Using Sentiment Analysis for Exploring the ‘Voice of Customer’ for AIB Banking.

Brona McKnight [Thesis]

Technological University Dublin

Follow this and additional works at: https://arrow.dit.ie/scschcomdis

Part of the Computer Engineering Commons, and the Computer Sciences Commons

Recommended Citation


This Theses, Masters is brought to you for free and open access by the School of Computing at ARROW@TU Dublin. It has been accepted for inclusion in Dissertations by an authorized administrator of ARROW@TU Dublin. For more information, please contact yvonne.desmond@dit.ie, arrow.admin@dit.ie, brian.widdis@dit.ie.

This work is licensed under a Creative Commons Attribution-Noncommercial-Share Alike 3.0 License
Using Sentiment Analysis for Exploring the ‘Voice of Customer’ for AIB Banking.

Brona McKnight

D12126371

A dissertation submitted in partial fulfilment of the requirements of Technological University Dublin for the degree of M.Sc. in Computer Science (Data Analytics)

2019
I certify that this dissertation which I now submit for examination for the award of MSc in Computing (Data Analytics), is entirely my own work and has not been taken from the work of others save and to the extent that such work has been cited and acknowledged within the text of my work.

This dissertation was prepared according to the regulations for postgraduate study of the Technological University Dublin and has not been submitted in whole or part for an award in any other Institute or University.

The work reported on in this dissertation conforms to the principles and requirements of the Institute’s guidelines for ethics in research.

Signed:    

[Signature]

Date: 24 September 2019
ABSTRACT

Allied Irish Bank and their ‘Voice of Customer’ programme has accumulated a rich source of digitally available textual sources in the form of survey responses from their customers. This large collection of unstructured content contains relevant customer input, however, humans can only deal with a limited amount of unstructured content in a given time. To address this problem, the concept of sentiment analysis has been established to help human's process large amounts of unstructured content.

Sentiment analysis or opinion mining is the methodology of computationally identifying and categorizing opinions expressed in a piece of text, in order to determine whether the review writer's attitude towards a particular topic or product is a polarity of positive, negative, or neutral.

This research focuses on sentiment analysis to evaluate AIB’s digital banking for older generation customers referred in literature as ‘Baby Boomers’ and ‘Traditionalist.’ Digital banking in AIB includes mobile, internet and in-branch self-service banking.

For realizing this study, a collection consisting of c.70,000 customer records with ‘Net Promoter Score of 0-10’ alongside a textual description of the customer experience is extracted from the bank’s enterprise data warehouse. Based on the textual data, Supervised or Machine-Learning algorithms will attempt to predict the known NPS scores and age group to inform customer satisfaction improvements. Descriptive statistics and feature extraction methods will identify negative features and actionable insights from the sentiments, in order to make the customer journey more inclusive and sympathetic for this cohort of older customers.

The evaluations find promising opportunities, but also limitations for sentiment analysis. The machine learning classifiers supported the known VOC programme NPS results and produced similar classification results with reasonable accuracy. Predicting a customer age based on the sentiment produced unreliable results and showed less promise. This is possibly due to factors that could be addressed in future studies such as the effects of overfitting as a consequence of limited input.
Key words: Voice of the Customer, Sentiment Analysis, Supervised and Unsupervised Machine Learning, User Characteristics, specifically older Customers.
ACKNOWLEDGEMENTS

I would like to express my sincerest thanks to my project supervisor Dr. Diana Carvelo who kept me motivated and provided invaluable direction through this process. You are an excellent tutor and I was lucky to have you as my guide.

I would also like to thank my AIB colleagues Frank Kendlin and Ade Ali for their support in the formulation of the experiment. Their in-depth knowledge of text mining and machine learning was invaluable. To Karon Kells and Conor O’Neill for supporting my professional and academic development.

Lastly, to my family, especially my husband Conall who had to mind the kids whilst I attended evening and weekend classes, and numerous Saturday’s whilst I wrote this dissertation. You provided me with the strength and courage to endure this journey to the end.
# TABLE OF CONTENTS

## CHAPTER 1: INTRODUCTION ................................................................. 1

1.1. BACKGROUND ............................................................................. 1
1.2. RESEARCH PROBLEM ................................................................. 2
1.3. RESEARCH AIM ......................................................................... 4
1.4. HYPOTHESIS ............................................................................. 4
1.5. RESEARCH OBJECTIVES .............................................................. 4
   1.5.1. MAIN OBJECTIVE .............................................................. 4
   1.5.2. SPECIFIC OBJECTIVES ....................................................... 5
1.6. RESEARCH METHODOLOGIES .................................................... 5
1.7. SCOPE AND LIMITATIONS ......................................................... 6
1.8. DOCUMENT OUTLINE ............................................................... 7

## CHAPTER 2: LITERATURE REVIEW ......................................................... 8

2.1. INTRODUCTION ........................................................................... 8
2.2. VOICE OF THE CUSTOMER AND NET PROMOTER SCORES .......... 8
   2.2.1. VOICE OF CUSTOMER SURVEY’S THROUGH THE MEDIUM OF EMAIL 10
2.2.2. SUMMARY ............................................................................ 11
2.3. SENTIMENT ANALYSIS ............................................................... 11
   2.3.1. EXPLORATION OF METHODOLOGIES SUITABLE FOR SENTIMENT ANALYSIS 13
      2.3.1.1. SIMPLE LINEAR REGRESSION ........................................... 13
      2.3.1.2. SUPERVISED, UN-SUPERVISED AND HYBRID METHODS .......... 14
      2.3.1.3. SUPERVISED OR MACHINE-LEARNING METHOD ....................... 14
      2.3.1.4. NAIVE BAYES .................................................................. 15
      2.3.1.5. SUPPORT VECTOR MACHINE ............................................. 15
      2.3.1.6. DECISION TREES ............................................................ 16
      2.3.1.7. UNSUPERVISED LEARNING OR LEXICAL METHOD .................. 17
      2.3.1.8. SENTIMENT LEXICON GENERATION APPROACH ...................... 18
      2.3.1.9. ISSUES OF SENTIMENT LEXICONS ....................................... 18
2.3.2. SUMMARY ............................................................................ 19
2.4. EXAMINING THE USER ............................................................... 19
   2.4.1. SOCIAL CULTURAL FACTORS .............................................. 19
   2.4.2. THE IRISH CULTURAL PROFILE ........................................... 20
   2.4.3. THE RELATIONSHIP BETWEEN CULTURE AND SURVEY RESPONSE STYLES .... 21
   2.4.4. NEEDS OF USER- HEDONIC VERSUS PRAGMATISM ...................... 21
   2.4.5. THE OLDER USER .............................................................. 22
   2.4.6. POPULATION OF OLDER PEOPLE IN IRELAND .......................... 24
   2.4.7. CHARACTERISTICS OF DIFFERENT GENERATIONS OF CUSTOMERS ........ 26
   2.4.8. SUMMARY ........................................................................... 27
2.5. SIMILAR RELATED WORK IN LITERATURE REVIEW- DISCUSSION ...... 28

## CHAPTER 3: EXPERIMENT DESIGN AND METHODOLOGY ...................... 29

3.1. CRISP-DM CROSS INDUSTRY STANDARD PROCESS FOR DATA MINING) ........ 29
3.2. BUSINESS UNDERSTANDING ....................................................... 29
3.3. DATA UNDERSTANDING .............................................................. 30
   3.3.1. DESCRIPTIVE STATISTICS OF VOICE OF CUSTOMER DATA ................ 31
   3.3.2. DIGITAL BANKING JOURNEY DESCRIPTION ................................ 35
   3.3.2.1. IN BRANCH SELF SERVICE OR QUICK BANKING MACHINES .......... 36
   3.3.2.2. INTERNET BANKING ......................................................... 37
CHAPTER 4: IMPLEMENTATION AND RESULTS ........................................... 55

4.1. SUPPORT VECTOR MACHINE MODEL ALGORITHM .............................. 55
4.2. NAÏVE BAYES MODEL ALGORITHM .................................................. 56
4.3. DECISION TREES DISCUSSION .................................................... 57
4.4. IMPLEMENTATION METHODOLOGY ................................................ 58
4.5. RESULTS ANALYSIS ....................................................................... 59

4.5.1. RESULTS TO PREDICT THE NPS SCORE OF A CUSTOMER BASED ON THE TEXTUAL DESCRIPTION ............... 60
4.5.1.1. OVERALL ACCURACY OF SUPPORT VECTOR MACHINE .................................................. 60
4.5.1.2. SVM CONFUSION MATRIX FOR PROMOTER .................................................. 60
4.5.1.3. SVM CONFUSION MATRIX FOR PASSIVE .................................................. 61
4.5.1.4. SVM CONFUSION MATRIX FOR DETRACTORS ........................................... 61
4.5.1.5. SUMMARY OF SVM’S ........................................................................ 62
4.5.2. NAÏVE BAYES ALGORITHM RESULTS TO PREDICT THE NPS SCORE OF A CUSTOMER BASED ON THE TEXTUAL DESCRIPTION .................. 62
4.5.2.1. NAÏVE BAYES CONFUSION MATRIX DETRACTORS, PASSIVE, PROMOTER .................................. 62
4.5.2.2. NAÏVE BAYES SUMMARY .................................................................. 63
4.5.3. RESULTS TO PREDICT THE CUSTOMERS AGE GROUP BASED ON DIGITAL BANKING SENTIMENTS ............... 63
4.5.3.1. SVM GENERATION X, Y, Z, BABY BOOMER, TRADITIONALIST ......................... 63
4.5.3.2. NAÏVE BAYES GENERATION X, Y, Z, BABY BOOMER, TRADITIONALIST ... 64
4.5.3.3. CLASSIFYING GENERATION X, Y, Z, BABY BOOMER, TRADITIONALIST RESULTS 65
4.5.3.3.1. TECHNIQUES TO IMPROVE THE MODEL’S ACCURACY AND LIMITATIONS OF TRAINING DATA SET .................................................. 65
4.5.3.4. PRODUCT FEATURE EXTRACTION FOR AGE FRIENDLY BANKING .......................... 66
4.5.4. RECOMMENDATIONS FOR DIGITAL INCLUSION FOR AGE FRIENDLY BANKING 67
CHAPTER 5: CONCLUSION AND FUTURE WORK ......................................................... 68

5.1. INTRODUCTION .......................................................................................... 68
5.2. RESEARCH OVERVIEW ............................................................................ 68
5.3. PROBLEM DEFINITION ........................................................................... 69
5.4. EXPERIMENTATION, EVALUATION & RESULTS .................................... 70
5.5. CONTRIBUTIONS AND IMPACT .............................................................. 71
5.6. FUTURE WORK & RECOMMENDATIONS .............................................. 72

6. REFERENCES .................................................................................................. 73

7. APPENDIX ...................................................................................................... 77

7.1. APPENDIX A: PRE-PROCESSING SCRIPTS AND SAMPLE OUTPUTS ........ 77
7.2. APPENDIX B: TRAINING AND TEST SCRIPTS AND SAMPLE OUTPUTS .... 81
LIST OF FIGURES

Figure 2.1: Net Promoter Score Scale 0-10 (Internal AIB Document) ........... 8
Figure 2.2: (Kenningham, 2017) Net Promoter results not matching actual
behavior. ............................................................................................................. 9
Figure 2.3: (Mikroyannidis et al., 2006): Polarity of Opinion in Sentiment
Analysis ............................................................................................................. 12
Figure 2.4: (Silverman, 2003) Simple Linear Regression ............................... 13
Figure 2.5: (Jain et al., 2016) Naive Bayes Algorithm ................................. 15
Figure 2.6: (Arhippainen et al., 2003) User experience influences ............... 19
Figure 2.7: (Von Wilamowitz et al., 2006) Hierarchy of UX needs ............... 22
Figure 2.8: (Caprani., 2012) Age related changes that have implications with
technology ......................................................................................................... 23
Figure 2.9: (CSO, 2016) Population of older people in Ireland (360,000 in the
year 2016) ........................................................................................................ 24
Figure 2.10: (CSO, 2016) % of older people in Ireland compared to Europe. .. 25
Figure 2.11: (CSO, 2016) % of older people in Ireland compared to Europe. .. 25

Figure 3.1: (Botzenhardt et al., 2011) Cross Industry Standard Process....... 29
Figure 3.2: Gender of total respondents .......................................................... 32
Figure 3.3: Distribution of total respondents by age ...................................... 32
Figure 3.4: Distribution of total respondents by county in Ireland ............. 33
Figure 3.5: Customer journey’s sampled ....................................................... 34
Figure 3.6: Average NPS score by age .............................................................. 34
Figure 3.7: Depicts 7,534- Digital Banking Customer journeys surveyed ..... 35
Figure 3.8: Depicts 7,534- Digital Banking Customer journey average or mean
NPS scores ......................................................................................................... 35
Figure 3.9 Example verbatim comments from detractors of digital banking
......................................................................................................................... 36
Figure 3.10: Depicts Quickbanking machine available in branches nationwide.
......................................................................................................................... 37
Figure 3.11 In-Branch Self Service Journey NPS and summary reasons for
NPS score .......................................................................................................... 39
Figure 3.12 Free online word cloud generator of In-Branch Self Service
Positive and Negative words ............................................................................. 40
Figure 3.13 Depicts In-Branch Self-service age profile of respondents ...... 41
Figure 3.14 OTC Journey NPS and summary reasons for NPS score ........ 41
Figure 3.15 Free online word cloud generator of OTC positive and negative
words .................................................................................................................. 42
Figure 3.16: Age profile of OTC respondents .................................................. 43
Figure 3.17 Internet Banking NPS and summary reasons for NPS score ...... 43
Figure 3.18 Free online word cloud generator of Internet Banking positive
and negative words ......................................................................................... 44
Figure 3.19 Internet Banking age profile of respondents .............................. 45
Figure 3.20 Mobile Banking NPS and summary reasons for NPS score ...... 45
Figure 3.21 Free online word cloud generator of Mobile Banking positive
and negative words ......................................................................................... 46
Figure 3.22: Age profile of mobile banking respondents ............................... 47
Figure 3.23: All other journeys ....................................................................... 47
Figure 3.24 Age profile of all other journeys respondents ......................... 48
Figure 3.25 (Tharwat, A. 2018) Accuracy measurement ............................. 53
LIST OF TABLES

TABLE 3. 1: VoC Data Set Description .......................................................... 31
TABLE 3. 2: In Branch Self Service Description ........................................... 36
TABLE 3. 3: Internet Banking Service Description ....................................... 37
TABLE 3. 4: Mobile Banking Service Description ......................................... 38
TABLE 3. 5 Calculations By Age of Detractors, Fence Sitters and Promoters of
In Branch Self Service or Quick-banking .................................................. 40
TABLE 3. 6 Calculations of Detractors, Fence Sitters and Promoters of Over
the Counter ................................................................................................. 42
TABLE 3. 7 Calculations of Detractors, Fence Sitters and Promoters of
Internet Banking ....................................................................................... 44
TABLE 3. 8 Calculations of Detractors, Fence Sitters and Promoters of
Mobile Banking Respondents ..................................................................... 46
TABLE 3. 9 Calculations of Detractors, Fence Sitters and Promoters of All
Other Journeys Respondents ...................................................................... 48

TABLE 4. 1 Overall Accuracy Support Vector Machine % .............................. 60
TABLE 4. 2 Confusion Matrix for Promoter .................................................... 60
TABLE 4. 3 Confusion Matrix for Passive ....................................................... 61
TABLE 4. 4 Confusion Matrix for Detractor .................................................... 61
TABLE 4. 5 Naïve Bayes Confusion Matrix ..................................................... 62
TABLE 4. 6 SVM Confusion Matrix generation Z of Digital Banking ............ 63
TABLE 4. 7 SVM Confusion Matrix generation Y of Digital Banking ............ 63
TABLE 4. 8 SVM Confusion Matrix generation X of Digital Banking ............ 64
TABLE 4. 9 SVM Confusion Matrix Baby Boomers of Digital Banking ........ 64
TABLE 4. 10T SVM Confusion Matrix Traditionalist of Digital Banking ....... 64
TABLE 4. 11 Naïve Bayes Confusion Matrix Age Bands of Digital Banking ... 64
TABLE 4. 12 Summary of Frequent Negative Features Extracted From Customer’s
Opinion ....................................................................................................... 66
TABLE 4. 13 Frequent Negative Features Recommendations ........................ 67
**LIST OF ACRONYMS**

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AIB</td>
<td>Allied Irish Bank</td>
</tr>
<tr>
<td>CRISP-DM</td>
<td>Cross Industry Standard Process for Data Mining</td>
</tr>
<tr>
<td>CSO</td>
<td>Central Statistics Office</td>
</tr>
<tr>
<td>DT</td>
<td>Decision Tree</td>
</tr>
<tr>
<td>EDW</td>
<td>Enterprise Data Warehouse</td>
</tr>
<tr>
<td>FN</td>
<td>False Negative</td>
</tr>
<tr>
<td>FP</td>
<td>False Positive</td>
</tr>
<tr>
<td>NLTK</td>
<td>National Language Toolkit</td>
</tr>
<tr>
<td>NB</td>
<td>Naïve Bayes</td>
</tr>
<tr>
<td>NPS</td>
<td>Net Promoter Score</td>
</tr>
<tr>
<td>OTC</td>
<td>Over The Counter</td>
</tr>
<tr>
<td>OVR</td>
<td>One-vs-the-rest</td>
</tr>
<tr>
<td>PoS</td>
<td>Part of Speech</td>
</tr>
<tr>
<td>ROC</td>
<td>Receiver operating characteristics</td>
</tr>
<tr>
<td>SSE</td>
<td>Sum of Squares of Errors</td>
</tr>
<tr>
<td>SVM</td>
<td>Support Vector Machine</td>
</tr>
<tr>
<td>TP</td>
<td>True Positive</td>
</tr>
<tr>
<td>TN</td>
<td>True Negative</td>
</tr>
<tr>
<td>UX</td>
<td>User Experience</td>
</tr>
<tr>
<td>VoC</td>
<td>Voice of Customer</td>
</tr>
</tbody>
</table>
CHAPTER 1-INTRODUCTION

1.1. Background

In AIB, the ‘Voice of Customer’ (VoC) programme has been measuring customers’ experience of banking products through ‘Net Promoter Scores’ (NPS), ranking customer experience from 0-10 as ‘positive’, ‘negative’ or ‘neutral’ and gathering customer sentiments in the form of short opinion statements.

The net promoter score is a management tool designed to gauge the loyalty of a company’s customer relationships. It is determined by surveying customers with one simple question: “How likely is it that you would recommend this service or company to a friend or colleague?” Based on their scores, respondents are grouped into Promoters (score 9-10), Passives (score 7-8) and Detractors (score 0-6).

Those who score 9 and 10 on the scale can be described as promoting a brand. A promoter can be a company’s best advocate and can influence many other potential customers. Those who score 0 to 6 are detracting from a brand. Detractors can be a company’s worst critics and can influence many other current and potential customers away from your company. Passives or fence-sitters, scoring 7 to 8, can be a company’s biggest uncertainty and may influence other current and potential customers away from your company.

Sentiment analysis is a subtask of text mining. The aim of sentiment analysis is to automatically extract knowledge from user-generated contents determining the subjectivity, polarity (positive or negative) and polarity strength (weakly positive, mildly positive, strongly positive, etc.) of a piece of text. Besides direct customer contact, which opportunity is limited and time demanding, specifically textual sources are of great importance to hear the customers’ voice. This has potential to provide a useful insight into a customer’s needs, desires, perceptions, and preferences.

The objective of the ‘Voice of Customer’ surveys in AIB is to use the information gathered to inform the loyalty of a customer and provide input into future designs
thereby closing the gap between customer expectations and the experience offered of the company.

This investigation was undertook because from a cursory observation of NPS results and sentiments of the various customer journeys surveyed in AIB, it appeared that:

1. The older generation customers ‘NPS scores’ and ‘Sentiments’ for digital banking journeys appeared to be low and somewhat negative when compared with other results.

2. In addition, it appeared that the NPS score did not always correlate, i.e. a positive sentiment did not always correlate with a high NPS score and a negative sentiment did not always correlate with a low NPS score.

This research will investigate the user experience of both younger and older customers of AIB’s banking journey’s. It will extract net promoter scores relating to these user groups and investigate the customer sentiments to determine, a) if there is a significant difference in customer dissatisfaction relating to digital banking journeys specifically and b), if there a difference in dissatisfaction, to what extent is it attributed to a customer’s age? In addition to these question, it will investigate, using different sentiment analysis techniques, if the NPS scores provided by the customers, are supported by their positive, negative and neutral sentiments as expressed in their feedback documents.

1.2. Research Problem

AIB is realising that the best way to deliver banking products is digitally, wherever and whenever a customer needs the utility of a Bank. Because AIB can sell products and services remotely, the digital transformation of banking has led to a decline in AIB branches in Ireland. The branches that do remain do not look anything like the transactional, teller spaces of the past. Tellers have been replaced by machines and branches have bright lights, music, colour, flashing images, advertising for products targeted at Generation X, Y and Z.
Wiedmer, (2015) defines the following five age groups into cohorts of customers with certain attributes. These attributes, which include social and cultural factors will be discussed in a later chapter.

- Traditionalists are customers born 1900-1945 and comprise the oldest generation of customers.
- Baby Boomers are customers born in a baby boom following World War II, 1946-1964.
- Generation X are born after the Western Post-World War II Baby Boom; birth dates range from 1961-1981.
- Generation Y is also referred to as Gen Y, Echo Boomers, Millennials or Millennials, Generation We, Internet Generation, Connect 24/7, and Leave No One Behind. Born 1980-1990s.
- Generation Z is the latest generation and will be leading the world in several decades and are born 1995-2015.

The digital transformation of banking may make sense from a commercial point of view but it may create discomfort for many baby boomers and traditionalists who may prefer the experience of physically meeting a bank official to carry out their transaction or deal with their problem in person.

This research will explore through appropriate analysis techniques the ‘Voice of Customer’ for AIB’s older customers (baby boomers and traditionalists) and determine if they have specific usability problems of AIB’s digital banking.

In order to remain competitive, a business must continuously improve product quality and performance for all customer cohorts including older generation customers. Despite the brighter features, speed and convenience of AIB’s digital banking, it may trigger many complexities and usability issues for the over older customers that warrants research attention.

The overall research will focus on applying text mining methods for customer voice extraction as input for improving customer satisfaction for older generation customers.
of digital banking. The following research questions are key drivers for the research work presented in this paper:

**RQ1** “Can text mining provide useful insights into data available to the bank through ‘Voice of Customer’ surveys?”

**RQ2** “To what extent is age a meaningful factor in influencing customer sentiments and NPS scores of AIB’s digital banking?”

### 1.3. Research Aim

The aim of the research is to explore the usefulness of sentiment analysis to inform possible product and service design affecting satisfaction of older customers using digital banking. The research findings will point to areas of interest for improving systems with regard to meeting older customers’ expectations. If the research suggests that AIB’s digital banking’s performance is sub-optimal at meeting older customers’ needs and wants, then recommendations on how to improve these systems will be made. In addition, explore if negative features extracted from the negative feedback could be utilised to inform how the needs and desires of their older generation customers are met.

### 1.4. Hypothesis

H1: Sentiment Analysis has the potential to usefully classify customer text-based feedback into positive, negative or neutral sentiments which, in turn, can inform customer satisfaction improvement initiatives for AIB’s digital banking journeys.

### 1.5. Research Objectives

#### 1.5.1. Main Objective

The main objective of this research is to assess the usefulness of machine learning methods including text mining and decision trees, to determine if automation can be applied to customer feedback collections to provide useful insights which may help AIB improve customer experience of the products offered.
1.5.2. Specific Objectives

In relationship to the problem statement, this research aims to:

- To assess if text mining and other machine learning methods are comparable to NPS scoring and could act as a more cost efficient way of assessing what the customer likes and dislikes using the bank’s digital journeys.
- To find features in the data set that warrant further attention in order to improves AIB’s product offering to assist its older customers.

1.6. Research Methodologies

The research conducted in this project is secondary as it relies on a data set collected by AIB’s ‘Voice of Customer’ programme. The research will be conducted by using quantitative research methods. Carr, (1994) explains that quantitative methodologies test theory deductively from existing knowledge, with a view to explaining the hypothesized relationships and proposed outcomes for subject of study. Quantitative research is used to quantify the problem by way of generating numerical data or data that can be transformed into usable statistics. It is used to quantify attitudes, opinions, behaviours, and other defined variables. For the purpose of this research, this will be achieved by conducting data analysis of the AIB ‘Voice of Customer’ programme data set in order to identify meaningful insights which could lead to actionable recommendations. Quantitative data collection methods include various forms of surveys. In this instance data is gathered systematically by AIB through online surveys from customers across fourteen customer journeys. More specifically, the sample comprises 70,000 AIB service users across different age groups and demographics. Customer sentiments will be analysed and categorised according to user characteristics, social and cultural variables and any other relations in the data identified.

The research will be deductive from the more general to the more specific, the top down approach. The research will begin with a general hypothesis and work down to a specific conclusion by examining the evidence gathered. The assumption is that older generation face greater difficulties using digital banking in AIB in relation to other age groups and this results in negative sentiments and low NPS scores. This study will aim to provide evidence of this, deduce that this is the case and restate the premise.
Secondary research is ongoing to complete a comprehensive literature review of the previous research already completed on Sentiment Analysis, Voice of Customer and NPS Scores and User characteristics, specifically of older generation customers. By conducting an extensive literature review using a number of sources, it will play an important role in getting an insight into the topics related to this research.

1.7. **Scope and Limitations**

The scope of this work is strictly limited to the use of the ‘Voice of Customer’ data set and is limited in size due to the nature of the research question. Although there are 70,000 records in the full data set, the focus of this research is to compare the full data set with a sub set of 7,534 returned surveys for specifically digital banking to determine if there are differences in results. Popular use cases, when conducting sentiment analysis, text mine large data sets such as Twitter or IMDB movie reviews for user opinions. This research is not based on such a social media data set. It is a real business problem with a finite set of customers who are not all going to respond to the survey. In the case of ‘Voice of Customer’ surveys for AIB, the response rate is typically 30%.

It should be noted that for the machine learning sentiment analysis method in this project, the modelling is performed on pre-labelled data set, obtained from the ‘Voice of Customer’ project in AIB. The labelled data is classified into promoter, detractor or passive based on the NPS score as calculated. Therefore the accuracy of the results is dependent on the quality of the labelling done by humans. In addition, freeware tools such as Python libraries like NLTK and SCIKIT learn were also utilised for this experiment. The libraries are not domain specific and this could have a bearing on quality of results generated. In summary, the accuracy of the results obtained depend to a large extent on the accuracy of the pre-labelling of the data and the tools used to perform the task.

Future studies should consider to utilise broader data sets such as twitter feeds of AIB in general. Benefits of this approach would include the larger size of the data set and reach to different types of customers, limiting certain bias in the surveys, which is discussed in a later chapter.
1.8. Document Outline

- Chapter 2 – Literature Review: This chapter provides a critical review in which the various papers on ‘Voice of Customer’ and ‘Net Promoter Scores’, ‘Sentiment Analysis’ and examining ‘User Characteristics’ of various generations of customers are compared. A summary will be provided at the end of each review describing the background/context of the research. It will also highlight the state-of-the-art techniques and methods used as well as their strengths, weaknesses while also pointing the way forward.

- Chapter 3 - Experiment Design and Methodology: This chapter summarizes the project approach in terms of design, experimental set-up, and describes the steps taken and the execution. Specifically, this chapter provides the data description step, the data exploration step as well as the preparation and pre-processing steps used to conduct the experiment. It also points out the relevant data quality issues. This chapter includes the methodology adopted for designing the solution and for evaluating it (e.g. errors, performance measures, accuracy)

- Chapter 4 – Implementation and Results: This chapter describes the model implementation and results. More precisely, the implementation and results of popular sentiment analyses methods are presented in this chapter. It will describe the actual work carried out. It will focus on the presentation and discussion of the findings in light of what is already known from the literature conducted in chapter 2. It will show how our findings confirm or refute the research hypothesis and how they differ with previous work in the literature.

- Chapter 5 - Conclusion: This chapter will provide an account of the results achieved and any insights gained, emphasising mainly what is new in relation to the problem addressed. It will also state the limitations of the research at this stage and suggest what further work might be done as well as providing a self-assessment of the research.
CHAPTER 2-LITERATURE REVIEW

2.1. Introduction

This chapter provides a critical, in-depth evaluation of research already undertaken on ‘Voice of Customer’ and ‘Net Promoter Scores’, ‘Sentiment Analysis’ and ‘User Characteristics’ It states where the weaknesses and gaps are, contrasting the views of particular authors, or raising questions. It will evaluate and show relationships, so that key themes emerge.

2.2. Voice of the Customer and Net Promoter Scores

Zaki et al, (2016) explains that the single-question customer metric has become very popular as a tool to measure customer loyalty. Many organizations use the Net Promoter Score (NPS) asking a single question “How likely is it that you would recommend this service to a friend or colleague?” to assess overall customer loyalty based on the customer’s intention to refer others. Customers provide an answer on an 11-point rating scale, ranging from 0 (not at all likely) to 10 (extremely likely). Customers are then grouped based on their chosen ratings into three segments: promoters, fence sitters or passives, and detractors. Promoters are defined as customers having the highest referral intention, with a rating of 9 or 10, Fence sitters or passives with ratings of 7 or 8 and detractors ranging from 0 to 6 on the NPS scale.

![Net Promoter Score Scale 0-10 (Internal AIB document)](image)

Figure 2.1: Net Promoter Score Scale 0-10 (Internal AIB document)
Although the NPS measure can be used as a loyalty indicator, it does not offer an explanation of the root causes of a low score. Relying solely on a simple single customer metric is risky and so companies are encouraged to adopt a more multidimensional approach to better predict customer behaviour. Text-mining for example can determine the complaint status and emotions of each customer using text-mining textual feedback that divides customers into groups of complainers, neutral/fence sitters or satisfied. The integration of the verbatim comments helps firms to understand why customers are leaving or inactive. (Zaki et al., 2016)

There is criticism of using a single metric to measure customer loyalty. This is evident in the literature by Aksoy, (2013) and Kenningham et al, (2007) who question the claim of Zaki et al, (2016) that the Net Promoter Score (NPS) is the “single most reliable indicator of a company’s ability to grow”. In Kenningham et al, (2007) the study found no real indication that average levels of attitudinal loyalty metrics significantly correlate with the relative change in revenue. Furthermore, single metrics alone cannot predict customer loyalty and consequently are unlikely to deliver actions. The paper concludes that customer’s loyalty-based behaviours are multidimensional and therefore other evaluation methods are required such as sentiment analysis.

Figure 2.2 Kenningham et al. (2007) shows the results of their study concluding that NPS scores did not always match actual customer behaviour.

![Perceived NPS vs Actual Behavior](image)

Figure 2.2: (Kenningham, 2017) Net Promoter results not matching actual behavior.
2.2.1. *Voice of Customer Survey’s through the medium of email*

Many organisations choose to deliver their ‘Voice of Customer’ survey online through the medium of email. Wright, (2017) explains that surveys though email is a fruitful medium for conducting survey research. As the cost of computer hardware and software continues to decrease, and the popularity of the Internet increases, more segments of society are using the internet and email for communication and information. Other advantages include:

- Access to groups and individuals who are difficult, if not impossible, to reach through other channels.
- Reach thousands of people with common characteristics in a short amount of time.
- Use of online surveys eliminates the need for paper and other costs, such as those incurred through postage, printing, and data entry.

Wright, (2017) also points out the disadvantages associated with online survey research which includes:

- Sampling issues such as inaccurate demographic or characteristics information or not all customers allow their email addresses to be listed. This makes accurately sizing an online population difficult.
- Problems such as multiple email addresses for the same person, and invalid/inactive email addresses makes for a problematic method of delivery in some circumstances.
- Response rates in email surveys are equal to or better than those for traditional mailed surveys. In AIB, the ‘Voice of Customer’ survey response rate is at 30%.
- Self-selection bias is another major limitation of online survey research. Undoubtedly some individuals are more likely than others to complete an online survey. Specifically detractors of a service who may feel that they have been let down maybe more inclined to respond. A bias that was not explored in the literature was the traits of people from different countries. Specifically, in Ireland for example, there is a perception that as nation people do not complain normally. There are a lot of skits on social media of Irish customers not enjoying a meal or
experiencing terrible customer service, but when the waiter asks how was the service, the Irish person responds that it was lovely! This may skew the survey results somewhat and is worth investigating. Other social and cultural factors may also contribute to erroneous NPS scoring. This will be discussed later in this chapter.

2.2.2. **Summary**

What is evident from the literature research is that the NPS score alone, may not be sufficiently strong enough to accurately measure customer loyalty. If organisations want to understand their customers, the answer could come from the verbatim comments provided. Survey as a medium to extract this information is very popular, but bias may exist. One reason for bias may be due to the perception of Irish people and the tendency not to complain. Other factors such as age and social cultural characteristics explored later in this chapter may also result in bias. Potentially due to these factors, customers may rank a service high unintentionally. Consequently, the NPS score considers the customer as loyal. Without exploring the sentiment, potentially vulnerable customers requiring intervention strategies will not be identified.

In addition, the response rates for the AIB VoC survey is typically 30%. Although better than traditional postal surveys, it makes accurately analysing the entire targeted population impossible.

2.3. **Sentiment Analysis**

Mikroyannidis et al. (2006) explains that sentiment analysis is a subset of text mining which is the process of analysing textual information in an attempt to discover structure and implicit meanings “hidden” within texts. It is a relatively recent technological development using techniques from data mining, machine learning and natural language processing.

The difference between regular data mining and text mining is that in text mining the patterns are extracted from natural language text rather than from structured databases of facts.

Wilson, (2005) explains that sentiment analysis or opinion mining is the process of identifying and extracting the polarity of positive, neutral and negative opinions, emotions, and evaluations from a piece of text. Sentiment analysis not only tells a
company whether they like the product or not, it can also tell how to make it better for them.

The available data from customers reviewed over the internet is growing exponentially. Huge amounts of texts expressing opinions are available on review sites, forums, blogs and social media. Sentiment analysis can help craft this unstructured information into structured data on public opinions on companies, products and brands. It is very useful for gaining insight into customer opinion. Once a company understands how a customer feels, it can identify what they like and dislike, and the company can build a recommendation system and have more targeted market campaign for their customers.

Sentiment analysis can be applied at different levels of scope: (Pang et al, 2008)

- Document level sentiment analysis obtains the sentiment of a complete document or paragraph.
- Sentence level sentiment analysis obtains the sentiment of a single sentence.
- Sub-sentence level sentiment analysis obtains the sentiment of sub-expressions within a sentence.
2.3.1. Exploration of Methodologies suitable for Sentiment Analysis

2.3.1.1. Simple Linear Regression

According to Silverman, (2003) the purpose of simple regression analysis is to evaluate the relative impact of a predictor variable on a particular outcome. This is different from a correlation analysis, where the purpose is to examine the strength and direction of the relationship between two random variables.

Regression allows us to model mathematically the relationship between two or more variables using algebra. The goal of simple linear regression is to create a linear model that minimizes the sum of squares of the residuals/errors (SSE). Simple Linear regression is a comparison of two models. One is where the independent variable does not even exist and the other uses best fit regression line. If there is only one variable, the best prediction for other values is called the residual or error. The residuals are squared and then added together to generate sum of squares or SSE. Simple linear regression is designed to find the best fitting line through the data that minimizes the SSE.

An option for this business problem is to define a dependent and independent variable and build a model to determine:

1. Is there any relationship between age and NPS score?
2. Is there any relationship between NPS score and sentiments?

By visualising the existing data on X and Y axis the NPS score can be used for predicting the positive, neutral or negative sentiments by using the mean of all NPS scores for
digital customer journeys. The best prediction or estimate would be the mean or best fit line for each NPS score for each age group.

2.3.1.2. **Supervised, Un-Supervised and Hybrid methods**

According to Gonçalves *et al.*, (2013) there are three types of methods of Sentiment Analysis 1. Supervised or Machine-Learning method. 2. Unsupervised Lexical-based method or 3: Hybrid Analysis method.

2.3.1.3. **Supervised or Machine-Learning method**

A Supervised or Machine-Learning method does not rely on manually created rules but on machine learning. By modelling as a classification problem, a given algorithm is fed text and returns a polarity of sentiment. The main advantage of this learning-based method is the ability to adapt and create trained models for a specific purpose or domain. The main challenge of the supervised approach is that it requires sentiments to be labelled either positive, negative or neutral. The labelled data is required to train the model.

The challenge of acquiring labelled data means the low applicability of the method on new data. Labels can be manually created by humans, but this is time consuming and costly. A method to manually label data includes crowdsourcing for example. The quality and consistency of the labelled data is also a consideration. According to the literature, accuracy of this method is 60%-80%. (Jyoti *et al.*, 2016)

The first step in a machine learning text classifier is to transform the text into a numerical representation, usually a vector. This process is known as feature extraction or text vectorization.

In the training phase, negative and positives sentiments created from a Bag of Words will be given to the machine along with their polarity labelled as input. The machine learns how to distinguish between polarities with the help of the training data. For the testing phase, the rest of the unseen data is then given to the machine. With the help of previous experiences, the machine can now determine the polarity of the unseen test data, thus following the supervised learning pattern. (Asghar *et al.*, 2018)
An early study on sentiment analysis was done by (Pang et al, 2008) on the classification of movie reviews data set. This study was interested in polarity detection of movie reviews in predicting IMDB reviews as positive, negative or neutral.

The first step in their study was to use an approach most frequently found in text mining systems called tokenization. This involves breaking the natural language into keywords, phrases, symbols and other elements called tokens. Characters like punctuation marks were discarded and the tokens become the input for text mining. The different classifiers used in their study included Naïve Bayes, Support Vectors machines and Decision trees, details of which are illustrated below.

2.3.1.4. Naïve Bayes

Naive Bayes classifier is especially popular for text classification. The algorithm predicts a class, given a set of features using probability. According to McCallum et al., (1998) the Naïve Bayes algorithm is a part of supervised learning process based on the Bayes theorem. It holds an assumption of independence among predictors. Jain et al, (2016) states that Naïve Bayes classifier assumes that the presence of a particular feature in a class is unrelated to the presence of any other feature. Naive Bayes model is easy to build and particularly useful for very large data sets. Along with simplicity, Naïve Bayes is known to outperform even highly sophisticated classification methods.

Jain et al, (2016) states that Naïve Bayes algorithm has an accuracy of 83%.

\[
P(A|B) = \frac{P(B|A)P(A)}{P(B)}
\]

Figure 2.5: (Jain et al, 2016) Naive Bayes Algorithm

2.3.1.5. Support Vector Machine

Joachims, (1998) explores the use of Support Vector Machines (SVMs) for learning text classifiers from examples. Support Vector Machine (SVM) is a type of supervised machine learning algorithm with high generalisation ability used for both classifications and regression. SVM takes a set of input data and predicts, for each given input, which of two possible classes forms the output. For classification, SVM performs classes by
finding the separating hyperplane between two classes in a high dimension feature space that maximises the separation margin between the two classes. The vectors that define the hyperplane are the support vectors. (Joachims, 1998)

Below illustrates this distance from the decision boundary to the nearest training instance. This is known as the margin. The dashed lines on either side of the decision boundary show the extent of the margin, and are known as the margin extents.

Figure 2.6: (Joachims, 1998) Distance boundary in SVN

2.3.1.6. **Decision Trees.**

Decision Trees is a supervised machine learning method for producing classifiers from data. It produces a flow-chart-like tree structure, where each node denotes a test on an attribute value, each branch represents an outcome of the test and tree leaves represent classes. In addition, it is used in determining the optimum course of action, in situations having several possible alternatives with uncertain outcomes. The algorithm begins with tuples in the training set, selecting best attribute yielding maximum information for classification. It generates a test node for this and then a top down decision trees induction divides current tuples set according to current test attribute values. Classifier generation stops when all subset tuples belong to the same class or if it is not worthy to proceed with additional separation to further subsets.

Tree pruning identifies and removes subtrees within a decision tree that are likely to be due to noise and sample variance in the training set used to induce it. (Jotheeswaran *et al*, 2013)
2.3.1.7. **Unsupervised Learning or Lexical method**

Unsupervised learning is a lexical or rule-based method making use of a predefined list of words, where each word is associated with a specific sentiment. These manually created rules identify subjectivity and polarity utilising a lexicon dictionary. The lexicon dictionary contains positive and negative opinion words and expressions with attached sentiment values. Numerical scores/ratings, probabilities and ranks. Sentiments tend to be ranked between 0 and 1. For example a 0.9 is positive and 0.3 is negative. Asghar *et al*, (2018) states that the accuracy of this method is 80%.

There are numerous popular free to use lexical libraries such as the SentiWordNet lexicon which is described by Gonçalves *et al* (2013) as a tool that is widely used in opinion mining, and is based on an English lexical dictionary. SentiWordNet associates three scores with a synset from the WordNet dictionary to indicate the sentiment of the text: positive, negative, and objective (neutral). It works by counting the number of negative and positive words, based on their membership of a synset, that appear in the text. If the number of positive words is greater than the number of negative words, it returns a positive sentiment, else if the number of negative words is greater, it returns a negative sentiment. If it’s the same amount it returns a neutral sentiment. (Gonçalves *et al*, 2013)

Other popular dictionaries include SentiStrength, Opinion Finder lexicon, AFINN lexicon for emotional ratings such as NCR lexicon which has been created by crowd sourcing tagging. Domain knowledge based lexicons can also be used in feature extraction. For example there are domain based lexicons for Healthcare, which are more focused for matching the words or phrases for this particular domain. (Thelwall, 2017)

An example of a how a rule-based implementation works is where two lists of polarized words either positive or negative in a Sentiment lexicon is given text. The number of positive and negative words that appear in the text is counted. If the number of positive word appearances is greater than the number of negative word appearances return a positive sentiment, conversely, return a negative sentiment. Otherwise, return neutral. (Gonçalves *et al*, 2013)
In summary, lexical analysis uses a dictionary of pre-tagged lexicons. Using tokenization, matches are found in the library and scores with corresponding values of the sentiment word. The accuracy of this method is 80% (Gonçalves et al 2013)

2.3.1.8. **Sentiment Lexicon generation approach:**

Sentiment lexicons are generated and maintained through manual or automated approaches. Manual lexicons require experts such as linguists, social scientists and psychologists which evolve over the years and can be costly and time consuming. Recently lexicons are automatically generated through crowd sourcing or online labour markets, where the crowd is asked to help tag or generate the text for some particular corpus. Sentiment lexicon’s can be also tailored to certain domains. This is useful because for different domains, often words have different meanings. For example camera reviews, the word ‘long’ is positive because a sentiment may include a statement that ‘the battery life is long’ and in this instance the word ‘long’ is positive. But this would not be the case for restaurant reviews, because the word ‘long’ can be deemed negative, because for example ‘the waiting times are long’. (Velikovich et al, 2010)

2.3.1.9. **Issues of Sentiment Lexicons**

- Lexicons require regular maintenance. The user needs to be sure it is up to date, otherwise the performance of using sentiment lexicons in different types of text processing could be compromised.
- Sentiments can vary by domains, as per the example above, the word ‘long wait’ versus ‘long battery’
- Sentiments words don’t always imply sentiments, like the sentence ‘who is mad?’ This sentence does not imply a negative sentiment. But because the word mad is in the sentence, the lexicon will imply it as negative sentiment.
- Sarcastic sentences for example “can the battery life be longer than 2 days? This is making fun of the poor battery life. But the lexicon will pull out the word ‘long’ as a positive for battery life.
- Non-sentiment words, where none of the words in a sentiment express a sentiment. Where a sentence is not straight forward but there is a sentiment, but the lexicon cannot pick it up. (Velikovich et al, 2010)
2.3.2. Summary

This work has covered a wide range of research on sentiment analysis and concluded that the experiment will be based on the popular techniques described by Jain et al., (2016) to build a Supervised Machine Learning model. Machine learning based method is deemed the more accurate method, but requires pre-labelled sentiment training data sets which are sometimes costly to produce and hard to maintain. For this experiment, the training data is pre-labelled Promoter, Detractor or Passive as implied from the respondent NPS scores. Most common supervised techniques employed for sentiment classification includes Support Vector Machine and Naive Bayes and in some studies Decision Trees. An accuracy ranging from 60% to 80% is observed for classification using these supervised techniques. (Gonçalves et al, 2013)

2.4. Examining the User

2.4.1. Social Cultural Factors

There are numerous different kinds of people, products and environments that influence the experience and the response to a survey on customer experiences. Positive responses depend on the context, the user’s mood, social factors and cultural factors. Users may feel surprise, delight, fun and playfulness when using Digital Banking products. Alternatively they may feel overwhelmed, confused or irritated with too much digital in their environment. The below details interaction in the context of use, including social and cultural factors dimensions. (Arhippainen, 2003).

Figure 2. 6: (Arhippainen et al, 2003) User experience influences

2.4.2. The Irish Cultural Profile

Hofstede’s (2001) original four cultural dimensions of power distance, uncertainty avoidance, individualism-collectivism, and masculinity-femininity dimensions can be applied to the Irish people. The correlation between these dimensions and testing method such as surveys to extract the user experience is explored below. According to Hofstede’s (2001) high power distance countries are societies where users understand their place and accept unequal distribution of power. Ireland ranks low on this dimension. Countries with a low power distance are societies where members do not accept situations where power is distributed unequally.

The dimension of masculine countries versus feminine is equally interesting. In masculine countries, users are more assertive, as opposed to feminine countries where modesty is perceived as virtue. A high score (masculine) on this dimension indicates that the society will be driven by competition, achievement and success. A low score (feminine) on the dimension means that the dominant values in society are caring for others and quality of life. A feminine society is one where quality of life is the sign of success and standing out from the crowd is not admirable.

Ireland is described as a masculine society – highly success oriented and driven. The Irish are proud of their successes and achievements in life. Conflicts are resolved at the individual level and the goal is to win. The dimension uncertainty avoidance has to do with the way that a society deals with the fact that the future can never be known. This ambiguity brings with it anxiety and different cultures have learnt to deal with this anxiety in different ways. Ireland has a low score on uncertainty avoidance. In individualist societies people are supposed to look after themselves and their direct family only. In collectivist society’s people belong to ‘in groups’ that take care of them
in exchange for loyalty. Ireland is an individualist culture. Hofstede’s (2001) cultural dimensions in relation to the Irish people will be taken in consideration when analysing the results of the experiment. The rationale behind certain outcomes maybe explained by these factors for certain trends found in the data as explained below.

2.4.3. The relationship between culture and survey response styles

Johnson, (2005) study of relationship between culture and response biases analysed nineteen different countries. Results found that there was a greater tendency of some cultures to select the endpoints of a response scale when answering questions. The associations between culture and extreme responding behaviour can be correlated with Hofstede’s cultural dimensions. Respondents from more individualistic nations were less likely to provide agreeable answers. Agreeable answers was greater among respondents within less uncertainty avoidant countries. In addition, agreeable answering of surveys behaviour was lower among individuals in more masculine cultures. Because countries high in masculinity value decisive, assertive, and daring action (Hofstede, 2001). Of additional interest is the negative association between national power distance and favourable answering of surveys which was a finding that is contrary to Johnson (2005) original hypothesis.

2.4.4. Needs of User- Hedonic Versus Pragmatism

Fulfilment of self-actualisation needs has a significant impact on continuance intention and positive or negative customer experience. (Cao et al., 2013)

In the hierarchy of UX needs, hedonic qualities such as pleasure are perceived to be at the highest level following by ease of use, reliability, functionality and purpose.

Von Wilamowitz et al., (2006) states that time is an important factor that may disrupt this hierarchy. Hedonic qualities are the psychological needs and emotional experience of the user. The research shows pragmatism as a contributing factor to usability improvements over time. Pragmatism relates to practicality and functionality. This is driven by familiarity and a better understanding of the product. Hedonic needs lessen over time as the product loses its wow factor and beauty. These factors including the
hierarchy of user experience needs may be considered as a cause for negative sentiments and low NPS scoring which will be explored in a later chapter.

Figure 2. 7: (Von Wilamowitz et al., 2006) Hierarchy of UX needs

2.4.5. The Older User

Caprani, (2012) explains that age classification is not a straightforward process as many variables impact the rate at which a person ages. Chronological age markers are the most common measurement used, for example, determining the age of retirement. However this does not take into account biological, psychological and social factors. Usability of digital computing in particular to older people is a gap that is widening. Frequently applications are designed without considering the specific characteristics and requirements of older users. There are lots of older people who can and do use Digital Banking. There are some older users that are not really separate from mainstream users, some who have fairly mild disability in one or more areas and who can cope with digital banking but there is a third grouping of older users who are sufficiently disabled to make learning or use of digital banking too hard.

The characteristics of the older generation will be considered in this experiment and evaluation of the negative sentiments, these include perceptual, psychomotor, cognitive and physical changes (see figure 2.8 for summary).
For vision and hearing, for example all men over the age of 65 and 30% of women suffer hearing loss, and most people notice visual problems around the age of 40 (Fisk et al., 2009). Cognitive functioning can be seen to stem from the reduction of cognitive resources available, impairing older adults’ ability to carry out cognitively demanding processes. Many systems rely on a person’s ability to keep information active however this is unrealistic for older users unless they are proficient users.

Older adults show changes in their physical abilities, due to loss of muscle mass and flexibility. Physical problems can also occur as a result of accidents and falls, creating difficulties grasping a mouse, and positioning and controlling the cursor. Arning et al., (2007) argue that it is necessary to focus on this user group as the number of senior citizens in western societies is growing rapidly. A gap in the research is that it would have been useful to explore further the significance of this user group. For example statistics on what share of the market is elderly (65 or over), how economically active they are? How many years’ are men and women at the age of 65 expected to live? This would give the reader a better perspective on the value of this user group as a target market. Carr et al., (2013) provides these important data insights into the aging population and predicts that with the declining birth rate and increased longevity, by 2036 the number of seniors will be greater than the number of children in Canada. This demographic shift is not only in Canada but worldwide. It will require more thoughtful
design of digital technology to better meet the needs of older persons. In the UK for example, Asmi et al (2012) states that there is a digital divide amongst older users for online banking and there is more than 9.2 million users still perceived resistant to getting involved in the digital age. Hawthorn, (2002) argues that the realities of designing for older users mean that designers cannot simply rely on guidelines but need to involve older users in the design process. In the below sections, descriptive statistics on the older generation in Ireland are taken from the Central Statistics Office. (2016). The size of the aging population in Ireland is significant enough for companies including AIB to ignore.

2.4.6. Population of Older People in Ireland

![Population of Older People in Ireland](image)

Figure 2. 9: (CSO, 2016) Population of older people in Ireland (360,000 in the year 2016)

CSO findings predict that the number of old persons (65 years and over) will almost double in every region by 2050 with the most marked increases likely to occur in the Mid-East (+136.5%) and Midland (+95.1%) regions.
Figure 2. 10: (CSO, 2016) % of Older people in Ireland compared to Europe.

Figure 2. 11: (CSO, 2016) % of Older people in Ireland compared to Europe.
2.4.7. Characteristics of different generations of customers

There are five generations that are currently customers of AIB which include
Traditionalists; Baby Boomers; and Generations X, Y, and Z.

Traditionalists
Traditionalists are AIB’s oldest customers born 1900-1945. According to Kane (2015),
traditionalists respect authority and possess family values that keep their work and
family lives separate. Generally motivated by money and position like their baby boomer
children, traditionalists take pride in being self-sacrificing and thrifty. They tend to work
hard from a sense of pride and determination, consider debt or obligation to be
embarrassing, and acknowledge that change comes slowly.

Baby Boomers
Heathfield, (2015) defined baby boomers (boomers) as the generation of customers born
in a baby boom following World War II, 1946-1964. Typically, boomers were the first
to be educated in their families, and their education translated into upward mobility.
Money, power, and recognition are boomers’ primary motivators.
Kane (2015) states that boomers have had good health, constitute the wealthiest
generation, and optimistically view the world as improving over time. Today, the oldest
boomers in AIB are considering their retirement options and are seeking ways and
opportunities to make their elder years personally meaningful.

Generations X
Is the generation born after the Western Post-World War II Baby Boom; birth dates
range from 1961-1981.
Schroer, (2015) noted that they are sometimes referred to as the “lost” generation—the
first generation of kids exposed to day-care and divorce. Gen X, for the most part, had
parents who were boomers and, as children, Gen X often experienced broken families
and absentee parents. Schroer, (2015) noted that Gen X individuals form families with
higher levels of caution and pragmatism than their parents demonstrated.
Gen X are engaging in financial planning, avoiding broken homes, and ensuring that
children grow up with a parent available.
Generation Y.

Born 1980-1990s, Generation Y is also referred to as Gen Y, Echo Boomers, Millennials or Millennials, Generation We, Internet Generation, Connect 24/7, and Leave No One Behind. (Schroer, 2015).

This generation is web savvy and highly connected via social media. They have been raised at a time of constant information about world events as they connect with technology, including computers, the Internet and mobile phones. Gen Y are attracted to companies and businesses that embrace technological advancements that have changed the way of doing business globally. (Gibson, 2015)

Generation Z.

Generation Z is the latest generation and will be leading the world in several decades. Schroer (2015) stated that Generation Z members were born 1995-2015. Traits that defines this generation include customers who are highly connected to having the lifelong use of communication and technology such as the World Wide Web, instant messaging, text messaging, mobile phones, Tablets and pervasive computing wearable devices. Such technological connections have earned this generation the name “digital natives” (Horovitz, 2012).

Cross-Bystrum (2015) noted that, thanks to Facebook, Skype etc. Gen Z are communicating in a “real” way with individuals who do not occupy the same physical space. This type of communication across the country or around the world includes constant updates, texts, and real-time dialogue with visual connections and interactions.

2.4.8. Summary

What is evident from the literature review is that in there is a community of different stakeholders within AIB with a variety of ability and needs. Differing characteristics and social cultural factors will impact how a user navigates the same banking processes. Some generations may have greater resilience in coping with technology, some may struggle as they age with cognitive decline. The aging population in Ireland is a sector significant in size which AIB cannot afford to ignore.

Feedback from different customers on their user experience is dependent on the same characteristics and cultural dimensions as described. Potential bias in the survey
responses must be a consideration of this study. The different user characteristics may skew the sentiments and NPS scoring from the different generations and this is something to be aware of when analysing the results of the experiment.

2.5. Similar related work in literature review - discussion

Sentiment analysis in general and its application in the financial domain in particular gained a lot of interest within the last decade. Unfortunately none of the mentioned papers in this literature review focuses specifically on the ‘Voice of Customer’ surveys of the older customer of digital banking. A lot of similar studies on sentiment analysis in banking industry use mainly annual reports, but also news stories or calls transcripts as their textual data sources. (Kogan et al. 2009)

One paper, Baumann et al., (2012) did study the factors that help to explain customers’ behavioural intentions, in the retail banking domain, but not specifically digital banking. This study did use the NPS score and text mining of sentiments to investigate the associations between customer satisfaction, perceived service quality, recent and current consumer behaviour and long-term intentions to remain a customer. The study found from their analysis that customer perceptions, market conditions and some customer characteristics are unique predictors of behavioural intentions.

Most of the authors reviewed have the same approaches for extracting sentiment from text and using similar methods for pre-processing of the data. This generally involves tokenization and removing nonessential stop words, spacing, punctuations, stemming and vectorization of words to numeric in which the model can understand. In addition they employ same machine learning algorithms, such as Naïve Bayes, Support Vector Machines and in some studies Decision Trees.

For labelling of the data, the related studies used freely available pre-labelled data, such as twitter feeds and IMDB reviews and other studies used crowd sourcing to label the data. The labelled data are then used for training machine learning algorithms.

Since the focus of this work lies on digital banking and the voice of the older customer, characteristics not employed by related studies will be investigated in the selected text as potential reasons for positive, negative or neutral returns in sentiments.
CHAPTER 3: EXPERIMENT DESIGN and METHODOLOGY

3.1. CRISP –DM Cross Industry Standard Process for Data Mining)

This section describes general structure of the study. The Cross industry standard process will be used to segment this text mining process into several phases: business understanding, data understanding, data preparation, modelling, evaluation and deployment. This research will use a similar approach used in (Botzenhardt et al, 2011).

![Cross Industry Standard Process Diagram]

Figure 3.1: (Botzenhardt et al, 2011) Cross Industry Standard Process

3.2. Business Understanding

The business problem is to determine from a database of customers, which cohort of customers have negative feelings and low NPS scores when using digital banking customer journeys in AIB. The goal is to use text mining methods to investigate the negative sentiments of AIB’s ‘Voice of Customer’ programme and determine whether age is a contributing factor to low net promoter scores and negative customer sentiments. The overall business objectives of this project include the potential to:

- Determine if particular customer segments feel more strongly about digital banking.
• Understand the differences between customer experiences of different age groups, along with some indication of why and how these differences exist.
• Use results of sentiment analysis to design better informed questions to ask on future surveys.
• Identifying if there are types of customers who feel strongly negatively towards the product or service.

3.3. **Data Understanding**

The data understanding phase starts with an initial data collection and proceeds with activities in order to get familiar with the data, to identify data quality problems, to discover first insights into the data. For data acquisition, gain access to the AIB ‘Voice of Customer’ programme data set from AIB’s Enterprise Data Warehouse (EDW). Data is gathered daily by AIB through surveys from customers across banking customer journeys. Data is integrated from multiple AIB systems for attitudinal and demographics data. The attitudinal data is collected from the customer survey, which includes structured (NPS rating) and unstructured data (verbatim comments). This is linked to demographic data from tables in the EDW which contains the regional locations of customers, names, sex etc.

EDW data and ‘Voice of Customer’ data is merged into one source. A preliminary exploration of the data is performed in Microsoft Excel 2013 to check the data types, fields, etc. Tableau was used to generate dashboards, bar charts, descriptive and visual statistics, using frequency counts, histograms, etc. Output of this analysis is presented as follows.
### 3.3.1. Descriptive Statistics of Voice of Customer Data.

Table 3.1 depicts the description of c.70,000 records in the full population of the ‘Voice of Customer’ data set. It contains 11 columns with customer records. The data includes a unique dummy id, as the real customer data was anonymised for the purpose of this experiment. Columns include the County, Age and Sex of the customer and which journey out of 14 was sampled. The customer type defines whether the customer was a business or personal customer. The year the customer was sampled and NPS score and verbatim commentary is also included.

<table>
<thead>
<tr>
<th>Data item</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>11 Dimensions</td>
<td>CNTY_CDE, DERIVED_AGE, SEX_CDE, JOURNEY_NAME, EVENT_LOCATION, INTERACTION_POINT, CUSTOMER_TYP, YEAR, QUARTER, NPS, VERBATIM</td>
</tr>
<tr>
<td>14 Customer Journeys or ways to bank</td>
<td>Personal Loan Successful, Credit Card Application, Deposit Account Opening, Change of Address, OTC, Mobile Banking, Loan Decline, Mortgage Drawdown, Internet Banking, Mortgage Decline, Mortgage Successful, Onboarding Personal, In Branch Self Service, QuickBanking, Personal New Current Account</td>
</tr>
<tr>
<td>NPS</td>
<td>Customers provide an answer on an 11-point rating scale, ranging from 0 (not at all likely) to 10 (extremely likely). Customers are then grouped based on their chosen ratings into three segments: promoters, passives, and detractors. Promoters are defined as customers having the highest referral intention, with a rating of 9 or 10, passives with ratings of 7 or 8 and detractors ranging from 0 to 6 on the NPS scale.</td>
</tr>
</tbody>
</table>

Table 3.1: VoC Data set Description
Figure 3.2: Gender of total Respondents

Figure 3.2 is a Tableau visualisation of the sex of the customer sampled. The population is nearly split evenly, with only 2% more males sampled.

Figure 3.3: Distribution of total respondents by age
Figure 3.3 was created in Tableau to show the distribution of age. It shows that the highest % of customers sampled is in 25-35 or the ‘Generation Y’ bracket. It shows that as the customer gets older they are less likely to respond to the surveys. This is interesting actionable insight in the data. Perhaps AIB should focus on marketing techniques to encourage older customers to provide feedback on AIB’s product offering.

Figure 3.4: Distribution of total respondents by county in Ireland

Figure 3.4 generated through Tableau, depicts the distribution of total respondents by where they live in Ireland. Not surprisingly the biggest % of respondents is from the two largest cities in Ireland. This is another actionable insight in the data. AIB could consider in future marketing strategies to target customers in more rural areas, to capture satisfaction as well as loyalty.
Figure 3.5: Customer Journey’s sampled

Figure 3.5 depicts the % of respondents according to the customer journey sampled. Here was the most surprising statistic as the vast majority of respondents 73.12% responded to surveys where the personal loan was successful. The medium to which the loan was granted was recorded through a dimension called touch point which includes, internet, phone and mobile. Approval online or in-app is guaranteed in minutes for loans from €1,500 up to €25,000. The large response rate shows that this service generates a huge measure of customer satisfaction and loyalty and it will be interesting to analyse the sentiments to extract pertinent information that can be used to garner same level of response for other journeys.

Figure 3.6: Average NPS score by age
Figure 3.6 depicts the average NPS score by age. Of Interest in this visualisation is the trend for the higher NPS scoring from 25-35 age group or Generation Y as we saw with the Figure 3.3 which shows this age category with the highest % response rate as well.

### 3.3.2. Digital Banking Journey Description

Over the past ten years, digital banking has matured and established its place in the financial services in Ireland. AIB in particular has experienced technological advancements and changing social trends, such as heightened customer proactivity and increased preferences for convenience banking through mobile, internet, in branch self-service. Figure 3.7 depicts the 3 digital banking journeys which is the focus of this work. Over the counter journey’s will also be sampled, because customers often make comparisons between real life humans and machines in their survey returns, and it will be useful to include these experiences in the experiment.

![Digital Banking Customer Journeys](image)

Figure 3. 7: Depicts 7,534- Digital Banking Customer Journeys surveyed

![Average NPS score](image)

Figure 3. 8: Depicts 7,534- Digital Banking Customer Journey average or mean NPS scores
Figure 3.8 returns an interesting insight with all age categories average mean are passives of digital banking. There is a slight decrease in the mean score as the customer ages. Albeit the score is still in the passive category.

Figure 3.9 Example Verbatim Comments from Detractors of Digital Banking

Figure 3.9 is a sample extract of verbatim comments from detractors of digital banking. Later in this chapter negative features extracted from the negative feedback could be utilised to inform how the needs and desires of their older generation customers are met.

3.3.2.1. **In Branch Self Service or Quick Banking Machines**

<table>
<thead>
<tr>
<th>Description</th>
<th>Services</th>
</tr>
</thead>
</table>
| Self-service quick banking machines are designed to help the customer to carry out Banking transactions securely in a fast and convenient way. Available in branches nationwide, a customer can use them to: | • Lodge cash and cheques  
• Check Account balance  
• View recent transactions  
• Pay bills and credit card  
• Transfer money between accounts  
• Send money abroad. |

Table 3. 2: In Branch Self Service Description
3.3.2.2. **Internet Banking**

<table>
<thead>
<tr>
<th>Description</th>
<th>Services</th>
</tr>
</thead>
</table>
| With AIB Internet Customers have the flexibility of banking at a time and place that suits them. It’s flexible banking, day and night, 365 days a year. | - Check your account transactions and balances  
- View your eStatements and eFee Advices  
- Pay bills and manage your standing orders**  
- Transfer money between accounts domestically and internationally**  
- Top up your mobile phone and access exclusive online products  
- Search for and place a stop on single and/or multiple cheques, as well as removing a cheque stop and view all stops in a Stop Cheque Log.  
- View all incoming payment details using the Incoming Payment History. |

Table 3.3: Internet Banking Service Description

3.3.2.3. **Mobile Banking**

<table>
<thead>
<tr>
<th>Description</th>
<th>Services</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mobile as a distribution network not only reaches a far broader</td>
<td>With mobile banking customers can:</td>
</tr>
<tr>
<td></td>
<td>- Open a Personal Current Account</td>
</tr>
</tbody>
</table>
audience, but also engages them at a very personal level. Customers now expect access to banking through their mobile phones, as a shift away from the Internet and other traditional banking channels. Recent research in AIB reported that the average mobile user uses their smart phones about 1,500 times a week – more than 200 times a day.

The AIB Mobile app has over 7.5 million logins and over 10 million Quick Balance views per month and is still growing approximately 2% every month.

<table>
<thead>
<tr>
<th>Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Quick Balance, view up to two balances without logging in</td>
</tr>
<tr>
<td>• Make payments</td>
</tr>
<tr>
<td>• Manage Direct Debits and Standing Orders</td>
</tr>
<tr>
<td>• Top up any prepaid mobile phone</td>
</tr>
<tr>
<td>• View up to 7 years of Statements</td>
</tr>
<tr>
<td>• Report AIB Credit/Debit Cards lost or stolen</td>
</tr>
<tr>
<td>• Get login details, or reset your PAC</td>
</tr>
<tr>
<td>• Apply for a Loan, Loan Top-Up, Overdraft or Click Credit Card</td>
</tr>
<tr>
<td>• Open an Online Saver. The secure, flexible and convenient way to save regularly online.</td>
</tr>
<tr>
<td>• Apply for Travel Insurance</td>
</tr>
</tbody>
</table>

Table 3.4: Mobile Banking Service Description

3.3.2.4. **OTC Tellers versus Machines**

Although not specifically a digital banking customer journey. A number of customers make comparisons when dealing with a human versus a machine in the branches. Therefore it was decided to include this data set of customers in the experiment. Over the counter is simply where customers go into a branch to receive a service from an over the counter Teller rather than other mediums to bank.

3.4. **Extracting Positive and Negative Features of Digital Banking**

The following section will detail the extraction of positive and negative features in relation to digital banking. This is done through Python using a noun phrase extractor called TextBlob. This library creates positive and negative words and their counts which can detail what the customer likes and dislikes in digital banking. Identifying negative and positive product features in the sentiments relies on the simple linguistic heuristic that features are usually expressed as nouns or noun phrases. Word cloud visualisations were created with words featuring most often. See appendix B for code and below for visualisations. In addition in chapter 4, a
summary of the most frequent negative features is provided and recommendations for more age friendly banking.

3.4.1. Using Industry Benchmarks to determine what is a good Net Promoter Score (NPS) goal

The NPS scores across the 4 digital banking journeys are compared below with other journeys. In addition a detailed breakdown of promoters, detractors and passives further split by age categories. To compare NPS scores against industry standards, a leader in benchmarking the ‘Temkin Group’ measured the NPS of almost 300 companies across 20 industries based on a survey of 10,000 U.S. consumers. The report determined that the banking benchmark average for Promoters is 56%. The 4 digital banking journeys analysed below actually fall below this average. All other journey’s result is 76% therefore comfortably above the industry standard. This finding supports the need for strategies to improve the customer experience of digital banking journeys and age friendly banking.

3.4.2. In Branch Self Service Analysis & Results

![In-Branch Self Service Journey NPS and summary reasons for NPS Score](image)

Figure 3.11 In-Branch Self Service Journey NPS and summary reasons for NPS Score
Figure 3. 12 Free online Word Cloud Generator of In-Branch Self Service Positive and Negative words

<table>
<thead>
<tr>
<th>Customer Journey</th>
<th>In Brach Self Service or Quickbanking</th>
</tr>
</thead>
<tbody>
<tr>
<td>Surveys sent</td>
<td>2406</td>
</tr>
<tr>
<td>Survey responses</td>
<td>802</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>NPS Scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>Detractors</td>
</tr>
<tr>
<td>0 to 6 on NPS scale</td>
</tr>
<tr>
<td>14% of Traditionalists are detractors</td>
</tr>
<tr>
<td>22% of Baby Boomers are detractors</td>
</tr>
<tr>
<td>18.5% of Generation X are detractors</td>
</tr>
<tr>
<td>8.9% of Generation Y are detractors</td>
</tr>
<tr>
<td>0% of Generation Z are detractors</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>fence Sitters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ratings of 7 or 8</td>
</tr>
<tr>
<td>25.7% of Traditionalists are Fence Sitters</td>
</tr>
<tr>
<td>34% of Baby Boomers are Fence Sitters</td>
</tr>
<tr>
<td>30% of Generation X are Fence Sitters</td>
</tr>
<tr>
<td>38.6% of Generation Y are Fence Sitters</td>
</tr>
<tr>
<td>45% of Generation Z are Fence Sitters</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Promoters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ratings of 9 or 10</td>
</tr>
<tr>
<td>57% of Traditionalists are Promoters</td>
</tr>
<tr>
<td>43.75% of Baby Boomers are Promoters</td>
</tr>
<tr>
<td>51% of Generation X are Promoters</td>
</tr>
<tr>
<td>52% of Generation Y are Promoters</td>
</tr>
<tr>
<td>55% of Generation Z are Promoters</td>
</tr>
</tbody>
</table>

Table 3. 5 Calculations by age of Detractors, Fence Sitters and Promoters of In-Branch Self Service or Quick-banking.
3.4.3. **Over the Counter Analysis & Results**

Figure 3. 14 OTC Journey NPS and summary reasons for NPS Score

Figure 3. 13 Depicts In-Branch Self-service age profile of respondents
Customer Journey | OTC-Over the Counter
--- | ---
Surveys sent | 7416
Survey responses | 2225

NPS Scores

| OTC Detractors | 9.3% of Traditionalists are detractors  
10.9% of Baby Boomers are detractors  
12.6% of Generation X are detractors  
10.8% of Generation Y are detractors  
12.3% of Generation Z are detractors |
| OTC Fence Sitters | 23.2% of Traditionalists are Fence Sitters  
26.5% of Baby Boomers are Fence Sitters  
22.7% of Generation X are Fence Sitters  
19.2% of Generation Y are Fence Sitters  
25.6% of Generation Z are Fence Sitters |
| OTC Promoters | 68% of Traditionalists are Promoters  
62% of Baby Boomers are Promoters  
64% of Generation X are Promoters  
69% of Generation Y are Promoters  
61% of Generation Z are Promoters |

Table 3. 6 Calculations of Detractors, Fence Sitters and Promoters of Over the Counter
3.4.4. **Internet Banking Analysis & Results**

![Age Profile of OTC Respondents](image1)

![Internet Banking NPS and summary reasons for NPS Score](image2)

Figure 3. 16: Age Profile of OTC Respondents

Figure 3. 17 Internet Banking NPS and summary reasons for NPS Score
Figure 3. 18 Free online Word Cloud Generator of Internet Banking Positive and Negative words

<table>
<thead>
<tr>
<th>Customer Journey</th>
<th>Internet Banking</th>
</tr>
</thead>
<tbody>
<tr>
<td>Surveys sent</td>
<td>6216</td>
</tr>
<tr>
<td>Survey responses</td>
<td>1865</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>NPS Scores</th>
<th>Detractors</th>
<th>Fence Sitters</th>
<th>Promoters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Internet Banking</td>
<td>20% of Traditionalists are detractors</td>
<td>33% of Traditionalists are Fence Sitters</td>
<td>46% of Traditionalists are Promoters</td>
</tr>
<tr>
<td>Detractors</td>
<td>11.7% of Baby Boomers are detractors</td>
<td>29% of Baby Boomers are Fence Sitters</td>
<td>58% of Baby Boomers are Promoters</td>
</tr>
<tr>
<td>0 to 6 on the NPS scale</td>
<td>11% of Generation X are detractors</td>
<td>32% of Generation X are Fence Sitters</td>
<td>56% of Generation X are Promoters</td>
</tr>
<tr>
<td></td>
<td>12% of Generation Y are detractors</td>
<td>31% of Generation Y are Fence Sitters</td>
<td>56% of Generation Y are Promoters</td>
</tr>
<tr>
<td></td>
<td>0% of Generation Z are detractors</td>
<td>21% of Generation Z are Fence Sitters</td>
<td>78.5% of Generation Z are Promoters</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Internet Banking</th>
<th>Fence Sitters</th>
<th>Promoters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fence Sitters</td>
<td>33% of Traditionalists are Fence Sitters</td>
<td>46% of Traditionalists are Promoters</td>
</tr>
<tr>
<td>Ratings of 7 or 8</td>
<td>29% of Baby Boomers are Fence Sitters</td>
<td>58% of Baby Boomers are Promoters</td>
</tr>
<tr>
<td></td>
<td>32% of Generation X are Fence Sitters</td>
<td>56% of Generation X are Promoters</td>
</tr>
<tr>
<td></td>
<td>31% of Generation Y are Fence Sitters</td>
<td>56% of Generation Y are Promoters</td>
</tr>
<tr>
<td></td>
<td>21% of Generation Z are Fence Sitters</td>
<td>78.5% of Generation Z are Promoters</td>
</tr>
</tbody>
</table>

Table 3. 7 Calculations of Detractors, Fence Sitters and Promoters of Internet Banking
Figure 3. 19 Internet Banking Age Profile of respondents

3.4.5. **Mobile Banking Analysis & Results**

Figure 3. 20 Mobile Banking NPS and summary reasons for NPS Score
Figure 3. 21 Free online Word Cloud Generator of Mobile Banking Positive and Negative words.

<table>
<thead>
<tr>
<th>Customer Journey</th>
<th>Mobile Banking</th>
</tr>
</thead>
<tbody>
<tr>
<td>Surveys sent</td>
<td>4880</td>
</tr>
<tr>
<td>Survey responses</td>
<td>1464</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>NPS Scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mobile Banking Detractors</td>
</tr>
<tr>
<td>100% of Traditionalists are detractors</td>
</tr>
<tr>
<td>4.2% of Baby Boomers are detractors</td>
</tr>
<tr>
<td>8.5% of Generation X are detractors</td>
</tr>
<tr>
<td>7.6% of Generation Y are detractors</td>
</tr>
<tr>
<td>3.4% of Generation Z are detractors</td>
</tr>
<tr>
<td>Mobile Banking Fence Sitters</td>
</tr>
<tr>
<td>0% of Traditionalists are Fence Sitters</td>
</tr>
<tr>
<td>25% of Baby Boomers are Fence Sitters</td>
</tr>
<tr>
<td>26% of Generation X are Fence Sitters</td>
</tr>
<tr>
<td>25% of Generation Y are Fence Sitters</td>
</tr>
<tr>
<td>22% of Generation Z are Fence Sitters</td>
</tr>
<tr>
<td>Mobile Banking Promoters</td>
</tr>
<tr>
<td>0% of Traditionalists are Promoters</td>
</tr>
<tr>
<td>69% of Baby Boomers are Promoters</td>
</tr>
<tr>
<td>64% of Generation X are Promoters</td>
</tr>
<tr>
<td>66% of Generation Y are Promoters</td>
</tr>
<tr>
<td>74% of Generation Z are Promoters</td>
</tr>
</tbody>
</table>

Table 3. 8 Calculations of Detractors, Fence Sitters and Promoters of Mobile Banking Respondents
3.4.6. **Comparison with All other Journey’s Analysis & Results**

![Figure 3. 22: Age Profile of Mobile Banking Respondents](image)

- **Gen Z aged 3-23**: 89
- **Gen Y aged 24-38**: 732
- **Gen X aged 39-53**: 479
- **Baby Boomers aged 54-74**: 163
- **Traditionalists aged 75 +**: 1

![Figure 3. 23: All other Journeys](image)

- **76%**: Fast & easy process, Great service, overall good, feel-value, Response quick, easy to use/apply/understand, extremely helpful, explained clearly, straightforward, information clear.
- **21%**: Bad experience, mistakes, declined, complicated, help required
- **3%**: Detractors

47
## Customer Journey vs. All other Journeys

<table>
<thead>
<tr>
<th>Surveys sent</th>
<th>c.240,000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Survey responses</td>
<td>61,864</td>
</tr>
</tbody>
</table>

### NPS Scores

<table>
<thead>
<tr>
<th>Journeys</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Detractors</td>
<td>10% of Traditionalists are detractors 6% of Baby Boomers are detractors 4% of Generation X are detractors 5% of Generation Y are detractors 4% of Generation Z are detractors</td>
</tr>
<tr>
<td>Fence Sitters</td>
<td>20% of Traditionalists are Fence Sitters 14% of Baby Boomers are Fence Sitters 13% of Generation X are Fence Sitters 13% of Generation Y are Fence Sitters 14% of Generation Z are Fence Sitters</td>
</tr>
<tr>
<td>Promoters</td>
<td>70% of Traditionalists are Promoters 80% of Baby Boomers are Promoters 83% of Generation X are Promoters 82% of Generation Y are Promoters 82% of Generation Z are Promoters</td>
</tr>
</tbody>
</table>

### Table 3.9 Calculations of Detractors, Fence Sitters and Promoters of All Other Journeys Respondents

#### Age profile of all Other Journeys

- Gen Z aged 3-23: 4000
- Gen Y aged 24-38: 34271
- Gen X aged 39-53: 19453
- Baby Boomers aged 54-74: 4025
- Traditionalists aged 75+: 115

#### Figure 3.24 Age Profile of all Other Journeys Respondents
3.4.7. Difference in Customer Satisfaction discussion

The analysis shows that there is a difference of approximately 27% in customer dissatisfaction of digital banking. The extent to which the ‘Traditionalist’ and ‘Baby Boomers’ dissatisfaction contributes to this is marginal. A limitation of this finding is that the sample size is small. This impacts the reliability of a survey’s results because it leads to a higher variability, which may lead to bias. Reasons for bias are detailed in the literature review such as cultural factors and self-selection bias. The survey accuracy may also suffer from non-response due to lack of capability (knowledge/competence/tools) of older customers. Therefore results from small samples representative of the overall population is a cognitive bias.

3.5. Data Preparation and Modelling.

Data preparation tasks include table, record and attribute selection as well as transformation and cleaning and quality analysis using functions in Python such as the NLTK library to tokenize, filter punctuation and remove stop words etc. Tokenization is most frequently found in text mining. This involves breaking the natural language in the sentiments into keywords, phrases, symbols and other elements called tokens. Characters like punctuation marks are discarded and the tokens become the input for text mining. Stop words in documents are meaningless for sentiment analysis and occur frequently. Words like 'IT', 'AND' and 'TO' can be found virtually in every sentence in English-based documents, these words make very poor index terms. Therefore it is best remove all stop word terms.

The objective is to resolve inconsistencies and address quality issues. Process the data to make it suitable for modelling. Weak data does not build strong models. It will also provide a general overview of the data health, its strengths and weaknesses to further understand the quality challenges. The preparation of unstructured data is usually more challenging than the preparation of structured data. (Botzenhardt et al, 2011)

After the documents have been transformed to a structured format which can be processed by the text mining application, the actual modelling of the data, like Feature Extraction and Categorization can begin. A sentiment analysis programme will be built.
Machine learning methods will use Python to classify customer opinions into positive and negative classes and Sentiment analysis. Python and Natural Language Tool Kit (NLTK) will be used to train and classify the sentiments. A pre-labelled data set from AIB will be used for training. Algorithms such as Naïve Bayes, which will list every distinct words ordered by frequency of their appearance. The Naïve Bayes classifier uses the prior probability of each label which is equal to the frequency of each label in the training set, and the contribution from each feature. Results will be analysed and any patterns found will documented in an attempt to answer the research question.

In future studies, in addition to machine learning methods a lexicon based sentiment analysis can be done using freely available out of the box sentiment libraries. The two popular libraries that can be used in future experiments are Textblog and vaderdsentiment. TextBlog is more of a natural language processing library, but it comes with a rule-based sentiment analysis library, specifically attuned to sentiments expressed in social media. In TextBlog, polarity and a subjectivity metric is provided. The polarity is the sentiment itself, ranging from a -1 to a +1. The subjectivity is a measure of the sentiment being objective to subjective, and goes from 0 to 1.

For VADERNet a positive sentiment: compound score >= 0.5 neutral sentiment: (compound score > -0.5) and (compound score < 0.5) and negative sentiment: compound score <= -0.5.

3.5.1. Data Preparation Machine Learning Method-Supervised

3.5.1.1. Pre-processing the full data set in Python

Data from surveys is especially prone to noise, missing, and inconsistent data due to onus on the customer to provide the data. Low-quality data will lead to low-quality results. Pre-processing is an important first step for serious, effective, real-world text mining. There are a number of steps in Python using the imported library called NLTK that cleans and prepare text ready for modelling with machine learning and getting it
ready for classification. The scripts for the below steps and sample outputs are detailed in Appendix A. In addition a full list of libraries imported to carry out the below steps are described.

- **Step 1: Tokenization -Split into Words.**
  Tokenization process is where a text document is split into a stream of words, phrases, symbols, or other meaningful elements called tokens. The list of tokens becomes input for further processing or text mining. NLTK provides a function called `word_tokenize()` for splitting strings into tokens (nominally words).
  It splits tokens based on white space and punctuation. For example, commas and periods are taken as separate tokens. Contractions are split apart (e.g. “What’s” becomes “What” “s”).

- **Step 2: Filter punctuation.**
  This step filters out all tokens with little benefit to the model, such as all standalone punctuation. This can be done by iterating over all tokens and only keeping those tokens that are all alphabetic. Python has the function `isalpha()` that can be used.

- **Step 3: Filter out Stop Words**
  Stop words are those words that do not contribute to the deeper meaning of the phrase. They are the most common words such as: “the”, “a”, and “is”. These words can be found virtually in every sentence and make up a large portion of most documents. The amount of information carried by these words is negligible. It makes sense to remove stop words. NLTK provides a list of commonly agreed upon stop words to remove for a variety of languages.

- **Step 4: Filter out Stem words**
  Stemming refers to the process of reducing each word to its root or base. For example “fishing,” “fished,” “fisher” all reduce to the stem “fish.” Some applications, like document classification, may benefit from stemming in order to both reduce the vocabulary and to focus on the sense or sentiment of a document rather than deeper meaning. There are many stemming algorithms, although a popular and long-standing method is the Porter Stemming algorithm. This method is available in NLTK via the Porter Stemmer class.
3.5.1.2. *Split available data set into Training and Testing Data.*

A good strategy in machine learning is to split the data into two parts, where part of the sample is used for training and the remaining sample is used for testing the model(s). Therefore, by training the model(s) with 75% of the training data, the remaining 25% will test the accuracy of the model because the model has not seen this 25% before. See Appendix A for script and sample output of splitting the data set into training and testing data.

3.5.1.3. *Summary*

In summary, cleaning and tokenization involved the task of removing punctuations, stop-words, numerics, symbols and other imprecise and improper language and words within the data set. This was performed in Python using the NLTK library involving:

- Split tokens on white space.
- Remove all punctuation from words.
- Remove all words that are not purely comprised of alphabetical characters.
- Remove all words that are known stop words.

The pre-processing of the data set removed the irrelevant and redundant information in order to make the training phase less difficult. The data preparation and filtering steps took a considerable amount of processing time. The aim of data pre-processing is the final training set. The overall size after the pre-processing reduced the original data set by 30% which is reflective of the amount of noisy data present and time took to prepare the cleaned up data set to perform sentiment analyses.

3.6. *Evaluation*

Evaluation needs to, specifically focus on the question if the achieved results create value and the most important the business issues are covered. The model results will be analysed in respect of the success criteria outlined in the business understanding phase of the project lifecycle. Similar to (Tharwat, A. 2018) assessment methods will be utilised in evaluating the classification performance and guiding
the classifier modelling. There are three main phases of the classification process, namely, training phase, validation phase, and testing phase. The model is trained using input patterns and this phase is called the training phase. These input patterns are called training data which are used for training the model. The goal of a learning algorithm is to learn from the training data to predict class labels for unseen data; this is in the testing phase. The validation phase is used for evaluating the performance of the trained model. The most well-known classification assessment methods are described below such as the Accuracy and Error Rate, Confusion Matrix and RoC Curve.

3.6.1. Accuracy and Error Rate:

Accuracy (Acc) is one of the most commonly used measures for the classification performance, and it is defined as a ratio between the correctly classified samples to the total number of samples as follows:

\[
Acc = \frac{TP + TN}{TP + TN + FP + FN}
\]

Figure 3. 25 (Tharwat, A. 2018) Accuracy Measurement

3.6.2. Receiver operating characteristics (ROC):

According to Tharwat, A. (2018) the receiver operating characteristics (ROC) curve is a two-dimensional graph in which the TPR represents the y-axis and FPR is the x-axis. It is a plot of the true positive rate against the false positive rate for the different possible cut points of a diagnostic test.

An ROC curve demonstrates several things:

1. It shows the trade-off between sensitivity and specificity (any increase in sensitivity will be accompanied by a decrease in specificity).

2. The closer the curve follows the left-hand border and then the top border of the ROC space, the more accurate the test.

3. The closer the curve comes to the 45-degree diagonal of the ROC space, the less accurate the test.
4. The slope of the tangent line at a cut point gives the likelihood ratio (LR) for that value of the test.

5. The area under the curve is a measure of text accuracy.

![ROC Curve](image)

**Figure 3. 26 (Tharwat, A. 2018) ROC Curve**

### 3.6.3. **Confusion Matrix**

 Visa, S *et al* (2011) states that a confusion matrix is a table that is often used to describe the performance of a classification model (or "classifier") on a set of test data for which the true values are known. It provides a tabular visualization of statistical classification per class. Each row of the matrix represents predicted value of a class while each column represents actual value of a class that a data instance belongs to. The diagonal entries are correctly classified instances while the remaining entries are mis-classified.

<table>
<thead>
<tr>
<th></th>
<th>Predicted Negative</th>
<th>Predicted Positive</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Actual Negative</strong></td>
<td>a</td>
<td>b</td>
</tr>
<tr>
<td><strong>Actual Positive</strong></td>
<td>c</td>
<td>d</td>
</tr>
</tbody>
</table>

**Figure 3. 27 (Visa, S* et al* 2011) Confusion Matrix**

### 3.7. **Deployment:**

Document results and findings into the dissertation. Also prepare a findings paper for the AIB ‘Voice of Customer’ programme.
CHAPTER 4: Implementation and Results

This chapter details the execution of the experiment conducted in this study. As the data description, exploration and preparation stages have already been covered in the previous chapter, this chapter deals with training and testing the model(s) to determine 1) if a model can be built to approximate the NPS score of a customer based on the textual description and 2) if a model can be built to indicate if customer age group is associated with any particular class of NPS.

The objective is to present a reliable model using various machine learning classifiers that can inform customer satisfaction improvements of AIB’s banking journeys, specifically digital banking. In this section, the classification algorithms, applied using the Python Scikitlearn library are Naïve Bayes (NB), Support Vector Machine (SVM) and Decision Tree (DT). Their performance against each other will also be compared.

The aim is to provide AIB with a system to classify customers according to their sentiments. The classifiers will learn from the training data set and perform the category assignments automatically.

Our findings will assess the research hypothesis that sentiment analysis has the potential to usefully classify customer text-based feedback into positive or negative sentiments which, in turn, can inform customer satisfaction improvement initiatives for AIB’s digital banking journeys.

4.1. Support Vector Machine Model Algorithm

By training the Support Vector Machine Model, based on the text from the training data set, can the model approximate the NPS score categories of the customer? Each of the 3 classes Promoters, Detractors and Passives will be applied in turn as target variables using an algorithm called One-vs-the-rest or OVR. This strategy consists in fitting one classifier per class. For each classifier, the class is fitted against all the other classes. One advantage of this approach is its interpretability. Since each class is represented by one and one classifier only, it is possible to gain knowledge about the class by inspecting its corresponding classifier. (Zimmert et al, 2017)
Similar to training the Support Vector, the Naïve Bayes will be used to predict the probability of three classes Promoters, Detractors and Passives. The model will learn to label the classes into Detractor (0), Passive (1) Promoter (2). A possible advantage of using Naive Bayes classifiers is that they work well in real business problems are robust to smaller and less rich data sets for training. Naive Bayes learners and classifiers can be extremely fast compared to more sophisticated methods. The decoupling of the class conditional feature distributions means that each distribution can be independently estimated as a one dimensional distribution. This in turn helps to alleviate problems stemming from the so called ‘curse of dimensionality’. (Jain et al, 2016)
4.3. **Decision Trees Discussion**

A decision tree is a flowchart-like tree structure where an internal node represents feature (or attribute), the branch represents a decision rule, and each leaf node represents the outcome. The topmost node in a decision tree is known as the root node. It learns to partition on the basis of the attribute value. It partitions the tree iteratively using a recursive partitioning technique. It's visualization like a flowchart diagram mimics the human level thinking. That is why decision trees are easy to understand and interpret. (Sharma, 2012)

A decision Tree was built in Python using Scikit-learn ‘Decision Tree Classifier'. Visualizing the Decision Trees was done through Scikit-learn's `export_graphviz` function.

The experiment followed the same methodology as per (Sharma, 2012) using Decision Trees for sentiment analysis by iteratively partitioning the document feature vector space into parts. At each step, the split improves the error of the overall tree on the training data. Through exploring decision trees in this experiment, it was discovered that the decision tree produced poor accuracies and, when used as a classifier based on text, did not produce interesting results. In this regard, it was not a good choice specifically for this data set with a high dimensionality. That is not to say decision trees have no use in text mining. Mitigating actions such as feature reduction, by reference to, for example, a ‘synset’ dictionary such as ‘Sentiword.Net where words of similar meaning are assigned an identifier. In this case terms in the data set can be reduced to fewer terms. Results were satisfactory from the other techniques, therefore it was decided to use the decision tree more conventionally in ‘R’ to explore the various features in the data set, such as age, sex, county code and type of customer to investigate if any feature is more likely to be a Promoter, Detractor or Passive Customer. This ‘Occam’s Razor’ approach, is one
which may appeal to a wider audience, such as the bank’s management team who need insights that are direct and easily explained.

Having trained a model using the ‘rpart’ library, it was found that the lack of response variety (e.g. large majority of ‘promoters’) meant that little improvement beyond standard descriptive techniques could be derived from trying to assess attitudes based on age or indeed other factors.

### 4.4. Implementation Methodology

Python Scikit-learn and Natural Language tool kit have been used to train and classify the reviews. 75% of the data set was used for training the models. The data is labelled as ‘promoter’, ‘detractor’ or ‘passive’ for the experiment to classify respondents based on the contents of their text statements. For the experiment to predict the age band, the data is labelled with the 5 different age bands.

![Flow of proposed methodology](image)

**Figure 4. 4 Flow of proposed methodology**
4.5. Results Analysis

Similar to Jain et al. (2016) the analysis is done using three parameters, Accuracy, Precision and Recall. In Python the imported SK Learn Metrics Library follows the following logic for Accuracy, Precision and Recall.

Accuracy
This is the total number of correct classifications divided by the total number of cases predicted whether correct or not.

Accuracy = \frac{Tp + Tn}{Tp + Tn + Fp + Fn}

Figure 4. 5 (Jain et al, 2016) Accuracy

Recall
This is the total number of true positives divided by the actual number of actual positives including false negatives out of the total.

Recall = \frac{Tp}{Tp + Fn}

Figure 4. 6 (Jain et al, 2016) Recall

Precision
Out of the total positive predicted results by the model, this is the percentage of actual positive results. It is given as:

Precision = \frac{Tp}{Tp + Fp}

Figure 4. 7 (Jain et al, 2016) Precision

Where
Tp= True Positive, Tn= True Negative, Fp= False Positive, Fn= False Negative
4.5.1. **Results to predict the NPS score of a customer based on the textual description**

The following details the performance measures of Support Vector Machine and Naïve Bayes based on the NPS categorisation of all customer journeys. Calculating a confusion matrix can determine what the classification model is getting right and what types of errors it making. The confusion matrix had been tested against 25% of the full data set, generated through the Metrics library in Python.

4.5.1.1. **Overall Accuracy of Support Vector Machine**

<table>
<thead>
<tr>
<th>Classifications</th>
<th>SVM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Promoter</td>
<td>84.60%</td>
</tr>
<tr>
<td>Detractor</td>
<td>95.46%</td>
</tr>
<tr>
<td>Passive</td>
<td>85.94%</td>
</tr>
</tbody>
</table>

Table 4.1 Overall Accuracy Support Vector Machine %

The classification accuracy of SVM’s is seen to be consistently high from 84.60% and reaching up-to 95.46% The results are in line with what previous research has shown for SVM’s and supports existing results collected through the ‘Voice of Customer’ programme. However, as with all machine learning endeavours, results showing high initial accuracy rates need to be treated with caution. There are possible problems related to the size and accuracy of the data set and the possibility of overfitting and bias. Accurate results are encouraging but many tests are required and different training samples as well as analysis of hidden bias before deployment could be considered

4.5.1.2. **SVM Confusion Matrix for Promoter**

<table>
<thead>
<tr>
<th></th>
<th>Predicted Not a Promoter</th>
<th>Predicted Promoter</th>
</tr>
</thead>
<tbody>
<tr>
<td>n= 10,358</td>
<td></td>
<td></td>
</tr>
<tr>
<td>True Not a Promoter</td>
<td>674</td>
<td>1304</td>
</tr>
<tr>
<td>True Promoter</td>
<td>291</td>
<td>8089</td>
</tr>
</tbody>
</table>

Table 4.2 Confusion Matrix for Promoter

Out of the promoter test cases, the classifier predicted true negative 674 times from the sentiments, that the customer was not a promoter and true positive 8089 that the
customer was a promoter. The accuracy of the promoter model is high at 84.60%. However, the model was not good at classifying ‘non-promoters’ and this could be a dangerous oversight as the bank could be under a false sense of security in relation to a possible problem. These false classifications can be very useful in further model development. An interesting insight also would be to analyse where the model misclassified the promoters. It will be worthwhile to look at the sample where the model predicted the 291 cases as not a promoter. It may support the literature that machine learning can discover specific trends and patterns that would not be apparent to humans. To improve the quality of the data these potential erroneous samples could be removed due to the clues from the confusion matrix.

4.5.1.3. **SVM Confusion Matrix for Passive**

<table>
<thead>
<tr>
<th>n= 10,358</th>
<th>Predicted Not a Passive</th>
<th>Predicted Passive</th>
</tr>
</thead>
<tbody>
<tr>
<td>True Not a Passive</td>
<td>8863</td>
<td>166</td>
</tr>
<tr>
<td>True Passive</td>
<td>1209</td>
<td>120</td>
</tr>
</tbody>
</table>

Table 4. 3 Confusion Matrix for Passive

Out of the passive test cases, the classifier predicted true negative 8863 times from the sentiments, that the customer was not a passive and true positive 120 times that the customer was a passive. Once again, the results highlight that investigation is required as to how the classification ‘passive’ could lull the bank into complacency regarding customer sentiment.

4.5.1.4. **SVM Confusion Matrix for Detractor**

<table>
<thead>
<tr>
<th>n= 10,358</th>
<th>Predicted Not a Detractor</th>
<th>Predicted Detractor</th>
</tr>
</thead>
<tbody>
<tr>
<td>True Not a Detractor</td>
<td>9677</td>
<td>113</td>
</tr>
<tr>
<td>True Detractor</td>
<td>357</td>
<td>211</td>
</tr>
</tbody>
</table>

Table 4. 4Confusion Matrix for Detractor

Out of the detractor test cases, the classifier predicted true negative 9677 times from the sentiments, that the customer was not a detractor and true positive 211 times that the customer was a detractor. This is an interesting insight into the detractor population. The data set contains mostly positive reviews for all banking journeys. This supports the data
exploration metrics in Chapter 3 where promoters >76% for all other journeys. Digital banking dissatisfaction is approximately 27% less. Therefore it would interesting to look at the 211 and discover what % of digital banking customers is in this small sample size of detractors.

4.5.1.5. **Summary of SVM’s**

The results show that SVMs have potential to provide insight into customer sentiment based on their text responses. SVMs achieved good performance on text categorization tasks for certain classes. The confusion matrices were very useful to point out weaknesses in the prediction of certain classes. At the very least, this provided a reduced data set for investigation. Such investigation may provide input into how a text based classifier may be improved by possibly, for example, building a domain specific dictionary where certain terms are weighted positively or negatively based on local analysis. SVMs show some promise and are an easy to use method for learning text classifiers from examples.

4.5.2. **Naïve Bayes algorithm results to predict the NPS score of a customer based on the textual description.**

Table 4.5 shows the performance measures of Naïve Bayes algorithm. The overall accuracy for this model is 82.81%. The confusion matrix had been tested against 25% of the full data set.

4.5.2.1. **Naïve Bayes Confusion Matrix Detractors, Passive, Promoter**

<table>
<thead>
<tr>
<th></th>
<th>Predicted 0: Detractor</th>
<th>Predicted 1: Passive</th>
<th>Predicted 2 : Promoter</th>
</tr>
</thead>
<tbody>
<tr>
<td>True 0: Detractors</td>
<td>171</td>
<td>88</td>
<td>309</td>
</tr>
<tr>
<td>True 1: Passive</td>
<td>41</td>
<td>118</td>
<td>1282</td>
</tr>
<tr>
<td>True 2: Promoter</td>
<td>22</td>
<td>100</td>
<td>8289</td>
</tr>
</tbody>
</table>

Table 4.5 Naïve Bayes Confusion Matrix
4.5.2.2. **Naïve Bayes Summary**

The NB algorithm was simple and fast to implement. Its high accuracy predicted class membership probabilities for detractors, passives and promoters with much the same results as SVM. The same caveats, of course, apply to the misclassified elements and the same potential value may be recovered by deeper analysis of same.

4.5.3. **Results to predict the customers age group based on digital banking sentiments**

The following is the performance measures of Support Vector Machine and Naïve Bayes algorithms based on the age bands of digital banking customer experiences. Calculating a confusion matrix can determine what the classification model is getting right and what types of errors it is making. The confusion matrix has been tested against 25% of the digital banking data set. The results will be summarised at the end of this section.

4.5.3.1. **SVM Generation X, Y, Z, Baby Boomer, Traditionalist**

Generation Z Accuracy: 98.09%

<table>
<thead>
<tr>
<th>n=1153</th>
<th>Predicted not a Gen Z</th>
<th>Predicted Gen Z</th>
</tr>
</thead>
<tbody>
<tr>
<td>Actual not a Gen Z</td>
<td>1131</td>
<td>0</td>
</tr>
<tr>
<td>Actual Gen Z</td>
<td>22</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 4.6 SVM Confusion Matrix generation Z of digital banking

Generation Y Accuracy 69.47%

<table>
<thead>
<tr>
<th>n=1153</th>
<th>Predicted not a Gen Y</th>
<th>Predicted Gen Y</th>
</tr>
</thead>
<tbody>
<tr>
<td>Actual not a Gen Y</td>
<td>762</td>
<td>60</td>
</tr>
<tr>
<td>Actual Gen Y</td>
<td>292</td>
<td>39</td>
</tr>
</tbody>
</table>

Table 4.7 SVM Confusion Matrix generation Y of digital banking
4.5.3.2. **Naïve Bayes Generation X, Y, Z, Baby Boomer, Traditionalist**

Accuracy 37.12%

<table>
<thead>
<tr>
<th>n=1153</th>
<th>Predicted Gen Z</th>
<th>Predicted Gen Y</th>
<th>Predicted Gen X</th>
<th>Predicted Gen BB</th>
<th>Predicted Gen TRAD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Actual (Gen Z)</td>
<td>0</td>
<td>7</td>
<td>5</td>
<td>10</td>
<td>0</td>
</tr>
<tr>
<td>Actual (Gen Y)</td>
<td>0</td>
<td>43</td>
<td>197</td>
<td>91</td>
<td>0</td>
</tr>
<tr>
<td>Actual (Gen X)</td>
<td>0</td>
<td>37</td>
<td>228</td>
<td>140</td>
<td>0</td>
</tr>
<tr>
<td>Actual (Gen BB)</td>
<td>0</td>
<td>21</td>
<td>188</td>
<td>157</td>
<td>0</td>
</tr>
<tr>
<td>Actual (Gen TRAD)</td>
<td>0</td>
<td>0</td>
<td>17</td>
<td>12</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 4. 11 Naïve Bayes Confusion Matrix age bands of Digital Banking
4.5.3.3. **Classifying Generation X, Y, Z, Baby Boomer, Traditionalist results**

The results show unreliable accuracies produced by the models. A poor result in predicting the customer’s age band for digital banking could potentially be due to the size of the data set. Supervised machine learning algorithms generally fare better with larger data sets. Results produced from each model are inconsistent and therefore unreliable - the results are only as good as the data. The models may simply not have had enough data to learn from. Training algorithms with data sets small enough to not be inclusive, results in biased predictions coming from a biased training set. The following techniques described may boost the limitations of the training data set.

4.5.3.3.1. **Techniques to improve the model’s accuracy and limitations of training data set**

The academic papers focus on new and improved models, with data usually chosen from known well known labelled data sets, tested multiple times for high accuracy, i.e. twitter feeds and IMBD reviews. This experiment is part of an actual real business problem where the accuracy of the models is only as good as the training data. The following techniques can be utilized to achieve accuracy improvements. (Sun, 2017)

1. Addition of more data could potentially result in better and accurate models. Getting more data from older customers, or expanding to other customer journeys, or even from other banks would allow the data to tell for itself, instead of relying on assumptions and weak correlations. Increasing the size of training data will reduce the challenges encountered in this experiment of working on a limited data set.

2. Analysing the error metrics from the confusion matrix will attempt to understand what the models were having the most problems with. It may be worthwhile to remove the words that are causing an issue. Watching how the numbers change during training by tweaking the data will show what categories the model is struggling to learn, and give areas to concentrate on when cleaning and expanding the data set.
3. As an alternative to pre-labelled data set based on what customers ranked as their NPS score, use crowd sourcing or online labour markets, where the crowd is asked to help tag the text. The results can be compared to pre-labelled data set to determine if there are differences in the ratings. Another option that may be worthwhile in the next iteration is to introduce training for labelers or hiring experts on banking sentiment analysis to label new training inputs.

4. It may be worth investing in a domain knowledge based lexicon. For example there are commercially available domain based lexicons for banking or finance which are more focused for matching the words or phrases for this particular domain.

4.5.3.4. **Product Feature Extraction for Age Friendly Banking**

In chapter 3 as part of data exploration, extracting negative features in the digital banking sentiments relied on the simple linguistic heuristic that features are usually expressed as nouns or noun phrases. This narrows down the words or phrases to be considered based on part-of-speech-tag patterns. POS-tagging method was done through NLTK via the TextBlob module.

Table (4.12) illustrates a summary of frequent negative features extracted from detractor’s opinions of digital banking which includes older customer’s sentiment.

<table>
<thead>
<tr>
<th>Humans replaced by Machines</th>
<th>Personal touch</th>
<th>Privacy</th>
<th>Queues long</th>
</tr>
</thead>
<tbody>
<tr>
<td>Support</td>
<td>Accessibility difficult</td>
<td>Navigation difficult</td>
<td>Error recoverability and timeout</td>
</tr>
<tr>
<td>Paper statements preference</td>
<td>Security/fraud worries</td>
<td>Passwords</td>
<td>Branch Closures</td>
</tr>
<tr>
<td>Loyalty</td>
<td>Fearful/Anxious</td>
<td>Customer centricity</td>
<td>Poor flexibility</td>
</tr>
</tbody>
</table>

Table 4.12 Summary of frequent negative features extracted from customer’s opinion.
4.5.4. Recommendations for digital inclusion for age friendly banking

An Age UK paper, which is the UK’s largest charity working with older people details the following recommendations that can address the frequent negative features highlighted in the POS tagging exercise.

<table>
<thead>
<tr>
<th>Negative Features</th>
<th>Recommendations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Customer centred service</td>
<td>Through improved policies and staff training, AIB should develop the ability to support their most vulnerable customers. Be seen as a helper rather than an additional source of pressure.</td>
</tr>
<tr>
<td>Preference for in-branch personal touch banking</td>
<td>With branches closing nationwide, AIB should consider expanding Mobile branches fleet, which have features such as high visibility markings on steps and handrails and open plan designs. Teller rather than machines can help customers with hearing impairments, and other cognitive challenges. This will help address the strong preference for talking with ‘live people’ rather than trying to negotiate automated menu systems.</td>
</tr>
<tr>
<td>Age-friendly design</td>
<td>New and existing technologies should be field tested with older people to make sure the full range of perceptual and motor needs are anticipated in the design process.</td>
</tr>
<tr>
<td>Paper statements</td>
<td>Many expressed a preference for paper statements and particularly did not like what were seen as ‘sneaky’ ways of moving people to electronic statements. The hard to find ‘tick box’ if they want to continue receiving paper statements should be made more evident.</td>
</tr>
<tr>
<td>Innovation in ID and security</td>
<td>To address how some older customers ‘fail security’ because they cannot recall passwords, alternative forms of ID’s checks, could be agreed by the banking industry as acceptable alternatives. In addition voice and face recognition can be introduced to in-branch, Internet and Mobile banking.</td>
</tr>
</tbody>
</table>
| Digital banking initiatives                | • Launch of a digital ambassador programme, which will proactively encourage and support older customers to embrace digital banking.  
• Supporting financial education, which will include initiatives to help customers improve their finances and capabilities.  
• Enhancing community engagement - demonstration days to be held in different branch locations. |

Table 4. 13 Frequent negative features recommendations
Chapter 5: Conclusion and Future Work

5.1. Introduction

This chapter provides conclusions inferred from this research. It summarises the research overview, problem definition, experiment design, evaluation and results as discussed in the previous chapters. It also discusses the contributions and impact of the experiment and future work and recommendations for further studies in this domain.

5.2. Research Overview

Single-question customer metrics such as the Net Promoter Score and Voice of Customer surveys are popular tools that are used widely in practice in an attempt to measure customer loyalty. The literature suggests that reliance on a simple single customer metric is a very risky trend and that companies should use sentiment analysis techniques to predict customer behaviour.

This project developed a system which uses supervised machine learning algorithms i.e. Naïve Bayes and Support Vector Machines to computationally identify and categorise sentiments expressed in a piece of text, in order to predict if a customer is a Promoter, Passive or Detractor of the AIB brand. The supervised classification approaches supported the known ‘Voice of the Customer’ programme NPS scores with reasonable accuracies of greater than 80%. Deeper analysis of the model results however showed that caution is required when relying on the machine learning output. Using the confusion matrices highlighted the fact that the model misclassified instances by classifying negative texts as positive sentiment. These errors are of importance as they can be expensive to the bank if operational decisions were made ignoring them. The accuracy of the models is only as good as the training data, therefore suggestions to improve the quality of the training set of the next iteration have been provided. A system, based on this research but following much testing of the data and controlled trials could have the potential to provide valuable insights for AIB Management. These insights may inform initiatives in the areas of marketing and customer engagement. Even small improvements in these areas can be crucial in a competitive environment.
Predicting a customer age based on the sentiment provided produced unreliable results and showed less promise due to the small sample size. This could be addressed in future iterations by expanding the data set to get more data on digital banking, potentially from other banks if possible.

In addition to machine learning, descriptive statistics with a blend of sentiment analysis techniques extracted positive and negative features to understand what customers likes and dislikes which can be examined for more age friendly banking.

It was discovered that decision trees are not a good choice for sentiment analysis in this experiment because they did not suit the high dimension data set. The training data performed the worst of all the algorithms with accuracies of less than 30%. The visualisation was far too ‘noisy’ to be of much use in telling a story about the data. Indeed, when it was decided to use the decision tree in ‘R’ using a much smaller dimension space instead of the full vocabulary, exploring features other than text sentiment, decision trees were found not to be a suitable technique with this particular data set.

5.3. Problem Definition

The research problem was defined by two questions: “Can text mining provide useful insights into data available to the bank through ‘Voice of Customer’ surveys?”

And also: “To what extent is age a meaningful factor in influencing customer sentiments and NPS scores of AIB’s digital banking?”

The research used different classification algorithms with a goal of classifying customer’s sentiment polarity to inform customer satisfaction improvements. This showed promising results but with indications for caution to be used. Predicting a customer age based on the sentiment provided produced unreliable results and showed less promise. Descriptive statistics analysis showed that there was a difference of approximately 27% in customer dissatisfaction of digital banking when compared to other banking journeys. The extent to which the ‘Traditionalist’ and ‘Baby Boomers’ contributed to this dissatisfaction was marginal and the research experiment did not shed further light on this. In addition, negative feature extraction of digital banking sentiments from detractors provided some interesting data insights in which the bank can consider for age friendly banking.
5.4. **Experimentation, Evaluation & Results**

The approach to perform the experiment was comprehensive as it included investigation of the data with actionable insights, machine learning and sentiment analysis techniques such as negative feature extraction. Python Scikit-learn was chosen to perform a blend of sentiment analysis and machine learning because it is a well-known resource for text mining. The classifiers were chosen based on exhaustive literature review. In addition, the feature extraction of negative sentiments was performed in order to provide recommendations on how to make digital banking better.

This results here provide enough evidence to accept the research hypothesis that using a blend of machine learning, descriptive statistics and sentiment analysis has the potential to inform customer satisfaction improvements of AIB’s digital banking journeys.

- The machine learning classifiers supported the known VOC programme NPS results and produced similar classification results with reasonable accuracy. Similar results were produced from each model suggesting a valid output. The machine learning classifiers that produced the best results in terms of highest percentage of accuracy was Support Vector Machines but Naïve Bayes was not significantly different.

- Even with the reasonable accuracies, the robustness of the models is only as good as the training data. Recommendations to improve the training data in the next iteration include 1) expanding the data set to include more data 2) investigate the error metrics from the confusion matrix and action any findings 3) using alternative labelling methods such as crowd sourcing or using experts to label the data set 4) Using domain specific lexicon for banking or finance.

- Decision trees are not always good choice for Sentiment Analysis because they do not suit high dimensionality. A smaller set of features which can define the data set in a much smaller dimension space was used instead of all the vocabulary. Exploring decision trees in ‘R’ using other features of the data set including age, sex, county code, business or personal customers did not provide more insights using this data set.

- Predicting a customer age based on the sentiment provided produced unreliable results and showed less promise, due to the small sample size. This could be
addressed in future studies by using a larger sample size and other recommendation described to improve the training data set.

- Descriptive statistics provided interesting insights into the data including:
  - Digital banking journey’s scored lower in NPS scores compared to other banking journeys.
  - Digital banking journey’s scored lower in NPS scores compared to industry standard.
  - Age is not a significant contributing factor to dissatisfaction of digital banking.
  - Sample size of traditionalists was small. Assumption that non-response is due to lack of capability (knowledge/competence/tools) of older customers.
  - Highest % of returned surveys was in the 25-35 or the ‘Generation Y’ bracket, thus supporting the assumption that as the customer gets older they are less likely to respond to the surveys.
  - Not surprisingly the biggest % of respondents is from the two largest cities in Ireland -Cork and Dublin.
- Negative feature extraction of digital banking journeys determined that detractors which include older customers are fearful of machines, prefer the personal touch, find systems difficult to navigate, dislike branch closures and prefer a customer centric approach. The negative feature extraction allowed for recommendations to be made for more age friendly banking.

5.5. **Contributions and impact**

This study may contribute to the body of knowledge of sentiment analysis as it combines sentiment analysis, voice of customer, user characteristics of older customer and the digital banking domain. It may be of extra value as it is focused on a narrow domain area, that of customer sentiment towards an Irish bank. The results may prove useful for AIB decision makers in support of their existing data analytics programmes.

Future developments and refinements, based on this research could be applied to models designed to compute the sentiment in other fields where ‘net promoter score’ plays an important role. It could be used, for example, to assist with misclassified NPS categories which, for example, (Kenningham et al. 2007) studies show nearly 25% of NPS scores
did not always match actual customer behaviour or (Johnson, 2005) extreme responding behaviour in surveys or (Wright, 2017) potential bias found in survey responses.

5.6. Future Work & recommendations

For future research, it would be interesting to validate the results by conducting the study on a larger scale, one could incorporate data from all European banks to widen the sample size.

The approach and techniques examined could also be used for other document types, for example AIB twitter texts or other kinds of user-generated content e.g., recordings of customer conversations or discussion groups. Regarding the methodology, it would be interesting to see how alternative algorithms would affect the results. In addition to machine learning methods, a lexicon based sentiment analysis could be performed using freely available sentiment dictionaries such as SentiStrength or Opinion Finder or commercially available lexicons. This research currently focuses on text only classification. In the future, the system could be trained to process emoticons such as emoji’s in reviews.

There are also several other methods for capturing older generation feedback rather than the online surveys which in this case had a small sample size and potential bias due to non-response, cultural and age related factors. Research suggests that interviews, observation and prototypes are the best way to elicit user experience from older customers. Evaluation is a key ingredient for a successful inclusion of this cohort of customers. A bank that improves its systems to assist its older customers and most vulnerable customers is likely benefit the rest of its customers.
6. References


7. Appendix

7.1. Appendix A: Pre-Processing Scripts and Sample outputs

This section presents code, imported libraries in which to execute the steps that was conducted as a part of the study but hasn’t been included in the chapters of this report.

Import various libraries in Python via Anaconda GUI- Jupyter

<table>
<thead>
<tr>
<th>Ref</th>
<th>Python Libraries</th>
<th>What do they do?</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td><code>import pandas as pd</code></td>
<td>Pandas is an open source, BSD-licensed library providing high-performance, easy-to-use data structures and data analysis tools for the Python programming language.</td>
</tr>
<tr>
<td>2</td>
<td><code>import seaborn as sns</code></td>
<td>Seaborn is a Python data visualization library based on matplotlib. It provides a high-level interface for drawing attractive and informative statistical graphics.</td>
</tr>
<tr>
<td>3</td>
<td><code>import numpy as np</code></td>
<td>NumPy, which stands for Numerical Python, is a library consisting of multidimensional array objects and a collection of routines for processing those arrays. Using NumPy, mathematical and logical operators on arrays can be performed. This tutorial explains the basics of NumPy such as its architecture and environment.</td>
</tr>
<tr>
<td>4</td>
<td><code>import matplotlib.pyplot as plt</code></td>
<td>Matplotlib is a Python 2D plotting library which produces publication quality figures in a variety of high quality formats and interactive environments across platforms. For simple plotting the <code>pyplot</code> module provides a MATLAB-like interface, particularly when combined with IPython.</td>
</tr>
<tr>
<td>5</td>
<td><code>from sklearn.ensemble import RandomForestClassifier</code></td>
<td>A random forest is a meta estimator that fits a number of decision tree classifiers on various sub-samples of the dataset and uses averaging to improve the predictive accuracy and control over-fitting.</td>
</tr>
<tr>
<td>6</td>
<td><code>from sklearn.feature_extraction.text import CountVectorizer</code></td>
<td>Word Counts with CountVectorizer. The CountVectorizer provides a simple way to both tokenize a collection of text documents and build a vocabulary of known words, but also to encode new documents using that vocabulary. You can use it as follows: Create an instance of the CountVectorizer class.</td>
</tr>
<tr>
<td>7</td>
<td><code>from sklearn.feature_extraction.text import TfidfVectorizer</code></td>
<td>TfidfVectorizer (TF-IDF) is an abbreviation for Term Frequency-Inverse Document Frequency and is a very common algorithm to transform text into a meaningful representation of numbers. The technique is widely used to extract features across various NLP applications.</td>
</tr>
<tr>
<td>8</td>
<td><code>from sklearn.svm import SVC</code></td>
<td>The objective of a Linear SVC (Support Vector Classifier) is to fit to the data you provide, returning a “best fit” hyperplane that divides, or categorizes, your data.</td>
</tr>
<tr>
<td>9</td>
<td><code>from sklearn import svm</code></td>
<td>A Support Vector Machine (SVM) is a discriminative classifier formally defined by a separating hyperplane. In other words, given labeled training data (supervised learning), the algorithm outputs an optimal hyperplane which categorizes new examples.</td>
</tr>
<tr>
<td>10</td>
<td><code>from sklearn.metrics import confusion_matrix, classification_report</code></td>
<td>Compute confusion matrix to evaluate the accuracy of a classification. By definition a confusion matrix C is such that C[i,j] is equal to the number of observations known to be in group i but predicted to be in group j. Thus in binary classification, the count of true negatives is C[0,0], false negatives is C[1,0], true positives is C[1,1] and false positives is C[0,1].</td>
</tr>
<tr>
<td>11</td>
<td><code>from sklearn.preprocessing import StandardScaler, LabelEncoder</code></td>
<td>Pre-processing refers to the transformations applied to your data before feeding it to the algorithm. In python, scikit-learn library has a pre-built functionality under sklearn.preprocessing.</td>
</tr>
<tr>
<td>12</td>
<td><code>from sklearn.model_selection import train_test_split</code></td>
<td>Split the data into training and Testing datasets.</td>
</tr>
<tr>
<td>13</td>
<td><code>get_ipython().magic('matplotlib inline')</code></td>
<td>%matplotlib is a magic function in IPython. %matplotlib inline sets the backend of matplotlib to the 'inline' backend. With this backend, the output of plotting commands is displayed inline within frontend like the Jupyter notebook, directly below the code cell that produced it.</td>
</tr>
</tbody>
</table>
Pre-processing steps code and sample outputs

- Step 1: Split into Sentences: A good useful first step is to split the text into sentences.

Script is as follows:
filename = 'Full_Data set.csv'
file = open(filename, 'rt')
text = file.read()
file.close()
# split into sentences
from nltk import sent_tokenize
sentences = sent_tokenize(text)
print(sentences[0])

Sample Output:
dummy customer id      VERBATIM
   o Very unhappy with the service

- Step 2: Split into Words.

Script is as follows:
filename = 'Full_Data set.csv'

file = open(filename, 'rt')
text = file.read()
file.close()
# split into words
from nltk.tokenize import word_tokenize
tokens = word_tokenize(text)
print(tokens[:100])

Sample Output:
['The', 'time', 'line', 'was', 'very', 'slow', ',', 'aib', 'very', 'good', 'and', 'sporting', 'There', 'was', 'an', 'absolute', 'lack', 'of', 'communication', 'from', 'branch', 'and', 'HQ', '.', 'The', 'mortgage', 'advisor', 'was', 'beyond', 'his', 'capabilities', 'There', 'were', 'no', 'solid', 'reasons', 'for', 'decisions', 'made', '.', 'It', 'is', 'not', 'branch', 'issue', 'it', 'is', 'across', 'the', 'country', '.', 'The', 'decision', 'making', 'process', 'has', 'been', 'drawn', 'out', 'now', '.', 'It', 'is', 'going', 'on', '6-8', 'weeks', '.', 'No', 'one', 'seems', 'to', 'know', '.', 'Bank', 'refused', 'to', 'top', 'up', 'a', 'loan', 'despite', 'my', 'good', 'ability', 'to', 'repay', 'it', '..', 'Not', 'accepted', ',', 'they', 'do', 'not', 'take', 'any', 'past']

- Step 3: Filter punctuation.

Script is as follows:
filename = 'Full_Data set.csv'
file = open(filename, 'rt')
text = file.read()
file.close()
# split into words
from nltk.tokenize import word_tokenize
tokens = word_tokenize(text)
# remove all tokens that are not alphabetic
words = [word for word in tokens if word.isalpha()]
print(words[:100])
Sample Output:
['The', 'time', 'line', 'was', 'very', 'slow', 'aib', 'very', 'good', 'and', 'sporting', 'There', 'was', 'an', 'absolute', 'lack', 'of', 'communication', 'from', 'branch', 'and', 'HQ', 'The', 'mortgage', 'advisor', 'was', 'beyond', 'his', 'capabilities', 'There', 'were', 'no', 'solid', 'reasons', 'for', 'decisions', 'made', 'It', 'is', 'not', 'branch', 'issue', 'it', 'is', 'across', 'the', 'country', 'The', 'decision', 'making', 'process', 'has', 'been', 'so', 'drawn', 'out', 'now', 'It', 'is', 'going', 'on', 'weeks', 'No', 'one', 'seems', 'to', 'know', 'Bank', 'refused', 'to', 'top', 'up', 'a', 'loan', 'despite', 'my', 'good', 'ability', 'to', 'repay', 'it', 'Not', 'accepted', 'they', 'do', 'take', 'any', 'past', 'history', 'into', 'account', 'its', 'all', 'based', 'on', 'future', 'predictions', 'After', 'years', 'in']

- Step 4: Filter out Stop Words

Script is as follows:
```python
from nltk.corpus import stopwords
stop_words = stopwords.words('english')
print(stop_words)
```

Sample Output:
['i', 'me', 'my', 'myself', 'we', 'our', 'ours', 'ourselves', 'you', "you're", "you've", "you'll", "you'd", "your", "yours", "yourself", "yourself", "he", "him", "his", "himself", "she", "she's", "her", "hers", "herself", "it", "it's", "its", "itself", "they", "them", "their", "theirs", "themselves", "what", "which", "who", "whom", "this", "that", "that'll", "these", "those", "am", "is", "are", "was", "were", "be", "been", "being", "have", "has", "had", "having", "do", "does", "did", "doing", "a", "an", "the", "and", "but", "if", "or", "because", "as", "until", "while", "of", "at", "by", "for", "with", "about", "against", "between", "into", "through", "during", "before", "after", "above", "below", "to", "from", "up", "down", "in", "out", "on", "off", "over", "under", "again", "further", "then", "once", "here", "there", "when", "where", "why", "how", "all", "any", "both", "each", "few", "more", "most", "other", "some", "such", "no", "nor", "not", "only", "own", "same", "so", "than", "too", "very", "s", "t", "can", "will", "just", "don't", "should", "should've", "now", "'d", "'ll", "'re", "'ve", "'d", "ain't", "aren't", "can't", "couldn't", "didn't", "doesn't", "hadn't", "hasn't", "haven't", "isn't", "ma", "mightn't", "mustn't", "oughtn't", "shan't", "shouldn't", "w asn't", "weren't", "weren't", "won't", "wouldn't", "wouldn't"]

Script is as follows:
```python
# filter out stop words
from nltk.corpus import stopwords
stop_words = set(stopwords.words('english'))
words = [w for w in words if not w in stop_words]
print(words[:100])
```

Sample Output:

Step 5: Filter out Stem words

Script is as follows:
```python
filename = 'Full_Data_set.csv'
file = open(filename, 'rt')
text = file.read()
file.close()
# split into words
from nltk.tokenize import word_tokenize
tokens = word_tokenize(text)
# stemming of words
```
from nltk.stem.porter import PorterStemmer
porter = PorterStemmer()
stemmed = [porter.stem(word) for word in tokens]
print(stemmed[:100])

**Sample Output**

['the', 'time', 'line', 'wa', 'veri', 'slow', '.', 'aib', 'veri', 'good', 'and', 'sport', 'there', 'wa', 'an', 'absolut', 'lack', 'of', 'commun', 'from', 'branch', 'and', 'HQ', '.', 'the', 'mortgag', 'advisor', 'wa', 'beyond', 'hi', 'capabl', 'there', 'were', 'no', 'solid', 'reason', 'for', 'decis', 'made', '.', 'It', 'is', 'not', 'branch', 'issu', 'it', 'is', 'across', 'the', 'countri', '.', 'the', 'decis', 'make', 'process', 'ha', 'been', 'so', 'drawn', 'out', 'now', '.', 'It', 'is', 'go', 'on', '6-8', 'week', '.', 'No', 'one', 'seem', 'to', 'know', 'bank', 'refus', 'to', 'repay', 'my', 'good', 'abil', 'to', 'repay', 'it', 'not', 'accept', 'they', 'do', 'n't', 'take', 'ani', 'past']

**Split the available data set into Training and testing script**

**Script is as follows**

x_train, x_test, y_train, y_test=
train_test_split X, y test size =0.25
7.2. Appendix B: Training and Test scripts and sample outputs

Support Vector Machines Script

'Support Vector Machines
clif_promoter = svm.LinearSVC(C=1.0,class_weight=None,multi_class='ovr',dual=True,Max_iter=3000)
clif_detractor = svm.LinearSVC(C=1.0,class_weight=None,multi_class='ovr',dual=True,Max_iter=3000)
clif_passive = svm.LinearSVC(C=1.0,class_weight=None,multi_class='ovr',dual=True,Max_iter=3000)

clif_promoter.fit(x_trainVect,y_train_promotor)
clif_detractor.fit(x_trainVect,y_train_detractor)
clif_passive.fit(x_trainVect,y_train_passive)

print(clif_promoter.coef_)
print(clif_detractor.coef_)
print(clif_passive.coef_)
print(clif_promoter.predict(x_testVect)
clif_promoter.score(x_trainVect)

Naïve Bayes Script

'Naïve Bayes
clif_NB = clf = MultinomialNB()
clif_NB.fit(xtrainVect,target)
clif_NB.predict(xtrainVect)
clif_NB.score(xtrainVect)

Bag of Words Script is as follows:

#!/usr/bin/python
file=open("Self Service data set negative.txt")
wordcount={}
for word in file.read().split():
    if word not in wordcount:
        wordcount[word] = 1
    else:
        wordcount[word] += 1
print (word,wordcount)
file.close();

Sample Output Negative:

1  OUTPUT: