Online Shopper's Purchasing Intention

AASHESH VONTIVILLU NARESHCHAND

Introduction

- Online shopping - Made our life easy
- But does the customer always have an intention to purchase whenever he visits an eCommerce website ?

Goal of the project

Analyze the factors that help in determining the visitor's purchasing intention

Predict if customer has purchasing intention or not given a new set of test data

Recommend the employers in targeting customers and help in improvising the marketing strategies

Dataset

Consists of feature vectors belonging to 12,330 sessions

Each session would belong to a different user in a 1-year period

Aggregated pageview data and logs captured during each session

Target Variable - Purchasing Intention (Yes/No)

Imbalanced Data with target variable proportion of 85:15. Sampling needs to be implemented

Algorithms used

Decision Tree

Feature Importance

Random Forest

Generalize Decision Tree

AdaBoosting

Improvise Decision Tree

Naïve Bayes

Attain Conditional Probability

SVM (Radial)

Nonlinear data

Clustering

Patterns and insights from data

Recall Score - Metric for evaluation

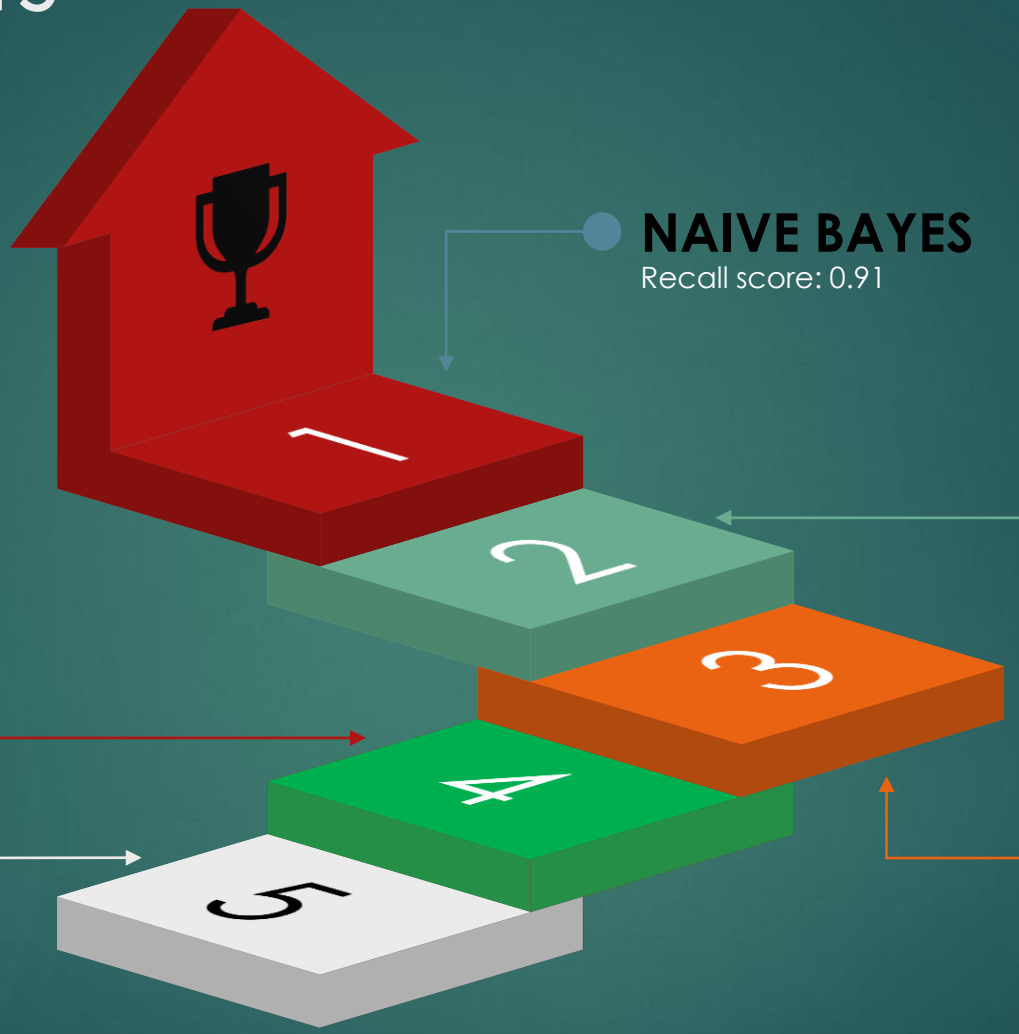
Determines how well the classifier was able to predict a specific target class

Our class of interest is to find the users having the intention to purchase

$

Need not worry much if user who has no intention to purchase is classified as interested user as its misclassification cost is very low

Model Results



NAIVE BAYES
Recall score: 0.91

RANDOM FOREST
Recall score: 0.802

DECISION TREE
Recall score: 0.788

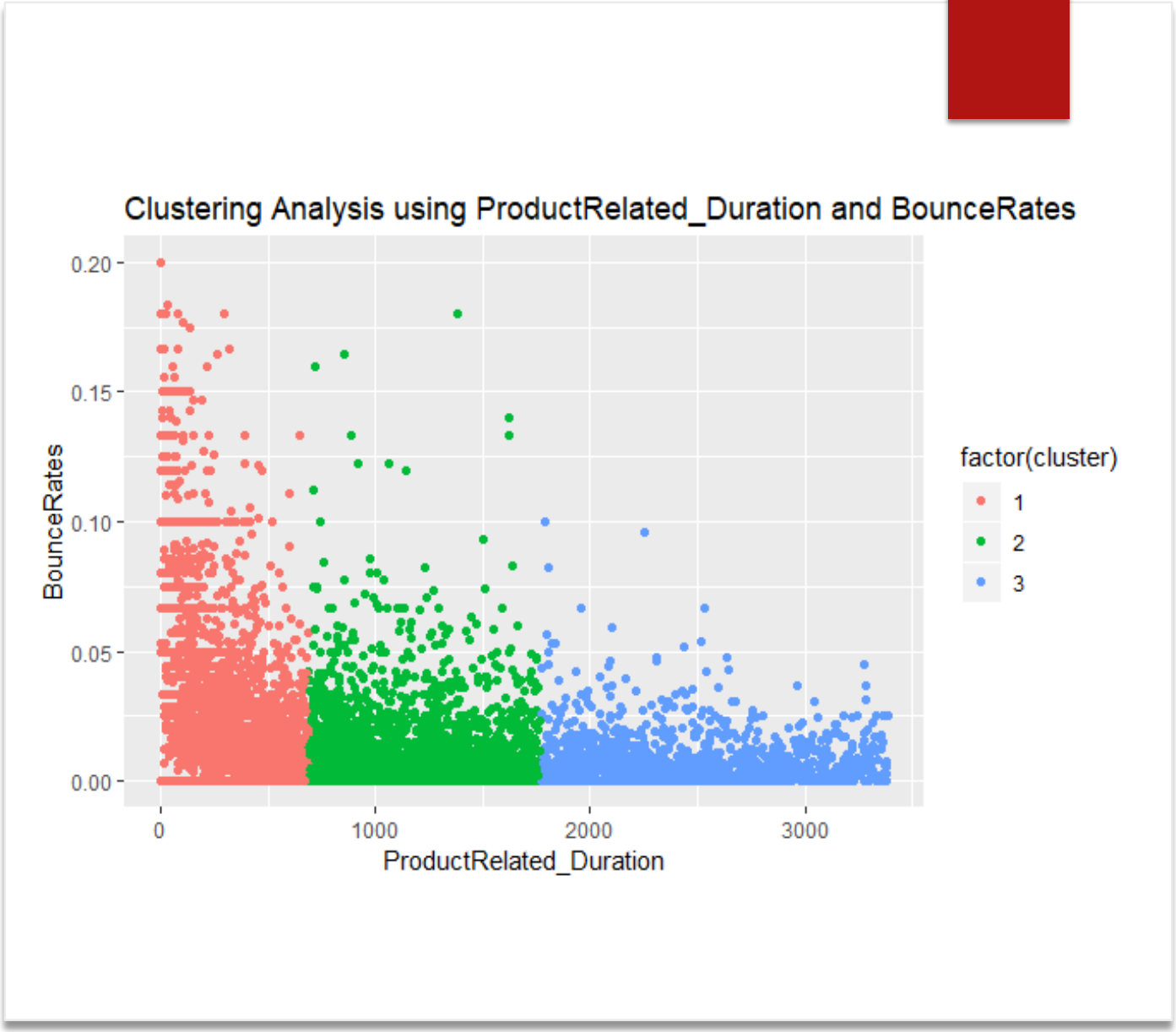
SVM-RBF
Recall score: 0.66

ADA BOOSTING
Recall score: 0.795

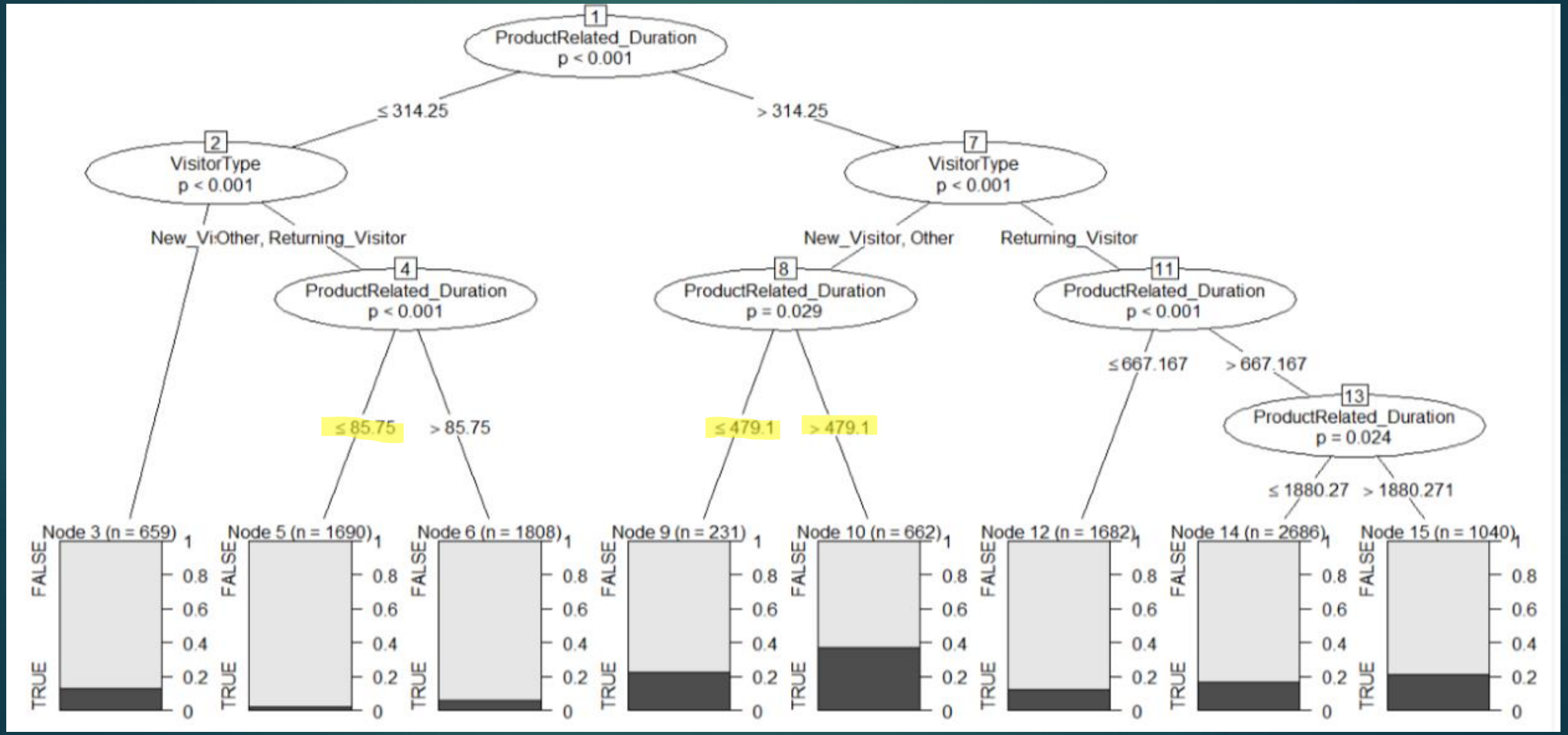
Performed hyperparameter tuning with cross validation to select the optimal hyperparameters for each algorithm and below are the results

K-Means Clustering

- **Elbow Plot** - Within Sum of Squares have reached its minimum saturation when the number of clusters was 3
- BounceRate was **high** for observations in cluster 1 – when the time spent on the product related page was **low**.
- BounceRate was **less** for observations in cluster 3 – when the time spent on the product related page was **high**.



Ctree Decision Tree



Conclusion and Recommendations

Based on Recall Scores, Naive Bayes algorithm performed the best. But cannot be concluded it as the best model for this data as we do not account for the other factors like misclassification cost, marketing cost & average profit when the users purchases an item

Page Value – Most Important Feature from Decision Tree

NaiveBayes - Probability of purchase is 62% when user visits a non-zero page value

Target new visitors, users who spent more time on the Product Related pages and users who visiting a non-zero page value

Questions?