Research Proposal / Proof-of-concept

Team B

Problem Statement

The client, Central Garden & Pet (Sponsor), wants to develop a reproducible analytical procedure to utilize customer reviews to extract useful insights such as customer satisfaction and product strategies.

Upon the initial examination of the customer review data, we identified that the dataset has 28,786 customer reviews about 11 brands of Sponsor, which were retrieved from 4 retailers - Homedepot, Walmart, Lowes, and Amazon. We also identified that the predominant sales revenue comes from Homedepot and 60% of the overall reviews were organic.

Given the qualitative aspect of the dataset regarding customer satisfaction, we would like to discover hidden business insights from the sentiment of customers towards Sponsor with the Natural Language Processing (NLP) model. The analysis aims to serve three main objectives - (1) Identify the overall sentiment from customers to provide insights to the Sponsor's marketing team in order to construct more efficient and targeted strategies. (2) Identify what aspects drive such sentiment- product, delivery, and so on. (3) Build a reproducible analytical foundation with an integrated database and coding set with qualitative data (customer reviews) for quantifiable metrics and findings.

Qualitative Objective/Scope

The objectives of implementing an NLP model on the given customer reviews are to tailor targeted marketing and customer strategies and to build appropriate product strategies to meet the needs of customers. Through this analysis, a strong list of use cases can be curated which can determine success metrics and pertinent KPIs for further study.

The ultimate goal of the model (discussed further in the *Plan* section) is to propose a solution to further understand (1) what sentiment customers are showing from their experience and (2) what keywords and key factors primarily drive customers' motivation to have a particular sentiment with their purchase whether it is positive or negative.

All of these would ultimately be used by the sponsor to plan and execute suitable marketing tactics. With this analysis, the marketing team and R&D team can incorporate the voice of customers more and further investigate problematic features for customers and eventually improve on the products mentioned.

I. **Investigate the overall sentiment of customer satisfaction and identify contributing factors to the sentiment**

Customers are at the core of any business and the sentiment analysis model can help assess the overall level of customer satisfaction and identify the key factors that drive positive and negative reviews from customers. For our project, we will examine factors including price, delivery speed, convenience, product effectiveness and quality, unique selling propositions, marketing outreach, etc. Accordingly, actionable strategies for the business can be to develop the product quality, undertake more deliveries and increase delivery speed, offer discounts and promotions, provide excellent customers, reach out to customers through unique marketing techniques and build a personal relationship with them.

II. **Investigate supporting insights for further strategies for individual brands**

Through sentiment analysis, we will surface insights on individual brands, which will frame analytical foundations to find significant insights to build further product strategies. Therefore, we will further conduct a detailed exploratory analysis to identify key factors that can potentially contribute to the success or failure of a brand amongst consumers. With the findings, Sponsor can leverage brand partnerships to position its products as unique and in demand, resonating with what customers are satisfied with and what not.

Metrics & Success Criteria & Risks

We have identified both intrinsic and extrinsic metrics that can be quantified.

I. Benchmark Measurements

These are benchmark analytics questions that we will use to measure increases in the effectiveness of our tool for future analytics questions.

1. Does **high price** link to negative reviews?
2. Does **delivery speed** contribute to positive or negative reviews?
3. Do products that customers find **easy to use** have positive reviews?
4. Does the **effectiveness** of products contribute to positive reviews?

II. Metrics and Success Criteria

Intrinsic Metrics - Focus on the performance of the models

1. **Extent** is to measure the extent of each topic assigned in our model (price, effectiveness, delivery, ease of use). It will be evaluated by the number of keywords of each topic matched and analyzed after training the dataset.
2. **Segmentation of reviews** is to measure how broad our analytics tool analyzed the reviews. It will be evaluated by the number of review clusters after training the dataset. These clusters refer to different segmentation of reviews.
3. **Prediction Error** is to measure how well our trained models can predict the test set. This is to see how effective our models can be in terms of being applied to a real dataset that is not used for training. This metric can be also a good measure to use for deciding which models to suggest in the end as the final reproducible codes for the Sponsor.

Extrinsic Metrics - Focus on the performance of the final outcomes

1. **The number of aspects needed to improve based on the analysis** is to measure the coverage of our suggested approach. We will initially have four pre-defined aspects - price, delivery, ease of use, and effectiveness. However, it is hard to promise that every review can be

grouped into the proper aspects, so this evaluation will help us decide whether we need to establish more labels with further analysis.

2. **The extent of objects achieved** is to measure the success of our model. We set up multiple objects at the beginning of the project and will examine how many of the total objectives we achieved (refer to the *Objectives/Scopes* section).

Plan

The following steps are derived from a data science process called CRISP-DM (Cross Industry Standard Process for Data Mining). From the steps of understanding Sponsor's given dataset to generating a reproducible model, we are going to follow this plan thoroughly.

While our initial exploratory data analysis (EDA) covers the understanding of the overall dataset and therefore has driven the modeling strategy of our project, it is imperative that we do some level of EDA on each brand every time we take a look at them because it may lead to different pictures that call out for different strategies. In addition, steps up until the final version can be iterative processes so as to improve the model as much as possible.

I. **Data understanding & preparation**

1. Explore and understand pre-processing steps

We will conduct a thorough EDA on the data and detect any issues that may impair the accuracy and integrity of the whole analysis. For instance, upon our initial examination, there are some issues with the raw dataset including mixes of upper & lower cases, review duplicates, and price recorded as 0. Understanding the pre-processing tools applicable to our case will impact how the dataset is going to be prepared for further analysis.

2. Feature Engineering (company-level data)

Regarding data quality issues found, we will take necessary actions accordingly. For example, we can unify category names, parse data, or lump them into smaller categories. This step is to reduce possible review cases that may not get analyzed in the process of NLP.

II. **Modeling**

1. Split data into train-valid-test using Cross-validation method (80%:10%:10%)

2. Execute the following on the train dataset ***(explained in detail in the Methodology section)***

   i) Zero-Shot text classification

   Categorize text data into several predefined labels and compute scores for each category to represent how relevant each review is to the categories.

   ii) Few-Shot text classification

   Improve upon the previous Zero-Shot classification by tuning with already labeled data and increasing topic relevance.

   iii) Sentiment Analysis

   Add a sentiment angle to the previous steps. Through this step, texts will be both classified into categories and labeled either negative, positive, or neutral.

   iv) N-gram (Bigram and Trigram)

   Tokenize two or three words together that are tailored to the Sponsor's business. This will enhance the model by being able to detect insights that are otherwise negligible.

III. **Performance monitoring & readjustments**

1. Evaluate initial Few-Shot text classification on the validation set, and measure performance using discussed metrics
2. Input different labeled data for different tuning with Few-Shot text classification
3. Repeat Sentiment Analysis, N-gram, and evaluate on the validation set to measure performance
4. Iterate these steps several times and pick the best resulting Few-Shot classification strategy

IV. **Development**

Apply the selected model to the test set and evaluate its performance using discussed metrics. This will be the final phase where we will be able to see how effective our final model can be when reproduced for future analysis needs.

V. **Final Deliverables**

After completing each modeling process, we will conclude the section with deliverables for the model that include research documents entailing details of the methodology, presentation of key findings, reproducible codes, and processed dataset.

VI. **Iteration for each Brand and each Retailer**

Individual brands and individual retailers will go through the steps from *I.3.'Simple EDA on prepared dataset' to IV. 'Deployment'*. Iteration by each brand is necessary to produce relevant modeling that matches each brand's uniqueness and its customers. Similarly, iteration by each retailer is important to incorporate distinct traits of the retailers and their customers into the modeling process. We can repeat tuning the model multiple times and deliver the finest model for this step.

VII. **Conclusion for each Brand and each Retailer**

We will show action plans that are tailored to each of or each group of 11 brands and each of 4 retailers as well as relevant KPIs that are trackable over time.

VIII. **Conclusion for the Project**

This part is to summarize key takeaways from analyzing data on a company level and by brand and retailer pillars. It will serve as a closing discussion with the best reproducible models for Sponsor.

IX. **Discussion on Possible Risks**

For further implementation and expandability, we outline possible roadblocks as follows to measure the performance of the final outcomes.

1. Ambiguity in clarification

The reviews included in the classification might not be consistent or clear because it might not be the case for one review to be matched to one classification and not to the other classifications. In other words, there might be overlapping sentiment categories and this issue can be more significant with generic models that have fewer numbers of predefined labels and do not reflect the uniqueness of the dataset. Therefore, we need to compare what different NLP models can do and if necessary, dig deeper into the unlabeled dataset to spot traits that can be used as an additional level that is specific to our dataset.

2. Innate emotional bias and limited implication

The bias of the dataset cannot be avoided as it consists of reviews resulting from customers' emotions. Incoherence in polarity, for example, can be observed for the same product depending on who leaves reviews. One customer might find the product highly effective while others do not, which can be due to the inconsistent product quality but also due to different expectations customers have or any biases that result from different personal backgrounds.

Methodology

I. **Background for Methodology**

1. What's the proportion of positive, negative, and neutral reviews aggregated on a company level and on a brand level?
2. What aspects of the reviews are contributing to positive/negative reviews? (price, effectiveness, delivery, ease of use)
3. What common words and phrases (n-gram) are used to categorize different aspects of reviews?
4. How tailored the sentiment (positive vs negative) is towards gardening and pet products?
5. What products return positive and/or negative reviews the most?
6. What unique value propositions are found in each product brand?

We will investigate it by breaking it down into price, delivery, ease of use, and effectiveness, as defined by Zero-Shot Classification Method. Furthermore, these factors would expand as we dive deeper into the Few-Shot Classification and N-gram model.

II. **Zero-Shot Text Classification**

To examine the overall sentiment of customer satisfaction towards the company, we will first categorize all the reviews into several labels using ***Zero-Shot Text Classification***. It is a methodology that adopts machines to learn with humanlike flexibility and efficiency for customer review analysis. The Zero-Shot Classification method takes an input of a phrase or sentence and appraises its relevance to predefined attributes. Furthermore, the method learns and trains itself from the infinite text data available online.

The method provides four basic category labels for classification - ***price***, ***delivery***, ***effectiveness***, and ***ease of use***. All reviews will get ratings for each category depending on the semantics related to each category. Applying this method to one sample review, the sentence was rated with a score of **0.97** in the attribute of ***price*** and **0.67** in ***effectiveness*** (example review below). We have arbitrarily chosen 0.7 as the cutoff score to label the reviews with some confidence.

```python
type = ['price', 'delivery', 'effectiveness', 'ease of use']
```
[12] ✓ 0.1s Python

```python
sample_review = "however, I should have done a price comparison before buying. \
    This product sells for less than half at the local store. Im sure its a wonderful product for plants and grass. \
        Ill be buying this at the local Home Depot/Lowes."

res = classifier(
    sample_review,
    candidate_labels = type,
    multi_label = True
)
res
```
[13] ✓ 2.5s Python

```
{'sequence': 'however, I should have done a price comparison before buying. This product sells for less than half at the local store. Im sure its a wonderful product for plants and grass. Ill be buying this at the local Home Depot/Lowes.',
 'labels': ['price', 'effectiveness', 'ease of use', 'delivery'],
 'scores': [0.9736257195472717,
  0.6739687323570251,
  0.3393574059009552,
  0.25122538208961487]}
```

While this is a great initial approach that allows us to use predefined attributes and reduce the amount of manual work of labeling data or text (Baranovskij, 2022), the findings sorely from zero-shot classification might have limits since the model predicts a large number of unseen object categories from few previously seen categories and thus there can be significant unseen categories (Dandu& Sharma & Bhandarkar, 2021). Therefore, we will take the analysis to the next step establishing more accurate labels based on the uniqueness of the dataset.

III. **Few-shot Classification**

While the Zero-Shot Classification worked well, it was unable to label 20% of the data into any of the predefined categories. However, when we observed some of the uncategorized reviews, we found out that some of these are still related to ***effectiveness.*** This suggests that the zero-shot classifications are returning some false negatives (example review below).

```
list(df_uncat['review_lower'])[1:3]
[36]  ✓ 0.4s                                                    Python
```

```
['followed directions; still have ants. no improvement. didnt work.',
 'the grass greened up nicely but i still have the weeds']
```

To improve this we can give the model with some labeled data to fine tune the zero-shot classification to fit our unique dataset better, which is, in other words, ***Few-Shot Classification***.

IV. Sentiment Analysis

Through zero-shot and few-shot classifications, we identified what those reviews are related to. With an added layer of sentiment analysis afterwards, we can identify if the reviews are positive, negative, or neutral.

For example, below is the same review as before in the zero-shot classification example. Zero-shot classification identified that this review is likely to be related to ***price***. Now the sentiment analysis identified that this review is a ***positive*** review with a score of **0.90.** Together we can conclude that this review is a ***positive*** review related to ***price.***

```
#sample sentiment analysis
from transformers import pipeline
sentiment_pipeline = pipeline("sentiment-analysis", model="cardiffnlp/twitter-roberta-base-sentiment-latest")

sample_review = "however, I should have done a price comparison before buying. \
    This product sells for less than half at the local store. Im sure its a wonderful product for plants and grass. \
        Ill be buying this at the local Home Depot/Lowes."

result = sentiment_pipeline(sample_review)

result
[37]  ✓ 4.4s                                                    Python
```

```
Some weights of the model checkpoint at cardiffnlp/twitter-roberta-base-sentiment-latest were not used when initializing
RobertaForSequenceClassification: ['roberta.pooler.dense.weight', 'roberta.pooler.dense.bias']
- This IS expected if you are initializing RobertaForSequenceClassification from the checkpoint of a model trained on another task or with another
architecture (e.g. initializing a BertForSequenceClassification model from a BertForPreTraining model).
- This IS NOT expected if you are initializing RobertaForSequenceClassification from the checkpoint of a model that you expect to be exactly identical
(initializing a BertForSequenceClassification model from a BertForSequenceClassification model).

[{'label': 'Positive', 'score': 0.9048624634742737}]
```

Combining the Zero-Shot/Few-Shot classification with Sentiment analysis we can intelligently label each of the reviews to be related to one or more of ***price, delivery, effectiveness,*** *and/or* ***ease-of-use*** with information on its sentiment.

	review_lower	price_m	delivery_m	effectiveness_m	ease of use_m	sentiment
0	this grass seed is excellent.	0.015903	0.014325	0.991545	0.133177	Positive
1	its bird food and the birds ate it...two thumb...	0.014241	0.067097	0.964720	0.495747	Positive
2	i am very happy with the results. i removed th...	0.384828	0.499681	0.921053	0.489307	Positive
3	i am a big fan of pennington bird seeds. i som...	0.022546	0.177406	0.719322	0.376021	Positive
4	came in convenient size bag. only problem the ...	0.023730	0.328921	0.255893	0.296941	Positive
5	this seed is great when your feeder is placed ...	0.153844	0.110225	0.929506	0.371779	Positive
6	i love this seed. it starts to sprout quickly ...	0.211188	0.289565	0.918216	0.980748	Positive
7	[this review was collected as part of a promot...	0.450796	0.129326	0.954556	0.727195	Negative
8	tbd but so far so good. planted 2wks ago. gras...	0.180277	0.166456	0.736569	0.365784	Positive
9	worked great, seeds sprouted within a week.	0.147515	0.458288	0.985626	0.962868	Positive

Reference

Baranovskij, A. (2022, January 9). Zero-shot text classification with hugging face. Medium. Retrieved October 4, 2022, from https://towardsdatascience.com/zero-shot-text-classification-with-hugging-face-7f533ba83cd6.

Cambridge University Press. (2009, April 7). Tokenization. Retrieved October 5, 2022, from https://nlp.stanford.edu/IR-book/html/htmledition/tokenization-1.html#:~:text=A%20token%20is%20an%20instance,containing%20the%20same%20character%20sequence.

Dandu, H., Sharma, K., and Bhandarkar, S. M.. "Exploring the Limits of Zero-Shot Learning - How Low Can You Go?," 2021. IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2021, pp. 1710-1719, doi: 10.1109/CVPRW53098.2021.00187.

Fedewa, D., and Holder, C. (2022, March 7). *Why business must heed customer reviews*. McKinsey & Company. Retrieved October 5, 2022, from https://www.mckinsey.com/capabilities/operations/our-insights/why-business-must-heed-customer-reviews.

Genoveva, G., & Siam, S. T. (2017). Analysis of Marketing Strategy and Competitive Advantage. International Journal of Economic Perspectives, 11(1).

Verma, Y. (2022, March 10). A complete tutorial on zero-shot text classification. Analytics India Magazine. Retrieved October 4, 2022, from https://analyticsindiamag.com/a-complete-tutorial-on-zero-shot-text-classification/.