Decision Trees

# Abstract

This report is on the Online Shoppers dataset, determining what shoppers might do online using a decision tree. The dataset is first loaded and explored. Then data cleaning techniques are used and one hot encoding as well. Finally, a decision tree model is applied on the dataset.

# Introduction

This analysis focuses on predicting revenue generation based on various features using a decision tree classifier. The goal is to explore the factors influencing online shopper intention and evaluate the performance of the decision tree model.

# Methodology

## Data Exploration

The dataset is loaded, and initial exploratory data analysis is done to understand it better. Descriptive statistics are examined, and visualizations are created to understand the distribution of key variables. In addition, data cleaning is done. Missing values are visualized using a heatmap, and then appropriate actions are taken, including dropping rows with missing values. Histograms and box plots are also used to better visualize the data.

## Decision Tree Model

A decision tree model is used through SK Learn. The dataset is split into training and testing sets, and model performance is evaluated on the testing set. The confusion matrix and recall score are calculated to assess the model's ability to correctly identify revenue-generating instances.

# Results

The analysis reveals insights into online shopper behavior. The decision tree model, after training and testing, achieves a recall score on the testing set. The confusion matrix provides additional information about the true positive, false positive, true negative, and false negative predictions. The scores were very poor however, probably due to the lack of training data.

# Discussion

The visualization of key features and the decision tree structure provides interpretability of the model. The recall score highlights the model's ability to identify instances of revenue generation. The model, while providing a reasonable accuracy, may benefit from more training and data to improve its accuracy.

# Conclusion

In conclusion, this report presents an analysis of the Online shopper’s dataset. The process includes data exploration, cleaning, and model training, with a discussion of the results and potential areas for improvement.