Online Shopping Intention Analysis

Gabriel Guerra

Problem to tackle and motivation

In recent years, e-commerce has grown massively, with an increasing number of users with Internet access turning to digital platforms to meet their needs by purchasing household products, entertainment, food, and more.

From a data science perspective, it would be highly valuable for many companies to predict whether a user's visit to their website will result in a purchase. With such information, businesses could identify which variables encourage purchases and design strategies to convert visitors into customers. In fact, I have seen job openings at renowned tech companies with product teams dedicated to improving processes and testing whether changes and strategies implemented on their online sales platforms lead to higher conversion rates.

Techniques to be applied

With historical data on whether a purchase was completed or not, along with related metrics, several techniques studied in the first part of the course could be applied, particularly Bayesian generative models such as Naive Bayes, LDA, or QDA, since this is a binary classification problem (0 → purchase not made, 1 → purchase made). Of course, the classical statistical model for binary classification, logistic regression, would also be relevant here, as well as its Bayesian version, which I have not yet implemented and would like to explore on this occasion. I find it interesting to see whether Bayesian generative models can compete with logistic regression in terms of classification performance for this kind of problem.

Data sources

With this motivation, I found a dataset called Online Shoppers Purchasing Intention Dataset[1] in the UC Irvine Machine Learning Repository, containing 12,330 observations and 18 columns (10 numerical and 8 categorical), which provides information about users and their interactions on a web portal. The dataset includes a column named 'Revenue', indicating whether or not a purchase was completed.

Fundamental references

I also came across a paper (Sakar, Polat, Katircioglu, & Kastro, 2018[2]) using this dataset, which I found particularly interesting, as it shows that more advanced techniques, that will be covered in the later sections of the course, can be useful for this problem. Therefore, I plan to make extended use of this dataset in future projects.

---

[1] Dataset retrieved from link

[2] Sakar, C. O., Polat, S., Katircioglu, M., & Kastro, Y. (2018). *Real-time prediction of online shoppers' purchasing intention using multilayer perceptron and LSTM recurrent neural networks.* Retrieved from link