

Final Project Report

Springboard – Capstone Two

An Analysis of Black Friday Sales

Tech Mart – a fictional big box store

Objective/Problem:

Use existing data to train a machine learning model that can predict purchase amounts from customer information. This will assist in making profit projections and determining the necessary inventory and staff requirements for the coming season.

Approach:

The existing data offers information about a customers location, marital status, age, and more. Initial data wrangling was preformed to ensure cleanliness.

Then the exploratory analysis began. A thorough analysis provided a better understanding of the data. And aided in finding oddities within the data that need to be altered.

Once the data was sorted, cleaned, and analyzed, the per-processing stage began. In this stage the data was encoded using one hot encoding. This ensures there is a column for every possible data type. And for each column there is either a 1 or a 0. Doing so allows the machine learning model to better fit the data and make predictions. Most machine learning models do not process objects\strings\categoricals. To finish up, the data was standardized. This will also improve model performance and reduce the effect of outliers.

Finally, multiple regression models were fitted and tested for their performance to select the best.

Additionally, the target column was binned so that the problem could be transformed to multi-classification.

From there multiple classification models were trained and tested on their performance to determine the best.

Findings:

The best performing model is a Linear SVR model, capable of predicting to within 0.37 standard deviations, or approximately \$2200. The classification models could not make it past 0.5% accuracy.

Some interesting notes: the majority of the customers were young, around 25 years old. Customers that had only lived in their current city for a year were also more likely to make purchases. The assumption would be that they are still purchasing basic necessities after having moved recently. Interestingly, all customers had similar purchase averages. With older customers(50-60) spending slightly more.

Ideas for Further Research:

In future predictions, it would be better to have descriptions of the product categories to improve the ability to make correlations. Perhaps an experiment that specifically analyzes the demographic of the stores customers could lead to some valuable insights. Allowing for specific marketing strategies to be employed.

Recommendations for Client:

1. Market towards the age groups that don't make as many purchases in hopes that you can increase overall profits. Also improve accessibility for the elderly so that they are more inclined to visit the store.
2. Capitalize on customers that have just moved. They will be on the look out for great deals to fill the voids left by moving. There could also be an opportunity in marketing to those that have been in the city for longer. Perhaps a push for a new aesthetic would bring more customers to the store.
3. A final option could be passing generated data to the machine learning model to determine the outcome of the specified data. This could aid in determining slumps in business due to various local changes.