Analysis of Online Shoppers' Purchase Intention

Problem Description

In the era of online shopping, known as e-shopping, people use online transactions to buy the items they need while exploring it online. This helps the buyers as well as sellers to understand the patterns, intentions, and behavior of various online customers. Thereby, helping businesses improve their revenue by focusing on customer experiences and marketing. Hence, the analysis of online shoppers' purchase intention has become an emerging field in data mining. Click-stream analysis refers to the online shoppers' behavior analysis as they invoke a sequence of web pages in a particular session. Therefore, analyzing this data is a primary goal for successful online businesses as they extract the clicks and behavior through web page requests. Our proposed solution is to provide a decisive and feasible recommendation algorithm that will allow us to predict the behavior of the shoppers'.

System Functionality

Firstly, we will build the model and analyze the performance of various classification algorithms such as Decision Tree, Random Forest, Support Vector Machine (SVM), and Naive Bayes. Then, the values of different evaluation metrics like Accuracy, Precision, Recall, F-score will be calculated to compare the performance of each of the algorithms. Lastly, we also plan on using these classification models to predict the shopper's intentions.

Requirements and Benefits

In today's economy e-commerce is becoming more extensive and businesses within this sector need to understand, the factors which come into play when a shopper ventures into a website to make a purchase. The benefit this holds is that it will enable the websites to better target ads or other factors which may lead to an increase in sales. These findings support the feasibility of accurate and scalable purchasing intention prediction for virtual shopping environments. This also helps in knowing the market capabilities of the brand when released in the new market while finding out the problems in the existing market and helps in relevant marketing strategies that can help in conquering the market.

Dataset Details and Core Algorithm

The dataset which is used is based on the "Online Shoppers Purchasing Intention" UCI dataset. It consists of numerical as well as categorical data. There are a total of 12,330 records where each row corresponds to the session data of the particular user. The total no. of records for which the session ended without any purchase is 10,442 which contributes to 84.5%.

The core algorithm which we will be implementing is a Decision Tree. Being a classification algorithm, it creates a tree-like structure by creating rules for breaking the dataset into small subsets in each step. We create a training model that is used to predict the class of the variable by simply learning decision rules deduced from the training set. At each step, a decision is taken to classify the data in the beneath classes. The leaf node holds the final results.

References:

[1] M. R. Kabir, F. B. Ashraf, and R. Ajwad, "Analysis of Different Predicting Model for Online Shoppers' Purchase Intention from Empirical Data," in *2019 22nd International Conference on Computer and Information Technology (ICCIT)*, Dec. 2019, pp. 1–6, doi: 10.1109/ICCIT48885.2019.9038521.

[2] C. O. Sakar, S. O. Polat, M. Katircioglu, and Y. Kastro, "Real-time prediction of online shoppers' purchasing intention using multilayer perceptron and LSTM recurrent neural networks," *Neural Computing and Applications*, vol. 31, no. 10, pp. 6893–6908, Oct. 2019, doi: 10.1007/s00521-018-3523-0.