The Customer Knows Best: An Evidence-Based NLP Framework for Analysing Vodafone NPS Survey Text

JCU Master of Data Science MA5853 Final Report

Prepared By:

- Nichola Christie
- Michael Couzens
- Daniel Evans
- Nikki Fitzherbert
- Matthew Moore

Prepared For:

- Vodafone Australia
- James Cook University

August 2021

Table of Contents

Executive Summary

(Michael Couzens and Nikki Fitzherbert)

Ensuring a high-quality customer experience has become a critical aspect of any organisation’s operating model, and a fundamental component of maintaining the existing customer base and attracting new customers. In fact customers will often pay a price premium simply due to the level of service received, and if a customer has a bad experience, they are very likely to let someone know about it - often not the company.

Net Promoter Score (NPS) and First Call Resolution (FCR) are two of the many existing approaches to monitoring and assessing customer experience. A NPS survey at Vodafone Australia (Vodafone) includes a quantitative rating component and a free-text field for customers to provide additional feedback. Compared to the ratings data, the extraction of insights from that free-text field has proven somewhat more challenging, in part due to the complex problem of how to automate the extraction of meaningful information from a regularly incoming stream of large amounts of unstructured text (around 4,000 surveys a week).

As a result, Vodafone engaged a James Cook University student consultancy team (the project team) to analyse the NPS and FCR data and design a NLP solution that could answer the following key business questions:

- What drives customers to become promoters or detractors?
- Which issues have the biggest impact if not resolved on the first call?
- What influences a customer to give a passive response?

However, due to legal complications, Vodafone was unable to provide the NPS survey data and so the project team, in consultation with Vodafone, developed an alternate data science solution that would provide a major contribution to achieving the original project objectives. It consisted of an in-depth review on NLP from a NPS/customer satisfaction analysis viewpoint, exploratory data analysis (EDA) of a similar, publicly-available dataset, and a proposed NLP framework as a roadmap for the next stage of the project.

The EDA component used a collection of around 2,000 online reviews of Vodafone scraped from productreview.com.au. Whilst the data was unsurprisingly highly biased with respect to the distribution of reviews by customer rating (79% had been given 1 out of 5), it proved adequate as a test-case for a number of different NLP techniques the project team thought would be useful in the final framework design. The EDA also indicated that there were differences in the language and terms used by detractor, passive and promoter customers from a number of different approaches, as well as the presence of correlations between text feature values and customer ratings and sentiment.

The literature review component examined how the problem of analysing free-text customer feedback had been approached by others, and the results of those approaches to ensure that the proposed NLP framework was within the scope of current advances and best-practice methodology.

The review found evidence that quantitative methods alone (such as the NPS) are insufficient to guide business strategy, and a mixed quantitative-qualitative approach is considered more useful, and that while the field of short-text topic modelling is still an area of active research, there are several viable approaches, with sentiment analysis in particular widely used.

The proposed NLP framework consists of a modular, four-part process - data pre-processing, NLP, modelling and analysis, and interactive visualisation. This design enables each individual component to be independently developed and modified to suit changes in

Vodafone’s requirements, systems and data sets without necessarily requiring a complete re-work of the other pipelines.

Whilst the compressed timeframes involved limited the ability of the project team to systematically cover all areas of research relevant to the Vodafone’s problem, to construct a proxy dataset that more closely resembled NPS survey data with a FCR component, and to test the proposed framework on real-world data, the work done provides a solid foundation for Vodafone to use in the future.

As a result, the project team recommends that Vodafone continue to pursue a NLP approach to analysing their NPS survey free text. In particular, they recommend that this work:

1. Explores the use of LDA and NMF for topic modelling, and Random Forest, Neural Networks, and Naive Bayes for predicting NPS ratings from free text.
2. Outputs key datasets in a format that can easily be incorporated into Vodafone’s current reporting mechanisms.
3. Includes key topic-related visualisations as part of the final output.

Introduction

(Nikki Fitzherbert)

Sam Walton, the founder of Walmart, reputedly once said: “There is only one boss. The customer. And he can fire everybody in the company from the chairman on down, simply by spending his money somewhere else.” (Ratcliffe, 2016). In an environment of increasing competition and consumer power, businesses in all industries including the telecommunications industry, have been turning toward customer experience as a source of durable competitive differentiation (McCall, 2015).

And there is evidence to suggest that customer satisfaction is a key influence on business success by, inter alia, reducing customer churn and the number of complaints, and increasing the number of referrals. For example, acquiring new customers is anywhere between 5 and 25[1] times more expensive than retaining existing ones (Gallo, 2014), 72% of happy customers will tell at least half-a-dozen others, 1 in 3 customers will leave a brand they love after just one bad experience, and only 1 in 26 unhappy customers will complain – the rest just leave (CommenceCRM, 2020).

Customer satisfaction surveys have become a common method of monitoring and benchmarking the customer experience by asking customers to provide one or more ratings and then often asking for an explanation. As a result, a business can get a sense of how customers feel about their product or service as well as insights into why (Qualtrics, n.d.).

The open text responses have historically been an underexploited source of business intelligence; partly because the traditional approach to a thematic analysis of qualitative data involved manually reading, coding and categorising the text (Kuckartz, 2019). Whilst a proven method in qualitative research, this is not a particularly viable or attractive solution when the sample size numbers are in the thousands and the analysis needs to be repeated on a regular basis. In contrast, text mining uses natural language processing (NLP) and other machine learning techniques to transform text data into a structured format that can be explored to uncover hidden relationships, new patterns and other meaningful insights (IBM Cloud Education, 2020).

[1] Depending on the industry

Stakeholder Requirements

(Nichola Christie)

Vodafone Australia (Vodafone) receives approximately 100,000 customer service contacts each week. From these contacts, about 4,000 surveys are filled out each week, and these are used to calculate an overall Net Promoter Score (NPS). The surveys also contain free-text fields. It is not feasible for a human to read all 4,000 responses each week and derive meaningful information from these. Vodafone engaged the James Cook University (JCU) student consultancy team to create a NLP solution to automatically extract actionable insights from the text responses. In particular, Vodafone wanted to answer the following business questions:

- What drives customers to become promoters or detractors?
- Which issues have the biggest impact if not resolved on the first call?
- What influences a customer to give a passive response?

The ideal solution to this data science task would have been a pipeline solution that processes the free-text responses and extracts topics and sentiments about those topics as a tagged dataset. This dataset would feed into an interactive visualisation tool which would allow Vodafone to monitor changes in customer sentiment and emerging topics of importance, as well as slice the data by NPS categories (promoters, detractors, neutral) and other connected demographic, geographical, and business context data. Additionally, a machine-learning model that predicts NPS from text would be desirable to Vodafone.

Unfortunately, Vodafone was not able to provide data to the JCU project team due to legal reasons, and so the project team, in consultation with Vodafone, developed an alternate data science solution that will be a major step towards achieving the final required outcome. This alternate solution consisted of a literature review on NLP within the NPS/customer satisfaction context, exploratory data analysis (EDA) of online Vodafone reviews, and a proposed NLP framework as a roadmap for the next stage of the project (that is, Project 2). The benefit of this solution to Vodafone is that they have insights and guidance from a thoroughly-researched NLP design, which has been further informed by EDA into a similar dataset. This work will make for a streamlined and effective implementation of the final NLP framework, and ultimately allow Vodafone to monitor and respond to issues that are most important to their customers, resulting in increased customer loyalty and decreased churn.

Data Collection and Methods

(Matthew Moore)

Data requirements:

So that the challenges and requirements for a solution to process Vodafone's 4,000 weekly surveys into actionable insights could be understood, a comparable dataset needed to be obtained. The data needed to contain free-form text expressing opinions about a product or service, coupled with a customer rating or similar measurement, and also provide a sufficiently large corpus to enable testing of data mining techniques. Online reviews were determined to have a high degree of similarity to survey data for testing and development purposes.

A review website, productreview.com.au, was selected as it contained a large number (currently over 2,000) of free-form customer reviews for Vodafone, accompanied with a rating between 1 and 5. While the customer ratings used in Vodafone's surveys range between 0 and 10, the data was deemed to be conceptually similar.

Feature Engineering:

In order to extract usable features, the raw review data was subject to a reasonably comprehensive normalisation and enrichment pipeline. Reviews were subjected to a subset[2] of the following processes in order to extract useable features:

1. Combining customer ratings into NPS categories to simulate actual NPS survey data[3].
2. Conversion of reviews and titles to lowercase.
3. Identification and correction of common misspellings and colloquialisms.
4. Expansion of common English word contractions; for example 'can't' became 'can not'.
5. Splitting of reviews into individual sentences and words.
6. Aggregation of review titles and body text into a single column so all the text could be processed and analysed simultaneously.
7. Sentiment scores calculated for each review and individual sentences using a sentiment lexicon.
8. Removal of punctuation, common English stop words and other features that tend not to provide additional information in a NLP analysis.
9. Word lemmatisation.
10. Application of Part-of-Speech and Named Entity Recognition algorithms.
11. Extraction of noun-phrases as well as bigrams and trigrams containing particular syntactic function patterns (such as trigrams beginning and ending with adjectives or nouns) to enhance the discovery of relevant insights.
12. Individual noun word extraction.

The final dataset consisted of 1,903 observations of 13 to 25 variables, including the review_id, review text, title and customer rating from the web scraping process, augmented with the sentiment score for each review. Of the remaining features, the sentences, bigrams, trigrams, noun lists and word lists were also tagged with the customer rating and the sentiment for the sentences they were extracted from for additional context.

---

[2] The EDA work was performed in parallel by two members of the project team due to incompatible Python environments, which meant that the pipeline followed by each was very similar, but not exactly the same.

[3] Reviews with a rating of 5 were categorised as 'promoters', reviews with a rating of 4 were categorised as 'passives' and the remainder were categorised as 'detractors'.

Solution and Analysis

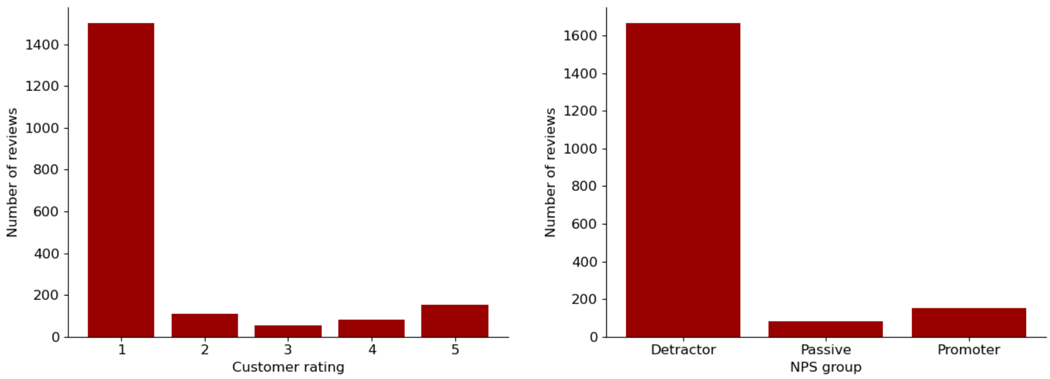
Exploratory Data Analysis

(Nikki Fitzherbert and Matthew Moore)

This section presents a summary of some of the main insights obtained from a preliminary exploration of the scraped data. See Appendix B for the full analysis.

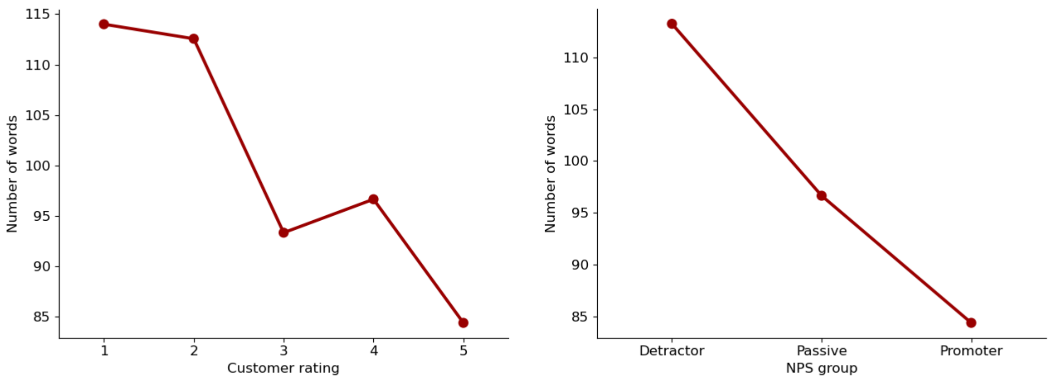
The dataset was highly biased as 78.9% had a customer rating of 1 (Figure 1). The mean length of the review text was 593 characters or 110 words, with individual reviews ranging from 3 to 715 words.

Figure 1: Number of reviews by customer rating (LHS) and NPS group (RHS)



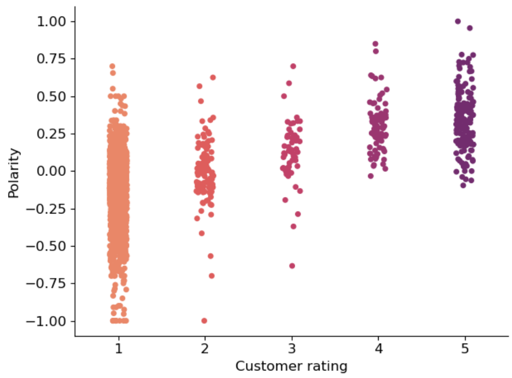
There was also a possible negative relationship between review length and customer rating/NPS group; that is, detractors tended to write the longest reviews and promoters the shortest (Figure 2).

Figure 2: Comparison of review lengths by customer rating and NPS group



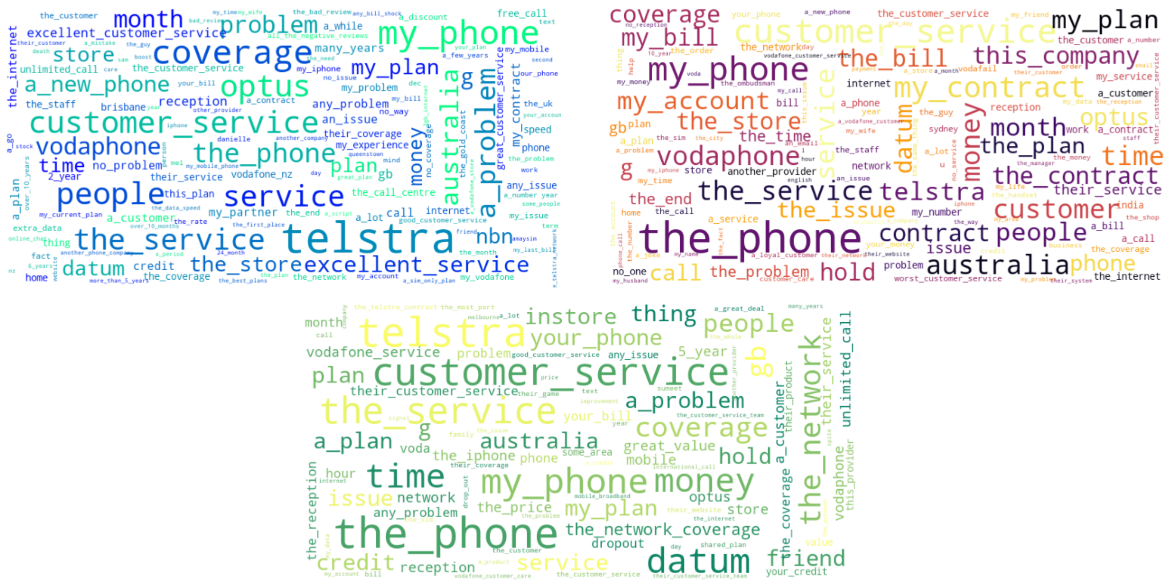
Sentiment lexicons were used to assess the relationship between customer ratings and review sentiment. As expected, there was a positive relationship between customer rating and sentiment score. However, a somewhat more interesting finding was that the range of sentiment observed across the reviews generally decreased as the customer rating increased (Figure 3 over the page).

Figure 3: Spread of lexicon-based sentiment scores by customer rating



Finally, noun phrase analysis indicated a difference in the terms and language used by the different NPS groups. For example, promoters tended to concentrate on the service received and the people, whereas detractors tended to mention money-related topics such as bills, accounts and contracts (Figure 4).

Figure 4: Noun phrase-based word clouds for promoters (LHS top), detractors (RHS top) and passives (bottom)



The text was also analysed for cardinality, with the original dataset found to contain 8,929 unique terms. To gain an understanding of variability of the data, the ‘spellchecker’ package was used to identify 1,972 unrecognised words. The most common recognised terms were either low meaning terms that would be removed as stop words, or references to Vodafone (Figure 5). The unrecognised terms consisted largely of industry terms such as ‘Vodafone’, ‘Telstra’, ‘optus’, ‘NBN’ and ‘3g’, as well as misspellings, colloquialisms, and non-text characters. The most common unrecognised terms appear in Figure 6, and a word cloud of common unrecognised terms is shown in Figure 7 over the page.

Figure 5: Common recognised terms

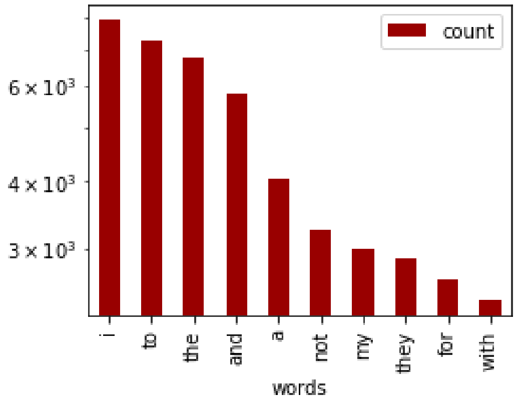


Figure 6: Common unrecognised terms

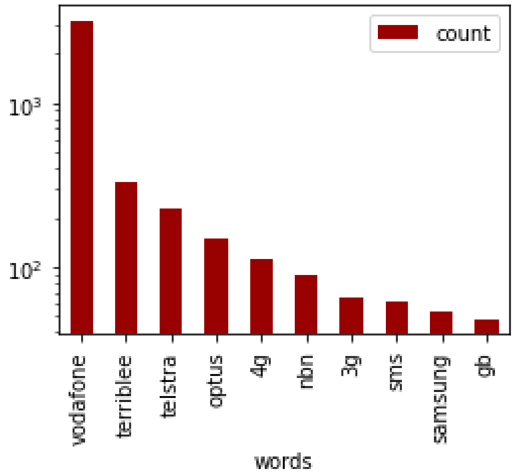
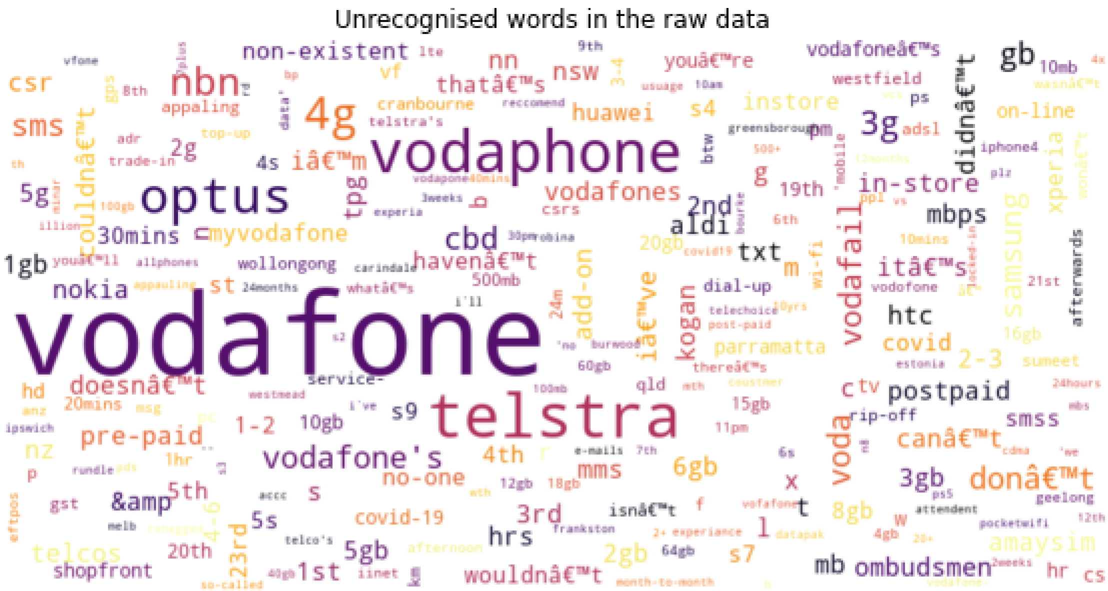
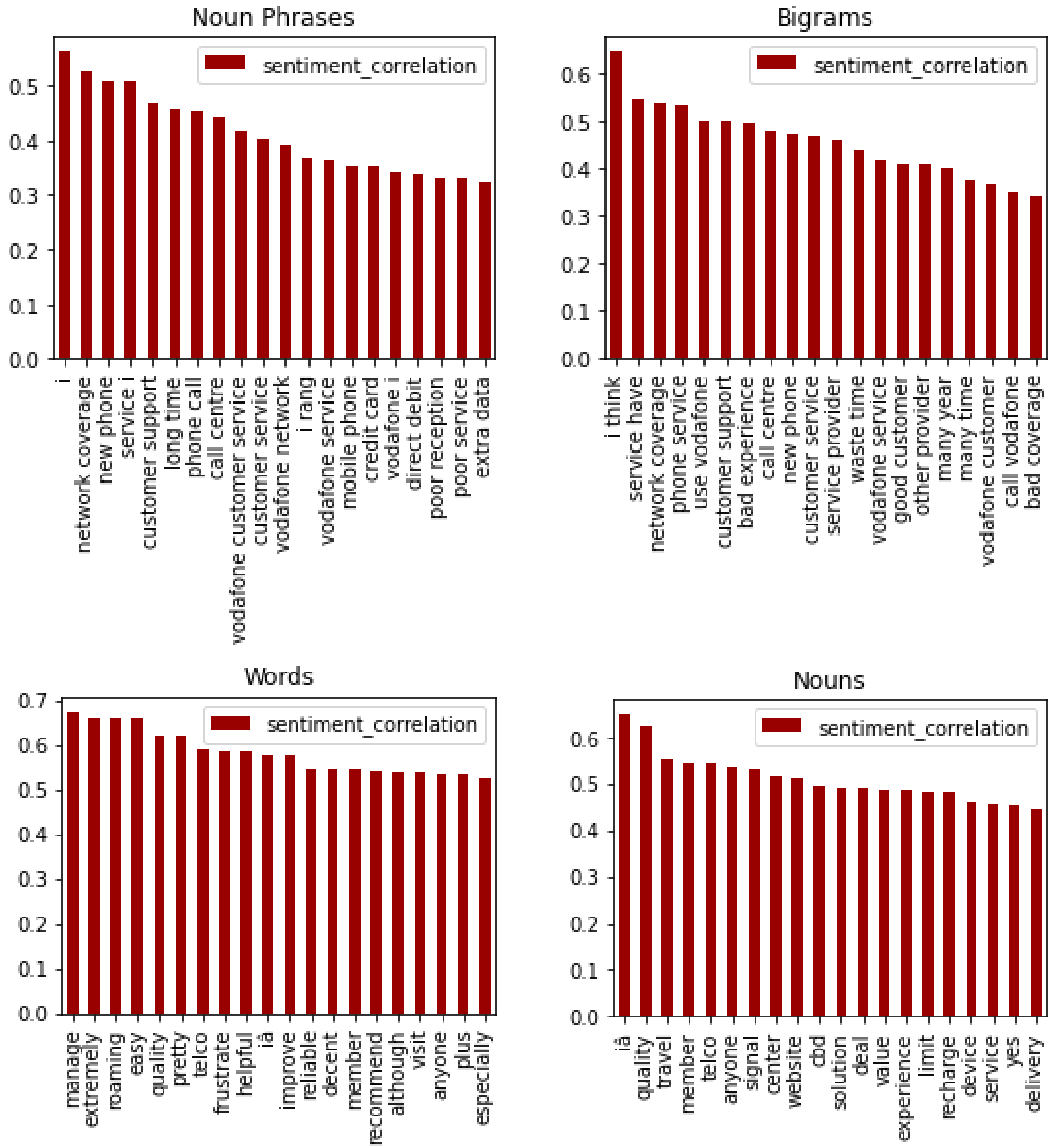


Figure 7: Unrecognised and misspelled terms



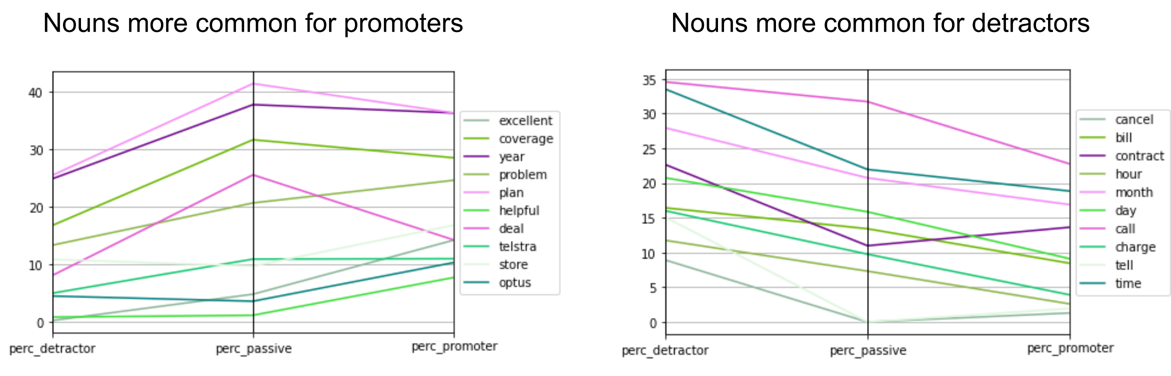
In order to assess the ability of the data to convey customer intent, the correlation between the sentiment with which a word, ngram or noun phrase was used and the rating was investigated. Figure 8 over the page shows examples of feature values with a strong Pearson correlation between the sentiment in which a term was used and the customer rating. The presence of this correlation for specific feature values suggests that review data can indicate customer attitudes towards descriptive terms.

Figure 8: Pearson Correlation between sentence sentiment and customer rating for feature values



To quantify the difference in language between promoter, detractor, and passive customers, the percentage of reviews in which each feature occurred was calculated for each group. The phrases with the largest change in frequency between detractors and promoters were then displayed in parallel coordinate plots. As an example, Figure 8 shows the frequency of nouns with the biggest change between promoters and detractors; reconfirming that detractors have a higher focus on billing, while promoters are more likely to discuss specific problems, plans, or competitors.

Figure 9: Nouns with the largest frequency change between promoters and detractors



Literature Review

(Nichola Christie and Daniel Evans)

A formative component of the NLP framework included a comprehensive literature review (Appendix A) of NLP methods and applications of these within the customer loyalty context. The review identified that quantitative methods (such as the NPS) alone are not sufficient to guide business strategy and increase customer loyalty and company growth (East, Romaniuk, & Lomax, 2011; Fisher & Kordupleski, 2019; Keiningham et al., 2007). Combining quantitative measures with qualitative text analysis is an effective method for understanding NPS ratings (Chaterjee, 2019). Customer experience was found to be positively related to customer loyalty (Barsky and Nash, 2002; Berry et al., 2002; Bhatti, 2020; Imbug, Ambad, & Bujang, 2018; Sahir & Situmorang, 2020;), with customer's "wireless disposition" (Eshghi, Haughton, & Topi, 2007), attractiveness of alternatives, search effort, and satisfaction (Calvo-Porral & Levy-Mangin, 2015) related to customer-switching intentions within the telecommunications sector. Network voice and data, tariff plan, billing and company website identified as important drivers of customer loyalty in one Greek telecommunications company (Markoulidakis et al., 2020).

Each company and customer base is different, so companies are turning to automated text analytics solutions to derive insights directly from the words of their own customers (Adams, 2020; Evelson, 2020; Tarnowska & Ras, 2021). There have been significant advances in NLP as evidenced by the increase in text analytics literature over the past two decades, with NLP applied to product/service reviews across many industries such as telecommunications, sales, hotels, business, restaurants, tourism, gaming, software, cars, banking, and airlines (Calaheiros et al, 2017; Kifetew et al., 2021; Kim and Lim, 2021; Kumar, 2021, Park et al., 2021; Piris & Gay, 2021; Xiangdong et al, 2022).

The literature review identified topic modelling, sentiment analysis and aspect extraction as key methods within customer free-text analytics. Early topic modelling methods such as Latent Semantic Analysis (LSA) (Deerwester et al., 1990), have given way to more reliable approaches such as Probabilistic Latent Semantic Analysis (PLSA) (Hofman, 1999), and Latent Dirichlet Allocation (LDA) (Blei et al., 2003). Short-text topic modelling is an active area of research with LDA and Non-Negative Matrix Factorisation (NMF) currently considered to be among the best approaches for short-text (Albalawi et al., 2020), and novel solutions such as short-sentence representation (SS-LDA) (Ozyurt & Akcayol, 2021) and semantic clustering of features using word embeddings (Gao et al., 2019) showing promise. Sentiment analysis is the measurement of opinions, sentiments, attitudes and perceptions towards topics, products, and services (Birjali, 2021), can be applied at the document, sentence or aspect level, may involve polarisation of negative/positive sentiment in addition to more complex emotional qualities, and has been used comprehensively over a wide number of applications (Birjali, 2021). The literature review identified two primary approaches of sentiment analysis: ML (Agarwal & Mittal, 2016) and lexicon-based methods using a predefined list of words (Jurek et al., 2015; cited in Birjali, 2021). Aspect extraction is the process of identifying individual aspect terms either by manually labelling terms (e.g. "internet speed" or "connection loss" within telecommunications) using domain knowledge or using semi-supervised topic modeling techniques (Anoop & Asharaf, 2018).

The literature review provided insight into notable NLP methods for short-text analysis, and in conjunction with the EDA summarised above, informed the proposed NLP framework to ensure the design was within the scope of current advances and best-practice methodology, with the end goal of producing a viable product solution for the client.

NLP Framework

(Matthew Moore)

Design Overview

Following a review into existing literature around text mining to understand customer promotor behaviour the project team identified a recommended architecture for a NLP and machine learning-based analytics framework.

The proposed framework consists of a modular four-part process - data pre-processing, NLP, modelling and analysis, and interactive visualisation, with each pipeline outputting data to file in a known format. The use of a modular process enables each pipeline to be independently developed and updated, and even written in different languages, without requiring re-work of the up- and downstream pipes.

Design Components

<u>Data Pre-Processing</u>

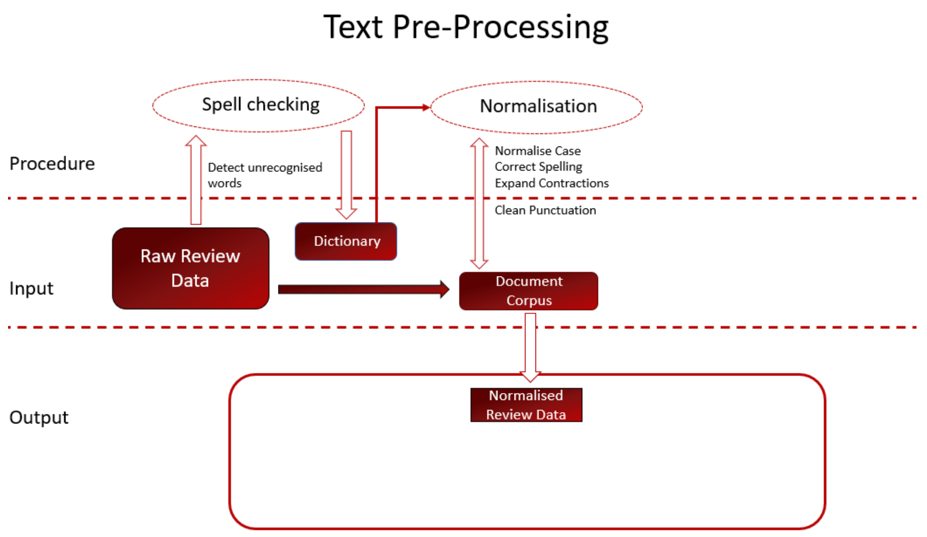
The data pre-processing component is designed to receive input data from one or more sources, and cleansing and normalising the data into a format that can be passed on to the NLP pipeline irrespective of origin or original formatting.

The pipeline executes the following steps producing a body of data in lower case, with irrelevant punctuation removed, and common misspellings identified and corrected as shown in Figure 10.

- NPS survey data ingestion
- Text normalisation
    - Case standardisation
    - Non-word character removal
    - Contraction expansion
    - Punctuation cleansing
    - Spelling and colloquialism correction
    - Cardinality reduction for selected synonyms

The cleansed and formatted output is saved to disk with the required data structure for ingestion into the NLP component, plus any additional metadata that has been provided with the customer surveys.

Figure 10: Data pre-processing module



Natural Language Processing

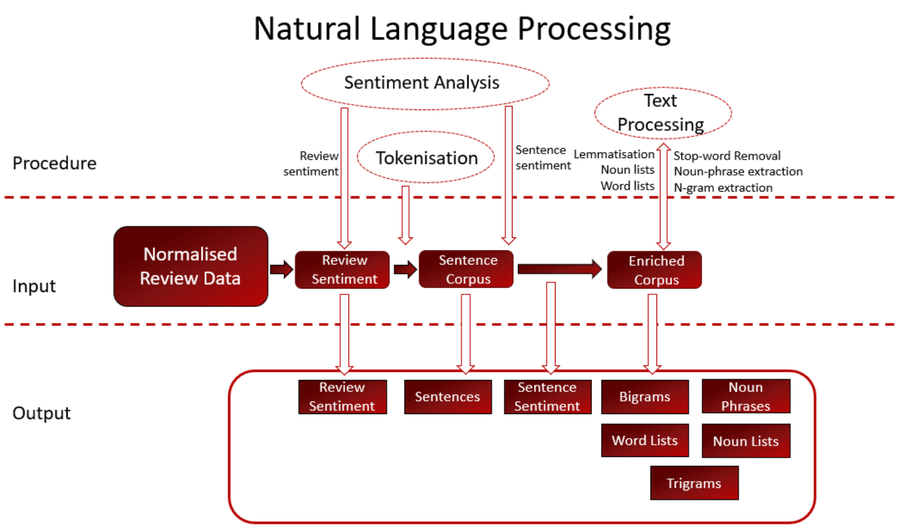
The NLP component is designed to extract low level features from the review text for analysis and modelling, whilst enriching and augmenting the data with lexicographic sentiment analysis.

The normalised and cleansed data from the pre-processing pipeline is ingested and the following processes are performed as outlined in Figure 11.

- Sentiment analysis of each review, and each sentence of the review
- Tokenisation of reviews into sentences
- Extraction of meaningful bigrams and trigrams
  - Bigrams of the form noun-noun, or adjective-noun
  - Trigrams of the form (noun or adjective) - anything - (noun or adjective)
- Extraction of noun phrases
- Cardinality reduction through lemmatisation and stopword removal
- Extraction of word lists and noun lists

The output of the NLP pipeline is a set of features including individual sentences, word and noun lists, noun phrases, bigrams and trigrams in a lemmatised form to reduce cardinality and tagged with their original survey score and the sentiment of their originating sentence. This data formatting was chosen to allow relationships between features and either their original rating or sentiment to be determined with reduced computational overhead in subsequent pipelines. The processed data is saved alongside any additional metadata in a format compatible with the modelling and analysis pipeline.

Figure 11: Natural Language Processing module



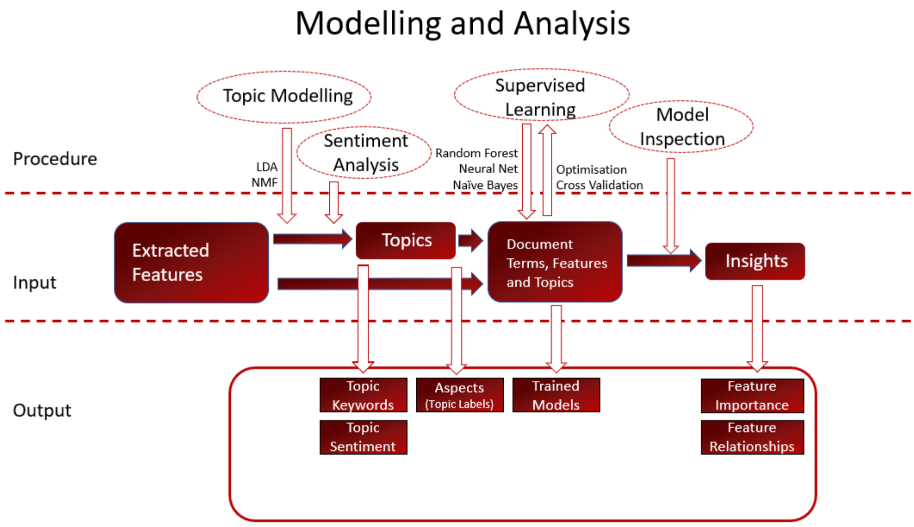
Modelling and Analysis

The modelling and analysis pipeline's purpose is to extract complex relationships and patterns from the features extracted in the NLP pipeline. The two key outcomes from the modelling and analysis pipeline are identification of common topics within the reviews, alongside their interpretable labels (aspects), and machine learning models trained on the extracted features to predict the survey score. The following methods are used as shown in Figure 12.

- Topic modelling using LDA and NMF
- Sentiment analysis of topics
- Supervised learning to classify feedback as a promoter, passive or detractor (Naive Bayes) or predict NPS as a continuous variable (Random Forest, Neural Nets).
- Hyperparameter optimisation and cross-validation of models
- Feature inspection to understand the relative importance of topics, sentiment, language and phrases, and the interaction between features.

The modelling and analysis pipeline output's the previously extracted features, augmented with the extracted sentiment scored topics and topic labels (aspects), as well as feature importance, relationship and probabilities extracted from the trained Random Forest and Naive Bayes models. The final set of engineered features is saved in a format compatible with the desired dashboarding platform.

Figure 12: Modelling and analysis module



Interactive Visualisation

The final body of features extracted from the review data is intended to be loaded into an interactive visualisation platform for dashboarding and analysis as shown in Figure 13 over the page. The final dataset is engineered with the intention of being able to provide nuanced answers to the key business questions:

- What drives customers to become promoters or detractors?
- Which issues have the biggest impact if not resolved on the first call?
- What influences a customer to give a neutral response?

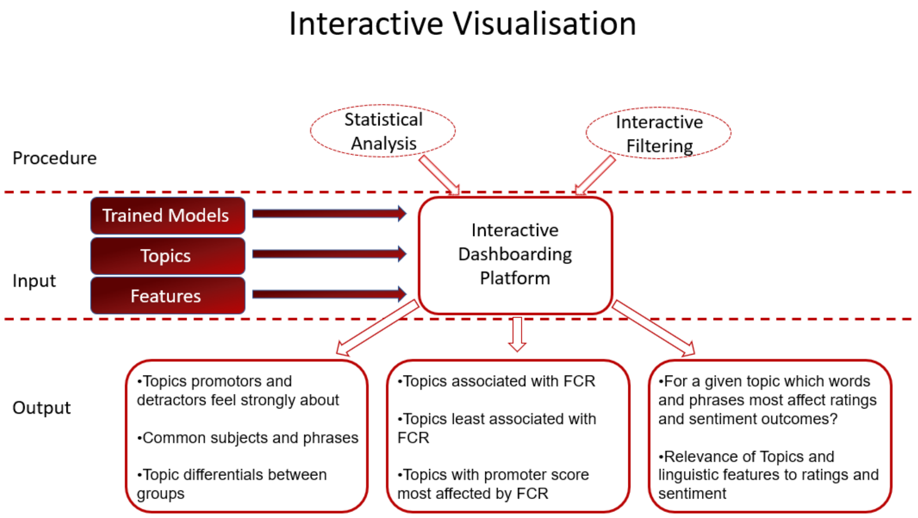
Statistical analysis of the extracted topics, alongside their sentiment of use, as well as extracted language elements is intended to provide insights into the prominent drivers behind detractors and promoters, their degree of influence and overall prevalence.

Understanding the difference in sentiment and survey score for topics based on First Call Resolution, as well as their prevalence provides an understanding of which areas are likely to provide actionable areas where NPS could be improved.

Inspecting and understanding the trained Random Forest and Naive Bayes models provides insights into which features and topics provide the greatest influence on survey scores, as well as providing a way to determine the likely effect of introduced topics, nouns and phrases on survey scores.

Displaying the data in an interactive format allows for a range of insights which can be filtered by any relevant customer metadata provided alongside the surveys.

Figure 13: Interactive visualisation



Discussion and Recommendations

(Nichola Christie and Daniel Evans)

Project objectives and achievements

The original objective was to develop a NLP framework to extract actionable insights from topics and sentiments expressed by customers in Vodafone's NPS survey free-text fields. Such a solution would allow Vodafone to monitor changes in customer sentiment and identify existing and emerging topics of importance, as well as answer key business questions around the drivers of NPS ratings and the impact of First Call Resolution on issues. Additionally the client was interested in developing a machine learning model that could predict NPS from text.

Unfortunately, due to legal issues around the release of data, Vodafone was not able to provide any data to the project team, and so an alternative scope and set of objectives were developed. The revised objectives were to:

1. Conduct a literature review on the usage of NLP for customer-generated text within the context of NPS/customer loyalty
2. Conduct an EDA of online Vodafone reviews; and
3. Design a NLP framework as a roadmap for the next stage of the project.

These revised objectives represent a major step towards achieving the final required outcome for the stakeholder, and demonstrate the flexibility and professionalism of the project team in responding "on-the-fly" to a real-world consultancy complication. The project team were able to achieve the revised objectives despite the uncertainty around the provision of data requiring adaptation and alternative definitions of project scope late into the project timeline.

The project achievements include an approximately 4,000 word literature review, and an in-depth EDA of 1,903 Vodafone reviews obtained independently by the project team from a product review website. The reviews contained both review text and customer star ratings, and so made a good proxy dataset for the original NPS free-text and ratings data. These two components (literature review and EDA) informed the NLP framework which formed the third deliverable for the project. The NLP framework provides an excellent roadmap for the next stage of this project, with a detailed process flow for the text pre-processing, natural language processing, modelling and analysis and interactive visualisation components. All the input, processing and analysis methods, and required outputs at the different stages of the pipeline are clearly outlined. The deliverables from this project will make for a smooth continuation of the JCU-Vodafone consultancy engagement, or provide guidance for Vodafone if they choose to implement the next stage with their own in-house data science team.

Limitations

Due to the compressed timelines following the project scope revision, there was not sufficient time to conduct a systematic literature review. This meant that while the literature review was extensive and thorough, there may be further relevant literature which was not captured by the review.

Regarding the exploratory data analysis, while the findings are undoubtedly relevant and informative to Vodafone, the dataset was heavily biased towards negative reviews, and may not be reflective of the content of the NPS survey free-text data.

The proposed NLP framework has not been tested on real-world text data, and will likely require some modification to accommodate for unforeseen problems that are often encountered only once data cleaning and analysis of the actual dataset is underway.

Improvements and recommendations

As suggested by the limitations above, improvements to the current project (if given more time) would include conducting a systematic literature review detailing all relevant NLP techniques applied to customer-generated text. Additionally, using a more balanced dataset (preferably from the NPS survey) would provide more insights on what drives promoters/satisfied customers to give high ratings. Finally, testing the NLP pipeline framework with NPS free-text or similar data would help uncover any weak points in the design.

Based on the work and findings outlined in the Solution and Analysis section, the project team recommends that Vodafone continues to pursue a NLP approach to analysing their NPS survey free-text data. In particular the project team recommends that this work:

1. Explore the use of LDA and NMF for topic modelling, and Random Forest, Neural Network and Naive Bayes supervised methods for predicting NPS ratings from free-text.
2. Output key datasets in a format that can easily be incorporated into Vodafone's current reporting mechanisms, allowing for monitoring of customer sentiment towards existing and emerging topics using in-house software (for example interactive visualisation software).
3. Include visualisations as a part of the final output which show:
    a. topics promoters and detractors feel strongly about
    b. common subjects and phrases
    c. topic differentials between groups
    d. topics most/least associated with FCR
    e. topics with promoter score most affected by FCR.

Conclusion

(Nichola Christie)

The findings from this project demonstrate that customer-generated free-text data is indeed a rich resource, and provides an evidence-based NLP approach to future work using the Vodafone NPS survey data. The unexpected obstacles to obtaining the Vodafone dataset demonstrate the challenges of real-world consultancy, but also provided the project team with an opportunity to demonstrate flexibility in the face of uncertainty, and to develop a creative alternative solution that was still able to address the client's requirements. The work conducted in this project will provide a solid foundation to deliver the outcomes Vodafone needs.

References

Adams, F. (2020). *The Forrester Wave^{TM}: Customer Feedback Management Platforms, Q1 2020*. 17.

Albalawi, R., Yeap, T. H., & Benyoucef, M. (2020). Using topic modeling methods for short-text data: A comparative analysis. *Frontiers in Artificial Intelligence*, *3*, 42.

Agarwal, B., & Mittal, N. (2016). *Prominent Feature Extraction for Sentiment Analysis* (1st ed. 2016). Springer International Publishing : Imprint: Springer.

Barsky, J. (2002). Evoking emotion: Affective keys to hotel loyalty. *The Cornell Hotel and Restaurant Administration Quarterly*, *43*(1), 39–46.

Berry, L., Parasuraman, A., & Zeithaml, V. (2002). Refinement and Reassessment of the SERVQUAL scale. *Journal of Retailing 67*.

Bhatti, H. S., Abareshi, A., & Pittayachawan, S. (n.d.). Factors that impact customers' loyalty for mobile telecommunication products and services in Australia. *International Conference on Information Resources Management (CONF-IRM)*, 13.

Birjali, M., Kasri, M., & Beni-Hssane, A. (2021). A comprehensive survey on sentiment analysis: Approaches, challenges and trends. *Knowledge-Based Systems*, *226*, 107134.

Blei, D. M. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research*, 993–1022.

Calheiros, A. C., Moro, S., & Rita, P. (2017). Sentiment classification of consumer-generated online reviews using topic modeling. *Journal of Hospitality Marketing & Management*. 26(7). 675-693.

Calvo-Porral, C., & Lévy-Mangin, J.-P. (2015). Switching behavior and customer satisfaction in mobile services: Analyzing virtual and traditional operators. *Computers in Human Behavior*, *49*, 532–540.

Chatterjee, Dr. S. (2019). Explaining customer ratings and recommendations by combining qualitative and quantitative user generated contents. *Decision Support Systems*, *119*, 14–22.

CommenceCRM. (2020, August 8). *What factors influence customer satisfaction?* Commence. https://commence.com/blog/2020/08/08/customer-satisfaction-factors/

East, R., Romaniuk, J., & Lomax, W. (2011). The NPS and the ACSI: A critique and an alternative metric. *International Journal of Market Research*, *53*, 327–346.

Eshghi, A., Haughton, D., & Topi, H. (2007). Determinants of customer loyalty in the wireless telecommunications industry. *Telecommunications Policy*. 31. 93-106.

Evelson, B. (2020). *The Forrester Wave^{TM}: AI-Based Text Analytics Platforms (People Focused), Q2 2020*. 15.

Fisher, N. I., & Kordupleski, R. E. (2019). Good and bad market research: A critical review of Net Promoter Score. *Applied Stochastic Models in Business and Industry*, *35*(1), 138–151.

Gallo, A. (2014, October 29). The value of keeping the right customers. *Harvard Business Review*. https://hbr.org/2014/10/the-value-of-keeping-the-right-customers

Gao, W., Peng, M., Wang, H., Zhang, Y., Xie, Q., & Tian, G. (2019). Incorporating word embeddings into topic modeling of short text. *Knowledge and Information Systems*, *61*(2), 1123–1145.

Hofmann, T. (1999). Probabilistic Latent Semantic Analysis. *ArXiv:1301.6705 [Cs, Stat]*.

IBM Cloud Education. (2020, November 16). *Text mining*. IBM. https://www.ibm.com/cloud/learn/text-mining

Imbug, N., Ambad, S. N. A., & Bujang, I. (2018). The Influence of Customer Experience on Customer Loyalty in Telecommunication Industry. *International Journal of Academic Research in Business and Social Sciences*, *8*(3), Pages 103-116.

Keiningham, T. L., Cooil, B., Andreassen, T. W., & Askoy, L. (2007). A Longitudinal Examination of Net Promoter and Firm Revenue Growth. *Journal of Marketing.* 71. 39-51.

Kifetew, F. M., Perini, A., Susi, A., Siena, A., Muñante, D., & Morales-Ramirez, I. (2021). Automating user-feedback driven requirements prioritization. *Information and Software Technology*, *138*, 106635.

Kim, J., & Lim, C. (2021). Customer complaints monitoring with customer review data analytics: An integrated method of sentiment and statistical process control analyses. *Advanced Engineering Informatics*, *49*, 101304.

Kuckartz, U. (2019). Qualitative text analysis: A systematic approach. In G. Kaiser & N. Presmeg (Eds.), *Compendium for Early Career Researchers in Mathematics Education* (pp. 181-197). Springer International Publishing.

Kumar, S., Kar, A. K., & Ilavarasan, P. V. (2021). Applications of text mining in services management: A systematic literature review. *International Journal of Information Management Data Insights*, *1*(1), 100008.

Markoulidakis, I., Rallis, I., Georgoulas, I., Kopsiaftis, G., Doulamis, A., & Doulamis, N. (2020). A Machine Learning Based Classification Method for Customer Experience Survey Analysis. *Technologies*, *8*(4), 76.

McCall, T. (2015, February 18). *Gartner predicts a customer experience battlefield*. Gartner. https://www.gartner.com/smarterwithgartner/customer-experience-battlefield/

Ozyurt, B., & Akcayol, M. A. (2021). A new topic modeling based approach for aspect extraction in aspect based sentiment analysis: SS-LDA. *Expert Systems with Applications*, *168*, 114231.

Park, S., Cho, J., Park, K., & Shin, H. (2021). Customer sentiment analysis with more sensibility. *Engineering Applications of Artificial Intelligence*, *104*, 104356.

Qualtrics. (n.d.). *Customer satisfaction (CSAT) surveys: Examples, definition & template*. https://www.qualtrics.com/au/experience-management/customer/satisfaction-surveys/?rid=ip&prevsite=en&newsite=au&geo=AU&geomatch=au

Ratcliffe, S. E. (Ed.). (2016). *Sam Walton 1919–92 American businessman* (4th ed.). Oxford University Press. https://www.oxfordreference.com/view/10.1093/acref/9780191826719.001.0001/q-oro-ed4-00011285

Sahir, S. H., & Situmorang, S. H. (2020). The effect of customer experience and emotional value on the net promoter score: Middle-class millennial climber category. *Jurnal Manajemen Dan Pemasaran Jasa*, *13*(2), 199.

Tarnowska, K. A., & Ras, Z. (2021). NLP-Based Customer Loyalty Improvement Recommender System (CLIRS2). *Big Data and Cognitive Computing*, *5*(1), 4.

Xiangdong, L., Youlong, H., Peng, Q., & Wei, L. (2021). A computer-aided approach for acquisition and importance ranking of customer requirements from the online comment mining. *Computer-Aided Design and Applications*, *19*(1), 132–151.

Appendix A - Literature Review

(Nichola Christie, Michael Couzens and Daniel Evans)

Understanding the drivers of customer loyalty: A review of natural language processing of customer-generated text to provide context to the Net Promoter Score

Introduction

The Net Promoter Score (NPS) is a widely used metric in telecommunications to represent customer loyalty (Bhatti et al., 2019; Fisher & Kordupleski, 2019), customer experience (Markoulidakis et al., 2020) and to manage customer relationships (Tong et al., 2017). Loyal customers are important for companies to retain a strong customer base willing to pay greater price premiums, increase word of mouth referrals and make long-term commitments (Bhatti et al., 2019). NPS classifies customers as promoters, neutral, or detractors (Keiningham et al., 2007; Reichheld, 2003) representative of the likelihood that they will recommend a company, product or service. However, quantitative measures alone do not provide sufficient information for qualitative analysis or decisive action based on descriptions of the customer experience (East, Romaniuk, & Lomax, 2011; Fisher & Kordupleski, 2019). Free-text responses provide a platform for the customer to express their individual perspective in relation to their needs. Customer feedback and text-based information can be generated on an immense scale making manual interpretation unscalable and impractical. Natural Language Processing (NLP) uses powerful methods for analysing large scale text-based information to extract sentiment and identify topics. The combination of quantitative and qualitative customer data has demonstrated significant improvements in many customer service domains (Chatterjee, 2019). The aim of this literature review is to identify known drivers of NPS and customer satisfaction, and to critically analyse applications of text analytics within the customer experience context.

Net Promoter Score

NPS was developed by Reichheld in 2003 as a single metric for measuring customer loyalty and predicting the future growth of a company. Customer loyalty had previously been defined in terms of repurchasing behaviour, for example Oliver (1997) defined customer loyalty as "A deeply held commitment to rebuy or re-patronise a preferred product or service consistently in the future, despite situational influences and marketing efforts having the potential to cause switching behaviour” (Oliver 1997,p.392, cited in Bhatti et al., 2019). However, Reichheld pointed out that customer loyalty is about more than repeat purchases, and argued that only a customer who feels intense loyalty will put their own reputation on the line by making a recommendation (Reichheld, 2003). The publication received high accolades in the Harvard Business Review promising a competitive advantage for revenue growth and ease of implementation. Businesses were quick to adopt the metric as a key performance indicator due to simplicity of design. The model records a single value between 0-10 based on how likely a customer would be to recommend the service/company to others. A customer is considered a detractor (0-6), passive (7-8) or promoter (9-10) depending on the score provided. NPS is calculated by subtracting the percentage of detractors from promoters indicating a net performance overall on a scale of -100 to 100

In the original context, promoters are considered customers likely to inform other customers of their positive experience with a company. Conversely, passives and detractors are either unlikely to express their opinion to others or reflect negatively on the experience. Essentially, NPS is an alternative way of measuring customer loyalty, and performs as well as traditional customer satisfaction metrics in predicting firm performance and customer behaviour (Lemon & Verhoef, 2016), for example NPS has a significant impact on customer retention, and has been demonstrated to be useful in discriminating between churners and retainers, identifying customers most likely to churn, managing customers and for assessing competitive positioning (de Haan, Verhoef, & Wiesel, 2015). Further NPS scores reflect real-world behaviour such as word-of-mouth, a form of unpaid advertising able to influence a company brand (Gremler & Brown, 1999), and electronic word-of-mouth (eWOW) messages, with detractors spreading negative and promoters spreading positive eWOW messages (Raassens & Haans, 2017).

Limitations of NPS

While NPS is an effective indirect way of measuring customer loyalty, the use of NPS as a single-metric model is not without limitations. Single-metric feedback is an internal measure of performance and does not describe how new customers can be acquired (East, Romaniuk, & Lomax, 2011). NPS does not include strategic guidance or improvement recommendations, is misrepresentative of negative word-of-mouth and does not mutually reciprocate loyalty from the company to the customer (East, Romaniuk, & Lomax, 2011; Fisher & Kordupleski, 2019). In addition, extensive research has questioned the relationship between NPS and company growth and suggests dual-metric or multi-metric models outperform single-metric predictors (Keiningham et al., 2007).

For example, Keiningham et al. (2007) replicated the analyses used in net promoter research and compared the findings of Reichheld (2003) and Satmetrix with the American Customer Satisfaction Index using the same industries employed in Reichheld's study. Their research rejects the claim that the NPS is the "single most reliable indicator of a company's ability to grow." In its macro-level analysis, the study found no real indication that average levels of attitudinal loyalty metrics significantly correlate with the relative change in revenue within the respective industry. Furthermore, single metrics alone cannot predict customer loyalty and consequently are unlikely to deliver actions to managers. Customers' loyalty-based behaviors are multidimensional and therefore a better measurement tool is required.

Drivers of customer loyalty

Research into antecedents of NPS, and more broadly customer satisfaction can help to bridge the gap between NPS and the drivers of NPS, and answer the question of *why* a particular customer (or groups of customers) give a particular NPS score. Bhatti and Hassan (2019) studied the Australian mobile telecommunications market and hypothesized that customer experience is a major (but not sole) contributor to customer satisfaction and subsequent customer loyalty, reflected in the resulting NPS scores. This hypothesis has been confirmed in other research that has found that customer experience has a significant positive relationship to NPS (e.g. Sahir & Situmorang, 2020).

In the follow-up to his first paper of 2019, Bhatti (2020) empirically concluded that performance expectancy and customer experience have the strongest direct influence on customer loyalty, which was in line with Imbut, Ambad, and Bujang's (2018) findings that customer experience positively influence customer loyalty in the telecommunication industry. These conclusions are further supported by Barsky and Nash (2002) and Berry et al. (2002).

While the literature above demonstrates that positive experiences are related to customer loyalty in the telecommunications industry, it is important for organisations to know which aspects of the customer experience are most important, and how different experiences may impact different groups of customers. For example, research has found that non-millennials and females are more often promoters than millennials and males (Lewis & Mehmet, 2020). Furthermore, recent literature has shown variation from the distinct classification of NPS groups. Many passive customers (7-8) also share similar appreciative values to promoters (9-10) suggesting positive affiliations. Tong et al (2017) conducted a large scale analysis of over 90 thousand NPS surveys and found six groups of distinct telecommunications customers based on their pattern of service usage (e.g. plan types, data usages, and calls), and concluded that different strategies were required to improve NPS in the different sub-groups.

Looking at specific drivers of customer satisfaction within the telecommunications industry, the literature identifies a number of attributes as important. For instance Markoulidakis et al. (2020) identified network voice and data, tariff plan, billing and company website as important drivers for a Greek telecommunications company, compared to less important drivers such as call centre, roaming and shops. Other research has found that the more wireless a customer's home is (defined as the propensity to add more cell phones and replace fixed phone lines with cell phones), the more likely that customer is to churn if not satisfied with the service (Eshghi, Haughton, & Topi, 2007), and that attractiveness of alternatives, search effort, and satisfaction are related to customers' switching intentions, while satisfaction itself is influenced by customers' perceptions of value for money and if the company has a positive corporate image (Calvo-Porral & Levy-Mangin, 2015).

While studies like those above give some general ideas of what drives loyalty/NPS scores within telecommunications, the reality is that each company will have their own unique set of issues which are important to their customers, and that these issues may change over time and differ by sub-groups of customers. Telecommunication companies often adopt a free-text response option for customers to describe their experience. Free-text responses provide companies with qualitative information about the customer perspective beyond the quantitative limitations of single values ratings tools. The combined use of quantitative scores and qualitative text has shown strong associations between customer satisfaction and NPS values to explain ratings (Chatterjee, 2019). Some companies are going as far as moving from systems that generate data-driven recommendations based on numerical and text data, to systems based on text only (Tarnowska & Ras, 2021). For companies not yet invested in text analytics, the underutilisation of these data presents a missed opportunity to connect with the issues that are most important to their customer base in order to increase customer loyalty. Therefore the best resource for understanding what is driving customer loyalty within a given company is the words of those customers themselves.

Description of text analytics

Text analytics is an important emerging area for customer experience professionals, with several established and new competitors turning their attention to text analytics software as a novel automated method of extracting customer insights (Adams, 2020; Evelson, 2020). Clarabridge, one of the leaders in text analytics software for business broadly defines text analytics as “the process of drawing meaning out of written communication” and go on to refine the concept within the customer experience context as unlocking meaning from unstructured text written by or about customers (e.g. online reviews, call center agent notes, emails, tweets, survey results) to uncover patterns and topics which reveal what customers want and need and provide early warning of trouble. It is a process which can be done manually, but is more efficiently done using text analytics software that “uses text mining and natural language processing algorithms to find meaning in huge amounts of text” (*What Is Text Analytics?*, n.d.).

NLP is a subfield of Artificial Intelligence as language is considered a part of human intelligence (Hagiwara, 2019). NLP algorithms are designed to take unstructured text as input, and convert the text to structured data which can be processed in a variety of ways to extract meaning. The foundational NLP methods include converting text to a ‘bag-of-words’, parsing text to identify and tag parts-of-speech, stemming and/or lemmatizing words to remove inflections, and constructing matrices from term frequency across a corpus of texts (for example, term frequency - inverse document frequency matrix). On this foundation more sophisticated machine learning techniques can then be employed to extract meaning, including both supervised and unsupervised methods such as artificial neural networks, tree-based algorithms, and unsupervised clustering algorithms. Two common tasks in text analytics are topic modelling and sentiment analysis, which will be described in more detail below.

Topic modeling

Topic modeling is a form of information retrieval used to find patterns of words in a corpus for the discovery of abstract topics (Likhitha, 2019). Topic modeling is used when manual labelling of documents is not feasible or unscalable for immense text data (Albalawi et al., 2020). Frequently occurring terms contained in large groups of documents form clusters of words, from which previously unrecognised information can be extracted (Salloum, 2018). The automatic extraction of “topics” from associated word terms has become a useful feature for semantic analysis.

Several topic modelling methods with varying degrees of accuracy have been developed over the years. Latent Semantic Analysis (LSA) developed by Deerwester et al. (1990) was among the first methods to introduce topic modelling as a viable method of understanding semantic relationships in raw text data. Probabilistic variations with more reliable results include Probabilistic Latent Semantic Analysis (PLSA) introduced by Hofman (1999). Blei et al. (2003) later developed Latent Dirichlet Allocation (LDA) by using generative Bayesian modeling presented as a finite mixture of the underlying set of topic probabilities. Each topic is modelled as an infinite mixture of topic probabilities to represent a document. Another method is Non-Negative Matrix Factorisation (NMF) which performs both dimension reduction and clustering simultaneously (Berry & Browne, 2005).

Topic modeling is often ineffective on short text generation of sparse text representations (Likhitha, 2019). The advantages and disadvantages of different topic modelling approaches when applied to short text are outlined in a comparative analysis by Albalawi (2020), suggesting that LDA and NMF were among the best models for representing short text descriptions. Issues encountered in some models include difficulty labelling a topic, human judgement required for number of topics, aggregation of short messages to avoid data sparsity, and inability to model relations or semantic incorrectness. The correct choice of model is still largely dependent on the structure of the text in the corpus and the number of desired topics.

Recent publications have focused on developing accurate representations on short text data used by many social media platforms containing word limits. Ozyurt and Akcayol (2021) developed a variation of LDA for aspect-based sentiment analysis called Sentence Segment LDA (SS-LDA). The novel approach considered the number of words contained in sentence segment generation to significantly improve grouping success. SS-LDA considers the coexistence patterns of words and associated relations between them. Key words found in sentences are therefore more likely to be included in other text documents that also contain associated words. Goa et al. (2019) developed Conditional Random Field Regularised Topic Model (CRFtopic modelling) which incorporates word embeddings into topic modelling to overcome data sparsity issues. The model allocates semantically similar words into the same topic assignment and is able to extract more coherent results.

Sentiment analysis

Customer feedback is a rich source of positive and negative attitudes towards a company and their products and services. The immense scale of online service availability in recent years has required adaptation to the volume of customer reviews, and resulted in a growing interest in the field of sentiment analysis. The earliest publications investigating the computation of subjectivity in sentences was published in the early 90's (Wiebe, 1990). The frequency of the google search term "sentiment analysis" has increased exponentially since 2004 and is now equivalent to "customer feedback" (Mäntylä et al., 2018). The growth in popularity has seen an increase in domain specific applications as well as variations in methods design from polarisation of negative/positive sentiment to identifying more complex emotional qualities.

Sentiment analysis is the task of extracting and analysing opinions, sentiments, attitudes and perceptions towards topics, products, and services (Birjali, 2021). The polarisation method aims to predict the negative, positive, or neutral perspectives of a customer review to develop deeper business insight and assist decision-making. Birjali et al. (2021) identified three levels of sentiment analysis performed at the document level, sentence level and aspect level. A document may have a net positive sentiment while also containing a combination of both positive and negative statements for different aspects or entities. Sentiment analysis has been used comprehensively over a wide number of applications, including for business intelligence, recommendations systems, government intelligence and in the healthcare and medical domain (Birjali, 2021). Two primary approaches of sentiment analysis include machine learning (Agarwal & Mittal, 2016) and lexicon-based methods using a predefined list of words (Jurek et al., 2015) identified by Birjali (2021).

Aspect extraction

An aspect is a term or combination of terms used to represent product attributes, features or services contained within a text document or sentence. The telecommunications industry may identify aspects such as “internet speed” or “connection loss” as parts of service delivery. Aspect extraction is the process of identifying individual aspect terms either by manually labelling terms using domain knowledge or using semi-supervised topic modelling techniques (Anoop & Asharaf, 2018). Figure 1 is an example of a single sentence containing two aspects. The overall sentiment of the sentence does not accurately represent the polarised sentiment of the aspects identified. Aspect extraction aims to isolate these key attributes for independent analysis.

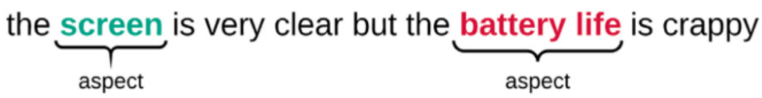


Figure 1. Example of two aspects in one sentence from Augustyniak (2021)

There is an emerging body of literature to support the use of aspect-based sentiment analysis (ABSA) for negative/positive polarity at an aspect level. The first solutions for aspect term extraction used a supervised learning method Conditional Random Field (CRF) (Toh & Wang, 2014). Other methods introduce a sequence tagging scheme (Jakob and Gurevych 2010) and hand-crafted features including lexicons, syntactic and semantic features (Toh & Wang, 2014).

Modern aspect extraction techniques use deep learning frameworks such as recurrent neural networks (Wang et al., 2016) and deep convolutional neural networks (Poria et al., 2016). Augustyniak et al., (2021) provide a comprehensive overview of aspect extraction techniques using a combination of CRF and deep learning frameworks that use long short-term memory (LSTM) and trained word embeddings.

Applications of NLP and customer experience

Given that correctly identifying and appropriately responding to customer needs and complaints has a positive impact on customer loyalty (Andreassen, 1999; Gelbrich & Roschk, 2011), in industries with major financial incentives (Press et al, 1997), text analysis has received increasing attention as a fundamental process for large scale interpretation of textual information to understand customer’s needs. A recent systematic review of the literature on text mining applications in the service and management field identified 125 research papers, of which the most dominant themes were machine learning, customer satisfaction, social media analytics, user-generated content, online review, and sentiment analysis (Kumar, 2021). The review found that text analytics methods such as sentiment analysis and topic modelling have demonstrated improved customer satisfaction and loyalty in sales, hotels, business, restaurants, tourism, and airlines (Kumar, 2021).

While there is an ever growing body of research in this area, much of the research is narrow in focus and designed to prove a specific concept or analytical approach. More mature implementations of text analytics exist (Adams, 2020; Evelson, 2020), however companies use these strategies competitively. Companies using customer experience text analytics have developed and refined their NLP methods and models over many years, and businesses pay a premium to access these software packages which are frequently deployed as Software-as-a-Service (SaaS) on cloud platforms. For many organisations both the cost of such bespoke software, and the security implications of moving sensitive customer feedback data from their own secure servers to a cloud platform for analysis presents an obstacle to using these software packages. For these reasons organisations may rely on in-house data analytics teams, or hire contractors to develop custom solutions.

Approaches to NLP text analysis within the customer experience literature illustrate the variety of methods, and ways of implementing those methods that exist and are continually being developed and improved upon in an emerging field. For example, two recent studies both employed topic modelling and sentiment analysis in relation to online customer reviews for a mobile game (Kim and Lim, 2021) and for online car reviews (Park et al, 2021), however the implementation of the approaches demonstrates how flexible these methods are.

Kim and Lim (2021) analysed online customer reviews of the mobile game Angry Birds 2 using NLP techniques. They created a service feature hierarchy with keywords related to each feature, and used valence aware dictionary and sentiment reasoner (VADAR) sentiment analysis to identify complaints about features. They demonstrated that these methods could be combined with statistical process control (SPC) to allow for real-time monitoring of customer complaints against specific service features. The customer complaints charts that they developed via SPC identified distinct time periods where the customer complaints index (CCI) exceeded their upper control limit (UCL), which was traced in subsequent analysis of the reviews to a specific compatibility issue with iPhone X following a game update. Their case study tested different time period and sensitivity parameters in their calculation of the UCL and found that their proposed method was reliable and robust, and can be applied to different organisational contexts.

Park et al (2021) proposed an improved method of sentiment analysis using machine learning to expand and refine word sentiments specific to an industry or domain. They demonstrated their proposed method using Word2Vec to construct a word graph from online car reviews, and then using graph-based semi-supervised learning (GSSL) to propagate sentiment labels for unlabelled words from a smaller selection of sentiment-labelled words within the graph. They used latent dirichlet allocation (LDA) to derive topics from the reviews, and then calculated two measures “controversy” (how often topics are raised) and complaint (how severely the topic is complained about). Complaint was derived from the co-occurrence of topic keywords and sentiment words. Using this method they were able to quantify customer dissatisfaction (calculated as controversy x complaint) with specific product features, as well as identify context specific sentiment words, including non-grammatical Internet language, industry-specific jargon, and local dialect (Park et al, 2021).

Other applications of NLP within the customer satisfaction context include topic analysis of a bank’s NPS survey free-text responses which found a significant negative correlation between NPS recommendation scores and the number of topics customers mentioned (satisfied customers raised fewer topics) (Piris & Gay, 2021); automated user-feedback driven requirements prioritization where NLP techniques including sentiment analysis were employed to bridge the gap between how users and developers viewed and formulated text about a software system’s requirements and bugs (Kifetew et al, 2021); extracting and ranking customer mobile phone requirements from online comment mining (Xiangdong et al,2022); modelling topics from reviews of a Portugese eco-hotel using a lexicon approach combined with LDA (Calaheiros et al, 2017); and identifying customer satisfaction dimensions from air travel reviews using LDA to identify dimensions and a Naive Bayes Classifier for sentiment analysis (Lucini et al, 2020).

These studies demonstrate the variety of ways that NLP methods can be employed, and the flexibility that exists for data scientists in designing novel approaches to NLP tasks in this space.

Conclusion

Text analytics using NLP is a valuable addition to quantitative methods such as the NPS in empowering companies to understand, monitor and improve their customers' experience and ultimately increase customer loyalty and decrease churn. The literature suggests that while use of single metrics indicators may represent customer loyalty, it is unable to provide feedback for decision making in response to customer reviews. The combined use of quantitative performance scores and qualitative text analysis provides businesses with more actionable customer/service insights.

Topic modelling is an important tool for identifying abstract words from a large corpus of documents. Short text topic modelling is an active area of research due to data sparsity implications and difficulty representing features in short texts. Novel solutions use short sentence representation (SS-LDA) and semantic clustering of features using word embeddings with improved results. Comparative analysis of topic modelling applications suggest LDA and NMF are among the best performing models.

Sentiment analysis is a widely researched area for free text analysis and customer reviews. The use of sentiment analysis can be used to label negative or positive sentiment at an aspect, sentence and document level. Approaches for sentiment analysis include machine learning, lexicon-based, and hybrid approaches among other methods.

Aspect extraction is an emerging field used to represent product attributes, features or services in a corpus of text. Aspect-based sentiment analysis (ABSA) is the use of aspect extraction applied to topic models and sentiment analysis to identify the sentiment of key attributes represented in the data.

As an exciting and growing field, text analytics using NLP is an active area of research and still has not reached maturity, as evidenced by the increasing number of publications in this domain over the past decade. This means there is a great opportunity for data scientists to creatively employ these techniques to add real value to business as well as continue to advance the field.

References

Adams, F. (2020). *The Forrester Wave™: Customer Feedback Management Platforms, Q1 2020*. 17.

Albalawi, R., Yeap, T. H., & Benyoucef, M. (2020). Using topic modeling methods for short-text data: A comparative analysis. *Frontiers in Artificial Intelligence*, *3*, 42.

Andreassen, T. (1999). What Drives Customer Loyalty with Complaint Resolution? *Journal of Service Research*. 1.

Agarwal, B., & Mittal, N. (2016). *Prominent Feature Extraction for Sentiment Analysis* (1st ed. 2016). Springer International Publishing : Imprint: Springer.

Barsky, J. (2002). Evoking emotion: Affective keys to hotel loyalty. *The Cornell Hotel and Restaurant Administration Quarterly*, *43*(1), 39–46.

Berry, L., Parasuraman, A., & Zeithaml, V. (2002). Refinement and Reassessment of the SERVQUAL scale. *Journal of Retailing 67*.

Berry, M. W., & Browne, M. (2005). Email Surveillance Using Non-negative Matrix Factorization. *Computational and Mathematical Organization Theory*, *11*(3), 249–264.

Bhatti, H. S., Abareshi, A., & Pittayachawan, S. (n.d.). Factors that impact customers' loyalty for mobile telecommunication products and services in Australia. *International Conference on Information Resources Management (CONF-IRM)*, 13.

Birjali, M., Kasri, M., & Beni-Hssane, A. (2021). A comprehensive survey on sentiment analysis: Approaches, challenges and trends. *Knowledge-Based Systems*, *226*, 107134.

Blei, D. M. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research*, 993–1022.

Calheiros, A. C., Moro, S., & Rita, P. (2017). Sentiment classification of consumer-generated online reviews using topic modeling. *Journal of Hospitality Marketing & Management*. 26(7). 675-693.

Calvo-Porral, C., & Lévy-Mangin, J.-P. (2015). Switching behavior and customer satisfaction in mobile services: Analyzing virtual and traditional operators. *Computers in Human Behavior*, *49*, 532–540.

Chatterjee, Dr. S. (2019). Explaining customer ratings and recommendations by combining qualitative and quantitative user generated contents. *Decision Support Systems*, *119*, 14–22.

de Haan, E., Verhoef, P. C., & Wiesel, T. (2015). The predictive ability of different customer feedback metrics for retention. *International Journal of Research in Marketing*. 32(2). 195-206.

Deerwester, S., Dumais, S.T., Furnas, G.W., Landauer, T.K. and Harshman, R. (1990), Indexing by latent semantic analysis. J. Am. Soc. Inf. Sci., 41: 391-407.

East, R., Romaniuk, J., & Lomax, W. (2011). The NPS and the ACSI: A critique and an alternative metric. *International Journal of Market Research*, *53*, 327–346.

Eshghi, A., Haughton, D., & Topi, H. (2007). Determinants of customer loyalty in the wireless telecommunications industry. *Telecommunications Policy*. 31. 93-106.

Evelson, B. (2020). *The Forrester Wave™: AI-Based Text Analytics Platforms (People Focused), Q2 2020*. 15.

Fisher, N. I., & Kordupleski, R. E. (2019). Good and bad market research: A critical review of Net Promoter Score. *Applied Stochastic Models in Business and Industry*, *35*(1), 138–151.

Gao, W., Peng, M., Wang, H., Zhang, Y., Xie, Q., & Tian, G. (2019). Incorporating word embeddings into topic modeling of short text. *Knowledge and Information Systems*, *61*(2), 1123–1145.

Gelbrich, K. & Roschk, H. (2011). A meta-analysis of organizational complaint handling and customer responses. *Journal of Service Research*. 14(1), 24-43.

Gremler, D. D., & Brown, S. W. (1999). The loyalty ripple effect: Appreciating the full value of customers. *International Journal of Service Industry Management*, *10*(3), 271–293.

Hagiwara, M. (2019). *Real-world natural language processing: Practical applications with deep learning.* Manning Publications.

Hofmann, T. (1999). Probabilistic Latent Semantic Analysis. *ArXiv:1301.6705 [Cs, Stat]*.

Imbug, N., Ambad, S. N. A., & Bujang, I. (2018). The Influence of Customer Experience on Customer Loyalty in Telecommunication Industry. *International Journal of Academic Research in Business and Social Sciences*, *8*(3), Pages 103-116.

Jakob, N., & Gurevych, I. (2010). Extracting opinion targets in a single cross-domain setting with condition random fields. *EMNLP 2010 - Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference*, 1035–1045.

Keiningham, T. L., Cooil, B., Andreassen, T. W., & Askoy, L. (2007). A Longitudinal Examination of Net Promoter and Firm Revenue Growth. *Journal of Marketing.* 71. 39-51.

Kifetew, F. M., Perini, A., Susi, A., Siena, A., Muñante, D., & Morales-Ramirez, I. (2021). Automating user-feedback driven requirements prioritization. *Information and Software Technology*, *138*, 106635.

Kim, J., & Lim, C. (2021). Customer complaints monitoring with customer review data analytics: An integrated method of sentiment and statistical process control analyses. *Advanced Engineering Informatics*, *49*, 101304.

Kumar, S., Kar, A. K., & Ilavarasan, P. V. (2021). Applications of text mining in services management: A systematic literature review. *International Journal of Information Management Data Insights*, *1*(1), 100008.

Lemon, K. N., & Verhoef, P. C. (2016). Understanding customer experience throughout the customer journey. *Journal of marketing: AMA/MSI Special Issue*. 80. 69-96.

Likhitha, S., S., B., & M., H. (2019). A Detailed Survey on Topic Modeling for Document and Short Text Data. *International Journal of Computer Applications*, *178*(39), 1–9.

Lucini, F. R., Tonetto, L. M., Fogliatto, F. S., & Anzanello, M. J. (2020). Text mining approach to explore dimensions of airline customer satisfaction using online customer reviews. *Journal of Air Transport Management*. 83.

Markoulidakis, I., Rallis, I., Georgoulas, I., Kopsiaftis, G., Doulamis, A., & Doulamis, N. (2020). A Machine Learning Based Classification Method for Customer Experience Survey Analysis. *Technologies*, *8*(4), 76.

Millard, N. (2006). Learning from the ‘wow’ factor—How to engage customers through the design of effective affective customer experiences. *BT Technology Journal*, *24*(1), 11–16.

Oliver R.L.(1997). Satisfaction: A Behavioral Perspective on the Consumer. Boston: McGraw-Hill.

Ozyurt, B., & Akcayol, M. A. (2021). A new topic modeling based approach for aspect extraction in aspect based sentiment analysis: SS-LDA. *Expert Systems with Applications*, *168*, 114231.

Park, S., Cho, J., Park, K., & Shin, H. (2021). Customer sentiment analysis with more sensibility. *Engineering Applications of Artificial Intelligence*, *104*, 104356.

Poria, S., Ofek, N., Gelbukh, A., Hussain, A., & Rokach, L. (2014). Dependency Tree-Based Rules for Concept-Level Aspect-Based Sentiment Analysis. In V. Presutti, M. Stankovic, E. Cambria, I. Cantador, A. Di Iorio, T. Di Noia, C. Lange, D. Reforgiato Recupero, & A. Tordai (Eds.), *Semantic Web Evaluation Challenge* (Vol. 475, pp. 41–47). Springer International Publishing.

Press, I., Ganey, R. F., & Hall, M. F. (1997). What's most important to customer satisfaction? *American Bankers Association ABA Banking Journal.* 89(9).

Raassens, N. & Haans, H. (2017). NPS and online WOM: Investigating the relationship between customers' promoter scores and eWOM behaviour. *Journal of Service Research*. 20(3):322-334

Reichheld, F. F. (2003). The one number you need to grow. *Harvard Business Review*, *81*, 46–55.

Sahir, S. H., & Situmorang, S. H. (2020). The effect of customer experience and emotional value on the net promoter score: Middle-class millennial climber category. *Jurnal Manajemen Dan Pemasaran Jasa*, *13*(2), 199.

Tarnowska, K. A., & Ras, Z. (2021). NLP-Based Customer Loyalty Improvement Recommender System (CLIRS2). *Big Data and Cognitive Computing*, *5*(1), 4.

Toh, Z., & Wang, W. (2014). DLIREC: Aspect Term Extraction and Term Polarity Classification System. *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, 235–240.

Tong, L., Wang, Y., Wen, F. & Li, X. (2017). The research of customer loyalty improvement in telecom industry based on NPS data mining. *China Communications: Networks & Security*. 260-268.

Wang, W., Pan, S.J., Dahlmeier, D., Xiao, X. (2016). Recursive neural conditional random fields for aspect-based sentiment analysis. In: *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics*, pp. 616–626.

What is Text Analytics? (n.d.). Retrieved 1 August 2021, from https://www.clarabridge.com/customer-experience-dictionary/text-analytics

Wiebe J, M. (1990). Recognizing subjective sentences: a computational investigation of narrative text.

Xiangdong, L., Youlong, H., Peng, Q., & Wei, L. (2021). A computer-aided approach for acquisition and importance ranking of customer requirements from the online comment mining. *Computer-Aided Design and Applications*, *19*(1), 132–151.

LITERATURE REVIEW APPENDIX - NLP process employed in key reviewed papers

Paper	Data	Techniques/models and packages/tools	Packages/tools	Process steps
Kim & Lim 2021	Online reviews for mobile game app	● Statistical Process Control (SPC) ● NLP pre-processing ● Feature hierarchy with keyword dictionary ● Sentiment analysis	● Valance Aware Dictionary and Sentiment Reasoner (VADAR) in Python ● Natural Language Toolkit (NLTK) in Python	● Pre-process reviews: ○ Tokenise into words ○ Remove stop words ○ POS tagging ○ Lemmatisation ● Construct service feature hierarchy ○ Service feature hierarchy - Feature A --> Feature A1 --> Keyword A11, Keyword A12… ○ Pre-processed data used to construct a service feature hierarchy with keyword dictionary, raw data identify customer complaints via sentiment analysis ○ basic hierarchy from experts' judgements and lit reviews then complements with text-mining ● Sentiment analysis to identify complaints ● Develop customer complaints charts (SPC) by customer complaints index (CCI) and by service features
Park et al 2021	Online car reviews	● Sentiment propagation using graph-based semi-supervised learning (GSSL) ● Latent dirichlet allocation (LDA)	● Word2Vec in TensorFlow	● Sentiment propagation ○ Use Word2Vec to create embedded word vectors ○ Construct word graph from embedded word vectors ○ Use graph-based semi-supervised learning (GSSL) with a small number of sentiment labelled words and then propagate sentiment labels based on the word graph for the other words ● Customer review analysis ○ Derive topics (parts of products) using latent dirichlet allocation (LDA) which gives *n* topics comprised of *m* keywords ○ Calculate controversy using the sum of frequencies of keywords and a sigmoid function to standardize frequency and prevent controversy from shifting when a topic has substantially higher/lower frequency than others ○ Complaint calculated by measuring the co-occurrence of topic keywords and sentiment words ○ Dissatisfaction = controversy x complaint ○ Derive controversy-complaint quadrant (similar to importance-performance analysis) to identify which topics are most frequently and strongly complained about

Piris & Gay 2021	NPS survey recommendation scores and free-text field	• LDA	• NLTK in python	• Data pre-processing
				  ○ Identification of multi-word items (e.g. compound words “give up”, fixed expressions “in the medium term”, context vocabulary “bank card”)
				  ○ Segmentation (tokenisation)
				  ○ Remove stop words
				  ○ Stemming
				• Data representation
				  ○ Bag-of-words (BoW)
				  ○ Term frequency - inverse document frequency (TF-IDF) weighting
				• Topic identification using LDA
				• Interpret topics using the notion of relevance (Sievert & Shirley, 2014)

Calaheiros et al (2017)	Reviews of a Portugese eco-hotel from various sources including TripAdvisor, Guest’s book, emails, evaluation website	• LDA • Sentiment analysis	• R “tm” package	Data pre-processing: • Used a lexicon which established the dictionary of relevant terms for sentiment analysis and hospitality terms • Used lexicon terms as input to identify similar terms (and n-grams) from the reviews, and used these as the sentiment and hospitality dictionaries • Rather than stem/lemmatize these dictionaries were used to reduce terms to a base term (e.g. ‘strong positive’ for words like brilliant, wonderful and excellent, and ‘decoration’ for decorative, interior design, and architecture) Topic modelling: • Number of topics set to half the number of terms initially (12, as they had 23 base dictionary terms across sentiment and hospitality) - then cyclic procedure to discover best number - evaluating the entropy measure • Topics presented as the most relevant hospitality and most relevant sentiment term per topic
Lucini et al (2020)	Air passenger reviews	• LDA • Sentiment analysis	• NLTK in python • Scikit-learn in python	Data pre-processing: • Non-English characters converted to ASCII code • Synonyms substituted by most frequent version in reviews • Text referring to specific periods of time, weights, sizes, prices, dates, airline names, airport names etc were replaced by a generic keyword - e.g. airport names replaced by “_airport_” (“_” at the beginning and end of key words identifies added words) • Sentence segmentation and tokenization • POS tagging - only adjectives and nouns retained (authors note there is some evidence for retaining adverbs also, but this did not hold true for their dataset) • Compound nouns adjective to become single tokens (e.g. “cabin_staff”)

				● Lowercase and stopwords removed ● Lemmatization ● Exclude low frequency (<2%) tokens ● TF-IDF matrix Dimension extraction (topic identification): ● Only nouns from records used to identify dimensions ● LDA used to identify latent topics ● Used Zhao et al (2015) perplexity-based method to determine ideal number of latent topics ● Looked at the distribution of customer satisfaction dimensions across different groups (e.g. first-class passengers, by different airlines, etc) Sentiment analysis: ● Only adjectives from records kept to calculate sentiment scores ● Used only reviews where customers had given a rating of one (classified as negative) or ten (classified as positive) ● Used Naive Bayes Classifier to predict the probability of the class (positive or negative) given the adjective used in the review Regression analysis to validate findings: ● All (pre-processed) text used ● Independent variables were the importance and sentiment scores of the dimensions identified using LDA, and dependent variable was if the customer did or did not recommend the airline

Appendix B - Exploratory Data Analysis

(Nikki Fitzherbert and Matthew Moore)

Code and data available at [GitHub - NikkiSarah/MA5853-Project1: Code repository for a group project performing a text analysis of Vodafone Australia customer satisfaction data](https://github.com/NikkiSarah/MA5853-Project1) (https://github.com/NikkiSarah/MA5853-Project1)

Exploratory Data Analysis For a Corpus of Online Vodafone Reviews

Introduction

Customer loyalty is an important consideration in running and growing a service-oriented business. In order to gain better insights into customer loyalty, and key drivers behind customer decisions, Vodafone Australia (Vodafone) gauges sentiment through customer surveys to obtain feedback and individual promoter scores. As this data contains high volumes of free-form text, methods are required to automate the mining of this survey data for actionable insights into what drives their promoters and detractors.

As part of the design process for an Natural Language Processing (NLP)-based analytics solution, a comparable dataset containing online customer reviews and recommendation scores was subject to exploratory data analysis (EDA). The purpose of this analysis was to gauge the effectiveness of cleansing and normalisation techniques and the relevance of data enrichment through techniques such as sentiment analysis. Beyond the composition of the data itself, the project team sought to confirm the existence of a relationship between language elements, customer ratings, and sentiment to gauge the likely effectiveness of using review data in lieu of actual NPS survey data to understand customer attitudes and drivers.

Overview of online review data

Vodafone customer review data was obtained from a popular online product review website as it contained a large number (currently over 2,000) of free-form customer reviews for Vodafone, with review scores between 1 and 5. While the customer ratings used in Vodafone's surveys range between 0 and 10, the data was deemed to be conceptually similar as customers gave their experience with Vodafone's product and service offering a score and then proceeded to explain that rating.

A corpus of 1,903 customer reviews was obtained as outlined below, and then processed to extract and normalise key linguistic features alongside as well as perform preliminary sentiment analysis.

Raw data

1,903 observations of 4 variables:

- Review_id - Categorical variable identifying the review
- Title - String variable containing the customers title for their review
- Review - String variable containing the body text for the review
- Score - The rating assigned to the review by the customer. Values 1-3 are assumed to be a detractor, 4 is assumed to be passive and 5 is assumed to be a promoter.

Treatments and feature engineering

In order to extract usable features, the raw review data was subject to a reasonably comprehensive normalisation and enrichment pipeline. Reviews were then subjected to a series of pre-processing and normalisation processes (outlined below) to extract useable features. Two members of the project team worked on the EDA component, and after determining that their respective Python environments were incompatible, it was performed in parallel. As a result, each team member followed a slightly different pipeline and utilised a subset of the processes in the following list.

- Combining customer ratings into NPS categories to simulate actual NPS survey data.
- Conversion of reviews and titles to lowercase.
- Identification and correction of common misspellings and colloquialisms.
- Expansion of common English word contractions; for example 'can't' became 'can not'.
- Splitting of reviews into individual sentences and words.
- Aggregation of review titles and body text into a single column so all the text could be processed and analysed simultaneously.
- Sentiment scores calculated for each review and individual sentences using a sentiment lexicon.
- Removal of punctuation, common English stop words and other features that tend not to provide additional information in a NLP analysis.
- Word lemmatisation.
- Application of Part-of-Speech and Named Entity Recognition algorithms.
- Extraction of noun-phrases as well as bigrams and trigrams containing particular syntactic function patterns (such as trigrams beginning and ending with adjectives or nouns) to enhance the discovery of relevant insights.

Treated data

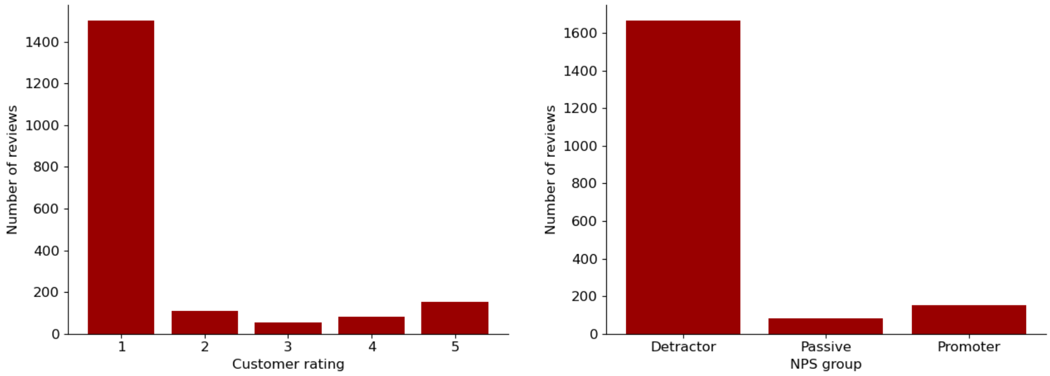
Dataset 1	Dataset 2
1,903 observations of 13 variables: ● review_id ● title ● review ● score ● review sentiment ● sentences ● lemmatised title ● lemmatised sentences ● noun phrases ● bigrams ● trigrams ● words ● nouns	1,903 observations of 17 variables: ● title ● review ● score ● nps category ● combined title & review text ● cleaned text ● words ● parts-of-speech ● named entities ● named entity labels ● pre-processed text ● pre-processed text parts-of-speech ● bigrams ● trigrams ● noun phrases ● cleaned noun phrases ● lemmatised noun phrases

Data set properties

In examining some basic properties of the data set, such as the number of reviews by customer rating or NPS category, it was determined that the distribution was highly skewed. Over 78.9% of reviews were associated with a customer rating of 1, which meant that most reviews were labelled as detractors.

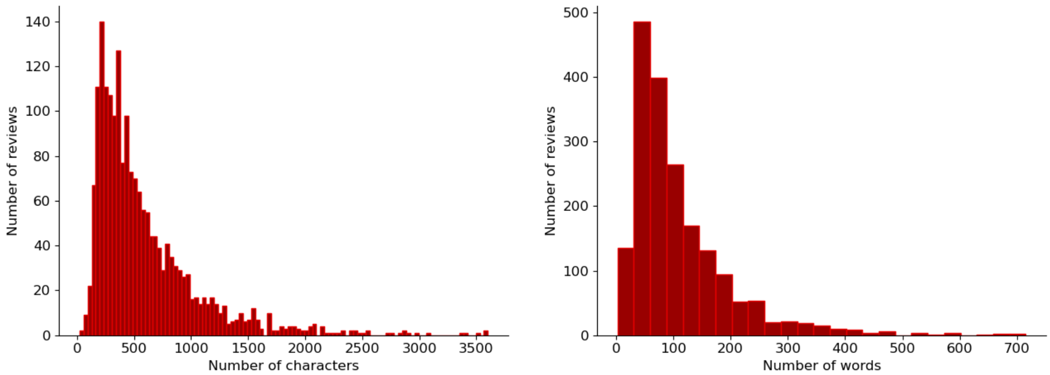
Whilst the collected data allowed for cleansing, processing and exploration techniques to be applied, the project team held the view that more effective analysis and exploration would result from the use of a larger and more evenly-distributed data set.

Figure 1: Number of reviews by customer rating (LHS) and NPS group (RHS)



Review lengths were also investigated to ultimately determine if there was any relationship between the review rating and the number of words or characters within a review. In general, it was determined that the distribution was somewhat right-skewed, with an average review length of 593 characters or 110 words, but reviews ranging from a minimum of 22 characters (3 words) to 2,602 characters (715 words) (Figure 2).

Figure 2: Distribution of review lengths by number of characters (LHS) and number of words (RHS)



Language and Vocabulary

The raw data contained 8,929 unique terms, including terms that contained digits and punctuation. Of the terms contained in the raw data, 1,972 terms were unrecognised by the ‘spellchecker’ module and were found to contain a combination of misspellings, colloquialisms, industry terms, and place names. The treated data contained 6,488 unique terms, including 1,345 unrecognised terms with a similar composition of misspellings, place names, websites and industry terms as shown in Figures 3 and 4 over the page.

Figure 3: Common recognised and unrecognised words before and after processing

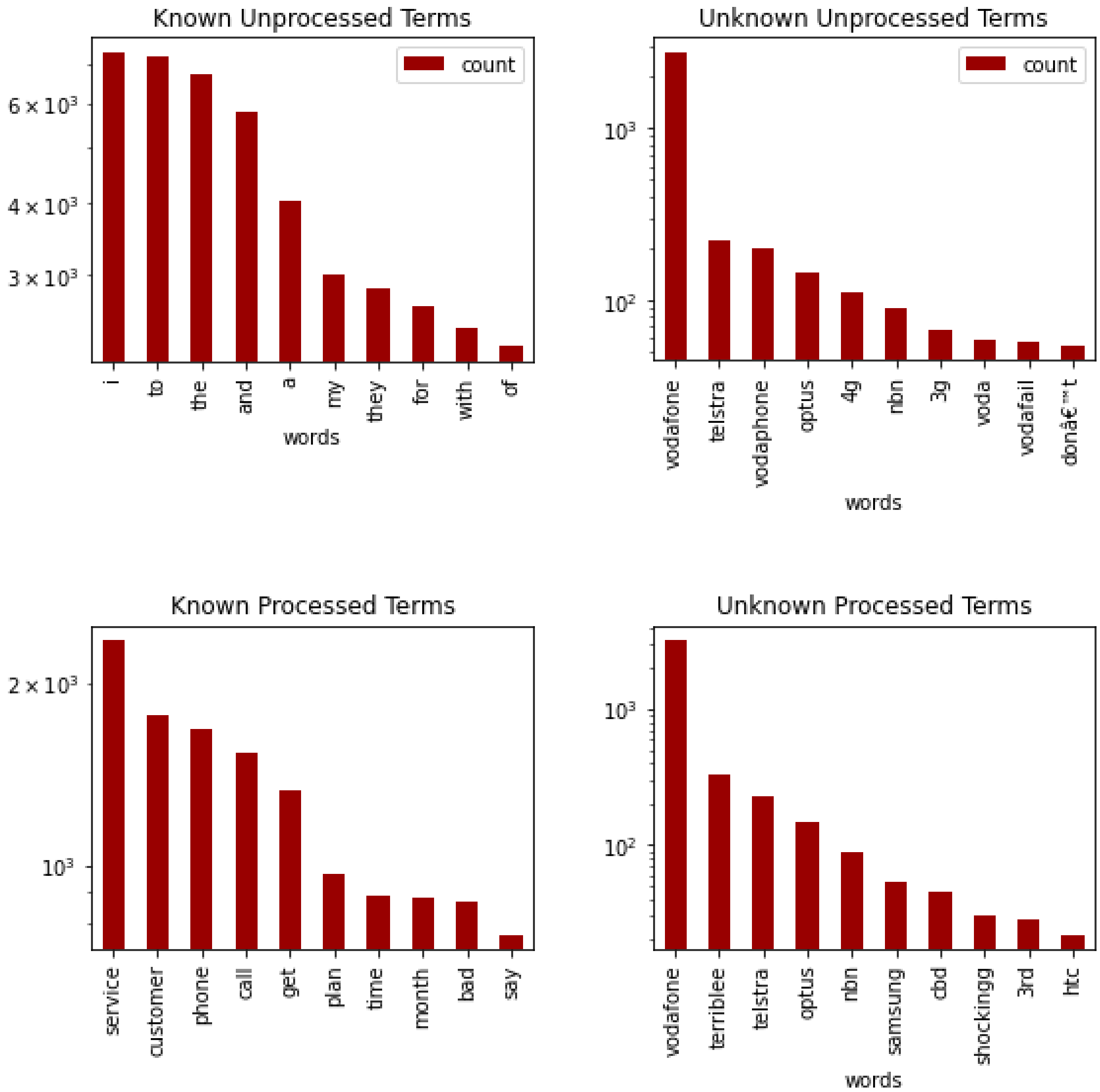
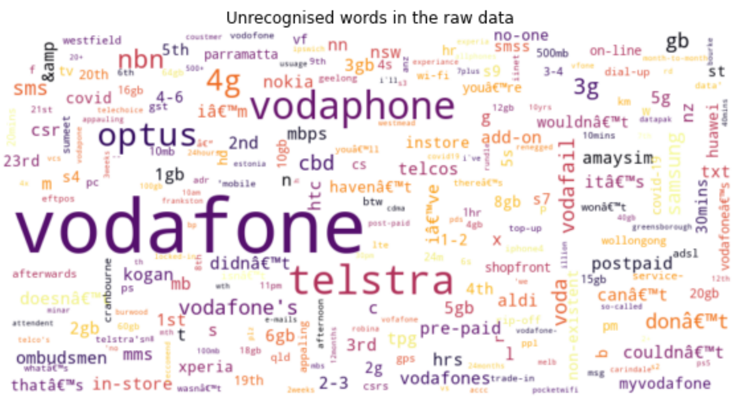
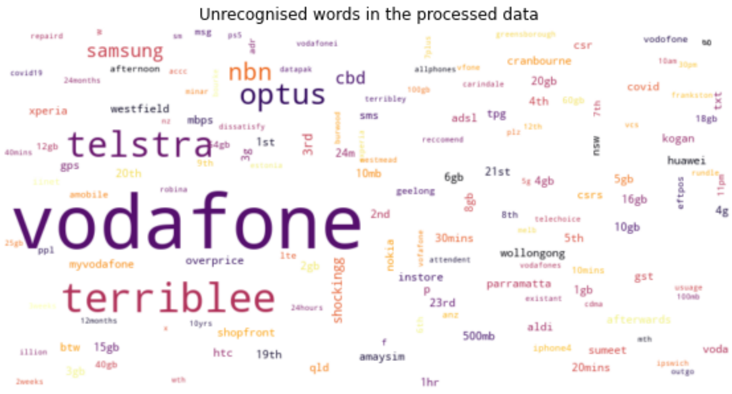


Figure 4: Word Clouds of unrecognised terms before and after processing





Parts-of-Speech and Named Entity Recognition analysis

The ‘spaCy’ module was used to parse and tag the review text that had undergone only preliminary cleaning at this stage (for example, conversion to lower-case and removal of punctuation) as well as the fully pre-processed text (such as removal of stopwords, digits and other features commonly found in text that the NLP literature indicated would usually add no information to text analysis). It indicated that nouns were the most common label assigned followed by verbs in the text that had undergone only the initial pre-processing, but that when the text had moved further along the normalisation pipeline, far-fewer parts-of-speech types remained and adjectives formed a larger proportion of the remaining corpus (Figures 5 and 6).

Figure 5: Frequency of course- and fine-grained PoS labels in the preliminary pre-processed text

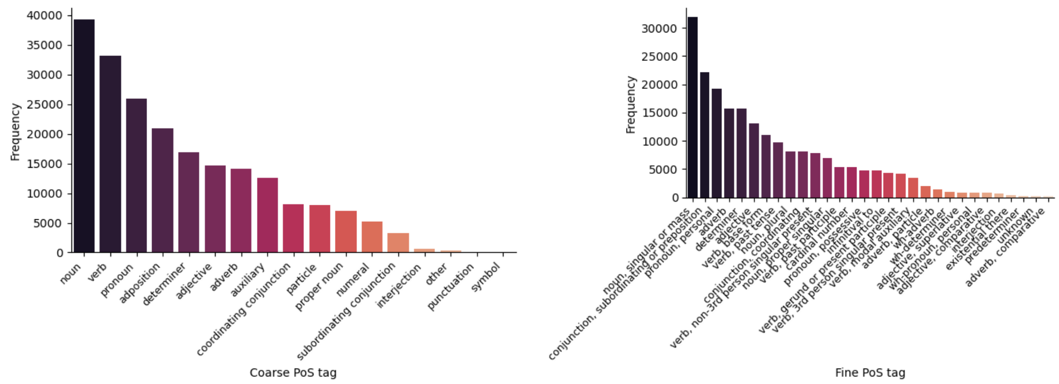
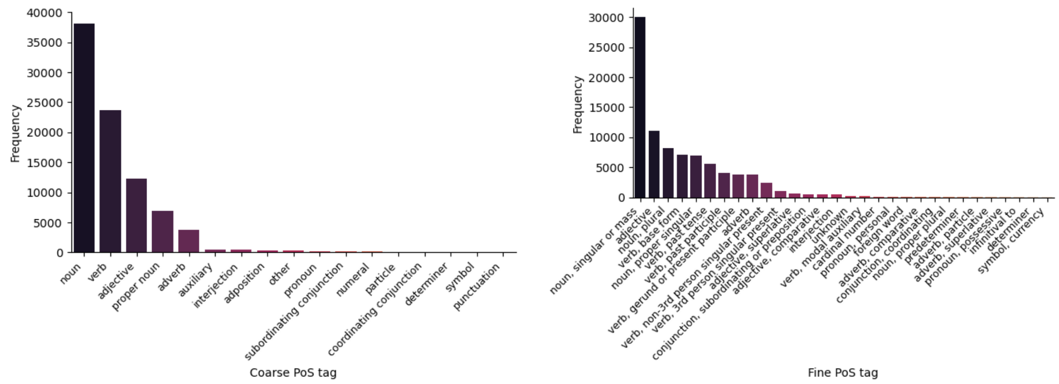
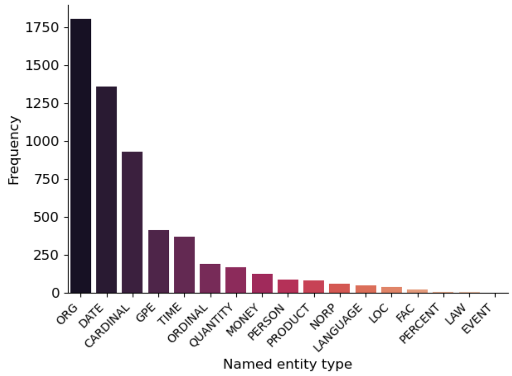


Figure 6: Frequency of course- and fine-grained PoS labels in the fully pre-processed text



SpaCy's statistical entity recognition system was also applied to the review text to identify the types and numbers of named and numeric entities present. Organisation names, dates and other cardinal entities were the top three types of named entities detected by the algorithm. However, it was interesting to observe that even without being able to use the information provided by capitalisation, the algorithm was able to recognise intentional misspellings of the client's name (such as Vodafail), as a type of organisation.

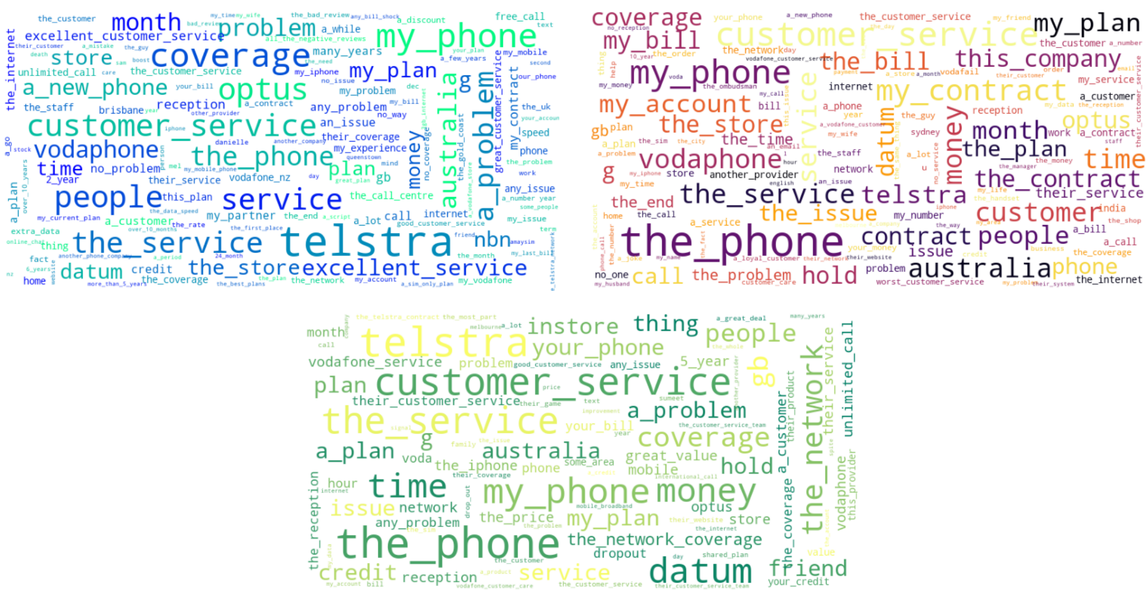
Figure 6: Frequency of detected Named Entity Recognition labels



Noun phrase and n-gram analysis

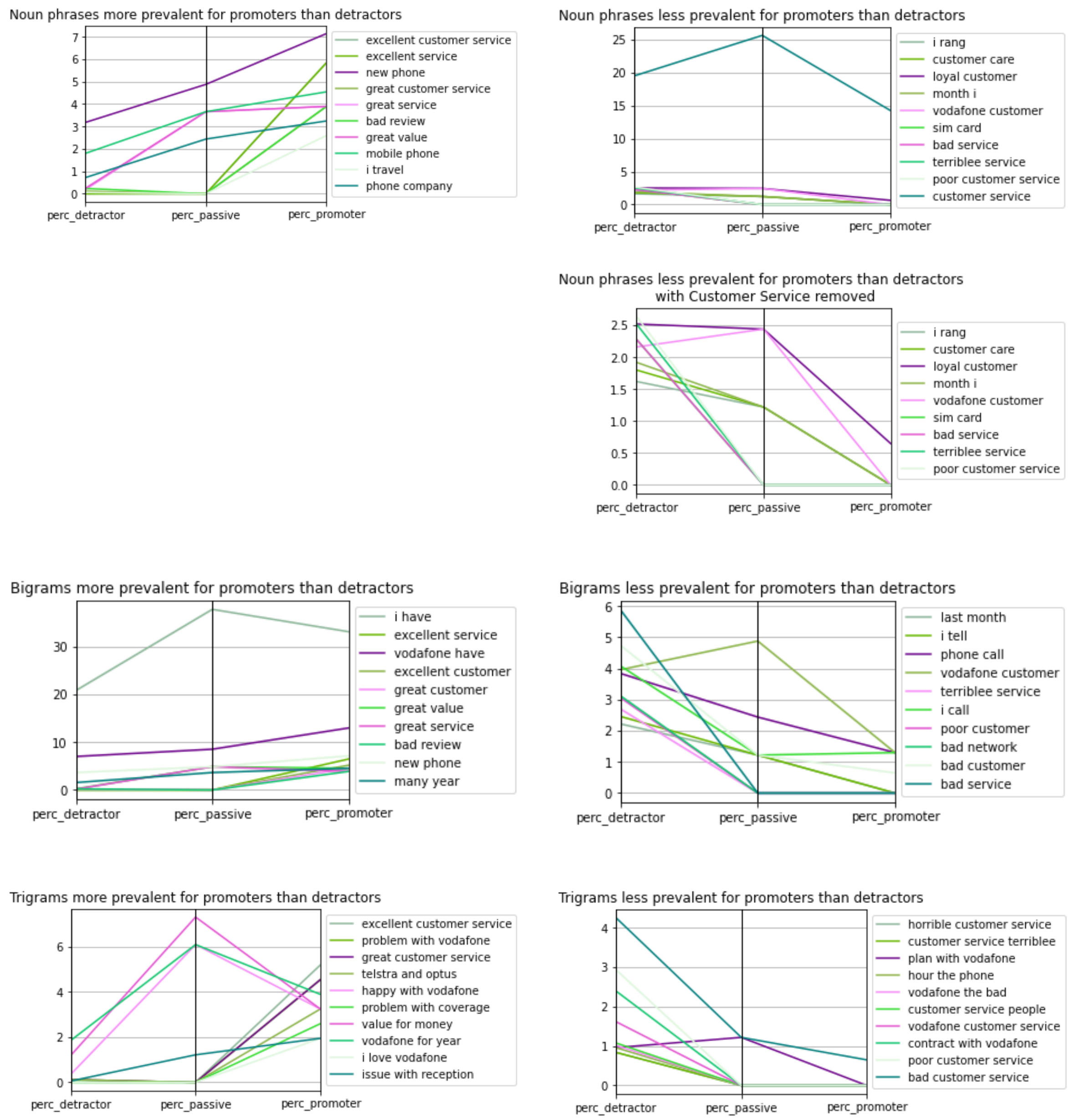
There is some evidence in the literature that selected extraction of words or phrases from sentences (such as only including nouns or adjective-noun combinations) can reveal more insightful and/or nuanced information about the contents of the text being examined. For example, use of this technique indicated a possible difference in the types of things and language used by detractors, passive and promoters. For example, in this dataset, promoters appeared to focus on the service they had received and interactions with people, whereas the detractors tended to focus on money-related issues such as billing, account problems and contracts (Figure 7).

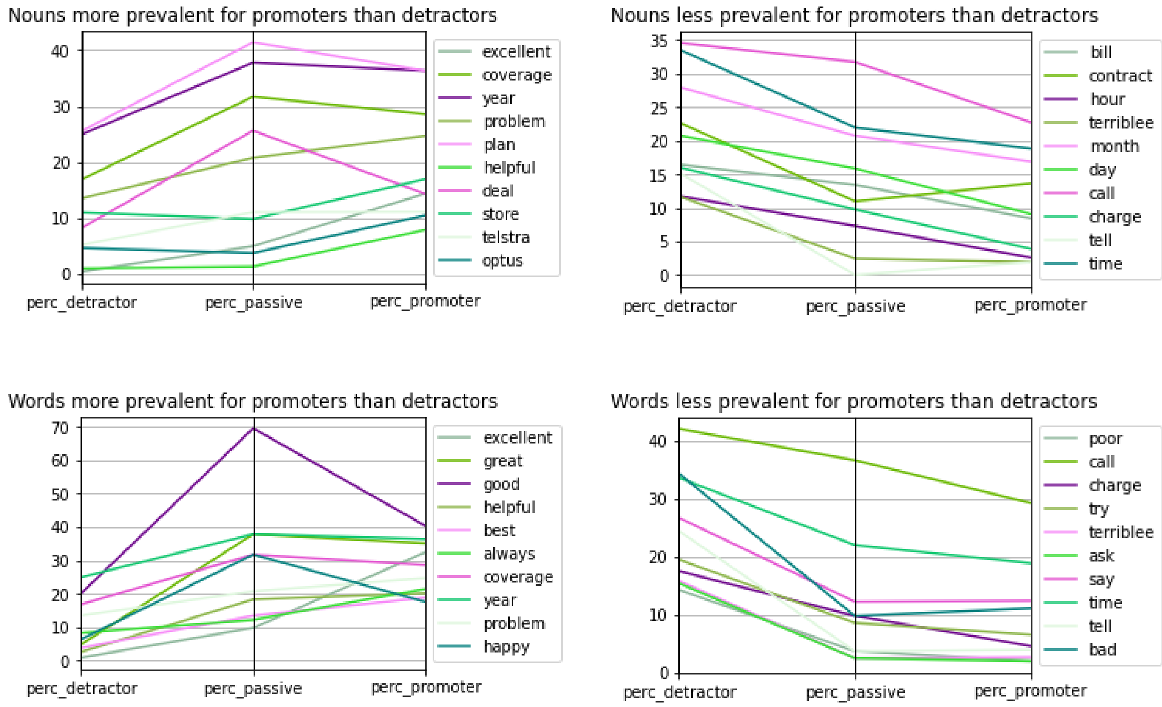
Figure 7: Noun phrase-based word clouds for promoters (LHS top), detractors (RHS top) and passives (bottom)



As the original project scope included investigating the extent to which the free text survey responses could differentiate detractor, passive and promoter customers, the extracted features were analysed to understand how their frequency changed between the three groups. To understand this, the percentage of documents each feature value appeared in for detractors, passive customers, and promoters was calculated for each feature. The features were then ordered by the difference in percentage between the detractor and promoter reviews, and the top ten values that increased or decreased between the two groups were plotted using parallel coordinates. The results indicated that not only was there an observed difference in feature expression prevalence between the groups, in the case of bigrams, trigrams, and noun phrases, different themes were also being expressed by promoters and detractors. Interestingly, the fact that some terms were more frequent for the passive group, for example “happy with Vodafone”, and “value for money” were more common in passive reviews than either promoters or detractors also suggested the presence of a nuanced relationship between language and review score.

Figure 8: Comparison of feature value expressions by NPS category

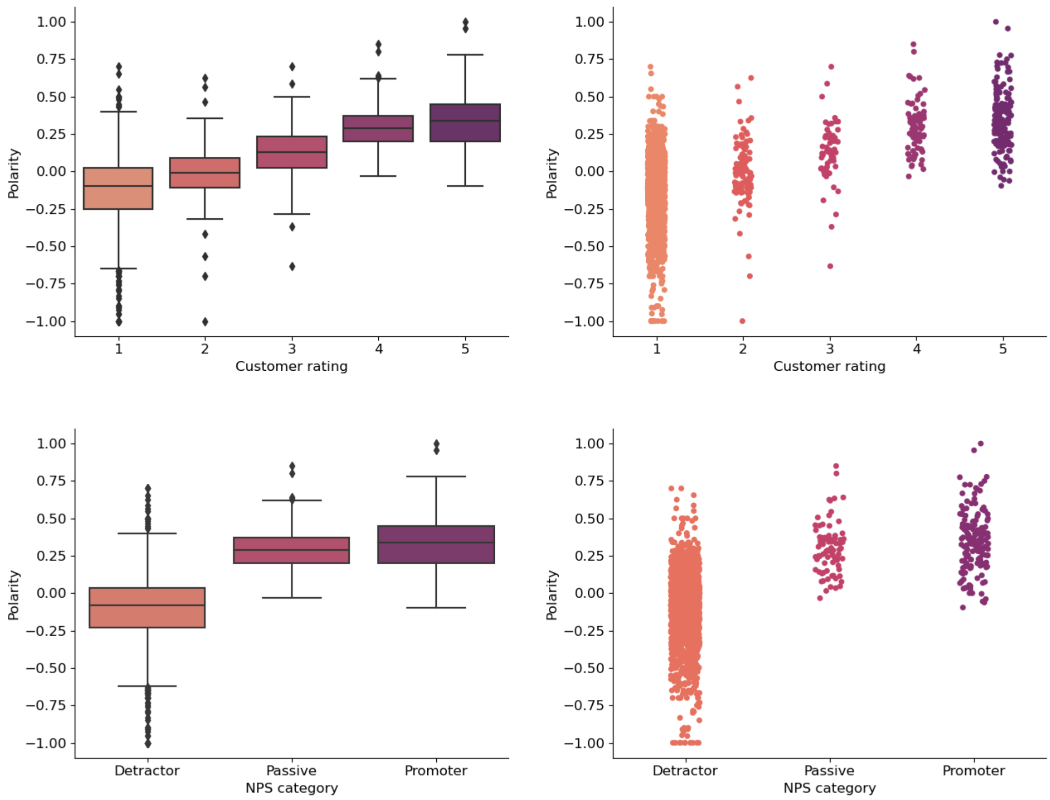




Sentiment analysis

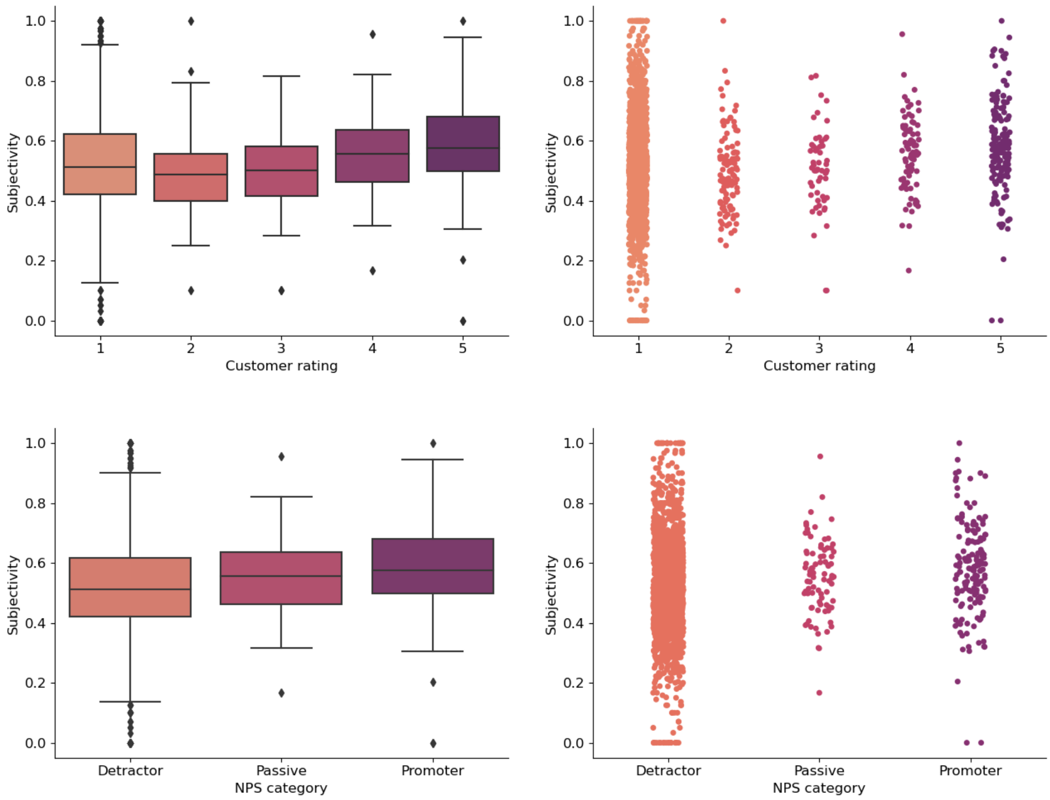
Two sentiment lexicons (VADER and TextBlob) were used to assess the relationship between customer ratings and review sentiment. As expected, there was evidence of a positive relationship between customer rating and sentiment score. However, a somewhat more interesting finding was that the range of sentiment observed across the reviews generally decreased as the customer rating or NPS category became more positive (Figure 9).

Figure 9: Comparison of the distribution of review sentiment by customer rating and NPS category



A similar analysis was conducted for review subjectivity (Figure 10), and whilst there was a slight indication of a similar relationship between the variance and customer rating, it was unable to be determined if this was in fact solely due to the uneven distribution of reviews across the ratings scale.

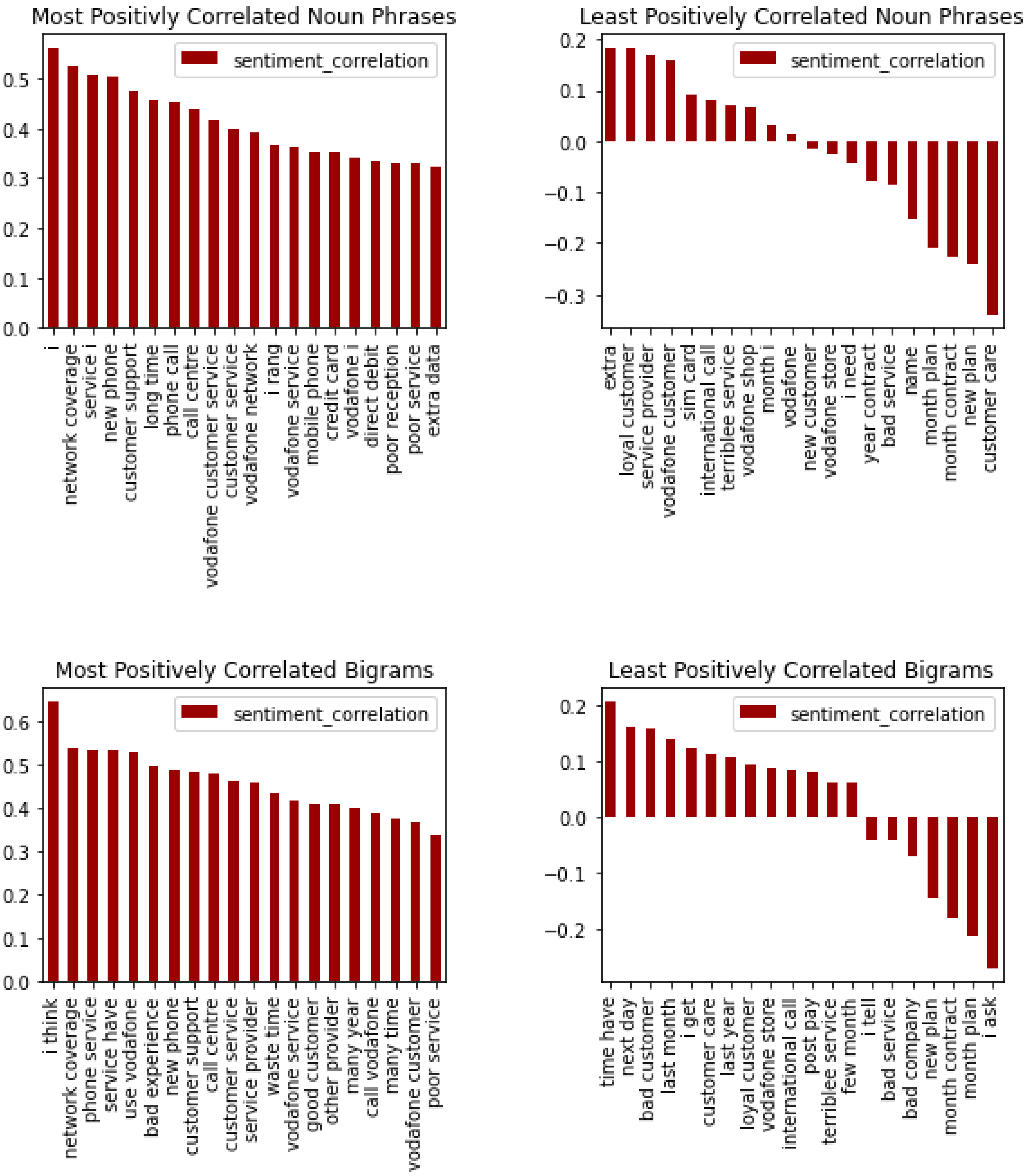
Figure 10: Comparison of the distribution of review sentiment by customer rating and NPS category

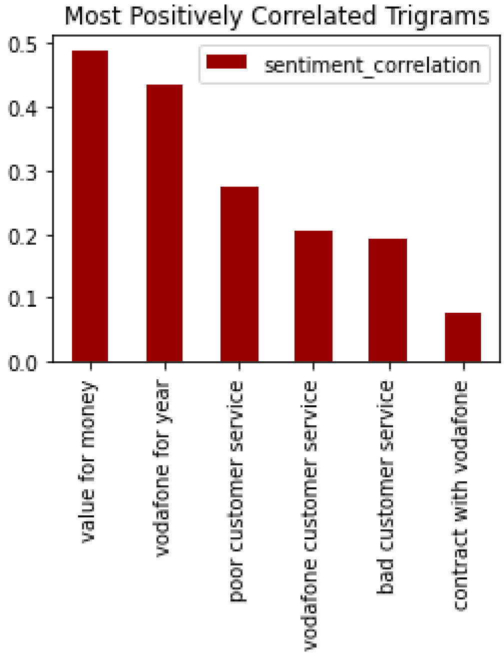


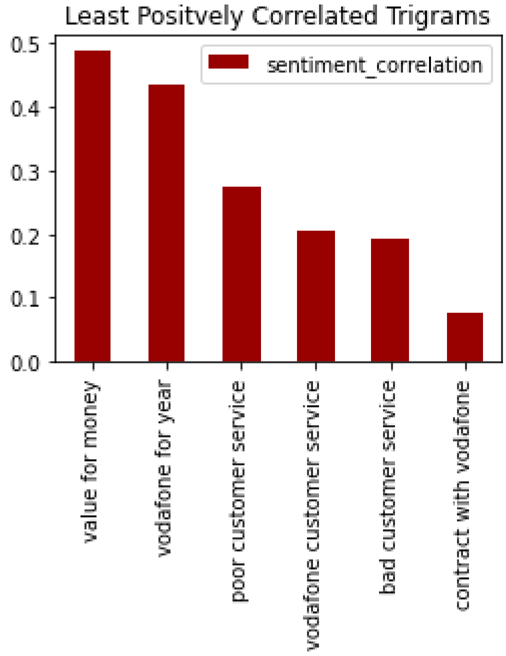
To further investigate how sentiment, review score, and language were related, each extracted feature was analysed to see the strength of correlation between the sentiment of the sentence in which it was used and the review's score. The plots over the next two pages (Figure 11) show the features with the most positive and least positive correlations between their sentiment of usage and sore outcomes. Note that not all features were able to have a correlation determined as the unbalanced data meant that some terms did not have sufficient variation in score.

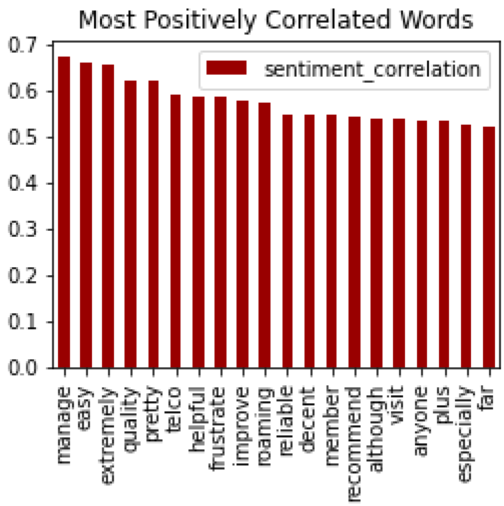
The analysis of sentiment correlation for specific terms showed that for this data, specific words and phrases did show a correlation between the authors sentiment and their review score. For the most part, the correlation was positive, meaning that positive sentiment was associated with higher scores than negative sentiment although to varying degrees of strength. However, there were a few terms more commonly for individual words and nouns where a negative correlation was seen; this indicated that for these terms, negative statements involving them were associated with higher review scores.

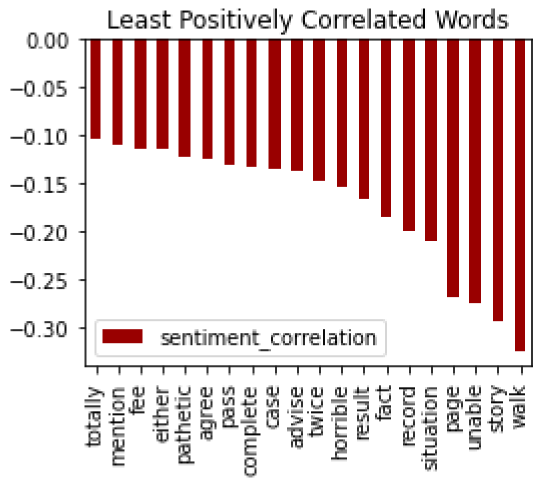
Figure 11: Pearson Correlation between sentence sentiment and customer rating for feature values




Most Positively Correlated Trigrams
sentiment_correlation
0.0
0.1
0.2
0.3
0.4
0.5
value for money
vodafone for year
poor customer service
vodafone customer service
bad customer service
contract with vodafone

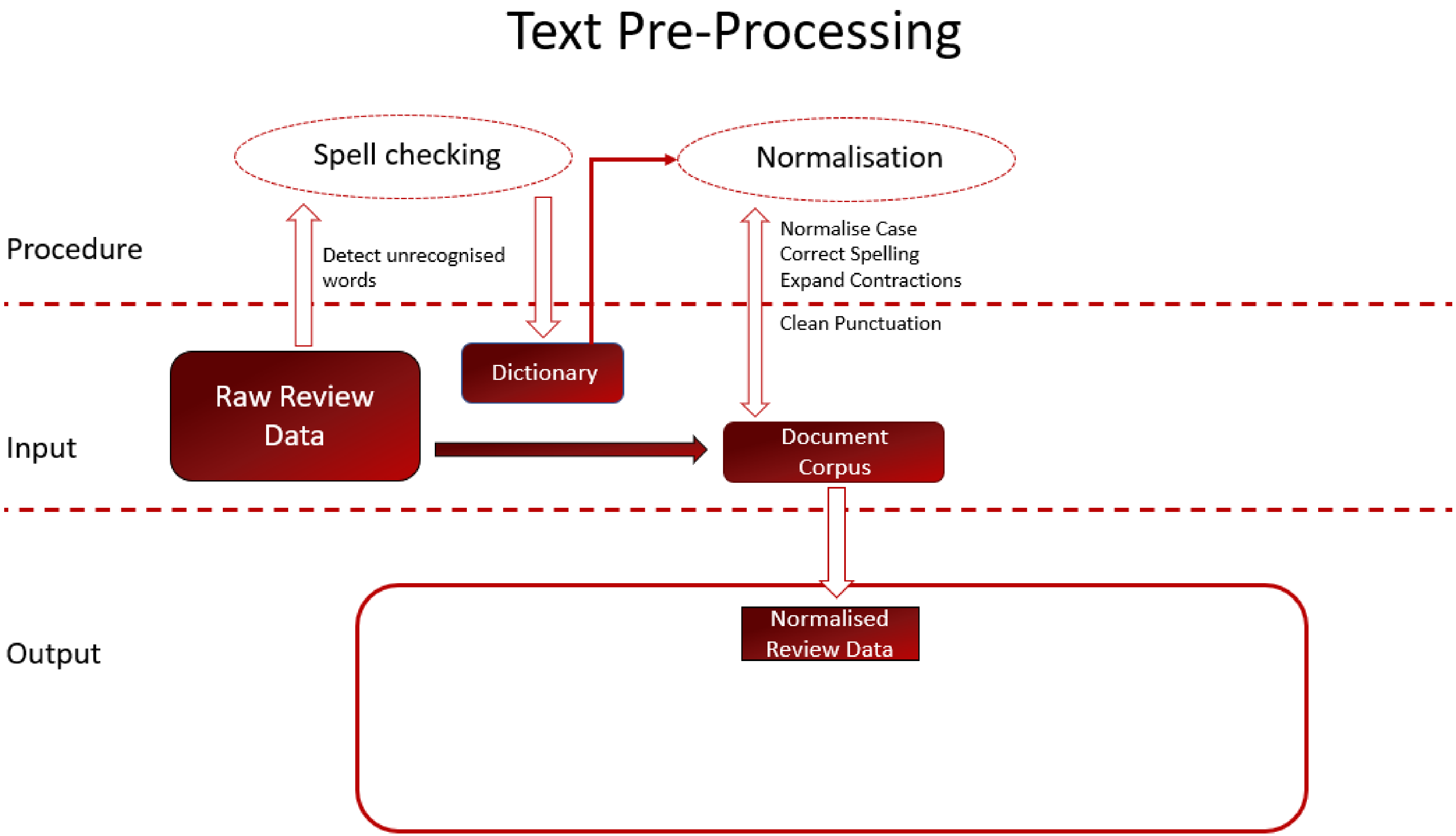

Least Positvely Correlated Trigrams
sentiment_correlation
0.0
0.1
0.2
0.3
0.4
0.5
value for money
vodafone for year
poor customer service
vodafone customer service
bad customer service
contract with vodafone


Most Positively Correlated Words
sentiment_correlation
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
manage
easy
extremely
quality
pretty
telco
helpful
frustrate
improve
roaming
reliable
decent
member
recommend
although
visit
anyone
plus
especially
far

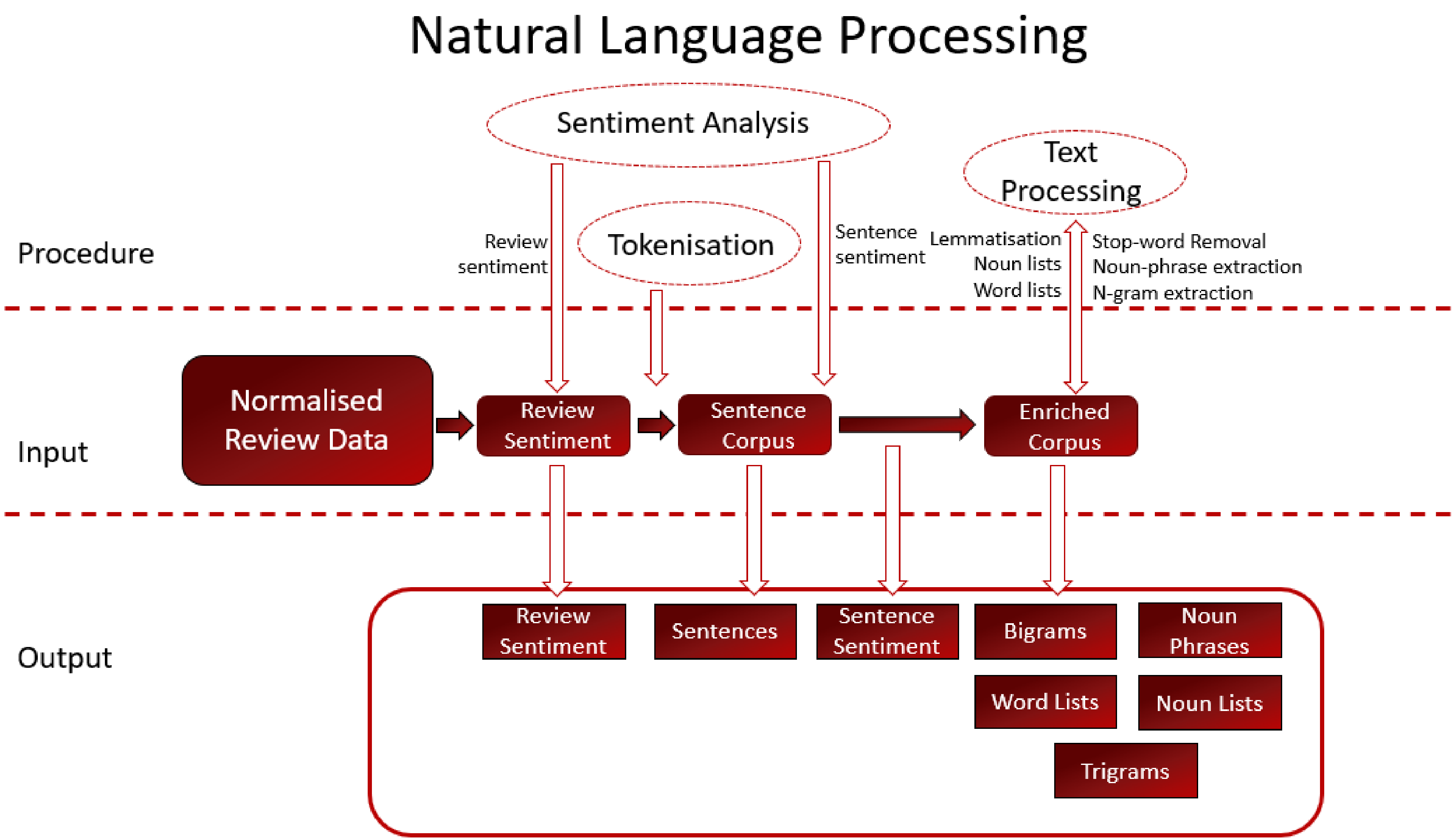

Least Positively Correlated Words
sentiment_correlation
0.00
−0.05
−0.10
−0.15
−0.20
−0.25
−0.30
totally
mention
fee
either
pathetic
agree
pass
complete
case
advise
twice
horrible
result
fact
record
situation
page
unable
story
walk

Appendix C - NLP Framework Diagrams
(Matthew Moore)

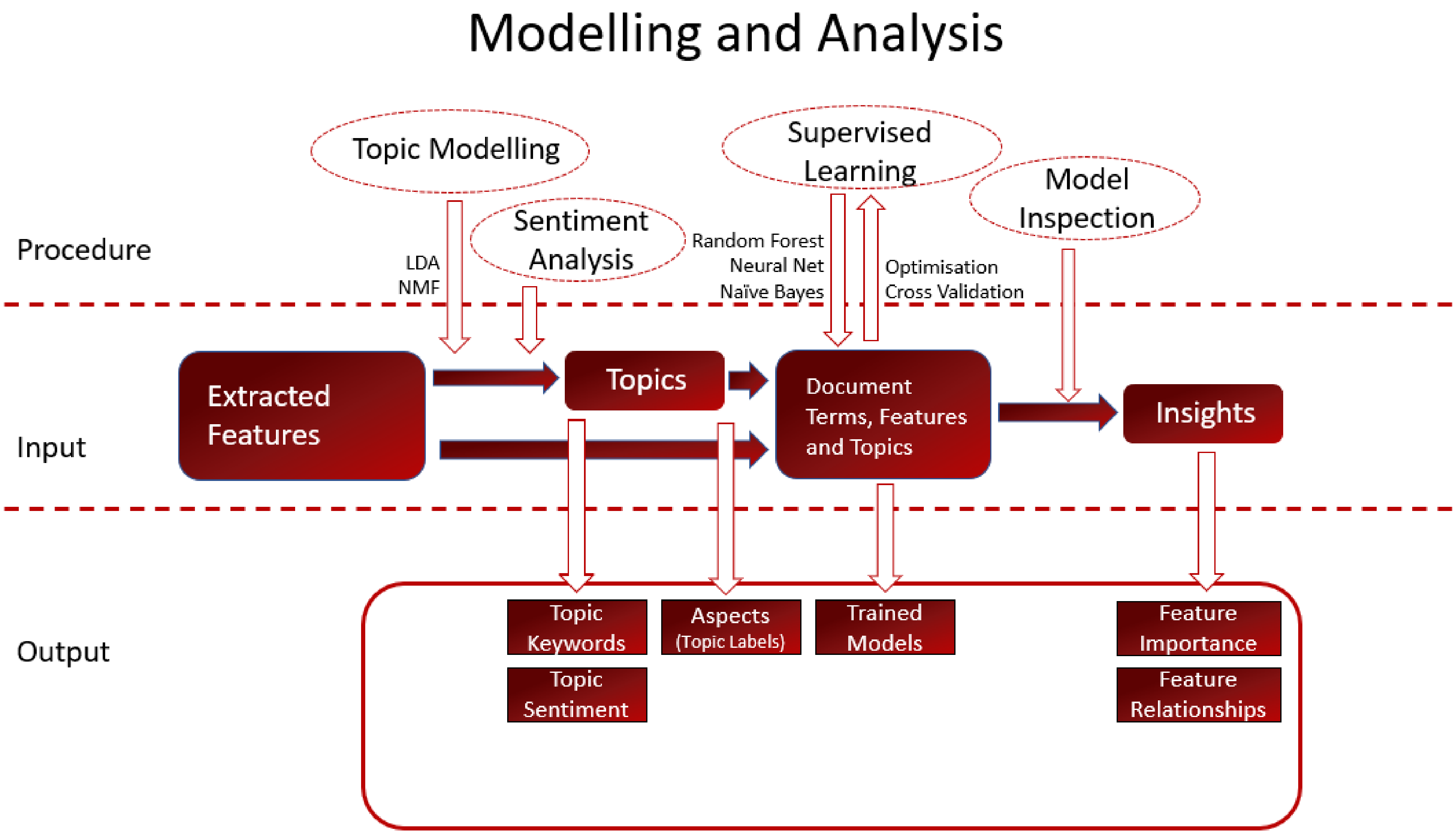
Data pre-processing module



Natural Language Processing module



Modelling and analysis module



Interactive visualisation module

Interactive Visualisation