

Intel Neural Network Task

Megan Joseph

April 2025

1 Introduction

Customer behavior can often be unpredictable, but as the leader of the analytics team at a startup, it's important to analyze nonetheless. In order to help our clients, we need to assess what influences a purchase and how to better entice the customers to do so. This drew me to the Customer Purchase Behavior Dataset¹ on Kaggle. This dataset contains simulated data on different characteristics of users, like annual income, time spent on the website, and loyalty membership. This is similar to the data we hope we collect on our client's customers.

With this data, I aim to create a binary classifier to predict whether a user will make a purchase or not based on all or a subset of the given features. This will be a stepping stone towards our goal of individualizing customer rewards by predicting what kinds of items will prompt them to make a purchase. I will also be able to determine whether other factors, such as socioeconomic ones, may come into play.

While I had hoped to find data from an actual vendor, it was difficult to find, especially with the constraints. Many of the insights may not be representative of real life and how people actually make purchases, but understanding and testing the coding and math behind classification and neural networks is essential before production. As a result, this analysis will be valid in the knowledge gained.

2 Phase 1

In this phase, I loaded the dataset and normalized the data. I chose to do a min-max normalization because the distribution of the data has few outliers and many of the features are roughly uniform. Then, I visualized the distributions of the data before and after normalization.

After looking at the distributions, I looked at the mean, median, minimum, and maximum for each feature. Some features were binary while the rest were continuous. Something interesting is that the maximum time spent of the website in 59.99 minutes. Finally, I checked the distribution of the target variable to make sure it's not imbalanced. It's about 57% to 43% so it's not imbalanced.

¹Rabie El Kharoua. (2024). Predict Customer Purchase Behavior Dataset [Data set]. Kaggle. <https://doi.org/10.34740/KAGGLE/DSV/8725150>

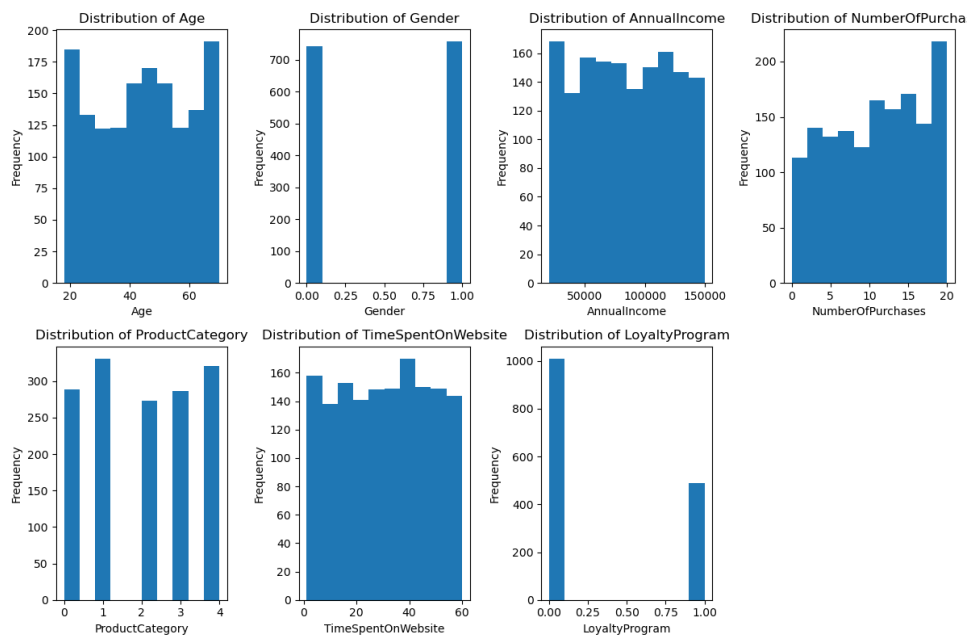


Figure 1: Distribution of features before normalization

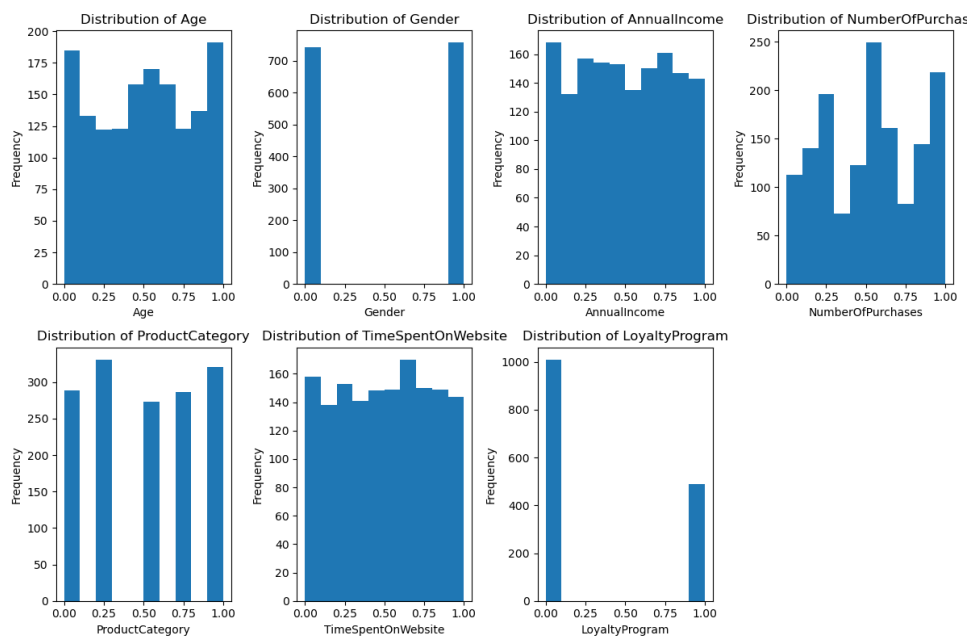


Figure 2: Distribution of features after normalization