Enhancing Net Promoter Score Analysis through Natural Language Processing: A Case Study on E.ON

1st Aadam Nadeem
School of Computation, Information and Technology
Technical University of Munich
Munich, Germany
adam.nadeem@tum.de

2nd Hung-Yu Yang
School of Management
Technical University of Munich
Munich, Germany
hung-yu.yang@tum.de

***Abstract*—This study presents a comprehensive approach to augmenting Net Promoter Score (NPS) analysis with advanced Natural Language Processing (NLP) techniques. We aim to close the gap between quantitative NPS measurements and the qualitative information concealed in customer feedback by concentrating on sentiment analysis and topic modeling. Our research provides a detailed insight into consumer opinions and preferences and identifies the primary reasons behind changes in NPS through the study of almost half a million customer evaluations. With the help of this dual strategy, NPS may be interpreted more intelligently, leading to focused increases in customer loyalty and happiness. NPS is a valuable metric, indicating how likely users are to promote the company/service they are using, and it can provide insights into customer loyalty.**

***Index Terms*—Net Promoter Score, Natural Language Processing, sentiment analysis, NPS correlation, topic modeling, loyalty drivers**

I. Introduction

Customer satisfaction and customer loyalty have always been crucial for EON. For this reason, customer surveys are conducted by EON's regional representatives every month across various European countries. In Harvard Business Review (HBR), Reichheld (2003) claimed that the Net Promoter Score (NPS) is "the one number you need to grow." Although Baehre et al. (2021) pointed out some concerns regarding the accuracy of NPS in predicting future growth sales, this paper does not focus on this aspect but rather on extracting additional insights from the free text section in the survey using the artificial intelligence approach.

To measure customer loyalty, Reicheld (2003) came up with the question, "How likely is it that you would recommend [company X] to a friend or colleague?" Which is the core concept of NPS. This allowed Reicheld (2003) to divide the customers into different customer segments and identify the loyal customers. Based on different NPS scores (0 – 10), customers can be divided into different segments, with higher scores indicating stronger loyalty. Respondents are categorized into "promoters" (rating 9 or 10), "passives" (rating 7 or 8), and "detractors" (rating 6 or lower). The NPS is derived by subtracting the percentage of detractors from the percentage of promoters obtained from survey responses.

II. Business Understanding

EON, a company focused on customer satisfaction, places significant importance on the Net Promoter Score (NPS) as a key metric for assessing customer loyalty. Through regular customer surveys conducted via paper letters, telephone calls, and email campaigns, E.ON collects the survey along with detailed feedback on various aspects of the customer journey. In contrast to the numeric NPS values, interesting insights are often provided in the free-text response in the survey. However, the manual analysis of textual feedback further creates challenges for EON in efficiently utilizing this data to support the decision-making processes within a reasonable amount of time and the possibility to improve further.

A. Challenges

The challenges provided by EON in the ReadMe file can be categorized into three goals.

- Sentiment Analysis:
  Determine the sentiment polarity of the customer utterances with a simple negative-neutral-positive classification, adjust the scale from 0 to 10, and further benchmark it with the NPS values.
- Inconsistency Flagging (NPS adjustment):
  Explore the relationship between the sentiment analysis and the NPS values to find out whether the textual comment actually supports the evaluation of the NPS.
- Topic Modelling:
  Derive a taxonomy of the loyalty drivers and cluster the loyalty drivers into meaningful categories to discover the weights in contributing to the NPS score. Furthermore, hidden topics should be identified.

B. Pain Points

We have identified some pain points for EON, and they are as follows. NPS comments are only analyzed manually and not too often, which results in underutilized textual feedback.

To the best of our knowledge, there are also no known sanity checks to see whether the NPS value aligns well with the comment. On the other hand, regarding the survey design, there is no standardized process across different countries - questionaries and the collection methods differ from region to region. There are some cases where certain countries are more advanced with added questions around topics that drive or lower the NPS score.

C. Work Package Overview

Our work package consists of three individual Python notebooks designed to support decision-making in shaping E.ON's customer experience journey. These notebooks leverage data exploratory analysis to provide various aggregated insights aimed at identifying interesting patterns within our dataset. Furthermore, our inconsistency flagging algorithm allows for the quick and robust identification of inconsistencies between the NPS value and the customer comments. This is crucial, particularly for customers who may not be familiar with the rating scale. By flagging such inconsistencies, we can accurately identify "real" promoters, enhancing the reliability of our analysis. Lastly, our topic modeling approach employs a powerful unsupervised learning algorithm to unveil the underlying structure of the data. By identifying hidden patterns and topics, we gain deeper insights into customer feedback, enabling us to make more informed decisions to shape EON's customer experience journey.

III. CRISP-DM

Our approach is mainly based on the CRISP-DM (Cross Industry Standard Process for Data Mining) (Schröer et al., 2021). CRISP-DM provided us with a standard approach to tackle the project (Schröer et al., 2021).

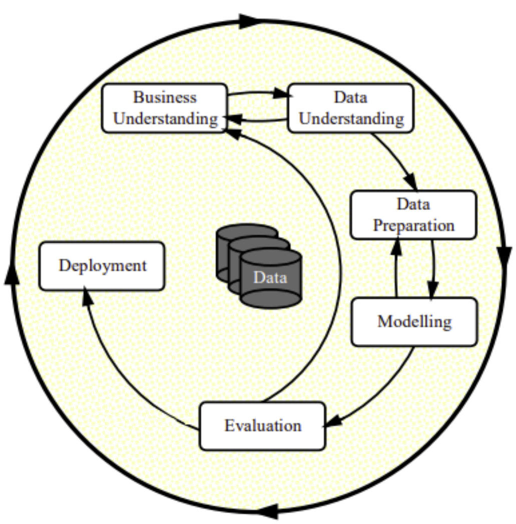

Fig. 1. CRISP-DM phases (Wirth et al., 2000)

The first phase begins with a business understanding of the overall problem. The second phase of the problem includes identifying data quality problems and understanding our data structure. The data preparation phase includes the cleansing and transformation of the data to be used as input to the model. In the modeling phase, several models are implemented. After the model has been implemented, the results are then evaluated and benchmarked against.

IV. Data Understanding

A. Original Data Structure

Fields	Explanation
ID	an artificial running ID for the survey
Interview data	the date of the customer interwies
Country	country in which the survey was conducted
NPS	the value of the index set by the customer
Comment	original textual response by the customer
Translated comment	comment translated in English

B. Data Preprocessing

To prepare our raw data for further analysis, we conducted several steps to achieve a cleansed dataset. Initially, we eliminated null values and standardized country names to address variations (e.g., "Czech" and "Czech Republic" mapped to the same country). Additionally, we filtered out NPS values exceeding 10. For text preprocessing, we first tokenized the text by splitting it into individual words or tokens. Next, we removed common words ("stopword removal"), such as "and" or "the," which may not provide informative insights. Furthermore, we eliminated emojis present within the text. Finally, we normalized the text by reducing words to their root form using lemmatization. These steps ensure our data is of good quality and a suitable format for further analysis.

C. Data Exploratory Analysis

In this section, we have selected some interesting graphs to discuss

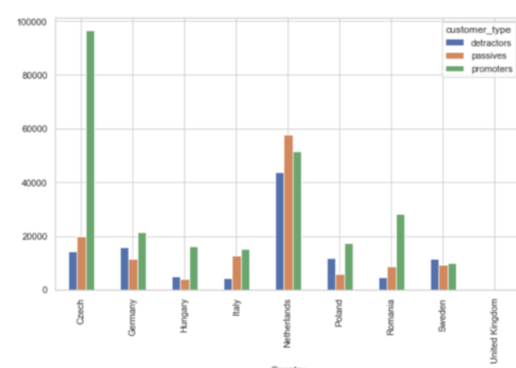

Fig. 2. Customer types for different countries

Figure 2 illustrates the distribution of different customer types across various countries. Upon closer examination of this graph, we observed a noticeable imbalance between promoters and other customer types in the data.

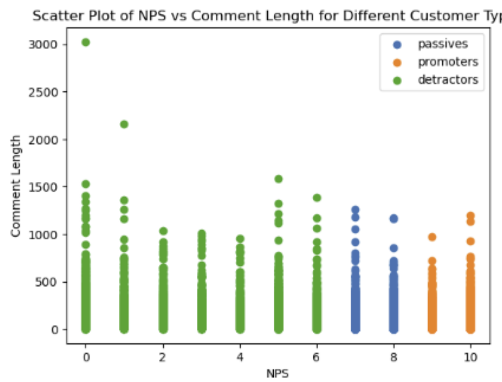

Fig. 3. Correlation between comment length and NPS score

We can see that there is very little correlation between the comment length and the NPS score in Figure 3.

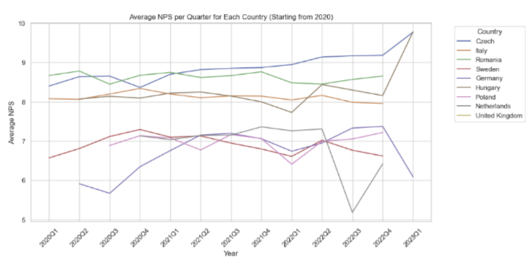

Fig. 4. Average NPS trend per country

Figure 4 shows the average NPS trend across different countries.

V. Methodology

A. Implementation Details

The implementation of our NPS analysis framework was structured around a robust NLP pipeline designed to process and analyze vast amounts of customer feedback data. We utilized Python as our primary programming language, using its extensive ecosystem of data science and NLP libraries. Key libraries included pandas for data manipulation, scikit-learn for machine learning tasks, NLTK and spaCy for natural language preprocessing, BERT, Vader, and TextBlob for sentiment analysis, and BERTopic, LDA, Top2Vec, and NMF for advanced topic modeling.

Our preprocessing pipeline encompassed noise removal, normalization, and tokenization, ensuring the textual data was primed for analysis. Note that not all the methods required the data to be preprocessed. Some would preprocess the data on their own and would discourage steps like stop word removal and lemmatization (Kurniasih & Manik, 2022). An example of this would be in the sentiment analysis notebook, where the Vader and Textblob required preprocessing for the input text. For this particular purpose, a new column was made in the dataset called 'translated comments preprocessed.' This column was developed through the original 'translated comments' column after subjecting it to the preprocessing techniques of converting the words to lower case, English stop word removal, removing punctuation, emoji removal through regex, email and phone number removals through regex, lemmatization, and replacing noise or incorrect words.

The last part included removing words like '#quot#' and replacing them with an empty character. Another example would be replacing 'essent' with 'eon' for a better understanding of just one subject in mind. Some of these preprocessing techniques were used in all algorithms as they are important, like the word replacement, but others, like stop word removal and lemmatization, were not used for BERT models or Top2Vec (Kurniasih & Manik, 2022).

The methodology was run in a Python framework to analyze the customer feedback. The output was a dashboard developed using Dash by Plotly, which provided insights into NPS drivers and trends with a separate table (through dropdown) for all different models used.

B. Sentiment Analysis

Three primary models were utilized to examine client attitudes as they were conveyed through textual feedback: BERT, VADER, and TextBlob. These models were selected due to their unique benefits. TextBlob is excellent at capturing emotional tones with little preprocessing, VADER offers a simple method for speedy sentiment analysis, and BERT enables deep contextual comprehension. By integrating the advantages of BERT, VADER, and TextBlob, we used an averaging strategy to guarantee a balanced sentiment analysis result.

1) BERT: Modern transformer-based language representation model- BERT was chosen because of its remarkable capacity to manage challenging sentiment analysis tasks. Because of its bidirectional nature, it can capture complex linkages and dependencies. It takes into account a word's whole context inside a phrase (Ambalavanan et al., 2020). BERT has a high degree of contextual knowledge and is able to comprehend the nuances and intricacies found in customer feedback, thanks to its training on a broad corpus of various texts. By ensuring that feelings are understood more accurately and nuancedly, this in-depth contextual analysis offers insightful information about the opinions that consumers have expressed about EON's services (Koroteev, 2021).

2) VADER: VADER was selected due to its ease of use and effectiveness in delivering rapid sentiment analysis. Based on an existing vocabulary, VADER rates words' feelings according to their valence and intensity. This vocabulary is very good at managing slang, casual language, and even emoticons that are frequently encountered in customer feedback since it contains terms with recognized sentiment polarity. (C. Hutto and E. Gilbert, 2014) VADER's speed and accessibility allow for quick comprehension of broad sentiment trends in the data, making it ideal for quick prototyping and preliminary research. (R. Chadha and A. Chaudhary, 2023) .

3) TextBlob: TextBlob, known for its simplicity and effectiveness, is employed to capture sentiment polarity in customer feedback. It utilizes a straightforward approach to determine whether a piece of text expresses a positive, negative, or neutral sentiment. TextBlob is particularly useful for its ease of use. It is a pragmatic choice for sentiment analysis tasks, especially when dealing with short and concise feedback. (I G. S. Mas Diyasa et al., 2021) TextBlob provides valuable insights into the general sentiment orientation of customer comments. (D. Hazarika et al., 2020).

4) Averaging Approach: An averaging approach that combines the advantages of BERT, VADER, and TextBlob is used to guarantee a thorough sentiment analysis result. This methodology recognizes the many benefits associated with each paradigm. A weighted average was taken with equal weights to BERT and VADER and half of that to Textblob. The reason for that was its higher deviation as compared to the other two. While VADER's lexicon-based method is excellent at handling informal language and emoticons, TextBlob's simplicity makes it useful for rapid sentiment analysis. BERT's extensive contextual analysis enables it to capture complex feelings. An analysis that provides a fair picture of consumer feelings is strengthened and rounded by averaging the sentiment ratings from different models (Y. Xie et al., 2019). The overall dependability of the sentiment analysis data is improved by this all-encompassing method. This helps to mitigate biases or limits that may be connected to certain models.

C. NPS adjustment

While analyzing data for key insights, a problem for data scientists is false labels and imbalances. A common problem that can arise with putting your trust in the NPS score is when the sentiment does not match the score. For example, a customer could have complained in the feedback form about bad customer service or high prices, but the given NPS score would surprisingly be 9. Another example could be the customer praised E.On and still gave an NPS of 3. This is an inconsistency in our data that we would like to point out. This inconsistency can arise through the customer's negligence, translation problems, or just sarcasm at the customer's end. The approach used in this report aims to flag such comments that have a weak alignment between NPS and sentiment score.

The methodology adopted builds on the sentiment analysis part. After calculating the weighted average sentiment of a comment from the BERT, VADER, and TextBlob models, a standard deviation (SD) and standard error (SE) of the difference between average sentiment and NPS were calculated. These values would serve to identify three categories of alignment between NPS and sentiment: Strong, Moderate, and Weak. Here, weak alignment is of interest, which indicates NPS and sentiment are far apart and hence identifies an inconsistency. With a low SE of 0.01 and SD of 2.3, the approach goes forward with an iterative approach of comparing how far the difference between NPS and sentiment score was for each comment in the dataset. Using statistical concepts of divergence from the mean, the three categories were assigned by examining how many SD away our difference. If the absolute difference had a deviation of half an SD, the comment can be said to have a strong alignment. If it was between half and one SD, it can said to be moderate alignment. Similarly, if it was more than an SD away, it is said to have weak alignment and gets flagged for review.

This approach was coupled with a 'frequent words' approach to increase its robustness. The frequent words approach utilizes the most occurring words in both ends of the sentiment and NPS spectrum. Firstly, the top 100 words that appear in positive/negative sentiments and promoters/detractors are identified. A union between similar sets of positive sentiment and promoters, as well as negative sentiment and detractors, is calculated to determine which words are most common between both these sets. The approach revolves around the idea that if the most common positive words like 'problem solved,' 'good communication,' 'great satisfaction,' etc., appear in a comment with a low NPS (detractor), there must be an inconsistency. Vice versa, if negative words or phrases like 'not getting answer,' 'no solution' etc., appear in comments with high NPS (promoter), there can exist an inconsistency.

Combining these two approaches leads us to more robust and complete results without flagging comments that do not need to be looked at.

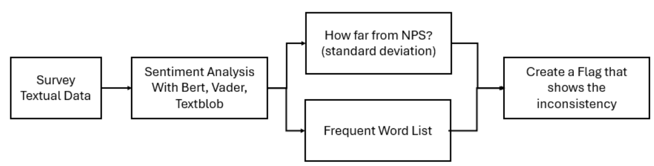

Fig. 5. NPS adjustment workflow

D. Identifying Loyalty Drivers

Applying feature engineering and noun extraction approaches to customer feedback allows for the meaningful extraction of insights that are necessary for the identification of loyalty drivers and service attributes. When we talk about feature engineering, we mean the process of obtaining pertinent data or characteristics that might help foster consumer loyalty (X. Zhang et al., 2018).

Isolating nouns and noun phrases from the textual feedback allows for the identification of important service attributes or components of the customer experience. A more methodical examination of the variables impacting customer loyalty is made possible by these extracted characteristics. Businesses such as EON may better understand the elements that generate client happiness and foster loyalty by recognizing and classifying these traits. This procedure establishes the groundwork for more research and the formulation of tactics meant to improve these loyalty-inducing elements.

Our methodology took advantage of the spaCy library from Python, which is best known for its industrial-strength

NLP capabilities. The process involved loading the English language model and writing a function to extract a noun phrase from a given comment. The function was then run on the whole dataset to extract all noun phrases and then keep the first one hundred most common ones to work on. Out of these hundred nouns and noun phrases, twenty-five were identified to be talking about service in one way or another. These noun phrases serve as service attributes of E.On that customers have most commented on. These service attributes can then also be categorized into certain loyalty drivers. Each loyalty driver can hold a number of service attributes of the same theme.

TABLE I
LOYALTY DRIVERS AND ASSOCIATED SERVICE ATTRIBUTES

Loyalty Driver	**Service Attributes**
Service Quality	speed, responsiveness, customer service, professionalism, efficiency, problem, no problem, service, satisfied service, satisfaction service
Product Quality	reliability, technology
Value	price, lower price, value, billing issues
Customer Experience	satisfaction, customer experience, accessibility, transparency, invoice
Communication	communication, good communication, information, good information

E. Topic Modeling

A natural language processing method called topic modeling looks for themes or subjects within a group of text documents. It may be applied to the study of customer feedback data as it is especially helpful in revealing hidden patterns and structures in huge textual datasets (C.B. Asmussen and C. Møller, 2019). Each of these models outputs a good amount (specified in some cases) of topics that are present in our data. They can also be ordered according to a given score or their overall occurrence in the data.

F. Chosen Topic Modelling Approaches and Key Parameters

Several topic modeling techniques were taken into consideration for the study of consumer feedback data to identify underlying themes. Latent Dirichlet Allocation (LDA), Non-Negative Matrix Factorization (NMF), Top2Vec, and BERTopic are the methods that have been selected. Every strategy has a unique set of benefits and constraints.

1) Latent Dirichlet Allocation (LDA): Assuming that documents are collections of themes and that each topic is a combination of words, LDA is a probabilistic model. The number of topics, alpha, and beta are important factors that affect the coherence and granularity of the subjects that are found (R.K. Gupta et al., 2022).

2) BERTopic: This topic modeling tool uses BERT (Bidirectional Encoder Representations from Transformers) contextual embeddings. In addition to giving details on topic size and sample words, it automatically counts the number of topics. Because of its low preprocessing needs and capacity for contextual comprehension, BERTopic is a useful tool.

3) Top2Vec: The method of finding significant topics in textual material is revolutionized by the ground-breaking topic modeling program Top2Vec. The combination of sophisticated BERT embeddings from transformer-based models with a density-based clustering strategy is what distinguishes Top2Vec. Top2Vec presents an automated approach to discover the best number of topics inherent in the dataset, in contrast to standard topic modeling techniques that frequently involve defining the number of topics beforehand (R. Egger and J. Yu, 2022). According to the Top2Vec documentation, Top2Vec does not require any text preprocessing and identifies the topic automatically. It has a built-in search function available to use. Top2Vec also works well on short text.

4) Non-Negative Matrix Factorization: The term-document matrix is factorized into two lower-dimensional matrices that capture the term-topic and topic-document associations using Non-Negative Matrix Factorization (NMF). The number of subjects and the selection of the update method are crucial factors that affect the interpretability and caliber of generated topics (X. Lin and P.C. Boutros, 2020).

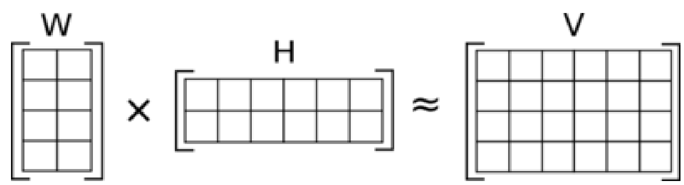

Fig. 6. Feature Matrix W and Coefficient Matrix H division in NMF

G. Applying Topic Modelling to Customer Feedback

1) Preprocessing Data: The textual data undergoes preprocessing, which includes tokenization, lemmatization, and stopword removal, before topic modeling. By doing this step, the format of the data is guaranteed to be appropriate for topic modeling methods.

2) Model Parameter Selection: The parameters for the selected topic modeling technique (BERTopic, NMF, or LDA) are chosen. These might be the number of topics, the alpha and beta values for LDA, the update method selection for NMF, or making use of BERTopic's autotopic count feature.

3) Using the Selected Model: The preprocessed customer feedback data is subjected to the chosen topic modeling technique. The models we apply are pretrained to recognize latent subjects in the text.

4) Interpreting Outputs: The topic modeling findings are explained, and to comprehend the thematic content of each selected issue, a review of the related words and documents is conducted. Deriving practical insights from the highlighted issues requires this interpretation (M. Maarif, 2022).

5) Iteration and Refinement: There is a chance to apply the algorithm again and refine the model parameters throughout this iterative procedure. The optimization of topic modeling findings is made possible by this iterative process, which also guarantees that the topics discovered correspond with the underlying themes in consumer feedback.

6) Optimal Number of Topics: In some models, it is important to identify how many topics to receive as output from the model beforehand. This can be a tricky challenge as you can never be sure of how many topics can be present in a dataset with over half a million comments. Coherence score is a method to identify an optimal number of topics for the model. Topic coherence measures the quality of the data by comparing the semantic similarity between highly repetitive words in a topic (Lau et al., EACL 2014). An iteration is run from a minimum number of topics to a certain maximum, each displaying a coherence score. At one point, the coherence score will reach a maximum, after which it will start declining and becoming steady.

H. Deployment

To display the results of our analysis, we utilized the Dash library in Python. Our deployment focused on a simple yet effective interface that allows users to view the outcomes of different NLP models used in our project. The core feature of this interface is a simple yet effective table displaying results from our NLP models for sentiment analysis and topic modeling. This deployment strategy was chosen for its simplicity and effectiveness in showcasing the project's findings without the need for extensive user interaction capabilities. The next iterative step for our deployment phase is a dropdown menu, enabling users to select between the results of topic modeling and sentiment analysis. Upon selection, the interface displays a table of results corresponding to the chosen model. Although the application's interactivity would be limited mainly to this dropdown feature, it provides a straightforward way for users to compare and examine the insights derived from each model's analysis.

VI. Results

A. Data Analysis Insights

The rigorous preprocessing and analysis of over half a million customer feedback entries illuminated the complex relationship between customer sentiment, specific feedback topics, and NPS scores. The variance in sentiment across different regions indicated that cultural and regional factors play a significant role in shaping customer expectations and perceptions. Moreover, it can also be the reason for E.On's performance in that specific area or, even more so, some political factors. The data was slightly skewed towards more promoters than passives and detractors combined, making it slightly biased.

One of the many hypotheses while looking at the data was the correlation between the comment length and the NPS score. The reasoning behind it came from the idea that dissatisfied customers would have a lot to say and complain about in the company feedback form. On the other hand, more satisfied customers would not have much to say as praises can tend to be shorter than complaints. After extracting the comment length in characters from the translated comments column, a scatter plot indicated that there is not much of a linear relationship between the NPS score and comment length. Upon calculating Pearson's and Spearman's correlation, the correlation turned out to be slightly positive but not enough to say they are correlated. The values of 0.25 and 0.36 were obtained using Pearson's and Spearman's methods, respectively. An individual value was calculated for each of the customer types as well, proving once again that there is no evidence to support our hypothesis. The Pearson coefficient values for promoters, passives, and detractors came out to be -0.07, 0.01, and -0.08, respectively, indicating no correlation.

Detailed examination of feedback through natural language processing techniques unveiled recurring keywords and phrases that were strongly correlated with NPS scores. This correlation was particularly evident in feedback related to product features, customer service quality, and overall customer experience.

B. Sentiment Analysis Results

Implementing three different sentiment analysis models gave us a well-rounded sentiment on the customer comments. The overall distribution of sentiment was towards the positive side, with more positive comments. This is due to the presence of a large number of promoters in the data. Each comment was categorized as either positive (sentiment score greater than 0), neutral (sentiment score close to 0) or negative (sentiment score less than 0). The sentiment scores for all models were then normalized from 0 to 10 to be compared with the NPS scores.

1) Positive Sentiment: Positive comments left by customers show that they are happy with and grateful for the services that EON has to offer. These opinions function as markers of EON's advantages and those instances in which the business has effectively fulfilled or surpassed client expectations. Positive parts of the customer experience are highlighted by instances where clients thank the personnel for their speedy service, laud them for their helpfulness, or praise them for solving problems quickly. The identification and reinforcement of areas of excellence in EON's services are greatly aided by these favorable attitudes (M. Wankhade et al., 2022).

2) Negative Sentiment: Customer reviews that are negative highlight situations in which clients were disappointed, irritated, or unsatisfied with EON's offerings. These opinions highlight potential areas for development or instances when consumers faced difficulties that resulted in unfavorable opinions. A few instances of unfavorable attitudes include grievances about the caliber of the services, annoyance over invoicing mistakes, and discontent with customer service encounters. By examining negative sentiment, EON may get important insights into particular pain spots, resolve problems, and improve service quality (M. Wankhade et al., 2022).

3) Neutral Sentiment: Customer feedback that is neutral in tone denotes a lack of strong emotional overtones and offers a well-rounded viewpoint that might not veer too much in either direction. Even while neutral feelings might not point out problems right away, they can provide important information about how customers see a product or service. Neutral feedback may be created by asking questions about product features, making factual remarks about service experiences, or just acknowledging the supply of services without using strong emotive language. EON can guarantee a full grasp of client feedback, pinpoint small areas for development, and uphold a holistic strategy to increase customer happiness by tracking and evaluating neutral attitudes over time.

C. Sentiment Analysis and NPS correlation

1) Positive Correlation: Customers who express satisfaction and optimism are more likely to promote EON's services to others, according to a positive connection found between higher NPS values and positive sentiment ratings. EON is able to determine important strengths and favorable drivers of customer loyalty by identifying situations in which good feelings coincide with high NPS ratings (C. Lewis and M. Mehmet, 2020). Customers who are pleased with EON's quick problem-solving, for instance, could be more likely to tell their friends and relatives about the company.

2) Negative Correlation: A negative correlation shows that consumers who express irritation or unhappiness are less likely to promote EON's services. This is indicated by lower NPS values and negative sentiment ratings. EON can identify areas that need improvement right away by looking into situations where low NPS levels and unfavorable feelings are present. Customers who express annoyance about inaccurate billing, for example, could be less inclined to suggest EON to others (T. Ho and V.-H. Nguyen, 2022).

3) Limitations: Although sentiment analysis provides insightful information, it is important to recognize its limits and the necessity for more in-depth research to obtain a complete picture of client sentiment.

Because sentiment analysis depends on pre-established linguistic patterns, it cannot accurately reflect the complex range of human emotions, particularly when expressed in ambiguous or sardonic ways. Sentiment analysis data must be carefully interpreted, taking into account the larger context of client feedback.

Sentiment analysis algorithms, even with their advanced sophistication, may have trouble understanding extremely context-dependent attitudes or industry-specific linguistic subtleties. To obtain a more comprehensive understanding of client attitudes and preferences, EON ought to supplement sentiment analysis with qualitative techniques like focus groups and customer interviews.

Sentiment analysis mostly concentrates on textual comments, leaving out other important data sources like demographics or numerical evaluations. To give customers a more comprehensive picture of their sentiment and behavior, EON should combine sentiment analysis with quantitative data.

D. Flagging Inconsistencies

In our study, we found that sometimes, customers' feedback in words didn't match the score they gave. For example, someone might have complained but still gave a high score or praised but gave a low score. This mismatch could be because of different reasons like mistakes, translation errors, or even sarcasm.

To solve this, we used a special method to check if the words in the feedback matched the score given. We used BERT, VADER, and TextBlob models to understand the sentiment of the feedback. Then, we looked at the difference between this sentiment and the NPS score. We classified this alignment into three categories: strong, moderate, and weak. We paid special attention to the weak alignment because it showed a big difference between what was said and the score given.

We also compared the most common words used in positive and negative feedback to the scores. This helped us find feedback that might look positive or negative but had a score that didn't match. Combining these methods helped us better understand when feedback and scores didn't line up, making our analysis of customer satisfaction more accurate.

E. Loyalty Drivers

Through the application of feature engineering and noun extraction techniques on customer feedback, we have successfully extracted meaningful insights critical for identifying loyalty drivers and service attributes. Utilizing the spaCy library for its advanced NLP capabilities, we isolated nouns and noun phrases from the feedback, which highlighted key components of the customer experience and service attributes.

Our analysis identified the top hundred most common nouns and noun phrases, from which twenty-five were directly related to service aspects mentioned by customers. These phrases provide a clear picture of the service attributes most valued by E.ON's customers. Further categorization of these attributes into loyalty drivers allowed us to pinpoint specific areas contributing to customer loyalty. They were further analyzed with promoters and detractors separately, as seen in the tables below. This categorization revealed that certain themes, encapsulated by the loyalty drivers, hold a significant influence over customer satisfaction/dissatisfaction and loyalty. The identification of these drivers lays the foundation for targeted strategies aimed at enhancing customer loyalty through improved service attributes.

TABLE II
BEST PERFORMING LOYALTY DRIVERS FOR PROMOTERS

Loyalty Driver	service attributes	Average NPS
Communication	communication, good communication, information, good information	7.8
Service Quality	speed, responsiveness, customer service, professionalism, efficiency, no problem, service, satisfied service	7.2
Product Quality	reliability, technology	7.2

F. Topic Modeling Findings

In our analysis, topic modeling emerged as a powerful technique to uncover the main themes present in customer feedback. Utilizing algorithms like LDA Top2Vec and BERTopic,

TABLE III
WORST PERFORMING LOYALTY DRIVERS FOR DETRACTORS

Loyalty Driver	service attributes	Average NPS
Customer Experience	satisfaction, customer experience, accessibility, transparency, invoice	4.05
Product Quality	reliability, technology	4.96
Value	price, lower price, value, billing issues	5.26

we could categorize the vast dataset into coherent topics that represent the key areas of customer concern and satisfaction. Even though each algorithm has different advantages and disadvantages, all of them have common findings. The findings revealed several dominant themes: service quality, pricing, product features, and responsiveness to customer support. Each theme was further broken down into sub-themes, providing a granular view of customer feedback.

Interestingly, BERTopic helped us identify nuanced topics that traditional methods might overlook, such as service interaction and detailed product feedback. This depth of insight allowed us to pinpoint exactly what aspects of our service and products were most impactful to our customers.

G. Gaining Strategic Insights through Topic Analysis

This paper aimed to shape customer experience strategies in one way or another. After having used sentiment analysis to understand customer sentiment and its relation with the NPS, flagging inconsistent comments, and loyalty driver extraction to understand which service attributes are performing well/worse in the target group, our exercise leads us to the identification of main and secondary driver topics through topic modeling.

Using the NMF model for topic modeling, we set the number of topics to 22, with the highest coherence score. The model outputs 22 topics with the most common words in each topic. For example, a topic would include the following words: *'answer questions didn question clear called received waiting got previous'*, or *'price reduce high increase compared increases offer year gouging best'*. These topics tell a story about what the user is indicating. In our examples, it would be customer service quality and pricing, respectively.

To understand the different topics for different groups of people, we ran the model on the promoters and detractors separately. For the purpose of this paper, the topics gained from the detractors will be focused on as they can provide critical actions and insights for E.On to take and increase the overall NPS score. With the output of 22 negative topics from the detractors and skipping duplicates, 12 topics were identified after manually decoding what each topic signifies. An example of these topics can be found in the table below. For the detractors, we will generate main and secondary drivers for NPS decrease and can work towards finding solutions to increase customer satisfaction. Similarly, we can investigate reasons for customer loyalty through these topics in the promoters.

With the output of precise topics related to customer difficulties in different countries and timeframes and LLM-generated actionable action items, our model can provide valuable insights to E.On to act upon. With this model, E.On can periodically identify which topics, through keywords or decoded with LLMs, are causing an increase in detractors for that region and, subsequently, a decrease in NPS score/customer satisfaction.

TABLE IV
TOPICS THROUGH NMF FOR DETRACTORS WITH THEIR KEYWORDS

Topic	Key words
Customer Service Quality	service, customer, poor, unfriendly, friendly, good
Pricing and Value Perception	price, reduce, high, increase, best, gouging

TABLE V
TOPICS THROUGH NMF FOR DETRACTORS AND ACTIONABLE INSIGHTS

Topic	Action
Customer Service Quality	Enhance training programs for customer service representatives to improve friendliness and efficiency
Pricing and Value Perception	Review pricing strategies and consider introducing more competitive pricing plans or loyalty discounts to address concerns about high prices

VII. BUSINESS IMPLICATIONS

The insights gained from data exploratory analysis and topic modeling present valuable opportunities for E.ON to optimize its strategies and enhance customer satisfaction. Moreover, the efficiency and reliability of these analyses provide E.ON with a streamlined approach to decision-making, minimizing manual effort.

The trend line depicting the past average NPS across different countries (as shown in data exploratory analysis) highlights variations in NPS scores among geographic regions. This indicates that a "one-size-fits-all" approach to pricing or policy may not be optimal for retaining customers. Therefore, E.ON should consider developing tailored strategies for each region to meet customers' unique needs and preferences. Moreover, the output of our topic models identified distinct topics associated with different customer groups. This insight offers valuable guidance for E.ON in identifying areas where improvements can be made to enhance customer satisfaction. By addressing these specific topics, E.ON can effectively prioritize initiatives to maximize customer satisfaction and loyalty.

VIII. CONCLUSION

In conclusion, this report has demonstrated the substantial benefits of integrating advanced NLP techniques with traditional NPS analysis to uncover more profound insights into customer loyalty and satisfaction. We have provided a more nuanced understanding of customer feedback through sentiment analysis, topic modeling, NPS score adjustments, and identification of loyalty drivers. This comprehensive approach highlights the complexity of customer experiences and offers actionable insights for E.ON to enhance service quality and customer loyalty. The findings underscore the importance of aligning customer feedback with business strategies to foster a customer-centric culture.

ACKNOWLEDGMENT

Special thanks to the teams at E.ON for providing the necessary data and insights that formed the backbone of our analysis. We are also indebted to the academic and technical advisors whose guidance was invaluable in refining our methodologies and approaches. Finally, we wish to thank the participants of the E.ON customer survey, without whom this study would not have been feasible. Their feedback is at the heart of our efforts to enhance customer satisfaction and loyalty.

REFERENCES

[1] Schröer, C., Kruse, F., & Gómez, J. M. (2021). A Systematic Literature Review on Applying CRISP-DM Process Model. Procedia Computer Science, 181, 526–534. https://doi.org/10.1016/j.procs.2021.01.199

[2] Wirth, R. (2000). Crisp-dm: towards a standard process model for data mining. Available at: https://api.semanticscholar.org/CorpusID:1211505

[3] Jacobs, I.S. & Bean, C.P. (1963). Fine particles, thin films, and exchange anisotropy. In G.T. Rado & H. Suhl (Eds.), Magnetism, Vol. III (pp. 271–350). New York: Academic Press.

[4] Elissa, K. (n.d.). Unpublished manuscript.

[5] Nicole, R. (in press). Title of paper with only first word capitalized. Journal Name Standard Abbreviations.

[6] Yorozu, Y., Hirano, M., Oka, K., & Tagawa, Y. (1987). Electron spectroscopy studies on magneto-optical media and plastic substrate interface. IEEE Transactions on Magnetics in Japan, 2, 740–741. (Digests 9th Annual Conference on Magnetics Japan, p. 301, 1982).

[7] Young, M. (1989). The Technical Writer's Handbook. Mill Valley, CA: University Science.

[8] Kurniasih, A., Manik, L.P. (2022). On the Role of Text Preprocessing in BERT Embedding-based DNNs for Classifying Informal Texts. Journal of Computer Science and Informatics, 29(4), 112–120.

[9] Ambalavanan, A.K., Devarakonda, M.V., "Using the contextual language model BERT for multi-criteria classification of scientific articles," Journal of Biomedical Informatics, vol. 112, pp. 103578, 2020. [Online]. Available: https://doi.org/10.1016/j.jbi.2020.103578

[10] Koroteev, M., "BERT: A Review of Applications in Natural Language Processing and Understanding," 2021. [Online]. Available: https://doi.org/10.48550/arXiv.2103.11943

[11] Hutto, C., Gilbert, E., "VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text," Proceedings of the International AAAI Conference on Web and Social Media, vol. 8, no. 1, pp. 216-225, 2014. [Online]. Available: https://doi.org/10.1609/icwsm.v8i1.14550

[12] Chadha, R., Chaudhary, A., "A Study Analyzing an Innovative Approach to Sentiment Analysis with VADER," Journal of Engineering Design and Analysis, vol. 6, pp. 23-27, 2023.

[13] Hazarika, D., Konwar, G., Deb, S., Bora, D., "Sentiment Analysis on Twitter by Using TextBlob for Natural Language Processing," pp. 63-67, 2020. [Online]. Available: https://doi.org/10.15439/2020KM20.

[14] Mas Diyasa, G.S., et al., "Twitter Sentiment Analysis as an Evaluation and Service Base On Python TextBlob," IOP Conference Series: Materials Science and Engineering, vol. 1125, no. 1, p. 012034, 2021. [Online]. Available: https://doi.org/10.1088/1757-899X/1125/1/012034

[15] Xie, Y., Zhu, Y., Cotton, C.A., Wu, P., "A model averaging approach for estimating propensity scores by optimizing balance," Statistical Methods in Medical Research, vol. 28, no. 1, pp. 84-101, 2019. [Online]. Available: https://doi.org/10.1177/0962280217715487

[16] Wankhade, M., Rao, A.C.S., Kulkarni, C., "A survey on sentiment analysis methods, applications, and challenges," Artificial Intelligence Review, vol. 55, pp. 5731-5780, 2022. [Online]. Available: https://doi.org/10.1007/s10462-022-10144-1

[17] Bordoloi, M., Biswas, S.K., "Sentiment analysis: A survey on design framework, applications, and future scopes," Artificial Intelligence Review, 2023 Mar 20:1-56. Doi: 10.1007/s10462-023-10442-2. Epub ahead of print. PMID: 37362892; PMCID: PMC10026245.

[18] Lewis, C., Mehmet, M., "Does the NPS® reflect consumer sentiment? A qualitative examination of the NPS using a sentiment analysis approach," International Journal of Market Research, vol. 62, no. 1, pp. 9-17, 2020. [Online]. Available: https://doi.org/10.1177/1470785319863623

[19] Ho, T., Nguyen, V.-H., "Customer Analytics Using Sentiment Analysis and Net Promoter Score," in Title of the Book or Conference Proceedings, pp. X-Y, 2022. [Online]. Available: https://doi.org/10.4018/978-1-7998-9220-5.ch062

[20] Zhang, X., Ding, X., Ma, L., Wang, G., "Identifying Factors Preventing Sustainable Brand Loyalty among Consumers: A Mixed Methods Approach," Sustainability, vol. 10, no. 12, p. 4685, 2018. [Online]. Available: https://doi.org/10.3390/su10124685

[21] Asmussen, C.B., Møller, C., "Smart literature review: a practical topic modeling approach to the exploratory literature review," Journal of Big Data, vol. 6, p. 93, 2019. [Online]. Available: https://doi.org/10.1186/s40537-019-0255-7

[22] Gupta, R.K., Agarwalla, R., Naik, B.H., Evuri, J.R., Thapa, A., Singh, T.D., "Prediction of research trends using LDA based topic modeling," Global Transitions Proceedings, vol. 3, no. 1, pp. 298-304, 2022. [Online]. Available: https://doi.org/10.1016/j.gltp.2022.03.015.

[23] Lin, X., Boutros, P.C., "Optimization and expansion of non-negative matrix factorization," BMC Bioinformatics, vol. 21, p. 7, 2020. [Online]. Available: https://doi.org/10.1186/s12859-019-3312-5

[24] Maarif, M., "Summarizing Online Customer Review using Topic Modeling and Sentiment Analysis," JISKA (Jurnal Informatika Sunan Kalijaga), vol. 7, pp. 177-191, 2022. [Online]. Available: https://doi.org/10.14421/jiska.2022.7.3.177-191

[25] Yazıcı, G., Çadırcı, T.O., "Creating meaningful insights from customer reviews: a methodological comparison of topic modeling algorithms and their use in marketing research," Journal of Marketing Analytics, 2023. [Online]. Available: https://doi.org/10.1057/s41270-023-00256-0

[26] Egger, R., Yu, J., "A Topic Modeling Comparison Between LDA, NMF, Top2Vec, and BERTopic to Demystify Twitter Posts," Frontiers in Sociology, vol. 7, p. 886498, May 6, 2022. [Online]. Available: https://doi.org/10.3389/fsoc.2022.886498. PMID: 35602001; PMCID: PMC9120935.

[27] Lau, J.H., Newman, D., & Baldwin, T. 2014. Machine Reading Tea Leaves: Automatically Evaluating Topic Coherence and Topic Model Quality. [Online] Available: https://aclanthology.org/E14-1056