing categories.

**- Strengths:** Interpretable, captures non-linear patterns, and handles both continuous and categorical variables well.

- **Limitations:** Prone to overfitting , especially without hyperparameter tuning (depth control).

**- Real-world Applicability :** Great for segmentation tasks where decisions can be based on clear thresholds, such as identifying customers who are likely to purchase again in the next month.

**K-Means Clustering:**

- Used for customer segmentation, determining the number of clusters via the Elbow Method.

The optimal number of clusters was chosen as 4.

- **Strengths:** Unsupervised learning, good for grouping customers based on behavior, and

discovering hidden patterns in the data.

**- Limitations** : Sensitive to the initial choice of centroids and requires pre-defined number of clusters (k). Doesn't capture complex relationships well without additional feature engineering.

- **Real-world Applicability:** Excellent for segmenting customers based on buying habits, product preferences, and other features. Ideal for targeted marketing campaigns.

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

Based on the results from all models, several actionable insights can be drawn for the electronicssection

- **Targeted Marketing**: Using the \*\*Decision Tree\*\*, we identified customer segments that are likely to make a purchase again in the next month. Marketing campaigns can be designed to target these customers with personalized offers.

- **Customer Segmentation:** K-Means Clustering provided meaningful clusters, allowing the business to segment customers based on behavior, such as spending, product preferences, and purchase frequency. Special offers or loyalty programs can be tailored to these segments.

- **Predictive Insights for Inventory Management**: The Regression model, while showing limited predictive performance, can still provide insights into spending trends. Further refinement of features or inclusion of additional data could improve predictive accuracy and aid in inventory forecasting.

- **Improvement Opportunities:** To improve model performance, especially for regression, it is recommended to explore advanced models (e.g., Random Forest, XGBoost) that handle non-linear relationships better. Adding more features like product category affinity or seasonality may also improve model accuracy.

**- Enhance Customer Experience:** Analyzing Recency and Purchase Frequency can help identify high-value, long-term customers, enabling targeted promotions or VIP programs. Additionally, understanding Dominant Brands and Top Product Categories can guide marketing strategies and inventory stocking decisions.