Overview

This machine learning software will analyze listing data to determine what is important when getting impressions,clicks, and sales on a product.

EBay is a platform where you can sell your products. Second only to the powerhouse Amazon, eBay has been a source of income for small business owners across the world.

The success of a listing is measure in three parts as mentioned earlier.

- Impressions
- Clicks (Click Through Rate)
- Sales (Sales off of Clicks)

Cleaning

Lets get down into the basics of how this is setup.

Popular sellers on the platform can recieve what is called a traffic report. It is given in csv and when read into pandas reads the following columns

There are several problems here when we clean data for a machine. While some of these are for optimization others are necessary for the machine to learn and provide a reasonable inference.

Listing title (obj)
> Words used to describe the product

eBay item ID (int64)
> The platforms identifier for a listing

Current promoted listing status (obj)
> Whether or not you pay for extra impressions on your listing

Quantity available (obj)
> The amount the sellers has availible for sale at the current time

Total impressions on eBay site (obj)
> The total amount of impressions from "Organic" and Promoted

Top 20 search slot impressions from promoted listings (obj)

How many impressions where from the first 20 listings of a search from a promoted rate

% change in top 20 search slot impressions from promoted listings (obj)

I am unsure of what the comparison of the dates are. I believe you can set the period when getting a traffic report.

This particular report is from January to February

Top 20 search slot organic impressions (obj)

How many impressions where from the first 20 listings of a search with no promoted help

% change in top 20 search slot impressions (obj)

This particular report is from January to February

Rest of search slot impressions (obj)

These impressions are from searches outside top 20

Non-search promoted listings impressions (obj)

These are impressions from eBay marketing

% Change in non-search promoted listings impressions (obj)

Percent change from non search impressions

Non-search organic impressions (obj)

Impressions these are from seller internal eBay marketing

% Change in non-search organic impressions (obj)

Percent change in impressions from seller internal eBay marketing

% Change in non-search organic impressions (obj)

Impressions these are from seller internal eBay marketing

Total promoted listings impressions (applies to eBay site only) (obj)

Total impressions on eBay site from promoted listings

Total organic impressions on eBay site (obj)

Total impressions on eBay site from organic views

Total page views (obj)

These are our clicks. So very important

Page views via promoted listings impressions on eBay site (obj)

Clicks based on promoted listings

Page views via promoted listings Impressions from outside eBay (search engines, affilliates) (obj)

Clicks via non-eBay affiliate advertising on promoted listings

Page views via organic impressions on eBay (obj)

Clicks for non-promoted listings

Page views from organic impressions outside eBay (Includes page views from search engines) (obj)

Clicks via non-eBay affiliate advertising on non-promoted listing

Click-through rate = Page views from eBay site/Total impressions (obj)

Self explainitory

Quantity sold (int64)

Amount Sold and our only other int

Sales conversion rate = Quantity sold/Total page views

Self explanitory

Unnamed: 24 (float64)

A random column. Probably some number it has on the side of the sheet

Shape

Using what we know from the data we are using their are 24 columns and 1560 rows in the data. For training data this is a pretty small dataset. Most AI have closer to 10,000 or 100,000 rows.

We used random sampling to increase our AI's ability to learn. Instead of the orginal 1560 rows there are going to be over 200,000 rows to work with. This is done by copying samples from the previous data set at random to increase the amount of training data to learn from.

Nulls

There are a few ways to handle nulls. The article "A Comprehensive Guide to Data Preprocessing" by Samadrita Ghosh published January 27th of 2023 goes into great detail about the ways we can clean data for the machine. We are going to do removal because there are only 35 rows that contain dangerous nulls.

Handling Outliers will also be taken care by removing nulls. Most of the outliers in this dataset are outliers because they have null values. Those with nulls are going to be removed. Imputing values would cause more harm than good in the model.

Encoding

The only encoding needed for this project is going to be for the column "Current promoted listing status" whose values are logistical instead of numerical. The other columns that had potential are going to be removed in dimesionality reduction.

The one-hot encoding method is going to be used here. This method makes a unique column where a boolean which is a true or false denoted by a 1 for true or a 0 for false. This makes the column numeric instead of a column with text.

Dimensionality Reduction (Making Our Dataset Smaller)

This can also be called Feature Selection

Model Selection

Performance

Video

Software Demo Video

Development Environment

Visual Studio Code
Python 3

Pandas

Numpy

Sklearn

Altair

Imblearn

Pickle

Useful Websites

- scikit-learn
- Neptune AI