SENTIMENT ANALYSES FOR MARKETING

TEAM MEMBER

810621104304: BALACHANDRAN.M

Project Title: Sentiment analyses for marketing

Phase 4: Development Part 2

Topics: Continue building the analyses for marketing model by feature engineering, model training , and evaluation.

INTRODUCTION:

- In this article we will use a marketing analytics project to perform data cleaning, feature engineering, model selection and hyper-parameter tuning and explain why it is significant to use Let's look at the marketing problem statement to understand about our dataset.

- The paper is concerned with the construction of predictive and parsimonious classification models to support decision making in marketing applications. Classification aims at approximating a functional relationship between explanatory factors (i.e., features) and a discrete target variable on the basis of empirical observations. The features characterize an object to be classified, whereas the target encodes its membership to a-priori known groups. In marketing contexts, the objects usually represent customers and the target variable encodes some behavioral trait. Exemplary applications include, e.g., the selection of responsive customers for direct-mailing campaigns or the identification of customers at the risk of churning.

- A categorization of customers into groups suffices to solve a focal decision problem, e.g., whom to contact within a campaign. However, the more general data mining objective of deriving novel and actionable knowledge from data is not necessarily fulfilled, unless the prediction model reveals its internal functioning in a self-explanatory manner. In other words, it is essential that decision makers can understand the nature of the relationship that has been discerned from historical data and what factors govern customer behavior.

- This, in turn, facilitates adapting customer-centric business processes to, e.g., prevent future customer defection. Consequently, models are needed which are accurate in terms of the predictions they generate and interpretable with respect to their use of customer attributes.

- The objective of this work is to present results of a large empirical study that examines the effect of three different FS strategies on predictive accuracy and attribute set size to better understand the trade-off between highly accurate but complex and interpretable, parsimonious models within a marketing context. In that sense, the results serve as a first step towards the development of a general standard how to organize FS if an assessment of attribute importance is given by some feature ranking mechanism.

Dataset:

- The dataset consists of feature vectors belonging to 12,330 sessions. The dataset was formed so that each session would belong to a different user in a 1-year period to avoid any tendency to a specific campaign, special day, user profile, or period.

Attribute:

- ✓ Revenue => class whether it can make a revenue or no
- ✓ Administrative, Administrative Duration, Informational, Informational Duration, Product Related and Product Related Duration => represent the number of different types of pages visited by the visitor in that session and total time spent in each of these page categories.

- ✓ Bounce Rate => percentage of visitors who enter the site from that page and then leave ("bounce") without triggering any other requests to the analytics server during that session
- ✓ Exit Rate => the percentage that were the last in the session
- ✓ Page Value => feature represents the average value for a web page that a user visited before completing an e-commerce transaction
- ✓ Special Day => indicates the closeness of the site visiting time to a specific special day (e.g. Mother's Day, Valentine's Day) in which the sessions are more likely to be finalized with transaction. For example, for Valentine's day, this value takes a nonzero value between February 2 and February 12, zero, before and after this date unless it is close to another special day, and its maximum value of 1 on February 8
- ✓ Operating system ,browser, region, traffic type => Different types of operating systems, browser, region and traffic type used to visit the website
- ✓ Visitor type => Whether the customer is a returning or new visitor
- ✓ Weekend => A Boolean value indicating whether the date of the visit is weekend
- ✓ Month => Month of the year.

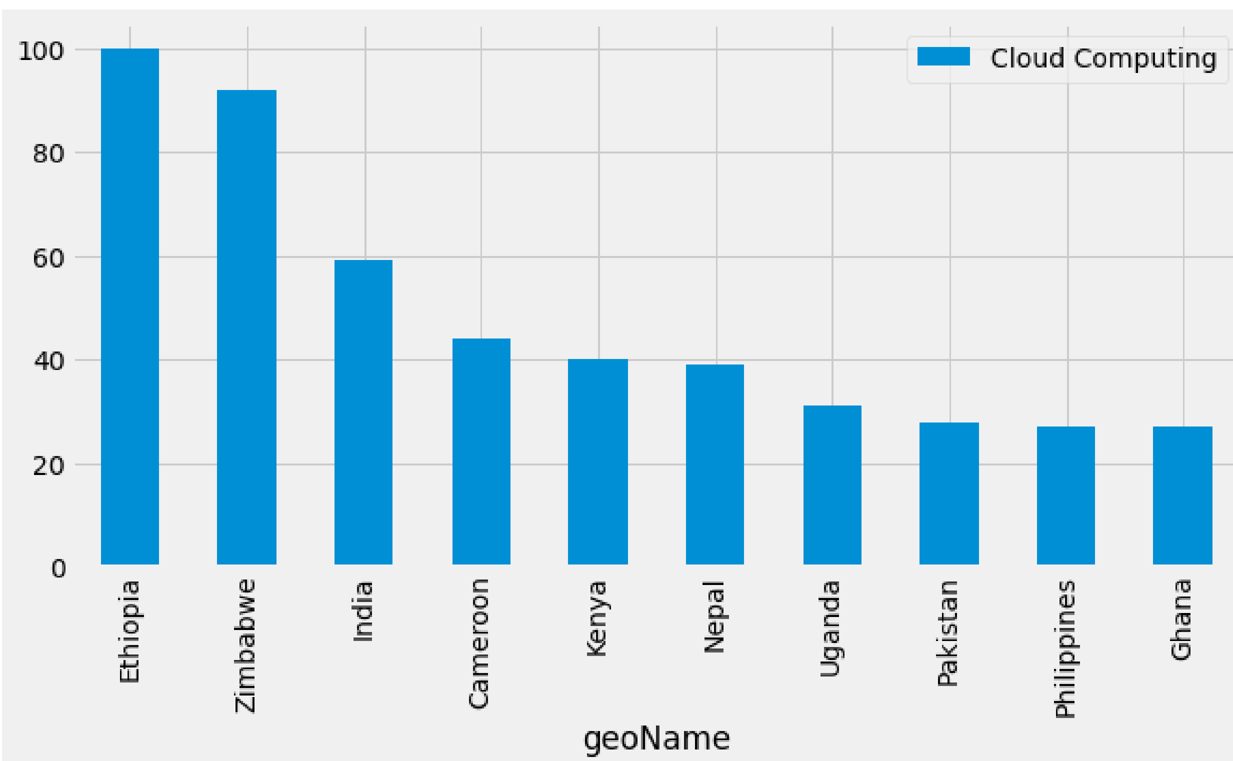
Overview of the process:

- ➢ According to Similar web, over 6 billion devices are connected to the internet, approximately 2.5 million terabytes of data are generated daily, and about 1.5 MB of data is created every second for each person. With this pile of data, marketers and business owners need to figure out how to personalize a product, offer it to a customer, and make buying easier. This process is possible with the help of data science.

Keyword Research Analysis

- The results of the first page of Google search get over 90% of the traffic. So it is becoming increasingly important for a business to appear at the top of popular search engines. Data scientists must constantly find new and innovative ways to improve search engine performance.

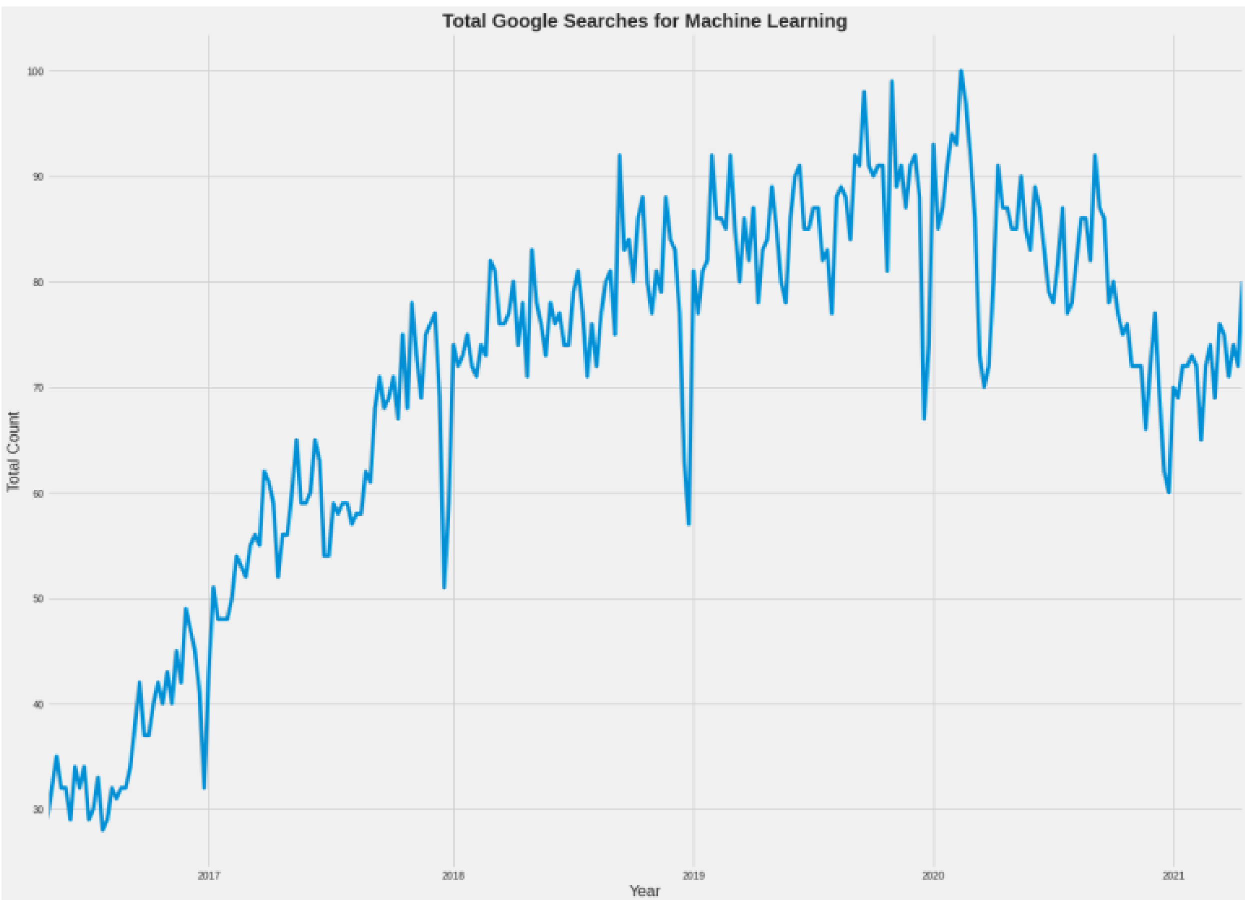
- In large tech companies, a data scientist can use Google Dataset Search to analyze the most searched keywords to find the best possible ones to use when promoting their company's products and services, thereby improving its search performance.



Keyword Research Analysis dataset visualization

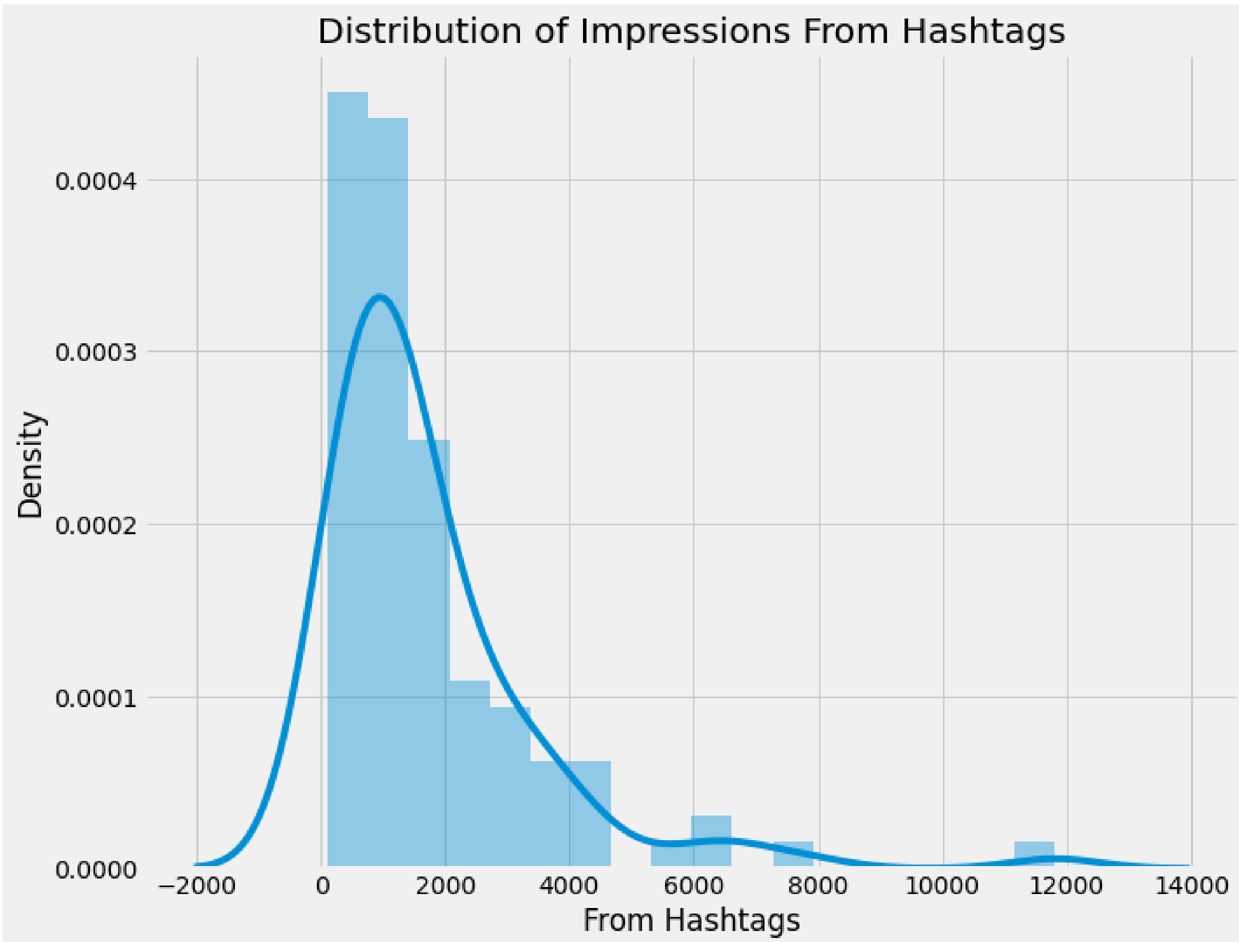
Google Search Analysis

- Google search analysis starts after you find the most used keywords. The aim is to analyze the search trends for that keyword on Google, which is very important for a business to reach more customers.



Social Media Campaign Analysis

- Social media platforms have become part of everyday life, and the number of tracking platforms is growing. However, people face significant challenges when properly harnessing these complex and noisy data streams.



MODEL TRAINING:

- The Pavlovian learning model is a powerful tool that can be used to understand how consumers make purchasing decisions. This blog explains the components of the learning model and why they are

important in making purchasing decisions.
components of the learning model and why they are important in making purchasing decisions.

Marketing Machine learning model:

- With R and R Studio installed and fired up, we'll start by creating a new script and loading the packages we need to build our model:

```
library(readr)
library(dplyr)
library(randomForest)
library(caret)
```

The first step of the process is to set a random seed:

```
set.seed(1234)
```

This should just ensure that our workflow and output is reproducible.

The proper first stage of the process is importing the data:

```
customer_data <- read_csv("repeat_customer.csv")
```

After that, we want to convert our gender variable from a character class of variable to a factor, and the output variable (returned) from an integer to a factor:

```
customer_data$gender <-
as.factor(customer_data$gender)
customer_data$returned <-
as.factor(customer_data$returned)
```

If we glimpse the data, we can have a quick look-see and make sure that the variables are as we want them, and dim will tell us how many rows and columns we are dealing with:

```
glimpse(customer_data)
dim(customer_data)
```

This is the output we see:

```
> glimpse(customer_data)
Observations: 340
Variables: 6
```

```
$ gender             <fct> male, male, male, male,
male, male, mal...
$ transaction_value    <int> 34, 45, 27, 57, 33, 28,
61, 13, 12, 7, ...
$ transaction_quantity <int> 4, 2, 6, 2, 7, 4, 2, 1,
1, 2, 4, 2, 3, ...
$ festive            <int> 0, 0, 0, 0, 0, 0, 1, 1, 1, 0, 0,
0, 1, ...
$ for_self           <int> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
0, 0, 0, ...
$ returned           <fct> 1, 1, 1, 1, 1, 1, 1, 0, 0, 0,
1, 1, 1, ...
> dim(customer_data)
[1] 340   6
```

<u>Model evaluation</u>:

- Market evaluation is the process of research and analysis of a specific market. This process involves gathering quantitative and qualitative data, such as the market size, statistics, average prices in the target market, customers' buying habits, and more.

How to Conduct Market Evaluation:

- A marketing evaluation is a simple process, but it takes a lot of dedicated research. Here, we will explore a step-by-step guide to teach you how to evaluate your market.

1. Determine the current state of the market.

The first section of your market analysis should include an overview or description of the market in which you want to operate. Take a look at some common factors that you should analyze

- Are there any current trends that affect the market?
- What is the growth rate of the market?
- What is the lifecycle of products and services in the market?

These are all questions your market overview should answer. This section is always at the top of your document, but it may be simpler to finish it after you have finished your research.

2.Analyze your target customer.

Analyzing your target audience is the next step of market evaluation. No matter how great your product is, you can not sell it to everyone. It would be a huge waste of time to try. Thus, the best strategy is to target your niche market, which will most likely buy your product.

You will need to define your segmentation strategy to help you find people in your target audience. Target segments enable you to segment your audience based on groups of shared characteristics, such as:

- Demographics: Age, gender, occupation, level of education, and religion
- Geographics: Where your target customer lives and works
- Psychographics: What your customers are interested in, what they are passionate about, or what personal values they are likely to have.

3. Analyze your competitors.

Another vital part of the market is your competitors. It’s very rare to find a market need that is not already being met by a business.

In this step of market evaluation, you will need to research the companies with whom you will be competing for customers. The following are some of the key topics you should cover:

- Identifying your direct and indirect competitors
- Market share of your competitors
- Their strengths and weaknesses

After you have done research on each competitor, it can be helpful to rank them from most dangerous to least dangerous to your business.

Feature Engineering

- Now let's work on the data little bit and jump into the feature engineering.

  - Handling Outliers

    Outliers increase the variability in your data, which decreases statistical power. Consequently, excluding outliers can cause your results to become statistically significant. Let's check out our numerical feature outliers through box-plot.

<u>Feature Scaling</u>

- Real world dataset contains features that highly vary in magnitudes, units, and range. Normalization should be performed when the scale of a feature is irrelevant or misleading and not should Normalize when the scale is meaningful.

- The algorithms which use Euclidean Distance measure are sensitive to Magnitudes. Here feature scaling helps to weigh all the features equally.

- Formally, If a feature in the dataset is big in scale compared to others then in algorithms where Euclidean distance is measured this big scaled feature becomes dominating and needs to be normalized.

  - Examples of Algorithms where Feature Scaling matters

    1. K-Means uses the Euclidean distance measure here feature scaling matters.
    2. K-Nearest-Neighbours also require feature scaling.
    3. Principal Component Analysis (PCA): Tries to get the feature with maximum variance, here too feature scaling is required.
    4. Gradient Descent: Calculation speed increase as Theta calculation becomes faster after feature scaling. Check this article to understand more.