



A data driven approach to analysing qualitative customer feedback - Understanding the sentiment of reed.co.uk customer's sales calls



**Tutor:** Thierry Chausalet

## Table of Contents

Abstract .....	3
Introduction.....	3
Background .....	3
Problem Domain.....	4
Objective .....	4
Deliverables.....	5
Further Applications.....	5
Approach .....	5
Classification techniques .....	6
Text view .....	6
Rating level.....	6
Resource & Tools .....	6
Risks & Potential Mitigations.....	7
Project plan .....	7
<b>Bibliography.....</b>	<b>8</b>
<b>Appendix .....</b>	<b>9</b>

## Abstract

In an ever growing competitive market, it is becoming increasingly vital for businesses to listen to their customers and attempt to action their feedback. This has been enhanced by the growing trend of customers basing their purchasing habits on opinion. It is the aim of this project to enable reed.co.uk to gain an understanding of the sentiment that their customers have of their products, services and business as a whole. This will be achieved by the creation and implementation of a sentiment analysis tool linked to the current CRM system, using sales call data as the input. I believe that by gaining this insight, the business is able to make better product decisions and action negative feedback promptly, thus encouraging better client relationships and retention.

## Introduction

Gaining an insight into the opinions of your customers has long been a strategy employed by the most successful and forward thinking businesses, in their quest to provide unrivalled customer service. Bill Gates was famously quoted as saying that for a business “your most unhappy customers are your greatest source of learning” [1], which is something that has allowed Microsoft to achieve their levels of greatness.

However, it is not just the most disgruntled customers that offer good insight for a business, where a good understanding of positive feedback is also important, allowing a business to understand what they are doing right and thus what they should continue to remain competitive. A good understanding of these “sentiments” can allow for [2]:

1. **Better customer loyalty and retention** – For a business, it is far more cost effective to retain customers, rather than having to spend to attract new business. Listening to customers will enable a better retention rate and reduce the churn of a business.
2. **Happy customers refer** – By utilising customer feedback and acting upon, customers are likely to feel as though they are being listened to and thus are more likely to recommend your services.
3. **Competitive advantage** – Understanding what your customers like and dislike provides the best proxy of what the wider market is demanding. Thus a strong customer focused strategy means a business can be one step ahead of competitors.

## Background

Sentiment analysis, which is the process of extracting the emotions, opinions or thoughts towards particular products, services, individuals or events [3] is a powerful tool, allowing for major insights to be uncovered. The mining of text to gain an insight into the sentiment of customers is not necessarily a new phenomenon, where work can be traced back to 1979. The work of **Carbonell, J.G.** [4], which focused on the sentiment of political opinion showed early on that there was value in trying to interpret people’s sentiment. In recent years, there has been an explosion in the interest of the practice, where its application has become far more prevalent. Its utilisation is usually applied within the realms of social media, where sentiment is typically drawn from sources such as Twitter and Facebook [5][6][7][8]. The reason social media presents such a good opportunity is the share volume and rapid pace at which data can be obtained.

Many have highlighted the great opportunities that exist from the use of opinion-rich resources such as online reviews, blogs and other user generated content [9]. This content allows for a readily available and often freely surrendered source of opinion. Whereas the major audience for sentiment analysis tools has been individuals, Pang & Lee recognise that there is a growing demand from “companies anxious to understand how their products and services are perceived” [9].

Sentiment analysis has evolved from simply being a tool used to give an insight of current opinion, to a tool that can be used on a more predictive basis [10]. The automotive industry has

taken to this predictive approach fairly rapidly, whereby companies such as Buick have utilised negative feedback received via social media to dictate the production of proposed car models [11].

Although numerous studies on sentiment analysis have focused on using social media as the primary source of opinion, the subject of recruitment and its related products have a very limited citing in social spaces. Searching twitter for the term reed.co.uk returns a very limited numbers of relevant tweets, highlighting the lack of opportunity that social media presents to understanding customer sentiment to a business of reed.co.uk's nature. Based on this, the use of call logs presents a robust alternative [12] to getting an understanding of customer opinions in a readily available format.

## Problem Domain

reed.co.uk, as the UK's number one traditional job board has a customer base of some 20,000 recruiters that span the UK. They represent a mix of recruitment and advertising agencies as well as directly purchasing clients. Like many online businesses, reed.co.uk has a number of touchpoints (**Fig. 2**) during the customer's journey, however, one of the most frequent interactions is with the sales team via the call centre.



**Fig.2: Customer touchpoints [13]**

As aforementioned, there are clear benefits to be had by a company utilising the feedback that their customers give, however it is not always easy to do, due to the manual nature of the task. For reed.co.uk, over 150,000 calls, emails and accompanying notes are logged in the CRM system (Salesforce) each month, with little to no analysis, meaning there is a wealth of data with no use to the business. Coupled with low response rates (less than 5% response rate in last 365 days) to customer surveys, reed.co.uk has a limited understanding of customer sentiment to the brand, products and services they provide.

To overcome this lack of insight into customer sentiment and insight, I propose to utilise a sentiment analysis tool which will draw on some of the most readily collected, but least used data the business has.

## Objective

I propose to create a sentiment analyser for commercial usage at reed.co.uk, which will allow the business to extract the sentiment of the sales calls that are logged by sales executives.

The use of data mining and sentiment analysis on the call/email logs will provide:

- A clearer idea of customer sentiment on an account by account basis, where I should be able to establish a sentiment for each interaction.
- Looking more holistically, aggregating and allowing for grouping by specific time periods, will make it possible for the account holder to easily ascertain the sentiment of the customer to reed.co.uk.
- Highlights into some of the key themes, products and services, so that the business has an insight into which are being discussed.

By accomplishing these objectives, it is my hypothesis that reed.co.uk will be able to have better customer satisfaction and retention rates. These are easily measureable metrics for the business, as they currently record the churn rate and NPS scores for their clients. Further, these objectives allow me to advance my skills in the essential tools for sentiment analysis, such as Python and R, as well as providing me the challenge of trying to find the best methods to employee to gain maximum accuracy.

The importance of this project to the business stems from a current inability to analyse the sentiment of customers. Where previous work in this field has focused on data that has been obtained from social media sources, my aim is to use call log data which has had minimal use within sentiment analysis, especially within the commercial context.

## Deliverables

Based on the aforementioned objectives, the key deliverables of the project will be:

1. A collection of individual sales interactions, with sentiment classifications (For those applicable)
2. An overall sentiment score for a given customer (Salesforce account)
3. A business wide sentiment score, based on active\* customer base
4. Topic areas, reflecting the conversations, categorised by sentiment.

\* Customers who have purchased in the last 12 months.

By delivering the anticipated deliverables, reed.co.uk is able to gain actionable insights into their customer's opinions of the business, their products and services they provide. This brings life to data that has been stored, but provided little to no use. Commercially, they are then able to gain a competitive advantage, placing them in a strong position in their market.

## Further Applications

Sentiment analysis of reed.co.uk's call data is simply the starting point of what is possible with the use of the proposed sentiment analysis tool. As previously stated, reed.co.uk has countless touch points with its customers and thus are able to gain a 360 perspective of their customers by integrating these sources. The OU (Organisational Unit) ID which is assigned to each customer as a unique identifier allows for this integration to be seamless.

Predicting the churn of customers based on the sentiment scores they are assigned will allow for the sentiment analysis to be used in a more predictive nature. This use in predictive analysis can even be furthered and applied to their marketing campaigns, where they are able to predict email response rates and engagements by sentiment groups.

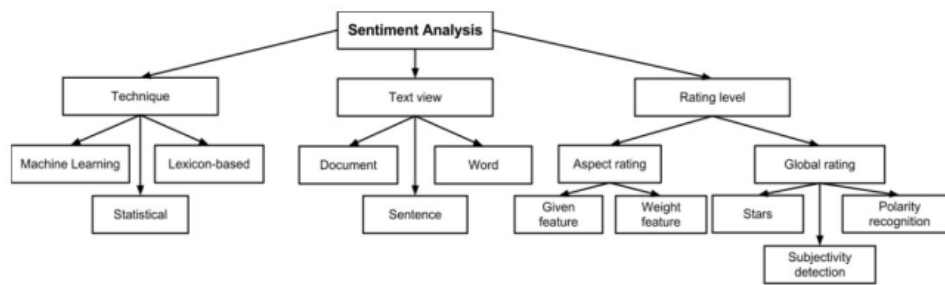
## Approach

Building a robust sentiment analysis tool is essential to the success of the project, where I will complete 5 key stages:



The initial stages involve the collections and pre-processing (text preparation), to ensure that the inputs into the sentiment analyser are optimal. Stages 3 and 4 are geared around the actual discovery and classification of sentiment that can be derived from the call log data. Finally, the last stage is focused around the presentation of the extracted data, which will take differing forms, depending on the chosen level. For some, a granular level is applicable and so sentiment can be seen on a document or account level. To cater to those more senior, an aggregated view of the data will present a more holistic view of customer sentiment.

The process of sentiment analysis can be classified from different perspectives, with the primary classifications being the technique employed, the view of the text being analysed and the “rating level” (**Fig.1**).



**Fig.1:** Differentiating sentiment analysis [14]

### Classification techniques

Producing a robust sentiment analysis model can be accomplished by employing a number of techniques. The main approaches that are available are machine learning, lexicon-based, statistical and rule-based approaches. These techniques were investigated by **Anais Collomb et al**, who explored their accuracy, concluding that there is still significant work that needs to be done in the field of sentiment analysis, where only a few of the techniques produced “a somewhat high level of accuracy” [14]. I will be exploring the options throughout this project, to see which technique produces the most accurate results.

### Text view

The chosen text view of a document is vital to the accuracy that can be obtained from a sentiment analyser. Analysing at word, sentence or even document level can produce differing sentiments. **Paltoglu G. et al** explains this difference well [15], where they use a film review extract to explain: ***“This film should be brilliant. It sounds like a great plot . . . a good performance. However, it can’t hold up.”*** Analysing this review at a sentence level, most classifiers would class this as positive sentiment, whereas taking the document as a whole, it becomes clear that although there are positive sentiments within the document, it is in fact a negative sentiment being conveyed.

I will be adopting a document level approach to sentiment analysis. Although there is the issues that the individual sentences may change the overall sentiment of the document, I believe that it is less likely in this scenario, due to the context of the conversations. With the use of human checking and evaluation, I will ensure that this is minimised.

### Rating level

When classifying sentiment, there are various approaches that can be adopted in terms of the rating system used. In many studies, implementers have opted for a more binary approach (polarity), which classifies each document into positive or negative sentiment [16]. However, others have taken it further [17], where they have implemented a scale classification for sentiment, producing outputs such as star ratings. These are usually achieved by using regression modelling, as the ratings are ordinal [18].

In order to keep the analysis of sentiment as simple as possible, I believe a binary approach is sufficient. This will allow the classification process to be more confident and negate the need for the tool to have to scale sentiment, which is an added complexity.

## Resource & Tools

In order to develop the sentiment analyser, there are a number of tools that will be employed which have been identified as the appropriate tools for the scenario [19]:

**Python** – Open source object-orientated programming language. Making use of modules including NLTK and MontyLingua.

**R** – Open source data analysis software

To complement the use of these tools, I will be making use of “Natural Language Processing with Python” [20], which provides a clear and concise way of implementing sentiment analysis tools.

## Risks & Potential Mitigations

During this project, there will be a number of risks that the project is exposed to. Their potential impact, likelihood and mitigations will vary.

Risk	Potential Impact	Likelihood	Mitigation
<b>Access to data</b> – Issues getting access to the required data	High	Low	Discuss project with the relevant stakeholders to gain permissions. (Already under way)
<b>Quality of data</b> – data may not be the best quality, making it hard to analyse	High	High	Usage of a specific group of sales executives, along with training. Also will use other sources of data as back-up.
<b>Open software integration</b> - Issues may arise when integrating the open source software to licenced software	High	Medium	Research integration techniques. Back-up will be to build a stand-alone program.
<b>Open software learnings</b> – There is limited support and documentation for open source software, which makes it harder to work with.	Medium	Medium	Liaising with software experts internally and externally of the business. Also, using the numerous online resources.
<b>Data security</b> – The data involved is very sensitive, so there is a fear that it may be at risk during the project.	High	Low	Liaise with the “Head of data and integration” to understand best-practices and how to ensure security of data.
<b>Scope creep</b> – There may be a temptation to continually add new features to the system, which could delay completion.	Medium	Low	Using a well-defined project timeline and clearly outlining the outputs will reduce this risk.
<b>Inappropriate hard/software</b> - Usage of low end hard/software will impact the performance of the system.	Medium	Low	Early testing of run-times and performance to ensure the final hard/software used are appropriate for the task.

## Project plan

For the project to run smoothly and by the anticipated end date, a details project plan has been constructed (**Appendix 1**) to highlight the key checkpoints in the project.



## Bibliography

- [1] **Gates, B. (1999)**, Business @ the speed of thought. New York, NY: Warner Books.
- [2] **Beard, R. (2013)**, 5 Reasons You Need To Be Tracking Customer Satisfaction. [online] Client Heartbeat Blog. Available at: <http://blog.clientheartbeat.com/tracking-customer-satisfaction/> [Accessed 8 May 2016].
- [3] **Liu, B. (2012)**, Sentiment analysis and opinion mining. San Rafael, Calif.: Morgan & Claypool.
- [4] **Carbonell, J.G. (1979)**, Subjective Understanding: Computer Models of Belief Systems (No. RR-150). YALE UNIV NEW HAVEN CONN DEPT OF COMPUTER SCIENCE.
- [5] **Tumasjan, A., Sprenger, T.O., Sandner, P.G. and Welpe, I.M., (2010)**, Predicting elections with twitter: What 140 characters reveal about political sentiment. *ICWSM*, 10, pp.178-185.
- [6] **Agarwal, A., et al (2011)**, June. Sentiment analysis of twitter data. In Proceedings of the workshop on languages in social media (pp. 30-38). Association for Computational Linguistics.
- [7] **Wang, X., et al., (2011)**, Topic sentiment analysis in twitter: a graph-based hashtag sentiment classification approach. In *Proceedings of the 20th ACM international conference on Information and knowledge management* (pp. 1031-1040). ACM.
- [8] **Ahkter, J.K. & Soria, S., (2010)**, Sentiment analysis: Facebook status messages. Unpublished master's thesis, Stanford, CA.
- [9] **Pang, B. & Lee, L., (2008)**, "Opinion Mining and Sentiment Analysis", *FNT in Information Retrieval*, vol. 2, no. 12, pp. 1-135
- [10] **Ostrowski, D.A., (2010)**, Sentiment mining within social media for topic identification. In *Semantic Computing (ICSC)*, 2010 IEEE Fourth International Conference on (pp. 394-401). IEEE.
- [11] **Berr, J., (2009)**, "GM scraps 'hideous' Buick after Twitter complaints", *The Sydney Morning Herald*. [Online]. Available: <http://www.smh.com.au/business/gm-scraps-hideous-buick-after-twitter-complaints-20090820-ervt.html>. [Accessed: 10- May- 2016].
- [12] **Overby, S., (2016)**, "Big Data Analytics Gold for the Call Center", *CIO*. [Online]. Available: <http://www.cio.com/article/2396132/customer-relationship-management/big-data-analytics-gold-for-the-call-center.html>. [Accessed: 11- May- 2016].
- [13] **Rajack, (2016)**, "Three steps to improve digital customer experience (CX) [APAC Case Studies]", *Econsultancy*. [Online]. Available: <https://econsultancy.com/blog/67201-three-steps-to-improve-digital-customer-experience-cx-apac-case-studies/>. [Accessed: 08- May- 2016].
- [14] **Collomb, A., Costea, C., Joyeux, D., Hasan, O. and Brunie, L., (2014)**, A Study and Comparison of Sentiment Analysis Methods for Reputation Evaluation
- [15] **Paltoglou, G. and Thelwall, M. (2013)**, More than Bag-of-Words: Sentence-based Document Representation for Sentiment Analysis.
- [16] **Wang, H., Can, D., Kazemzadeh, A., Bar, F. and Narayanan, S., (2012)**, A system for real-time twitter sentiment analysis of 2012 us presidential election cycle. In *Proceedings of the ACL 2012 System Demonstrations* (pp. 115-120). Association for Computational Linguistics.
- [17] **Pang, B. & Lee, L., (2005)**, "Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales," *Proceedings of the Association for Computational Linguistics (ACL)*, pp.115–124
- [18] **Liu, B., (2010)**, Sentiment Analysis and Subjectivity. *Handbook of natural language processing*, 2, pp.627-666.
- [19] **Miller, T., (2015)**, Modeling techniques in predictive analytics with Python and R. Upper Saddle River, NJ: Pearson Education
- [20] **Bird, S., et al., (2009)**, Natural language processing with Python. Beijing: O'Reilly



### Appendix 1: Proposed timeline for sentiment analysis project

# Project Planner

Plan

ACTIVITY	PLAN START WEEK	PLAN DURATION	Week													
			1	2	3	4	5	6	7	8	9	10	11	12	13	14
Gain business requirements	1	1														
Creat project proposal	1	1														
Brief business on proposed project	2	1														
Gain sign-off for proposal from business	2	1														
Sample data collection	3	2														
Data pre-processing including cleaning	5	2														
Development of sentiment analysis tool	6	4														
Testing & Evaluation	10	1														
Deliver findings back to business	11	1														
Write up of project	11	3														
Final project submission	14	1														