**Prediction of Online Customers’ Purchasing Intention**

Team Name: YZman

Team Member: Xiayi Ye, Chenxing Zhai

**Our web page:** <https://github.com/zcx10025/DM-Project>

**Project description:**

The rapid development of e-commerce is inseparable from the advancement of big data technology. Analysis of customers’ behavior, purchasing intention and their preference are useful for business decision. For example, after you buy a computer at Amazon, next time you enter Amazon, it will automatically recommend some products related to computer to you, such as keyboard, mouse and so on. After we grasp many big data technologies this semester, we have a new understanding of online shopping. So we choose this topic to explore how big data technology predict online shoppers’ purchase intention. Through this topic, we can not only understand the analysis method of e-commerce from the perspective of a consumer, but also evaluate which consumers have a strong willingness to purchase from the perspective of a merchant.

**Project goals:**

We want to build several models to predict whether a visit will be finalized with a transaction with this data set. If our model is precision enough, perhaps it can really be used in reality to predict the customers’ purchasing intention. Because this will be a classification problem, so I decide to use confusion matrix to evaluate our model.

**Data Set:**

We download the data set from UCI Machine Learning Repository. The link is : [https://archive.ics.uci.edu/ml/datasets/Online+Shoppers+Purchasing+Intention+Dataset#](https://archive.ics.uci.edu/ml/datasets/Online+Shoppers+Purchasing+Intention+Dataset)

The dataset consists of 10 numerical, 8 categorical attributes and 12,330 rows. Description of them is below:

|  |  |  |
| --- | --- | --- |
| Name | Description | Type |
| Administrative | Number of administrative pages visited by the visitor | Numeric |
| Administrative Duration | Total time spent on administrative pages by the visitor | Numeric |
| Informational | Number of informational pages visited by the visitor | Numeric |
| Informational Duration | Total time spent on informational pages by the visitor | Numeric |
| Product Related | Number of product related pages visited by the visitor | Numeric |
| Product Related Duration | Total time spent on product related pages by the visitor | Numeric |
| Bounce Rate | Bounce rate of the pages visited by the visitor | Numeric |
| Exit Rate | Exit rate of the pages visited by the visitor | Numeric |
| Page Value | Value of the page visited by the visitor | Numeric |
| Special Day | Measure how close the day of the visit is to a special day | Numeric |
| Month | The day of the visit is in which month | Categorical |
| Operating System | Operating system version of the visitor | Categorical |
| Browser | Browser type of the visitor | Categorical |
| Region | Region where the visit is located | Categorical |
| Traffic Type | Ways for visitors to visit the web page | Categorical |
| Visitor Type | Whether the visitor is a new visitor, returning visitor or other | Categorical |
| Weekend | Indicate whether the day of the visit is weekend | Categorical |
| Revenue | Indicate whether this visit become a transaction finally | Categorical |

**Tools:**

We plan to use python for our project. First, package “numpy” and “pandas” indispensable. This two packages can help us load the data and calculate some indicators of the data. Packages like “matplotlib” and “seaborn” are also important. Because it is a machine learning problem, package “sklearn” must be included. With this package, we can perform feature selection, make a prediction, calculate the confusion matrix and so on.

**Literature Review:**

[1] Yi Jin Lima, Abdullah Osmanb, Shahrul Nizam Salahuddinc, Abdul Rahim Romled, Safizal Abdullahe, 2015. Factors Influencing Online Shopping Behavior: The Mediating Role of Purchase Intention.

Abailable at: <https://www.sciencedirect.com/science/article/pii/S2212567116000502#bibl0005>

Summary: When customers have a good impression of the product and feel that it is useful, their willingness to buy it will increase significantly. However, customers' doubts about the standardization of products on the website can adversely affect purchase behavior. In addition if the customer subconsciously feels that the product is useful, then they do not care whether to buy online or offline. In a nutshell, the wishes of customers determine whether they will buy online.

[2] Dai, H., Wang, L., Li, Y., Nie, Z., Wen, J. R., & Zhao, L. (2010). *U.S. Patent No. 7,831,685*. Washington, DC: U.S. Patent and Trademark Office.

Abailable at: <https://patents.google.com/patent/US7831685B2/en>

Summary: The data extracted from the web browser or search behavior can be used to detect users' browsing or search intent. The article mentions that machine learning can automatically detect and classify online users' business intent based on these data. Therefore, the related advertisement can be matched with the user or potential user who has the purchase intention to increase revenue.