Final Presentation

Vodafone and JCU Master of Data Science Project Team

JAMES COOK UNIVERSITY
AUSTRALIA

Agenda

Project recap

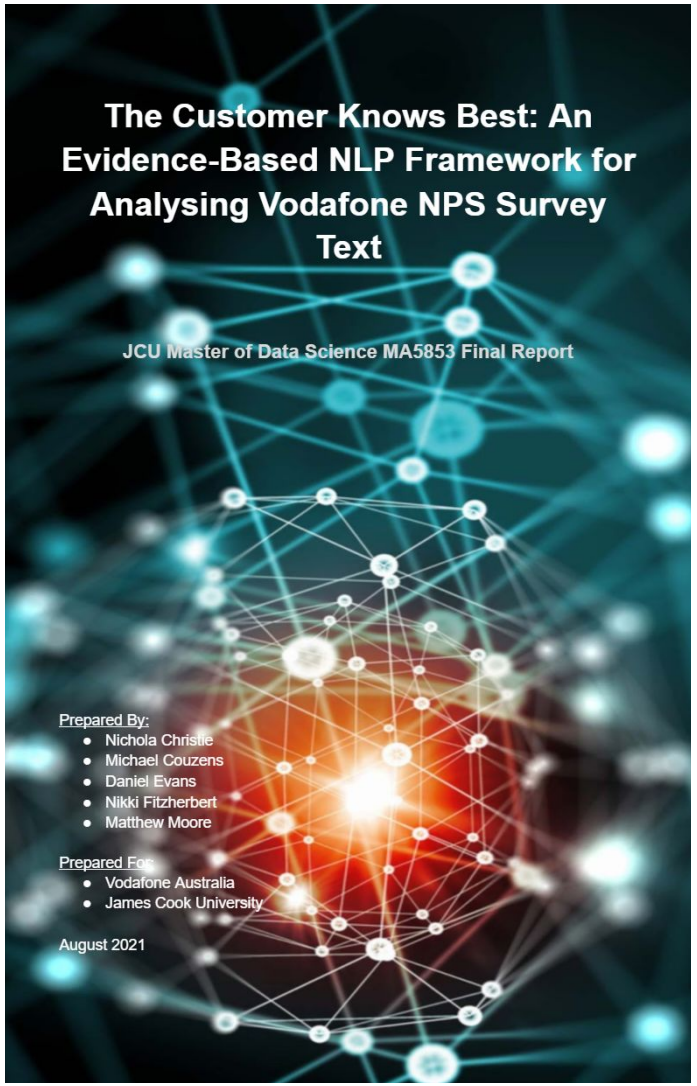
Key results and insights

NLP framework

Recommendations

Next steps

Q&A



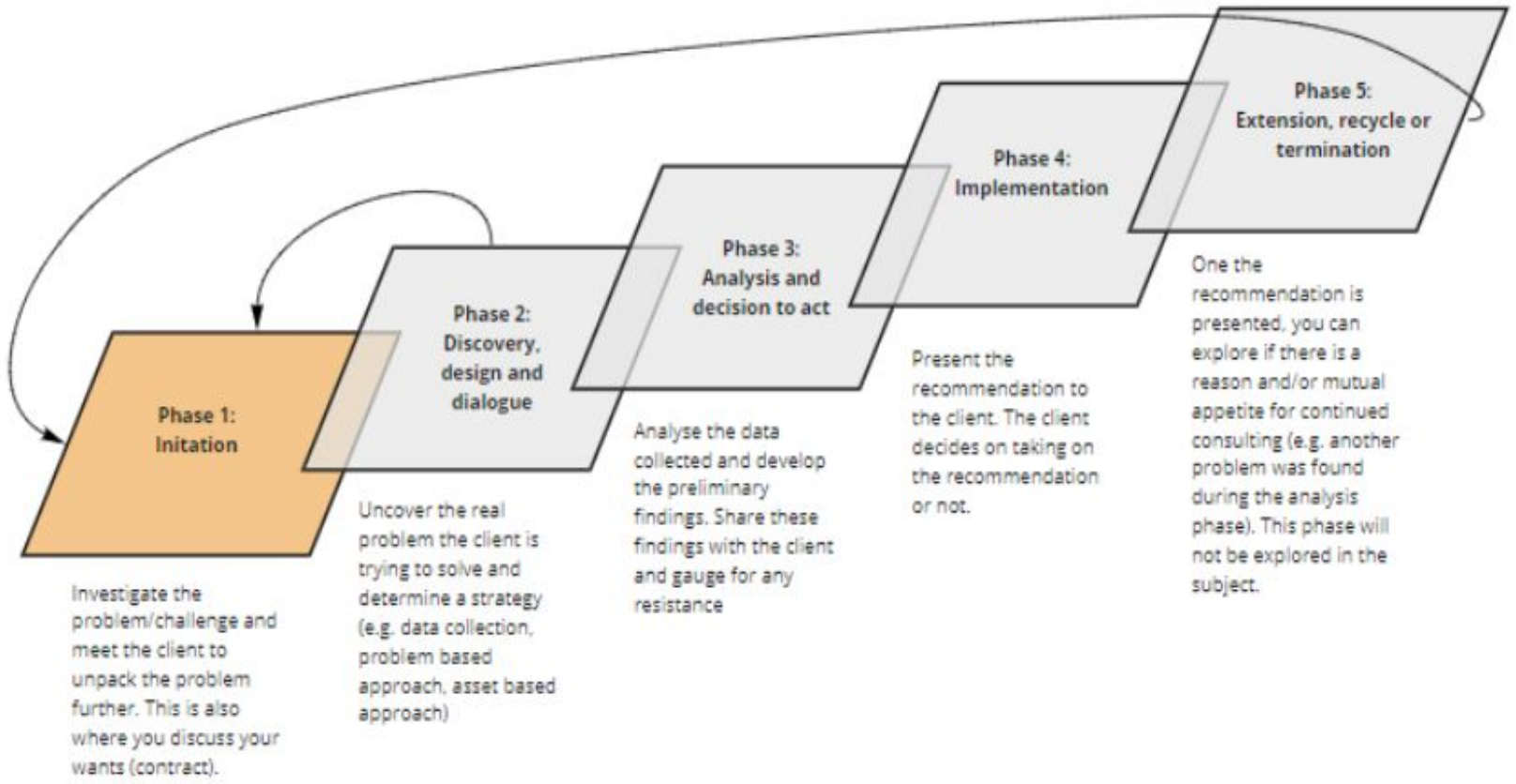


Figure 2.1. The five phases of a consultancy project: the consultant's view (JCU, 2019).

Project Overview

Phase 1: Initiation (Wks 1-2)

- Problem statement/stakeholder requirements
- Key business questions

Phase 2: Discovery (Wk 3)

- Data access issues and subsequent project re-scope
- Mid-point check-in

Phase 3: Analysis & Solution (Wks 4-6)

- Data collection, literature review, EDA & NLP framework

Phase 4: Recommendations & Implementation (Wk 7)

Exploratory Data Analysis

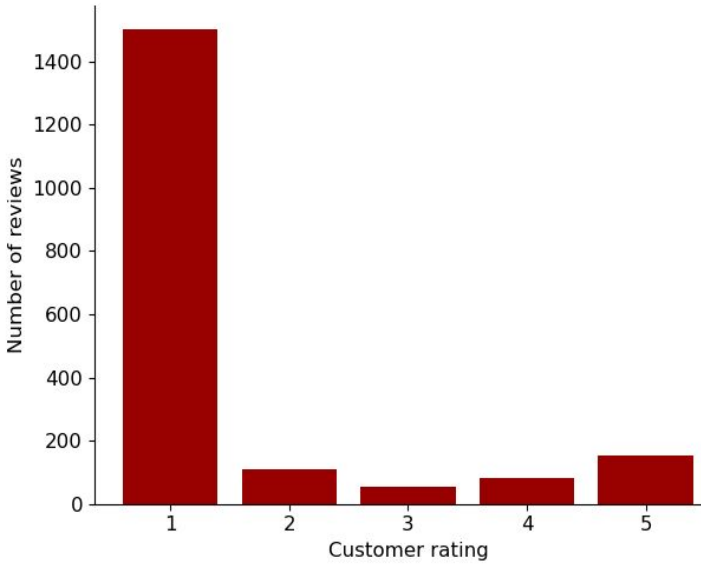
Data Collection

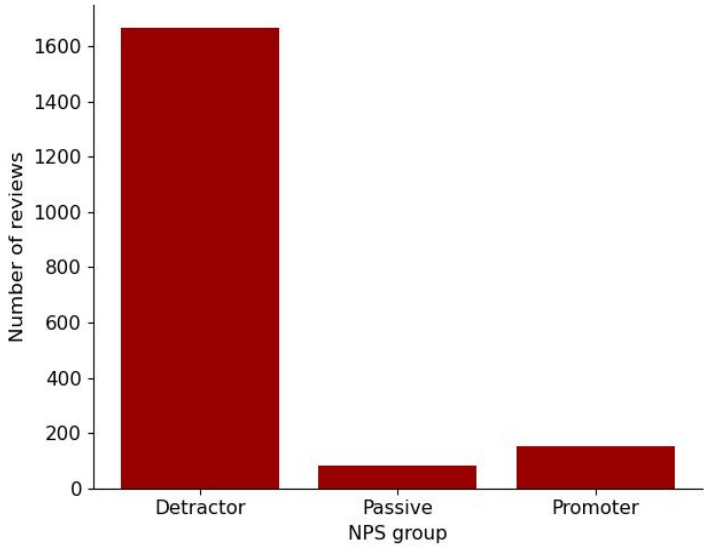
Raw Dataset Description

- ~2,000 reviews
- Review text
- Review title
- Rating/score out of 5
- Some metadata for recent reviews

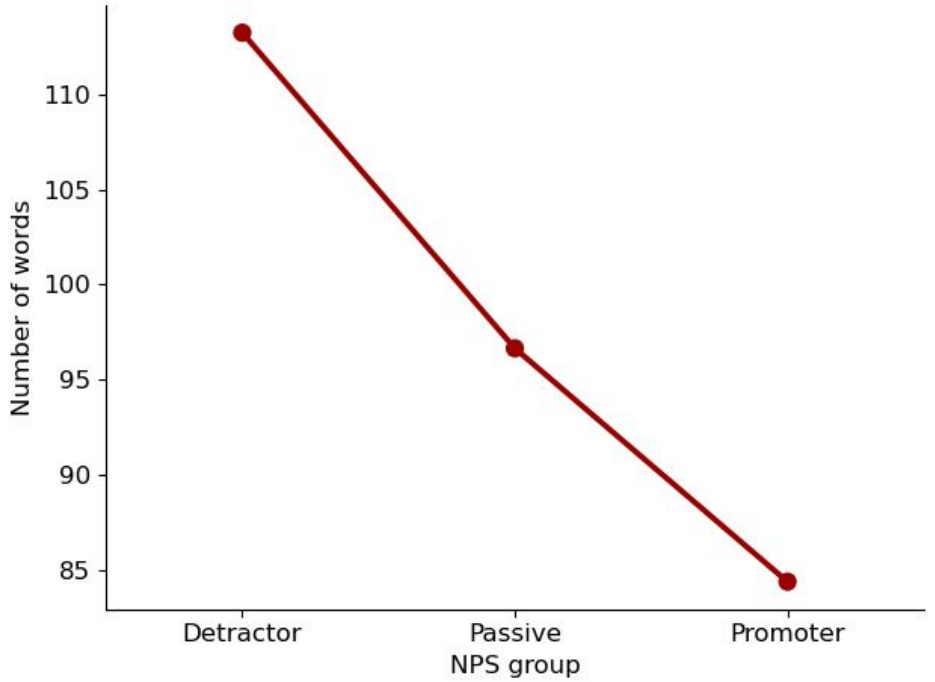
Key Findings – EDA

The dataset was unbalanced





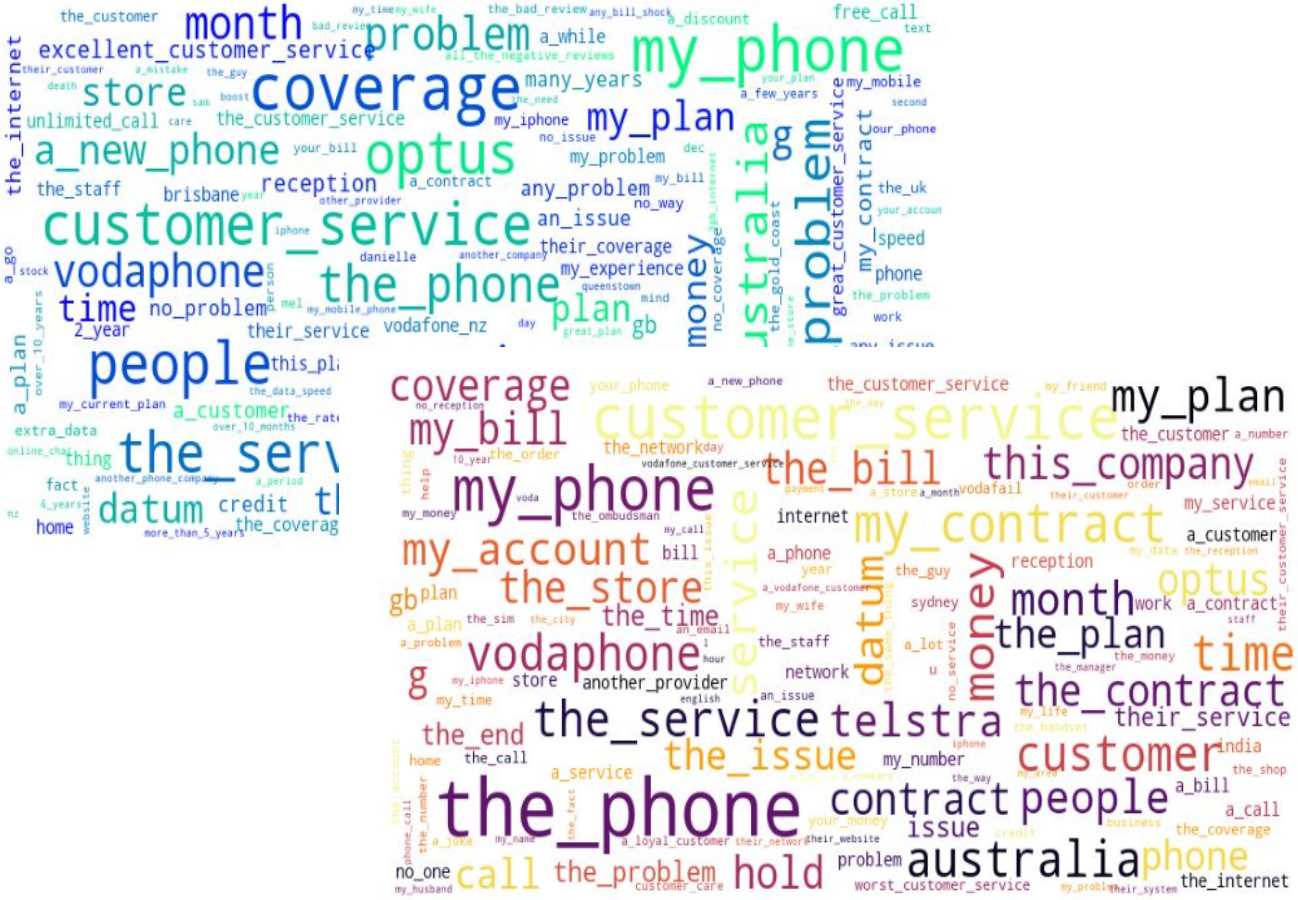
Unhappy customers tended to write more than happy customers



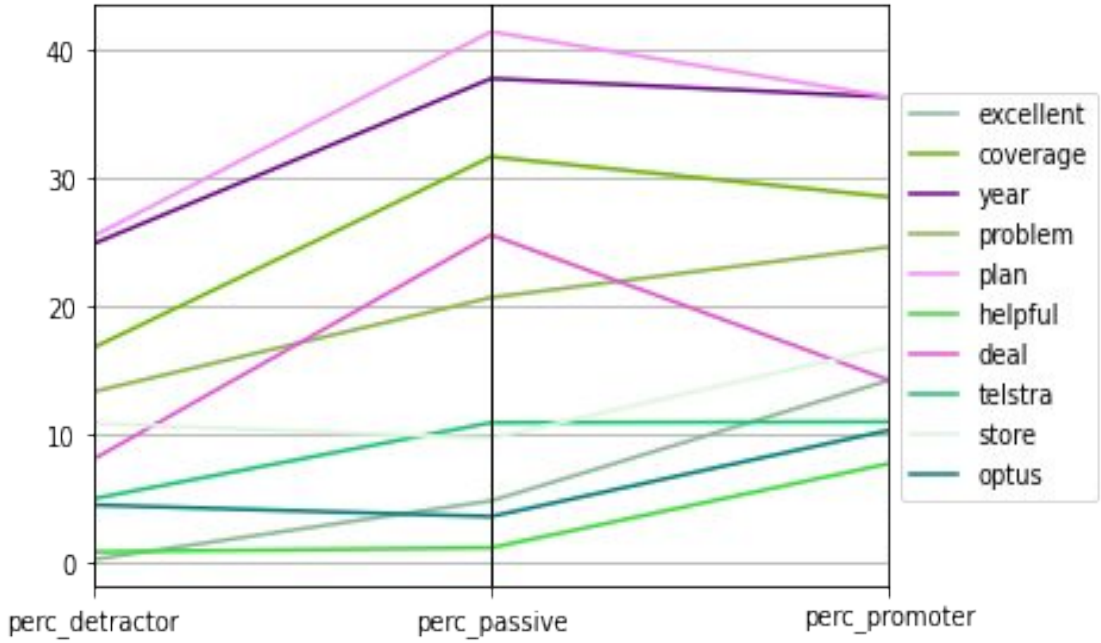
Images sourced from: (Christie et al., 2021)

Key Findings – EDA (cont'd)

Selected n-gram extraction can lead to more nuanced and useful insights



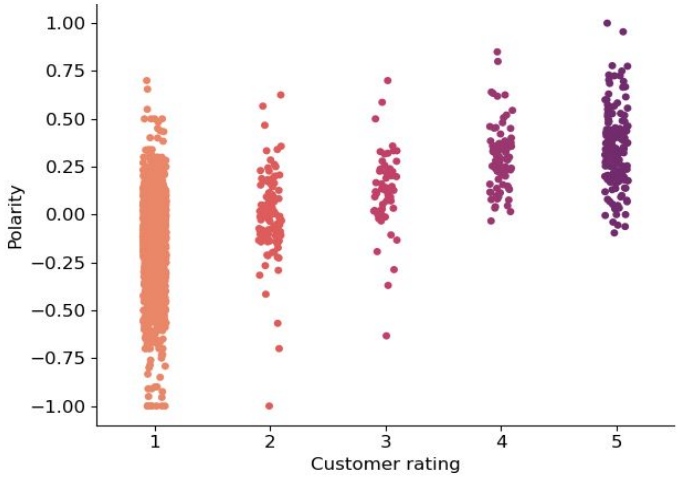
Thematic differences between detractors and promoters could be seen by observing term prevalence between the NPS groups

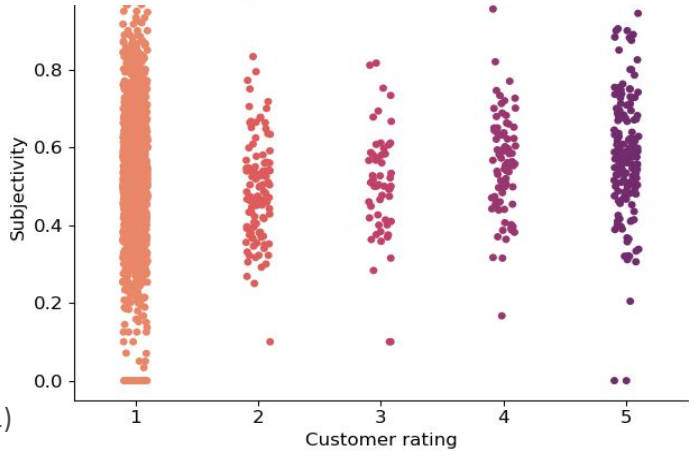


Images sourced from: (Christie et al., 2021)

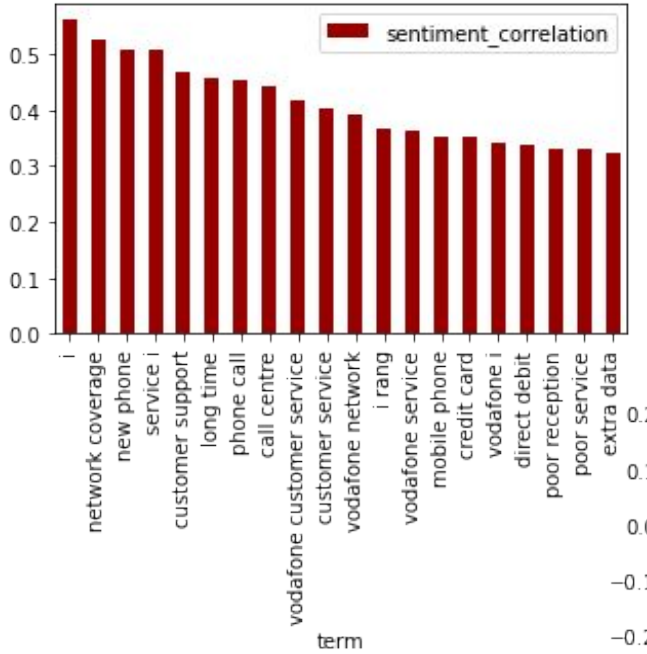
Key Findings – EDA (cont'd)

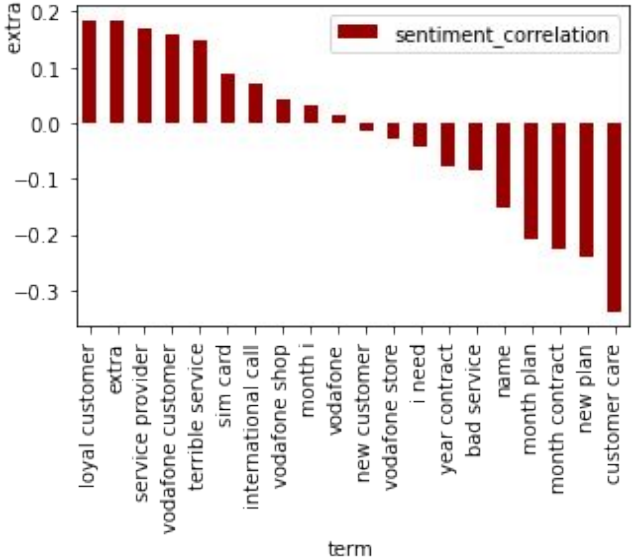
The amount of variability in sentiment and subjectivity tended to decrease as the review score increased.





There was a correlation between words/phrases and review score.





Images sourced from: (Christie et al., 2021)

Literature Review

Overview

- Discover:
  - Known drivers of NPS
  - Current advances and best-practice NLP methodology
- Covered NLP methods within the customer loyalty context, specifically:
  - NPS, and limitations of NPS
  - Drivers of customer loyalty
  - Relevant text analytics methods:
    - Topic modelling
    - Sentiment analysis
    - Aspect extraction
  - Applications of NLP to customer-generated text

Key Findings – Literature Review

- NPS alone not sufficient to guide business strategy
- Text analysis is effective for understanding NPS
- Positive experiences → Customer loyalty
- Attractiveness of alternatives, search effort, and satisfaction → switching intentions
- Huge flexibility in number of methods and ways of combining and implementing these
- In line with the EDA findings that positive reviews were shorter, the literature also identified that satisfied customers raised fewer topics

Key Findings – Literature Review (cont'd)

- Current best short-text topic modelling approaches include LDA and NMF; promising advances include SS-LDA and semantic clustering
- Sentiment analysis provide insights into customer loyalty:
  - highly active area of research
  - polar (positive or negative), detect emotions (e.g. disgust, trust, sadness)
  - applied to document/sentence/aspect
  - ML or lexicon-based
- Promising aspect-based sentiment analysis methods include:
  - conditional random fields
  - sequence tagging
  - hand-crafted features including lexicons, syntactic and semantic features
  - neural networks
- NLP methods have been applied to improve customer satisfaction and loyalty across many industries including telecommunications

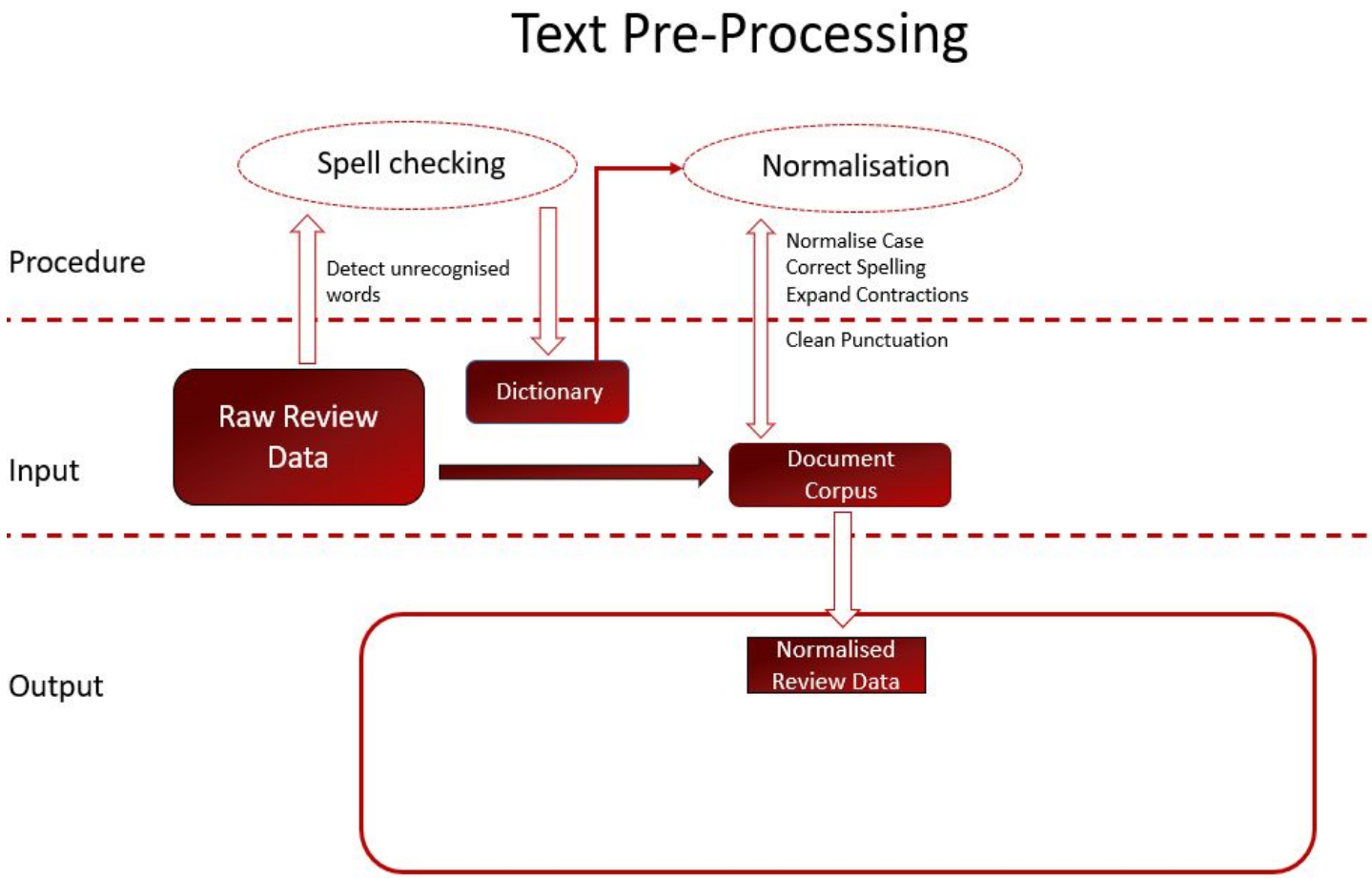
NLP Framework

Modules

Principals

NLP Framework – Module 1

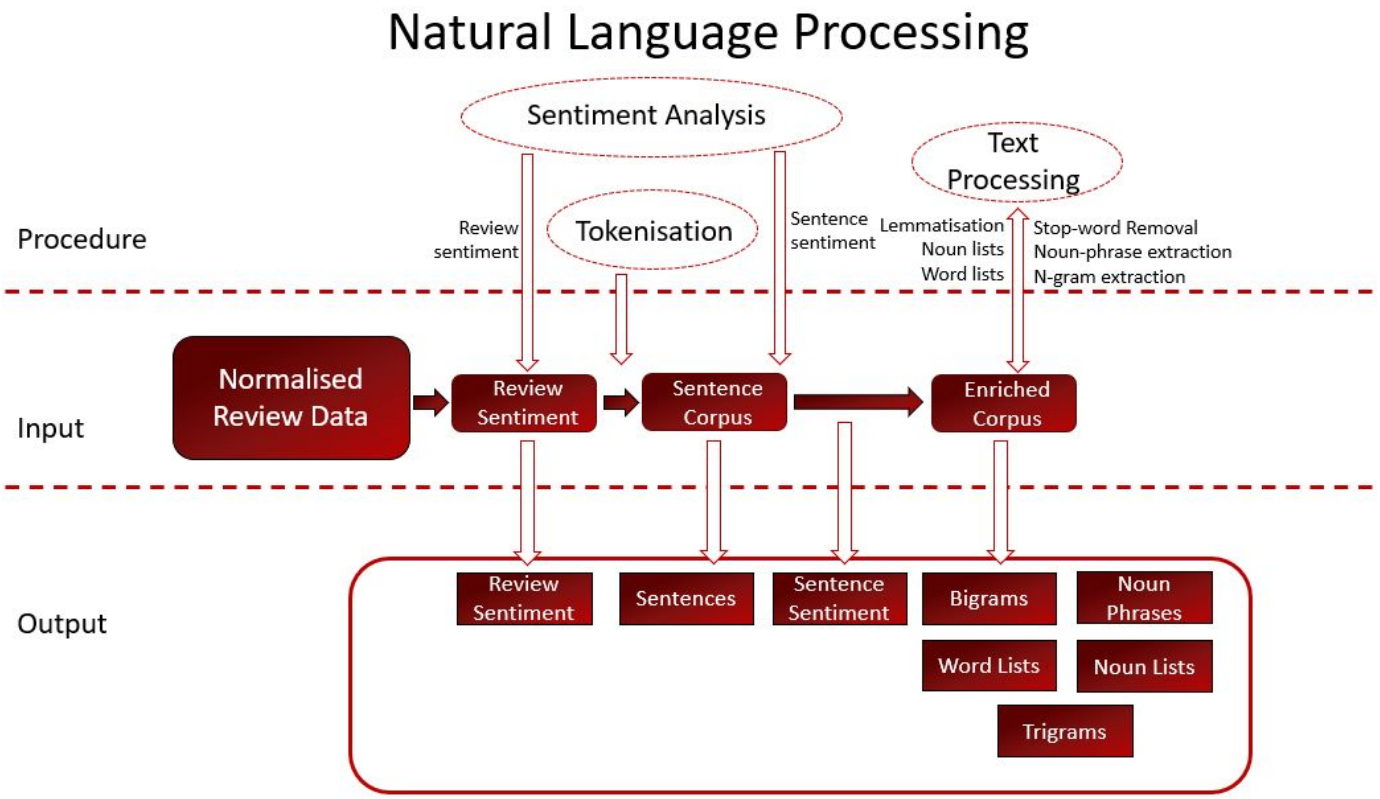
- Is often data/problem/task-specific.
- No single approach and frequently a compromise.



Images sourced from: (Christie et al., 2021)

NLP Framework – Module 2

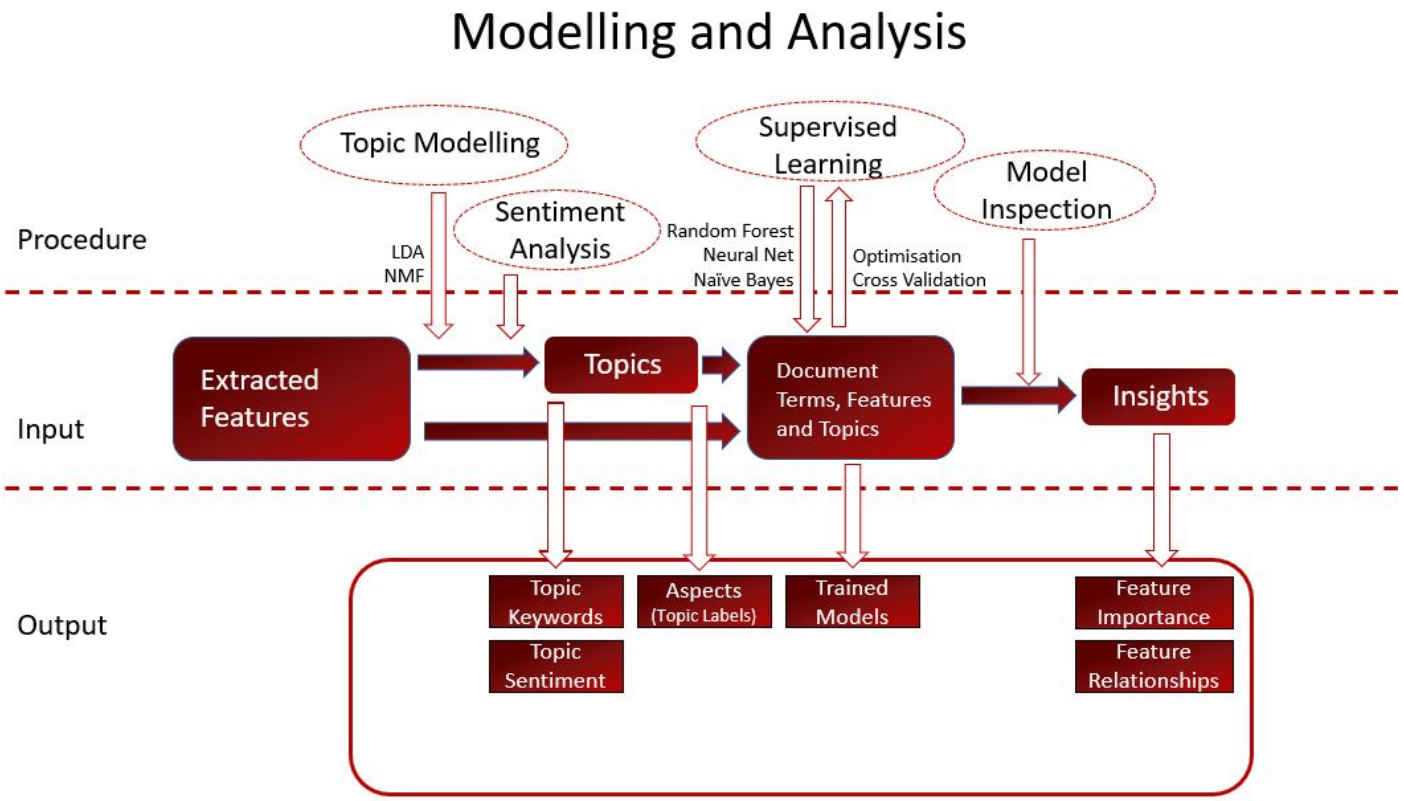
- Co-developed as part of the EDA
- Cardinality considerations
  - stopword removal
  - lemmatisation
- Spelling Correction and Dictionaries
- Sentiment scoring and context
- Dictionaries
- Actionable n-grams
- Output data-structures



Images sourced from: (Christie et al., 2021)

NLP Framework – Module 3

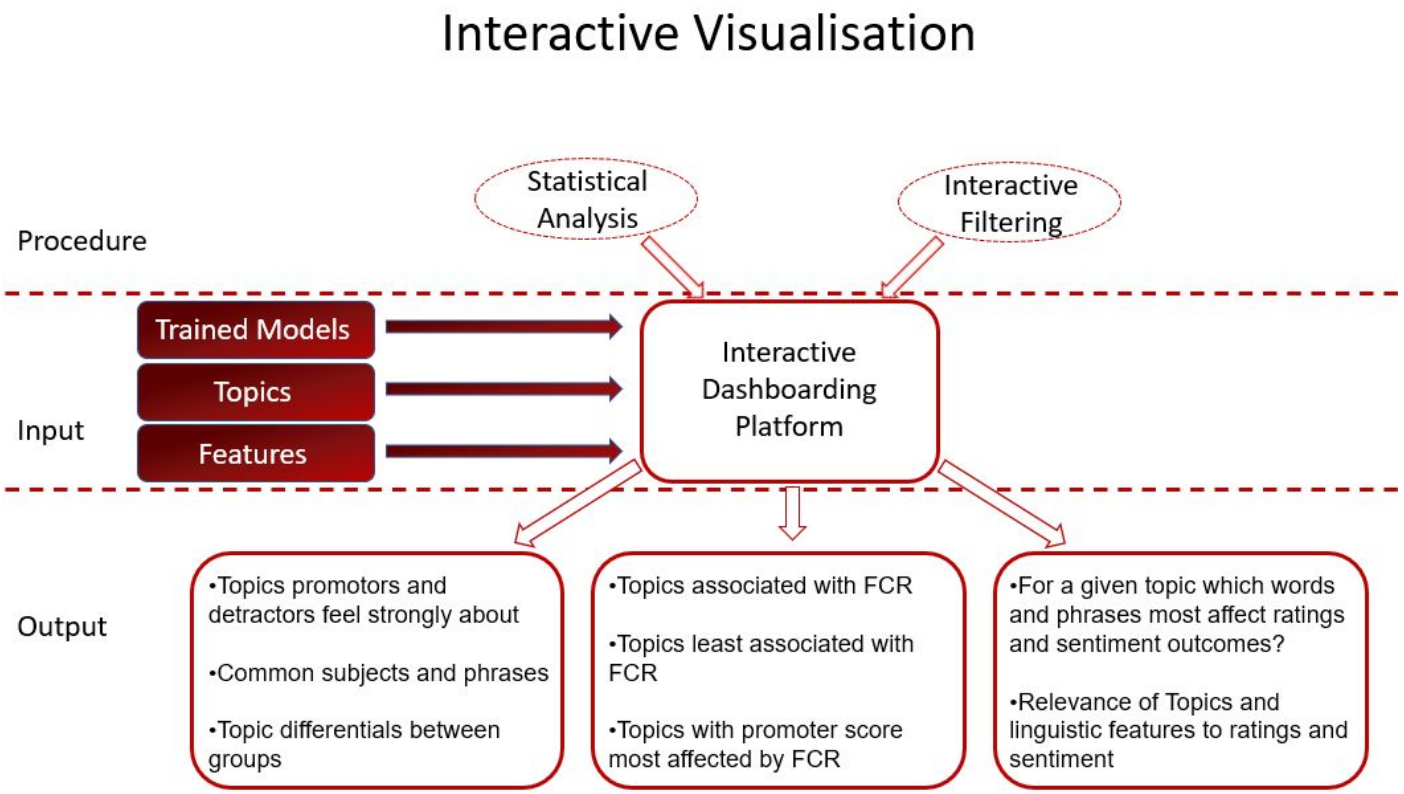
- Topic modelling using LDA and NMF
- Sentiment analysis of topics
- Hyperparameter optimisation and cross-validation of models
- Feature inspection to understand the relative importance of topics, sentiment, language and phrases, and the interaction between features.



Images sourced from: (Christie et al., 2021)

NLP Framework – Module 4

- Ingests the Sentiment tagged features from the Analysis pipeline:
  - words, nouns, and n-grams
  - Noun Phrases
  - Topics and Aspects
- As well as
  - Feature Importance
  - Class probabilities for each feature
  - Feature relationships
- And allows;
  - Questions to be asked
  - Filters to be applied
  - Insights to be gained



Images sourced from: (Christie et al., 2021)

Recommendations

The project team recommends that Vodafone continue to pursue a NLP approach to analysing their NPS survey free text

1. Explores the use of LDA and NMF for topic modelling in conjunction with sentiment analysis, 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.

Next Steps

Q&A

Thank You

References

Christie, N., Couzens, M., Evans, D., Fitzherbert, N., & Moore, M. (2021). *The customer knows best: An evidence-based NLP framework for analysing Vodafone NPS survey text* [Unpublished assignment submitted for MA5853]. James Cook University.

James Cook University. (2021). *MA5853: Project 1: Week 2 topic 1: Initiation*. LearnJCU. https://learn.jcu.edu.au/