Voice of the Policygenius Customer

Godwin Anguzu Lexin Lu Deekshita Saikia Michelle Van

Duke University 2023

1 Motivation

The importance of customer feedback cannot be overstated in business today. By listening to what customers have to say, companies can stay one step ahead of the competition and ensure that they are providing the best possible products and services. Identifying key topics and themes from this feedback can guide informed decisions and highlight areas for improvement. However, manual analysis of large amounts of customer feedback can be time-consuming and error-prone.

To tackle this issue, we have developed an automated text analysis pipeline that can efficiently identify key topics from customer feedback. The motivation behind this project is to provide businesses with a reliable and scalable way to analyze customer feedback, saving time and resources that would otherwise be spent on manual review. Moreover, manual analysis is often subjective and can vary greatly based on the reader's interpretation, whereas the automated pipeline uses a consistent and objective methodology. Leveraging advanced natural language processing techniques, the pipeline can identify key topics and themes that may not be obvious through manual analysis.

The output from the pipeline is then fed into a Tableau dashboard that displays various customer feedback analytics. By automating this process, we aim to significantly reduce the e ort required to identify key themes from customer feedback while improving the accuracy of the analysis. Ultimately, this project seeks to provide businesses with an e ective and efficient way to analyze customer feedback and gain valuable insights into their customers' needs and preferences.

2 Overview

Customer experience plays a significant factor in today's economy, with business models built around customers' needs and wants. The project aims to leverage structured and unstructured information to gain an understanding of the customers' reviews on Policygenius products and establish key themes that can be actioned to improve customer retention and overall satisfaction. Di erent unsupervised natural language models, including Latent Dirichlet Allocation (LDA), Latent Semantic Analysis (LSA), Non-negative Matrix-Factorization (NMF), and deep learning models - Bidirectional Encoder Representations from Transformers (BERT), were explored to achieve these goals.

Extensive model evaluations using coherence scores and intertropical distance maps were utilized to compare the di erent models and ultimately BERTopic model was chosen to identify the key themes from customer surveys. BERTopic was applied to the customer feedback data, which was divided into 4 segments based on varying customer sentiment levels: promoters, passives, detractors, and CSAT.

Results from our BERTopic models show that *great customer service* and *easy process* are the more prevalent themes amongst promoters, *communication issues* and *good service* in passives, *high quotes* and *poor organization* among detractors, and *great customer service* and *excessive contact* among CSATs. In addition, there are several overlaps between the di erent segments such as *long overall process* and *poor communication* in both passives and detractors. These analyses will allow Policygenius to gain insight into their customer sentiments. The findings of the key themes from this project matched those from Policygenius's manual theme extraction process results. Furthermore, yearly and monthly trends over the past 7 years were assessed, where certain trends supported Policygenius' past strategic product decisions over the years.

3 Background

The recent rapid proliferation of Internet technology has fueled a surge of interest in customer feedback among both academic researchers and business professionals. The ease with which customers can now easily share their experience or provide feedback on a product or service they have bought or experienced [1, 2], can a ect other customers' purchase decisions [3, 4]. Customers' statements are known as reviews (or product reviews), which are made readily available to a large audience via the Internet [5]. Both academics and business practitioners believe that the use of customer reviews strongly a ects the customer purchase process [6–8]. Research has shown that customer review is a vital, trustworthy source of information for people to gather product-related information [2, 9] that helps them avoid laborious search e orts and mitigate the risk of decision-making [6]. Businesses can leverage these reviews to assess customer satisfaction ratings and determine how to best improve or change their products and services.

Customer satisfaction is defined as a measurement that determines how happy customers are with a company's products, services, and capabilities [10]. The two widely used metrics for measuring customer satisfaction are NPS (Net promoter score) and CSAT (customer satisfaction) score [11]. NPS is considered a gold standard of customer experience metrics, designed by Fred Reichheld, Bain & Company, and Satmetrix Systems in 2003. Customers are divided into 3 categories according to scores collected from questions such as "How likely would you be to recommend...?" in surveys: "promoters" rated 9 or 10, "passives" rated 7 or 8, and "detractors" rated 6 or lower. The NPS is then calculated by subtracting the percentage of detractors from the percentage of promoters. On the other hand, the CSAT score measures the proportion of satisfied customers, who often are in the top-rated two groups from survey questions such as "How would you rate your overall satisfaction with...?". Companies often send out NPS surveys to customers who are 'in-force', i.e., actively pursuing or have successfully bought company products, while CSAT surveys are sent out to customers who have dropped out of the company application process [12].

However, with the considerable growth of text documents from surveys and on the Internet, manual analysis of this data is no longer feasible or scalable [2]. Efficient approaches to keyword extraction are necessary to identify the key elements of such documents [2]. Topic modeling methods have been established for text mining and can automatically extract topics from short texts [13] and standard long-text data [14]. These methods provide reliable results in numerous text analysis domains, such as probabilistic latent semantic analysis (pLSA) [15], latent semantic analysis (LSA) [16], latent Dirichlet allocation (LDA) [17] and non-negative matrix factorization (NMF) [18,19]. However, with recent advancements in the Natural language processing (NLP) field, emerging modeling techniques such as BERTopic [20] and Top2Vec [21] further complicate the process of big data analytics, pressing the need to evaluate the performance of di erent algorithms.

Currently, Policygenius relies on a manual approach to extract key topics. Our study aims to compare various topic modeling methods and identify an efficient model that can be used by Policygenius to extract key topics and track trends over time. This model will significantly reduce the time and e ort required to gain valuable insights and make informed decisions. Ultimately, the adoption of an efficient and automated topic modeling approach will result in significant time and cost savings for Policygenius.

4 Data

We utilized four data sources provided by Policygenius: a customer demographics dataset, a CSAT survey dataset, an NPS survey dataset, and an Application_id customer mapping table. The customer data contains over 62 million records of 12 demographic features including gender, BMI, etc, whereas the mapping table provides keys (*application_id, application_user_id, and product_type*) for merging the di erent datasets. 6,372 and 27,066 historical customer responses for CSAT and NPS surveys are recorded in these two survey datasets.

We began our data pre-processing by merging the above datasets to produce model-ready inputs. Our final datasets contain the NPS and CSAT scores, the open text field from the corresponding survey attributes, and customer demographics. To do this:

1. Only the survey comments, NPS scores, and the keys - *application_id* and *product_type* - were

retained in the NPS dataset.

2. The customer demographics were then mapped into this dataset by merging on *application_id* and *product_type*.

A visualization of the di erent data sources and their merges can be observed in Figure 1.

For the CSAT dataset, similar data merging was performed. Instead of merging on both *application_id* and *product_type*, the CSAT and customer demographics tables were joined only on *application_id*, as all the records in the CSAT data correspond to a single product: Life Insurance.

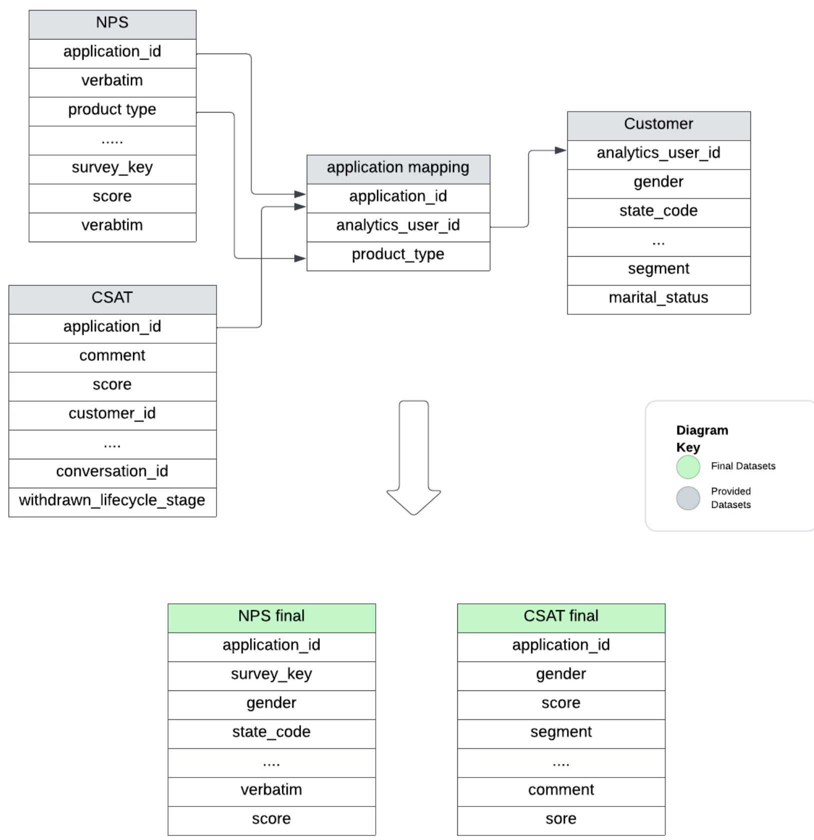

Figure 1: Data Pre-processing from sources

Since the open-ended comments in the NPS and CSAT surveys are optional, we observed a high number (approx. 45%) of missing survey responses in our dataset. These survey records were dropped from our analysis. We end up with 13,261 unique records in the NPS dataset corresponding to three products (Life, Home & Auto, and Disability Insurance), and 3,486 distinct records in the CSAT dataset, all corresponding to Life Insurance. The datasets are unique at the level of *survey_key*, which is a unique key produced for every survey sent out to the customers.

4.1 Data Pre processing

To prepare data for modeling, we performed the following steps on our text attribute:

Normalization: To convert words into lower cases, remove numbers, remove special characters, etc.

Tokenization: To separate text into words, or tokens, and creates the vocabulary we would feed into our model.

Removal of stop words: Stop words are defined as high-frequency words which do not hold any meaning by themselves, such as "the", "on", "she", etc. In our case, we also removed some specific words, like "Policygenius", which can appear in a majority of the comments, but does not contribute to identifying themes or topics.

Lemmatization and Stemming: Lemmatization helps convert words into their root, or lemma in linguistical terms. For example, "approve" and "approval" are both mapped to "approve". Stemming removes the derivational affixes. For instance, only "product" will be retained when the original word is "products".

After finalizing the data, we then subgrouped the whole NPS data into three customer segments by the score customers rated in the NPS surveys: 0-6 Detractors, 7-8 Passives, and 9-10 Promoters. Finally, we ended up with four model-ready data corresponding to Detractors, Passives, Promoters, and CSAT customers.

5 Methods

In this project, text analysis techniques were utilized to automatically identify patterns and extract valuable insights from large volumes of textual data. Text analysis provides people with useful machine learning and natural language processing (NLP) tools to quickly digest unstructured data; from categorizing text documents into predefined categories based on their content (Text Classification), determining the emotional tone of a piece of text (Sentiment Analysis), and extracting named entities from text data (Named entity recognition), to identifying the underlying topics or themes in a large collection of text documents (Topic Modeling).

For the purpose of this project, we employed a topic modeling approach to identify key topics from the raw text of customer reviews. However, since the nature of this project is unsupervised, there are no straightforward quantitative metrics to perform model selection and result reliability checks. Our solution is to implement various topic modeling methods and compare their results and performance by indirect evaluation metrics. We first fit di erent models on the NPS data using Latent Dirichlet Allocation (LDA), Structural Topic Modeling (STM), Non-Negative Matrix Factorization (NMF), and Bidirectional Encoder Representations from Transformers Topic Modeling (BERTopic). Then, we compared the topics results of di erent models - if one topic is extracted by most models, we are confident to believe that it is a huge talking point among customers and thus Policygenius should dig into it in depth. Otherwise, the results are less reliable and constructive if topics greatly vary across models. The evaluation metrics we utilized focus on whether the extracted topics are semantically meaningful. These include the coherence score, which measures the semantic similarity of words within a topic, and inter-topic distance maps, which visualizes the similarities and di erences between topics that have been identified in a corpus of text.

5.1 Topic Modelling Overview

Topic modeling is a powerful unsupervised NLP technique that is capable of scanning a set of documents, detecting word and phrase patterns within them, and automatically clustering word groups and similar expressions from large unstructured text data. It is widely used as a text-mining tool for discovering the hidden and abstract semantic patterns present in a set of text documents, or corpus. Topic modeling focuses on the relationships among words and co-occurrence, which are assumed to be driven by an underlying set of topics. In other words, models typically assume that the distribution of words within the corpus is at least partially a function of a set of latent topics that cannot be directly observed but can be understood through an examination of term co-occurrences. Statistical estimates of both term and topic frequency and distribution in the sample are expected outputs of a topic modeling algorithm. The most popular models include LDA, Latent semantic analysis (LSA), Correlated Topic Model (CTM) [22], and BERTopic [23], among others. Once topics have been identified, they can be used for a variety of purposes, such as document classification, information retrieval, and recommendation systems. Many real-world applications of topic modeling have added business value to companies, including social media analysis, market research, and content recommendation. Some of these methods have been discussed below.

5.2 Topic Modelling: Bag of Word Methods

Conventional Bag-of-Word methods of Topic Modelling, including LDA, LSA, CTM, etc, represent a document as a bag of words in matrix format and treat it as the model inputs. They all start with the conversion of a textual corpus into a Document-Term Matrix (DTM), a table where each row is a document, and each column is a distinct word. Then, with the assumption that each document is a mixture of latent topics, they apply di erent mathematical models to identify the hidden themes. Three methods were used in our analysis: LDA, STM, and NMF.

5.2.1 LDA

The most popular algorithm for topic modeling is LDA, which is a generative probabilistic model generating topics and corresponding word distributions based on word frequency and co-occurrence with other words in the observed text corpora using a three-level hierarchical Bayesian framework. It assumes that the topic probabilities provide an explicit representation of a document, while each topic is a distribution over words.

As the graphic representation of LDA shows in Figure 2, each document is generated as the following process:

Choose $N \sim Poission(\xi)$ and the topic proportion parameter $\theta \sim Dir(\ )$;

For each of the N words w_N:

- Choose a topic $z_n \sim \text{Multinomial}(\theta)$;
- Choose a word $w_N \sim p(w_N \mid z_N, \beta)$, a multinomial probability conditioned on z_N

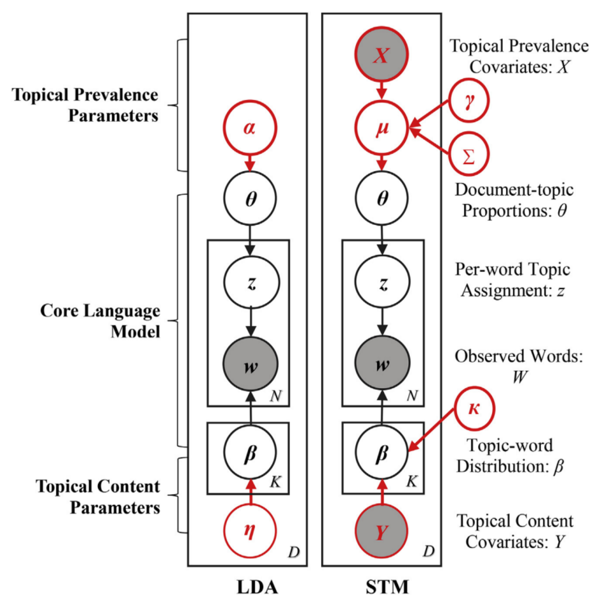

Figure 2: Graphic model representation of LDA Left) and STM Right).
Note Picture is cited from Hu et al. 2019) [24].

The empirical Bayes parameter estimation provides a generative probabilistic model of a corpus that assigns a high likelihood to the elements in the corpus and other similar documents. The approximate inference can be done by various methods, such as the EM algorithm, Laplace approximation, variational approximation, and Markov chain Monte Carlo.

5.2.2 STM

Structured Topic Modeling is a variation of LDA. As shown on the right of 2, the prominent improvement between STM compared to LDA is that STM allows for the inclusion of covariates of interest into the prior distributions to influence both topical prevalence (i.e. document-topic proportions) and topic content (i.e. topic-word distributions). In addition, unlike traditional LDA, topics (θ) can be correlated with each other through the distribution of μ. Moreover, each document has its own prior distribution over topics, defined by covariate X, rather than sharing a global mean. These di erences allow STM to require few a priori assumptions and discover topics that are associated with specific covariates to provide insights into how di erent factors influence the topics that are discussed in a text corpus.

5.2.3 NMF

NMF is a variation of LSA, one of the conventional and well-known topic modeling approaches as well. LSA uses Singular Value Decomposition (SVD) to reduce the dimensionality of the DTM and identify the underlying relationships between terms and documents. It assumes that words appearing in similar contexts have similar meanings, and it represents each document as a vector of weights for each term in the vocabulary. This allows LSA to identify the most important topics in a corpus of text based on the similarity between the vectors. However, the basis vector of SVD are orthogonal to each other, forcing some elements in the bases to be negative, which largely eliminate the interpretability. To output more interpretable and coherent topics, NMF decomposes the DTM into a topic-documents matrix U and a topic-term matrix V^T with the additional constraint that both two matrices can only contain non-negative elements.

5.3 Topic Modelling: Deep Learning Methods

5.3.1 BERTopic

Text embedding techniques have rapidly become popular in the natural language processing field. More specifically, Bidirectional Encoder Representations from Transformers (BERT) [23] and its variations [25–27], have shown useful results in generating contextual word and sentence vector representations.

BERTopic is a text embedding technique that utilizes transformers and c-TF-IDF to generate compact clusters, enabling the formation of easily interpretable topics while retaining significant keywords within the topic descriptions [20]. To make the underlying algorithm more accessible to a wider audience, BERTopic's main algorithm can be broken down into a set of steps, which are outlined below.

1. **Generate embeddings**: Documents are converted into numerical representations using sentence transformers, specifically "all-MiniLM-L6-v2" and "paraphrase-multilingual-MiniLM-L12-v2". The former, which we have utilized is an English language model that works well for most use cases and the latter is similar but works for 50+ languages. These models are optimized for semantic similarity and generate document or sentence embeddings.

2. **Dimensionality reduction**: After creating numerical representations of documents, dimensionality reduction is required as cluster models struggle with high-dimensional data. Uniform Manifold Approximation and Projection (UMAP) is the default choice in BERTopic, although other techniques such as principal component analysis (PCA) are also available. UMAP is

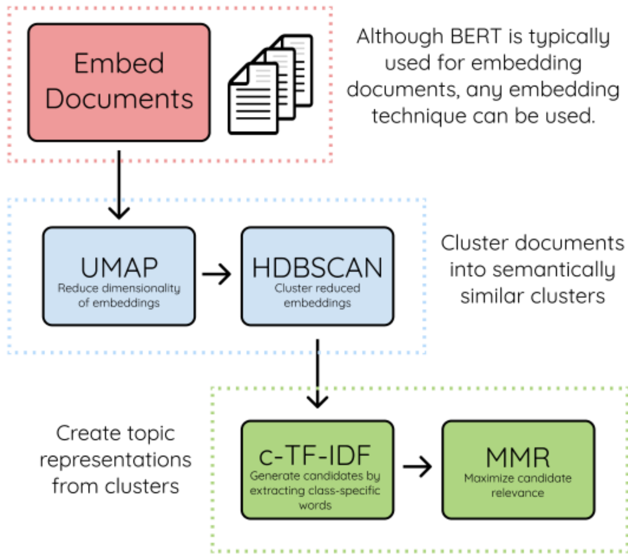

Figure 3: BERTopic visual overview

a non-linear technique that preserves the local structure of high-dimensional data when projecting it onto a lower-dimensional space. This structure is important for creating clusters of semantically similar documents.

3. **Cluster Documents**: The next step after reducing the embeddings is clustering the data, which is done using a density-based clustering technique called HDBSCAN. Unlike traditional clustering techniques, HDBSCAN can identify clusters of di erent shapes and also detects outliers. This results in a better topic representation as documents are not forced into clusters they do not belong to, resulting in less noise in the clustering.

4. **Bag-of-words**: The selection of an appropriate technique for modularity in BERTopic's algorithm is crucial, and when using HDBSCAN as a cluster model, a centroid-based topic representation technique might not be the best fit. Therefore, to create a topic representation that makes little to no assumption on the expected structure of the clusters, all documents in a cluster are combined into a single document, and a bag-of-words representation is generated by counting the frequency of each word in each cluster. This approach is cluster-level rather than document-level and allows for L1-normalization to account for clusters of di erent sizes.

5. **Topic representation**: To determine which words are typical for each cluster in BERTopic, TF-IDF is modified to consider topics instead of documents. By treating all documents in a cluster as a single document and applying TF-IDF, the importance scores of words within a cluster are obtained. The more important a word is within a cluster, the more representative it is of that topic. This modified model is called class-based TF-IDF, or c-TF-IDF:

   For a term x with class c:
   $$W_{x\,c} = \| tf_{x\,c} \| \quad log(1 + \frac{A}{f_x})$$
   ,where $tf_{x\,c}$ is the frequency of word x in class c, f_x is the frequency of word x across all classes, and A is the average number of words per class. The cluster is converted to a single document, and the frequency of words x in class c is extracted. Then, the logarithm of one plus the average number of words per class divided by the frequency of the word x across all classes are taken. The importance score per word in each class is obtained by multiplying tf with idf. This approach allows for a better representation of topics than the classic TF-IDF

algorithm.

6. **Fine-tune Topic representation optional)**: The c-TF-IDF method generates accurate topic representations quickly, but new methods are constantly being developed in the NLP world. To keep up with this, BERTopic o ers the possibility of fine-tuning the c-TF-IDF topics using various techniques such as GPT, T5, KeyBERT, and Spacy. The c-TF-IDF generated topics are considered candidate topics, with a set of representative documents for each topic that can be used for further fine-tuning. This reduces computation for large models, making them feasible in production settings, and takes less time than the dimensionality reduction and clustering steps.

5.4 Model comparison and evaluation

To determine which model is most suitable for this project, the 4 models mentioned above - LDA, STM, NMF, and BERTopic - were fine-tuned and applied to the entire NPS dataset. The model performances were evaluated using coherence scores, to understand the degree of semantic similarity between the high-scoring words in a given topic. We used c_v coherence metric that creates content vectors of words using their co-occurrences and calculates the score using normalized point mutual information (NPMI) and cosine similarity. The figure below shows the comparison in coherence scores between the models as the number of topics varies from 4 to 10 (too few topics may not cover all the themes that customers talk about, whereas too many topics may overlap with each other). There are 5 lines instead of 4 as the LDA model has one variation with TF-IDF and one without it.

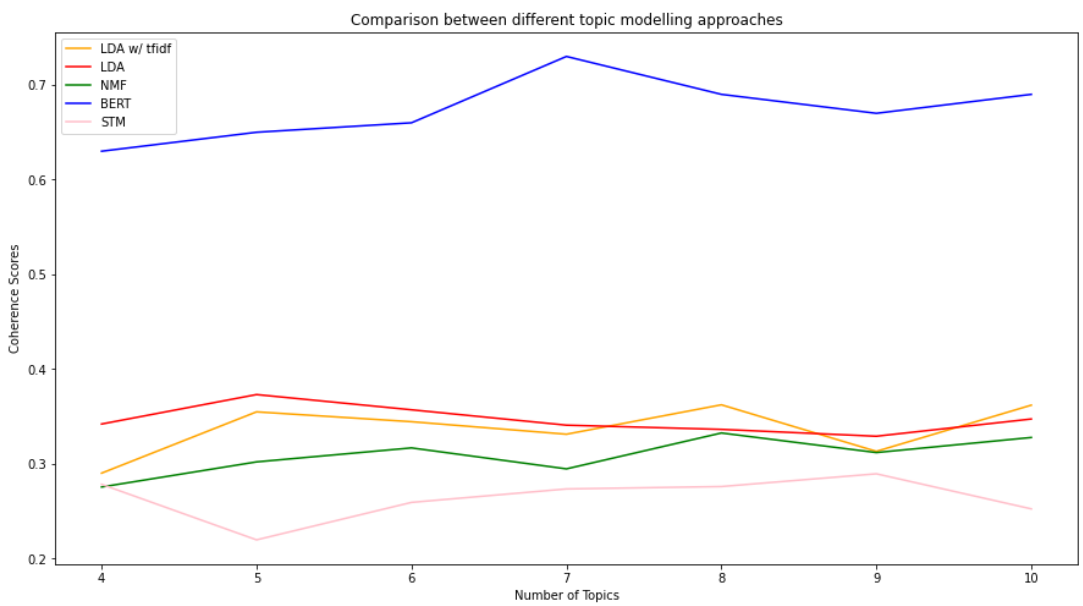

Figure 4: Model comparison using coherence scores

From the figure, it is clear that BERTopic continuously achieved higher coherence scores compared to the rest of the models. BERTopic gave a range of coherence scores from 0.62 to 0.73 while models such as LDA with TF-IDF and STM gave an average coherence of 0.35 and 0.32 respectively.

Due to the significantly higher coherence scores attained by BERTopic, BERTopic was chosen as the approach for this project. The high coherence score obtained by BERTopic can be explained by the structure of BERTopic that leverages transformers and c-TF-IDF to create dense clusters allowing for easily interpretable topics whilst keeping the important word topic descriptions. The embeddings in BERTopic allow the model to capture the contextual nature of the words, unlike the other models, which rely on a bag-of-words approach. Moreover, this structure allows for a more flexible algorithm that can adapt to new advancements in language models and clustering methods whilst handling noisy datasets.

6 Results

BERTopic was fit separately on the 4 di erent segments – promoters, passives, detractors, and CSAT, where each model was fine-tuned extensively via the parameters in the clusters and embeddings. Due to the unsupervised and qualitative nature of this work, a combination of human judgment and evaluation metrics was used to determine the number of topics for each segment.

BERTopic provided the following results:

Coherence scores

Topic-word score distribution - this maps out the top words in the topic and gives the ‘score’

Inter-topic distance maps - this shows the separability between the topic clusters. This helps minimize the overlap between the topics to ensure the separability between the topics.

Together with the comments corresponding to each topic, a general theme or topic was identified using human interpretation. Historic trends (yearly, quarterly, and monthly) over the past 7 years were plotted to show the proportion of each theme relative to the rest .

The Venn diagram in Figure 5 describes an overview of the key themes identified in each segment. It shows some overlap between di erent segments; for example, *Long overall process* and *Poor communication* are shared between passives and detractors, whilst *Excessive contact* and *High quotes* are shared between CSAT and detractors. This diagram can give Policygenius an indication of how the di erent segments’ sentiments relate to one another.

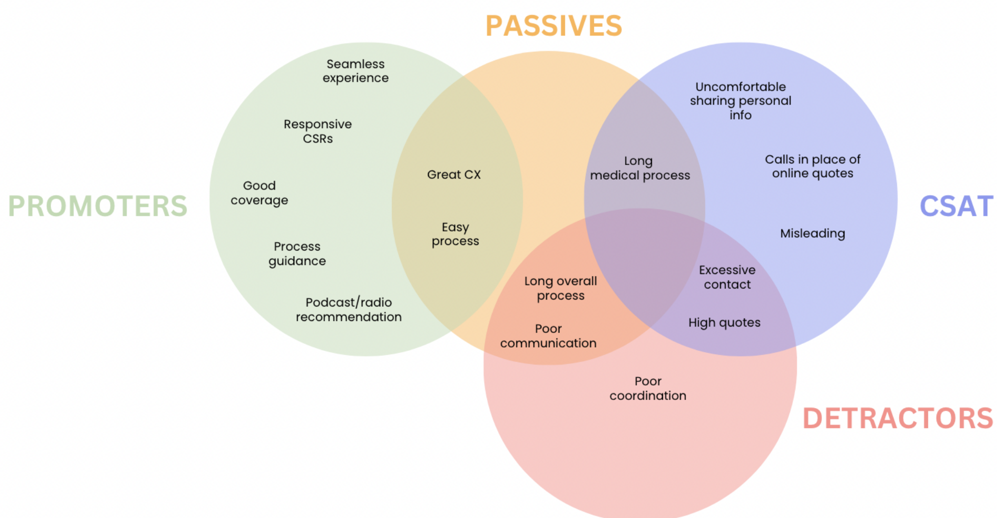

Figure 5: Overall topic results in segments

A further breakdown of the themes and trends for each segment is described in the next sections.

6.1 Promoters

Promoters are the segment of customers who are highly likely to recommend Policygenius to others. Here, the topics extracted by BERTopic are generally positive, highlighting some of the strengths of their insurance application process. The coherence score of the 7 topics obtained within this group is 60.11%. The topics have been visualized in the form of an inter-topic distance map as shown in Figure 6.

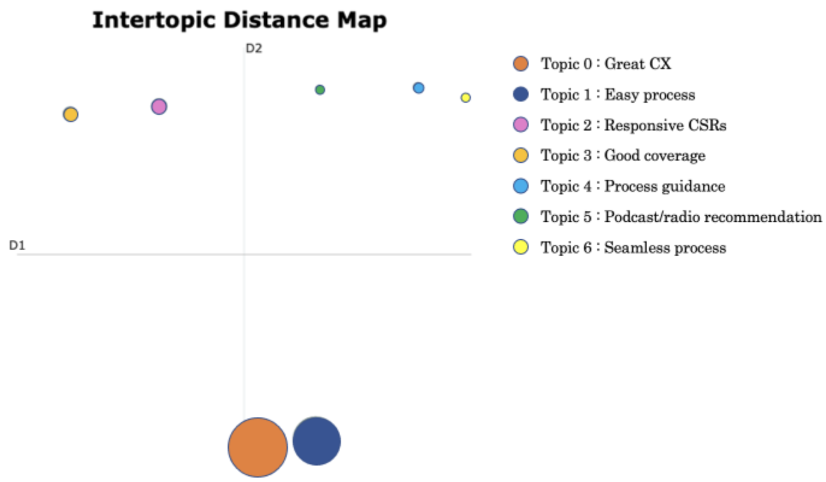

Figure 6: Topics extracted by BERTopic, Promoters

As can be observed, the topics are well distributed across the space. However, Topic 0: *Great customer experience* and Topic 1: *Easy process* are very close. This is expected, as customers often talk about these two topics simultaneously within their feedback. These topics are also of the biggest proportions among all other topics, suggesting that this segment's opinion is skewed toward these topics.

6.1.1 Topic word distributions

The distribution of the most frequently occurring terms is presented in Figure 7. The topics extracted within this segment are summarized below:

1. **Topic 0: Great customer experience** - The most frequent terms within this topic centered around *experience* and *service*. Customers usually spoke about their overall satisfaction and experience with the process.

   > "The experience was easy and I was given a variety of options. I was treated well by Harmony and was very pleased with the entire experience."

   > "Excellent customer service. Good follow-up. Delivered a solution to the problem we had. Very impressed."

2. **Topic 1: Easy process** - Within this topic, the recurring theme was the ease of the insurance application process.

   > "Working with you was fast and easy and I like the result."

   > "Easy process, multiple quotes with a recommendation provided, great UX/UI, and you can talk to a real human being located in the United States."

3. **Topic 2: Responsive Customer Service Representatives** - Within this topic, customers spoke about their good experience with the customer service representatives (CSRs), and how they are very responsive to the customers' queries and needs.

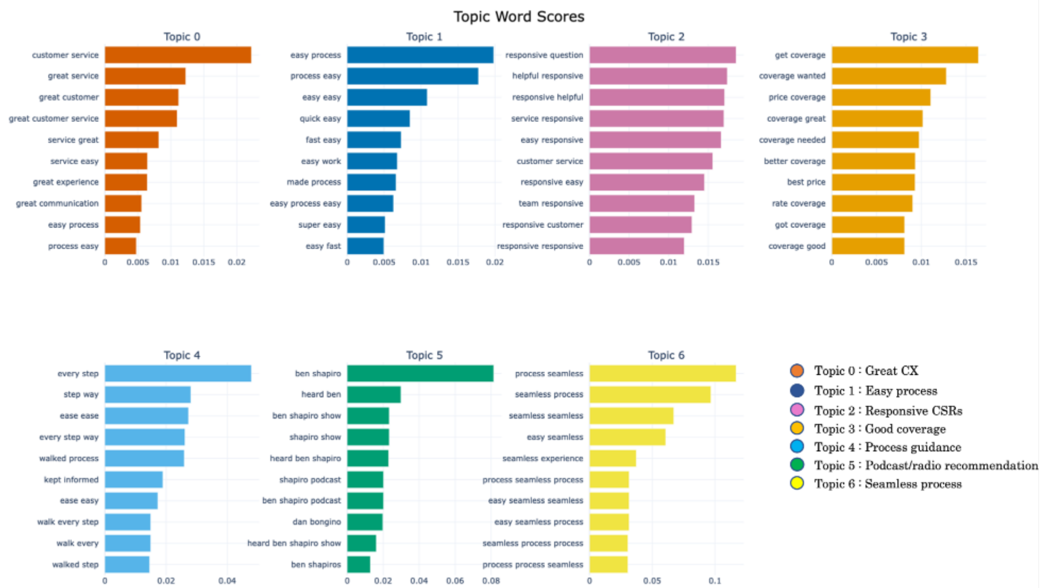

Figure 7: Most frequent words by topics, Promoters

> "Policygenius was very responsive and kept me up to date with the status of my insurance. A very customer-focused company that makes the process easy!"

> "Responsive team that worked diligently to move the process forward in a timely manner."

4. **Topic 3: Good coverage** - This topic highlights how customers feel that Policygenius provides good value for their money.

> "At the end of a 10-year term policy, my premiums more than doubled. I heard about Policy Genius from a radio ad and decided to try it. Found a policy for the same amount of coverage, but for 20 years instead of only 10 years, and quite a ordable for my budget. I would definitely recommend Policy Genius to others. Thank you."

> "We have been overspending for years on insurance and it feels amazing to be able to have even better coverage for less price."

5. **Topic 4: Process guidance** - Within this topic, customers talk about how there is guidance readily available for various stages of the application process.

> "Kept me updated every step of the way and when I needed help with something they were quick to assist and resolve the issue. Very friendly and courteous."

> "Policygenius agents walked me through the process of choosing the right policy for my family... and each step of the process thereafter. I also find the newsletters very informative as well."

6. **Topic 5: Podcast/radio recommendations** - Policygenius routinely sends out marketing campaigns via di erent podcast channels. In this topic, we see comments where people talk about how they were referred to the platform via podcasts.

> "I heard about Policygenius through the Ben Shapiro Show podcast. It was an easy

process, saved me money, and the representative that called me personally was kind and helpful."

"Heard about it on the Clark Howard podcast. I was very pleased with the ease of obtaining quotes and the follow-up afterward."

7. **Topic 6: Seamless process** - The most dominant term within this topic was *seamless*. Customers are appreciative of the smooth application process without any glitches.

    "Seamless. No hiccups. Everything digital."

    "The seamlessness of the process - takes the stress out of it completely! Thank you, PG!"

6.1.2 Trends of topic proportions

The trends of topic proportions within the Promoters group are displayed in Figure 8. As can be observed from the plot, Topic 0: *Great customer experience* and Topic 1: *Easy process* are the most frequent topics across all years. The trend corresponding to Topic 3, which we summarize as *Good coverage* has been steadily increasing since 2018. The trend corresponding to Topic 5: *Podcast/radio recommendations* fluctuates in tandem with their marketing campaigns.

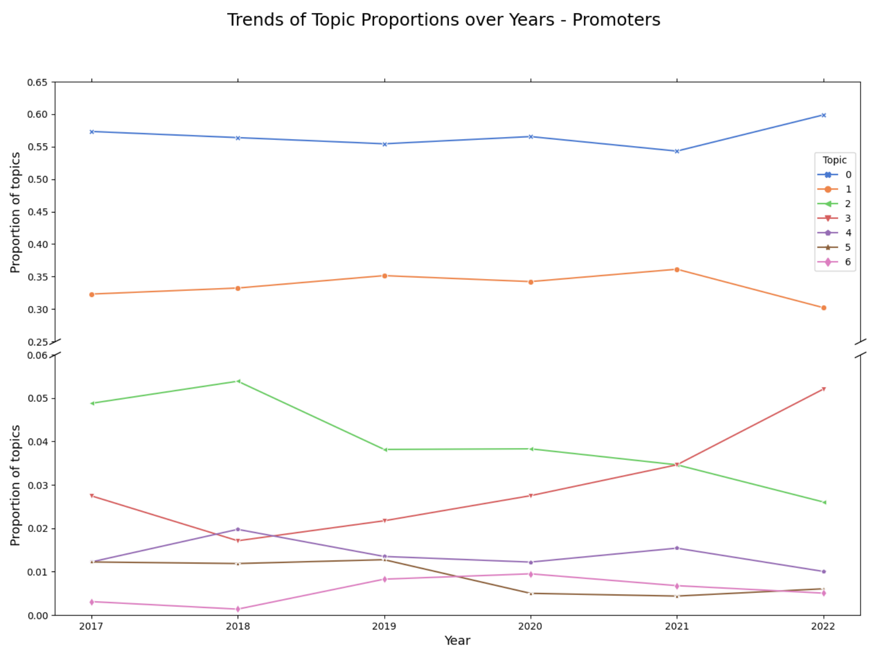

Figure 8: Trends of topic proportions, Promoters

6.2 Passives

Passive customers are those with neutral attitudes towards Policygenius' products and services (rate 7 to 8 out of 10). The fine-tuned BERTopic model found that there are five commonly mentioned topics, whose inter-topic distance map is shown in Figure 9.

As the above plot shows, the five topics that passive customers discussed are separated well. Topic 0: *Communication Issues* and Topic 1: *Good Service* account for the most proportions, which indicates that generally speaking, passive customers were satisfied with the overall customer service, but points were deducted for minor problems such as issues in communication and thus they were less

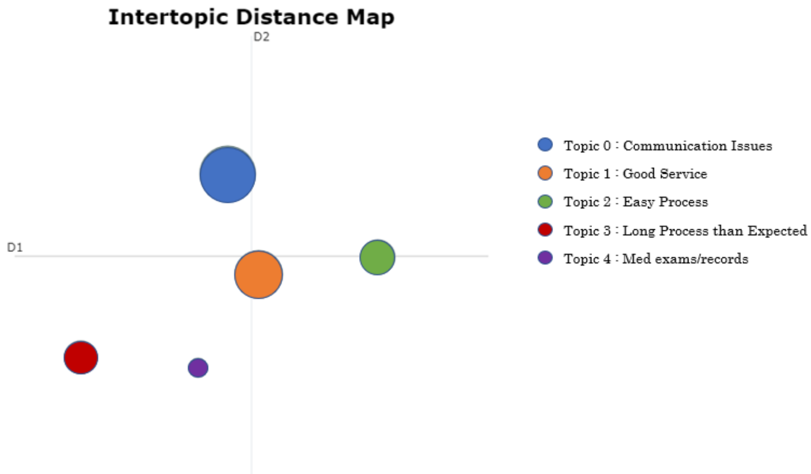

Figure 9: Topics extracted by BERTopic, Passives

motivated to recommend. The other problems include the process taking longer than expected, and delays in medical examinations.

6.2.1 Topic word distributions

The high-frequency words of each topic are visualized using a bar plot as shown in 10. With their help, as well as reading the raw comments, we summarized the topics of this customer segment as follows (with two sample original reviews followed):

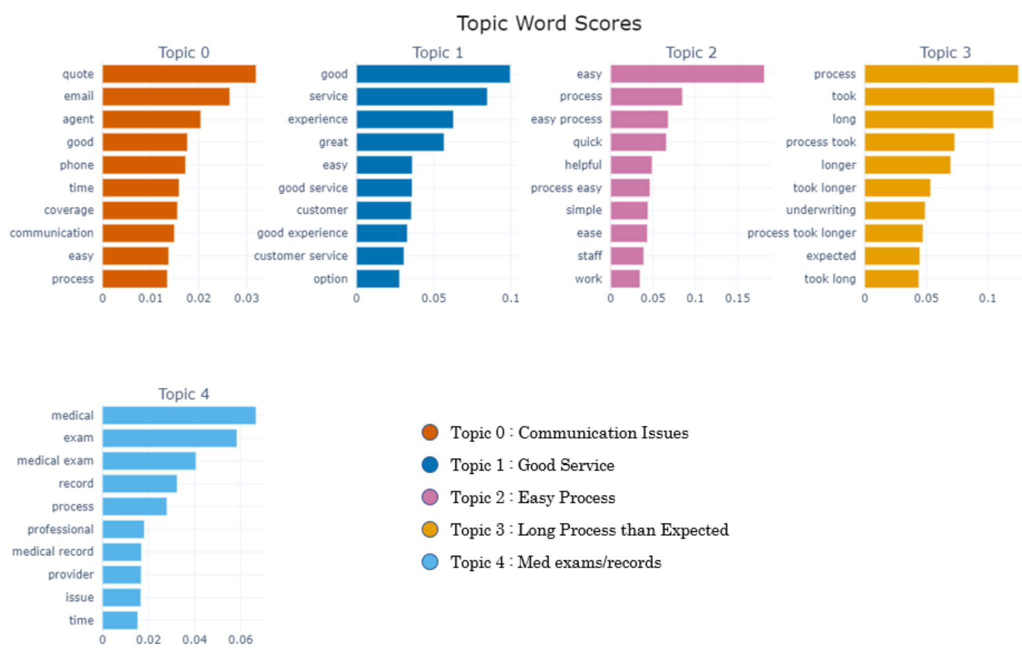

Figure 10: Most frequent words by topics, Passives

1. **Topic 0: Communication Issues** - Within this topic, people are frequently mentioned *email*, *phone*, and *communication*. Most passive customers complained about there are some communication issues during the application process.

"One complaint is that the agent assigned to me during the process would only communicate via email and text. Even when I asked the agent to call me, he did not. It would have been a better experience if the agent had been more easily accessible."

"Straightforward approach and easily reachable but too many emails/reminder texts."

2. **Topic 1: Good Service** - Similarly to promoters, passive customers were satisfied with the overall service that Policygenius provided. Here, we found that *good service*, *good experience*, and *great customer service* occur in high probability.

    "Customer service was e ective and helpful."

    "I'm new to the service but have enjoyed it so far."

3. **Topic 2: Easy Process** - Again, similar to the promoters, the recurring theme of this topic was the ease of the insurance application process of Policygenius.

    "It was a relatively easy process, and was just about right for us."

    "It's simple, it's easy, and you kept me informed along the way."

4. **Topic 3: Longer Process than Expected** - This topic reflects that passive customers were discontent about the time that the application process took.

    "Easy to use and search for the best price but underwriting took a bit long."

    "It's taking a lot longer to complete this process than I expected."

5. **Topic 4: Problems about Medical Exams/Records** - Within this topic, customers talk about how there is guidance readily available for various stages of the application process.

    "Generally I had a great experience. I knocked 2 points because I had difficulties getting my exam scheduled (I was rescheduled twice which was frustrating)."

    "I think Policygenius could have been more aggressive with the healthcare provider's delay in sending information to the underwriter. To be clear, the health provider was dragging their feet, and this was not due to any incompetency by Policygenius. But I still think there should've been more follow-up with the provider. I needed to step in after the provider did not submit the information for a couple of weeks to tell them they were not acting in a timely manner."

6.2.2 Trends of topic proportions

Figure 11 presents the yearly trends of topic proportions within the Passive customers from 2017 to 2022. We noticed some highlights from the plot: Policygenius has steadily received the praise of good services as Topic 1: *Good Service* grows over the year. However, Topic 0: *Communication Issues* has a generally increasing trend at the same time. Policygenius should draw more attention to dealing with communication barriers, such as the over-contacting problems in emails and texts. But it is reassuring to find that the topic proportions of the other two problems discovered - Topic 3: *Longer process than Expected* and Topic 4: *medical exam problems* - have consistently dropped, reflecting that people are positive about some improvements in the process of insurance application and medical records.

6.3 Detractors

Detractors are customers least likely to recommend Policygenius services to others. After fine-tuning the model, the number of topics for the detractor model attained was 5, with a coherence score of

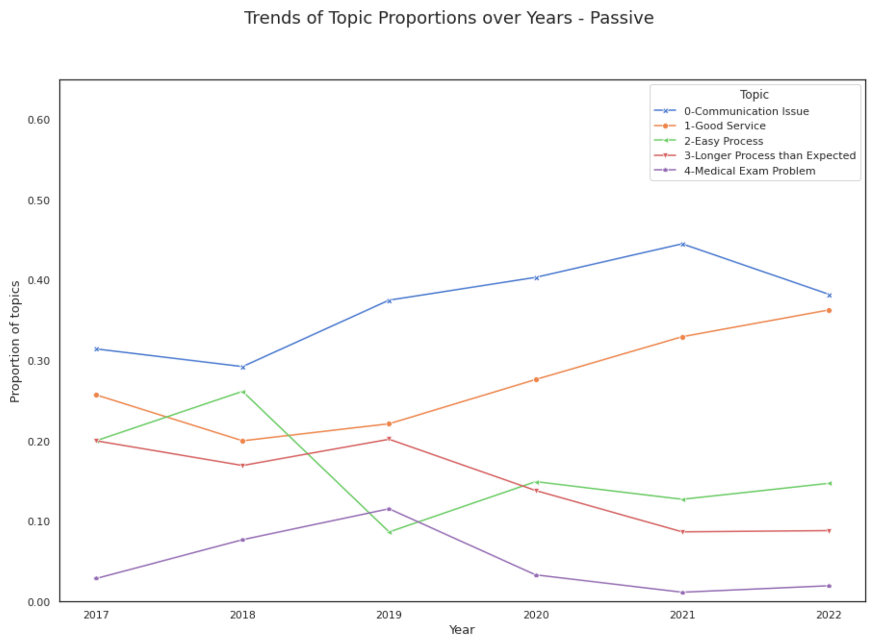

Figure 11: Trends of topic proportions, Promoters

0.62. The inter-topic distance map is depicted in Figure 12, showing a good separability of the topics of similar cluster size.

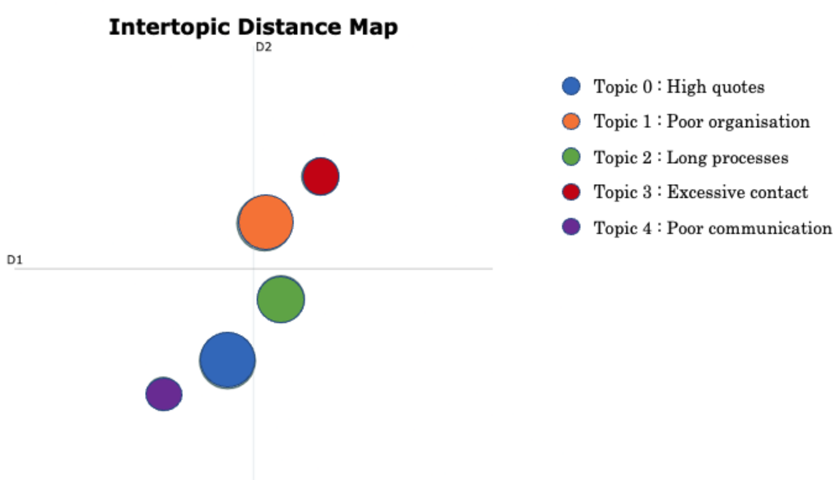

Figure 12: Inter-topic distance map, Detractors

6.3.1 Topic word distributions

The topic-word score distribution is shown in Figure 13, displaying the top frequent keywords per topic. Further explanation of each topic is described as follows:

1. **Topic 0: High quotes** - The top keywords in this topic are *customer service, felt like, bait switch, customer experience, long time*, suggesting that there is an issue with the customer experience. This topic relates to the high prices quoted by Policygenius. The customers felt

unsatisfied that the final quotes given to them were di erent and higher than the initial quotes given to them. Examples of the comments relating to this topic are shown below:

> "Final premium was almost 80% higher than quoted premium."

> "My rate was twice as much as what was quoted. No consideration for my healthy lifestyle and honesty."

2. **Topic 1: Poor coordination/organization** - The top keywords are *process took, took months, way long, long process*, with examples of associated comments shown below. The keywords and the comments suggest that poor coordination and organization issues exist in the Policygenius services. Problems with the agents handling customers' accounts include mishandling of account setup, and not knowing the correct coverage for canceled appointments.

> "There were SO many errors, miscommunications, canceled appointments, or never-scheduled appointments, appalling confirmations, and an inability to change things."

> "I had to go back and forth so many times with di erent employees about the confusion around my simple application. I'm honestly baffled that it was so difficult for your team to actually listen to me and sort out the application."

> "Policygenius needed at least 3 attempts to get my login info correct to access my policy to sign. Never got my blood work results."

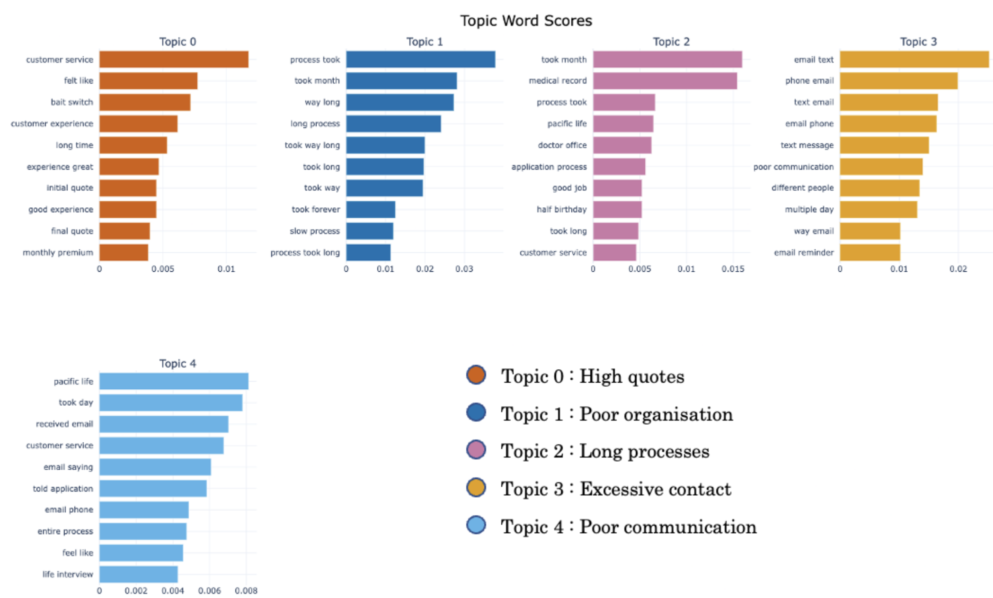

Figure 13: Topic word score, Detractors

3. **Topic 2: Long overall process** - The top keywords are *took month, medical record, process took, paci c life*. Within this topic, the comments clearly indicate that the process was too long with issues regarding slow response to the long application process. Sample raw comments are as followed:

> "Got reply after 1 month with so many reminders.. are you kidding?"

"Took me over a year to get insurance confirmed. That seems....long"

4. **Topic 3: Excessive contact** - The top keywords *text email text, phone email, text email, email phone, text message*. Customers complained about the excessive contact via emails, texts, and calls within this topic. Below are examples of the comments on this topic:

   "Too many phone calls and emails and texts. Harassment."

   "You all email and call way too much. It is overwhelming and ridiculous"

5. **Topic 4: Poor communication** - Within this topic, issues in communications from Policygenius are found. The company lacks communication in aspects ranging from customer service and pricing to the overall process, where customers were confused and unsure about the services provided due to poor communication. The top keywords of this topic are *paci c life, took day, received email, customer service, email saying* with examples of comments:

   "Lack of communication and lack of attention to detail. Amateur hour."

   "Lack of communication, delays in processing, scheduling of exam, lack of transparency in pricing."

6.3.2 Trends of topic proportions

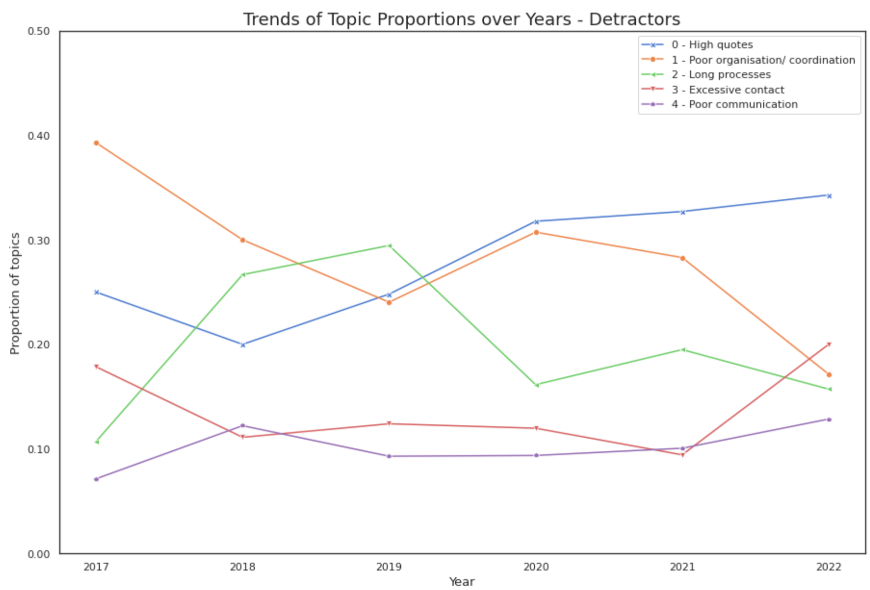

Figure 14: Yearly trend, Detractors

The graph above shows the trends of topic proportion over the past 7 years amongst the detractors. Generally, the sentiment of Topic 0: *High quotes* amongst customers has been growing over the past 6 years. In contrast, concerns of Topic 1: *Poor organization* and Topic 2: *Long processes* have been decreasing over the years. This can be supported by the fact that Policygenius changed its application process in 2019, which explains the reduction in the long process sentiment. Furthermore, Topic 4: *Poor communication* seems to be consistent throughout the years while the concerns about excessive contact show a sharp rise in 2021.

6.4 CSAT

The CSAT segment is customers who have previously used Policygenius' products but have dropped out. They are given a customer satisfaction survey to assess their level of satisfaction with the company's products and services. Reviews from these surveys are mostly negative as customers express their reasons for dropping out. Results from the BERTopic model show a coherence score of 76.85% obtained from the 7 topics extracted. The inter-topic distance map to visualize these topics can be found in Figure 15

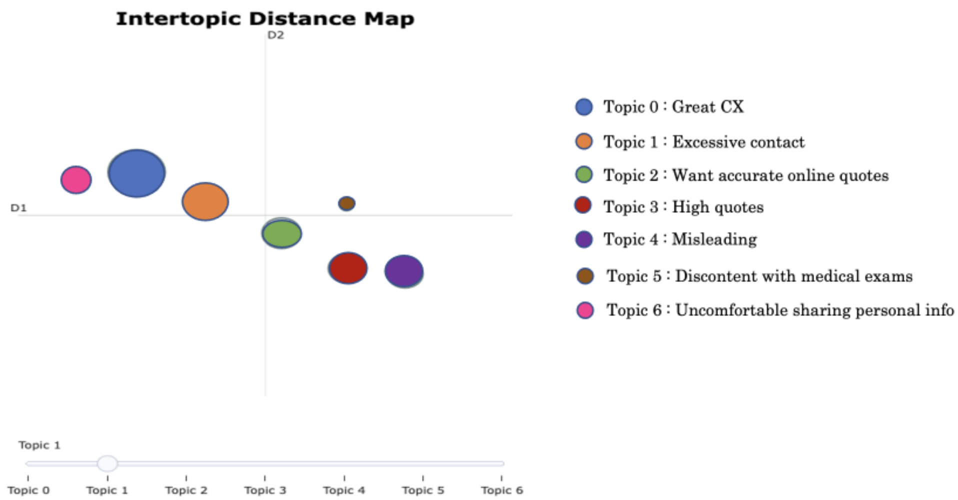

Figure 15: Topics extracted by BERTopic, CSAT

The visualization above indicates a clear separation between topics, with Topics 0 through 4 being relatively equal in size and distance. Topics 5 and 6, on the other hand, appear to be smaller in size. It is possible that the smaller size of Topic 5 is due to Policygenius' interventions aimed at reducing medical discontent among customers. Topics 0 - 4 remain of high interest among the CSAT segment.

6.4.1 Topic Word distributions

The distribution of the most frequently occurring words within the topics is presented in Figure 16. And the topics extracted within this segment are summarized below:

1. **Topic 0: Great customer service and experience** - The most frequent terms within this topic centered around customer experience and service. The majority of customers in the CSAT segment felt that the services provided were generally great, even though they may have had some concerns or discontentment around other issues.

   > "Great customer service, but would be better if the rep was trained to advise on term and whole life insurance."

   > "I experienced great service from rep A. Even though we didn't close a deal he was excellent."

2. **Topic 1: Excessive contact** - Customers have expressed concerns about receiving too many calls and texts from customer service representatives (CSRs), even after indicating that they do not wish to be contacted.

   > "Stop calling me so much and I did not appreciate being referred for lab testing when I never agreed to anything."

Figure 16: Most frequent words by topics, CSAT

> "As soon as my representative entered my phone number I started receiving non-stop calls from your call center, despite specific instructions to contact my rep. For this reason, I opted to go with a di erent company."

3. **Topic 2: Want accurate online quotes** - In this topic, customers have expressed their preference for accurate online quotes rather than being contacted by customer service representatives (CSRs). They would prefer that the entire process of application and receiving quotes be completed online, in order to simplify the process.

> "I'm no longer interested in doing business with Polygenius.I was tricked into signing up for a phone call and never asked whether I wanted a call. I just wanted online quotes."

> "I requested an email with quotes and was immediately taken o the list. I requested an online quote and did not receive an online quote. Do not want to call to get quotes, want them in writing."

4. **Topic 3: High quotes** - Customers have expressed their dissatisfaction with high quotes and prices of products, which often leave them with no choice but to seek out other, more a ordable alternatives.

> "Subsequent, the quote provided proved to be about 1/3 higher than the present policy I am under. So, great service, a lengthy interview, high price quote."

> "Price was just higher than the renewal of my current policy at least for next year by 300 dollars. Maybe you will be cheaper the following year. Plan to check back then."

5. **Topic 4: Misleading quotes** - Customers have expressed their discontent with dishonest and misleading information, particularly from customer service representatives. As a result, many customers feel that their expectations do not align with the reality of the services provided by the company.

> "They don't even know what you applied for and then change what you did apply for Initial quote was completely dishonest. Would not have wasted 40 minutes on the phone

if I had any indication the rate would double."

"Online quote which I received specifically mentioned that it doesn't require medical exam. But when I spoke to live person for 15 minutes and provided all the information, he mentioned that I will need one and rate was increased. It's not that I had issue with medical exam, but didn't like dishonesty, so disappointing."

6. **Topic 5: Discontent with medical exams** - Policygenius has faced criticism from customers who are unhappy with the mandatory requirement of medical exams. Many customers feel that this requirement is inappropriate and unnecessary, and as a result, some have sought out alternative providers that do not require medical exams.

   "I had everything lined up for a policy and then the person who was to give me a medical exam had to cancel at the last minute. Then, COVID shut everything down. I was reluctant to have a stranger come to our home to do an exam so I pursued a policy without one. I almost made it to the finish line, but then was told I would in fact need a medical exam. In that time I was pursuing other policies and found a company that would give me a fair rate and no medical exam."

   "I just found a provider that gave me same rates without a medical exam."

7. **Topic 6: Uncomfortable sharing personal info** - This topic relates to customers expressing their unwillingness and discomfort in providing personal information, particularly social security numbers, as a requirement during the process. Customers feel that there is no clear understanding of the need for social security numbers as requested by customer service representatives, and therefore, they are hesitant to provide this information.

   "He got very aggressive when I told him I was not comfortable giving him my social security number over the phone and asked if there was an alternative way."

   "I expected better before I provided my social security number. I wanted to view the application and have in writing how the number will be used. Unfortunately, that was not possible. I spent over 20 minutes answering questions that were already answered on the application online."

6.4.2 Trends of topic proportions

Figure 17 displays the trends of topic proportions within the CSAT group. The data shows that great customer service and experience, compared to other topics, has consistently had higher proportions throughout the years. Although customers generally feel the services are great, there are still some challenges that need to be addressed.

Between 2020 and 2022, there had been an increasing trend in Topic 1: *Great CX*. However, Topics 4, 5, and 6, which indicate the issues of Policygenius' service, have been following a downward trend, with Topic 5: *Discontent with medical exams* having the steepest slope. This steep decline could have been attributed to reforms in medical exam requirements that were implemented in 2020. Topics 2: *Excessive contact* and Topic 3: *Want accurate online quotes* have shown relatively constant slopes over the years.

To further analyze these trends, it is important to investigate the specific challenges that customers are facing in relation to each topic. This can help identify areas for improvement and inform strategies to enhance customer satisfaction. Additionally, tracking these trends over time can provide valuable insights into the e ectiveness of any changes or improvements made to customer service and experience.

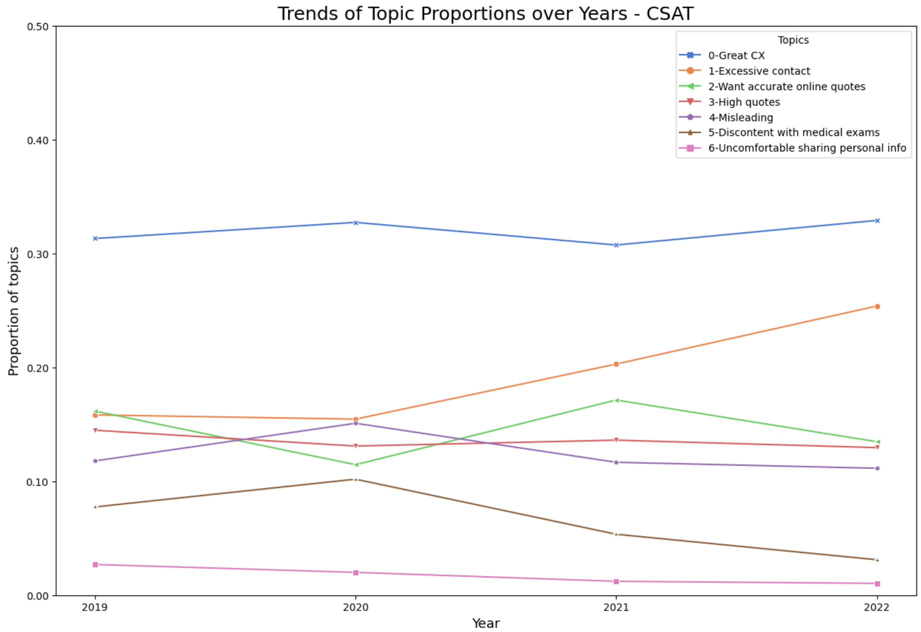

Figure 17: Trends of topic proportions, CSAT

7 Limitations

While such an automated text analysis pipeline can provide significant benefits for a business, there are also some potential limitations to consider, which are outlined below:

Quality of Customer Feedback: The e ectiveness of the pipeline is heavily dependent on the quality of the customer feedback. If the feedback is vague, incomplete, or inconsistent, the pipeline may not be able to accurately identify key topics and themes.

Limited data: Only about 50% of all surveys we worked with had recorded customer feedback in the open text fields. The quality of the topic model used by the pipeline is dependent on the quantity of training data. The model may not be e ective if the training data is not representative of the customer feedback being analyzed.

Biases: Some amount of sampling bias may be present as all customers may not be equally likely to receive a survey. The people who respond to surveys may also be di erent than the people who do not. This is especially the case with the CSAT dataset where we see people who drop out of the application process often do not like to leave detailed feedback.

Interpretability: Unsupervised learning techniques are dependent on pattern recognition and dimensionality reduction methods which may involve substantial information loss. As a result, final outputs may still be subject to human judgment for better interpretability.

8 Conclusion

In conclusion, the development of an automated text analysis pipeline using natural language processing techniques can provide Policygenius with a more objective and consistent methodology for analyzing raw text from customer feedback. With the natural text in the responses received from the surveys that Policygenius sends out, we have fit BERTopic models to extract underlying themes and customer sentiment. The integration of BERTopic and a Tableau dashboard can help Policygenius gain a more holistic understanding of its customer feedback and identify patterns and trends that may have gone unnoticed in previous manual analysis processes.

While there are limitations to this project, such as the potential for bias in the input data and topic modeling algorithm parameters, the automated text analysis pipeline represents a significant improvement over existing manual analysis processes. Overall, this tool has the potential to transform how Policygenius can e ectively approach customer feedback analysis, providing deeper insights and informed decision-making for improving its products and services.

9 Technical documents

todo

Currently, we are still working on the technical documents. We will put them in this section once we completed them.

References

[1] D. Godes and D. Mayzlin, "Using online conversations to study word-of-mouth communication," *Marketing science*, vol. 23, no. 4, pp. 545–560, 2004.

[2] C. Dellarocas, "The digitization of word of mouth: Promise and challenges of online feedback mechanisms," *JSTOR*, 2003.

[3] S.-C. Chu and Y. Kim, "Determinants of consumer engagement in electronic word-of-mouth (ewom) in social networking sites," *International Journal of Advertising*, vol. 30, pp. 47–75, 01 2011.

[4] S.-J. Doh and J.-S. Hwang, "How consumers evaluate ewom (electronic word-of-mouth) messages," *CyberPsychology & Behavior*, vol. 12, pp. 193–197, 2023/04/03 2008.

[5] T. Hennig-Thurau and G. Walsh, "Electronic word-of-mouth: Motives for and consequences of reading customer articulations on the internet," *International Journal of Electronic Commerce*, vol. 8, pp. 51–74, 12 2003.

[6] W. Duan, B. Gu, and A. Whinston, "The dynamics of online word-of-mouth and product sales: An empirical investigation of the movie industry," *Journal of Retailing*, vol. 84, pp. 233–242, 05 2011.

[7] K. Floyd, R. Freling, S. Alhoqail, H. Y. Cho, and T. Freling, "How online product reviews a ect retail sales: A meta-analysis," *Journal of Retailing*, vol. 90, no. 2, pp. 217–232, 2014.

[8] K. P. Gu Bin, Park Jaehong, "Research note: The impact of external word-of-mouth sources on retailer sales of high-involvement products," *JSTOR*, 2012.

[9] J. A. Chevalier and D. Mayzlin, "The e ect of word of mouth on sales: Online book reviews," *Journal of Marketing Research*, vol. 43, pp. 345–354, 2023/04/03 2006.

[10] R. Hallowell, "The relationships of customer satisfaction, customer loyalty, and profitability: an empirical study," *International journal of service industry management*, 1996.

[11] A. Baquero, "Net promoter score (nps) and customer satisfaction: Relationship and efficient management," *Sustainability*, vol. 14, no. 4, p. 2011, 2022.

[12] C. Bălan, "Net promoter score: Key metric of customer loyalty.," *Quality-Access to Success*, vol. 13, 2012.

[13] B. Cheng, I. Ioannou, and G. Serafeim, "Corporate social responsibility and access to finance," *Strategic management journal*, vol. 35, no. 1, p. 1–23, 2014.

[14] P. Xie and E. P. Xing, "Integrating document clustering and topic modeling," *arXiv preprint arXiv:1309.6874*, 2013.

[15] T. Hofmann, "Probablistic latent semantic indexing proceedings of the 22nd annual international acm sigir conference on research and development in information retrieval," *ACM Press: Berkeley, CA, USA;*, 1999.

[16] S. Deerwester, S. T. Dumais, G. W. Furnas, T. K. Landauer, and R. Harshman, "Indexing by latent semantic analysis," *Journal of the American society for information science*, vol. 41, no. 6, p. 391–407, 1990.

[17] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation," *Journal of machine Learning research*, vol. 3, no. Jan, pp. 993–1022, 2003.

[18] D. Maier, A. Waldherr, P. Miltner, G. Wiedemann, A. Niekler, A. Keinert, B. Pfetsch, G. Heyer, U. Reber, T. Häussler, H. Schmid-Petri, and S. Adam, "Applying lda topic modeling in communication research: Toward a valid and reliable methodology," *Communication Methods and Measures*, vol. 12, pp. 93–118, 04 2018.

[19] H. Chen, M. Gao, Y. Zhang, W. Liang, and X. Zou, "Attention-based multi-nmf deep neural network with multimodality data for breast cancer prognosis model," *BioMed Research International*, vol. 2019, p. 9523719, May 2019.

[20] M. Grootendorst, "Bertopic: Neural topic modeling with a class-based tf-idf procedure," *arXiv preprint arXiv:2203.05794*, 2022.

[21] D. Angelov, "Top2vec: Distributed representations of topics," *arXiv preprint arXiv:2008.09470*, 2020.

[22] D. M. Blei and J. D. La erty, "A correlated topic model of science," 2007.

[23] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," 2018.

[24] N. Hu, T. Zhang, B. Gao, and I. Bose, "What do hotel customers complain about? text analysis using structural topic model," *Tourism Management*, vol. 72, pp. 417–426, 2019.

[25] J. Li, X. Zhou, and Z. Zhang, "Lee at semeval-2020 task 12: A bert model based on the maximum self-ensemble strategy for identifying o ensive language," in *Proceedings of the Fourteenth Workshop on Semantic Evaluation*, pp. 2067–2072, 2020.

[26] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov, "Roberta: A robustly optimized bert pretraining approach," *arXiv preprint arXiv:1907.11692*, 2019.

[27] Z. Lan, M. Chen, S. Goodman, K. Gimpel, P. Sharma, and R. Soricut, "Albert: A lite bert for self-supervised learning of language representations," *arXiv preprint arXiv:1909.11942*, 2019.