# Chapter 1 – Introduction

To determine whether there is a link between the sentiment of a TV show and the viewer ratings. This project will explore areas of Sentiment Analysis (SA), different ways to perform sentiment analysis, why is this relevant and how will others benefit from this. This will involve some data analysis and manipulation to find out if there are any correlations. This chapter explores the background of the research as well as a justification for it. The aims and objectives are also considered.

## 1.1 – Background

Sentiment analysis has been defined as opinion mining (Ding, et al. 2008) and according to Feldman (2013), sentiment analysis is used to look at the “decision-making process of people”. The value of this is we can better understand them as consumers, voters, reviewers etc.

## 1.2 - Justification

(why? Practitioner and academic value)

Feldman (2013) states that by using sentiment analysis it “offers these organizations the ability to monitor the different social media sites in real time and act accordingly”. This would give companies a much better understanding of their customers and can benefit from this.

## 1.3 - Aims and objectives (100 words)

The aim of this project is to explore the areas of sentiment analysis and to create a script which will look at the sentiment of an episode of a TV show and the viewer ratings and see if there is a link between them.

* To perform a literature review of sentiment analysis.
* To investigate the sentiment of a TV show, per episode/season.
* To investigate the viewer rating of a TV show from reviewer websites.
* To investigate if there is a link between both results.

## Deliverables

* Dissertation
* R-Script which will be created to do the sentiment analysis.
* User manual which will be created to guide other people through the script.

# Chapter 2 – Lit Review

According to Pang & Lee (2008) sentiment analysis has also been referring to it as ‘brand monitoring,’ ‘buzz monitoring’ and ‘online anthropology,’ to ‘market influence analytics,’ ‘conversation mining’ and ‘online consumer intelligence’.

## 2.1 - What does it do?

Sentiment analysis is a method of analysis which looks at the emotion of a word with the positivity and negativity of the said word. This style of analysis is used in marketing to measure the reviews of a service or product with the product reviews which is also what Taboada, et al (2011) states.

## 2.2 - Types of sentiment analysis

There are multiple types of sentiment analysis, which looks at different types of entities within a data set. These different types are called: Document-level sentiment analysis, Sentence-level sentiment analysis, Aspect-based sentiment analysis, Comparative sentiment analysis and Sentiment lexicon acquisition Feldman (2013).

### 2.2.1 – Document-level sentiment analysis

The first type of sentiment analysis which will be explored is Document-level. This type of sentiment analysis is known as the simplest form of as it looks at the whole document as one attribute (Feldman 2013). For an example of this, we could look at different types of reviews from Amazon and would give you an overall rating. This type can also be done with machine learning which consists of supervised and unsupervised learning (Bibi 2017).

Supervised sentiment analysis considers such algorithms As Bibi (2017) pointed out, “Naive Bayes, Maximum Entropy classification and Support Vector Machines (SVM).”

* Naive Bayes – has real time prediction, is very fast algorithm.
* Maximum Entropy Classification
* Support Vector Machines

With unsupervised, the approach is a little bit different. As it would need to have been given a certain threshold for the semantic orientation (SO), this would be like a level of positivity to make is overall positive or under making overall negative.

Advantages

* Can easily look at a document and give it an overall sentiment score.
* Can be done quickly.

Disadvantages

* Difficult to learn.

FIND REAL EXAMPLES – USE Bibi (2017)

Conclusion

Whilst Document level would be great for individual reviews with a single viewpoint, but when there are multiple entities there could also be different types of reviews in a single document. Including others which will be explored later in the report, such as comparative analysis.

### 2.2.2 - Sentence-level sentiment analysis;

The second type of sentiment analysis is Sentence-level. This type looks at each sentence as an individual entity, so will break down each sentence into an ‘opinion’. Looking at the emotion of each sentence and will show the overall sentiment at the end and how much the sentiment can differ between sentences, from positive, negative or neutral. This type of sentiment analysis would usually be used for the subjectivity classification and the sentiment classification (Bibi 2017).

(Look more into these)

FIND REAL EXAMPLES

Advantages

* Can easily look at a document and give it an overall sentiment score.
* Can be done quickly.

Disadvantages

* Difficult to learn.

Conclusion

### 2.2.3 - Aspect-based sentiment analysis

Aspect-based sentiment analysis is also known as feature-based sentiment which as stated by Feldman (2013) and is used to identify the sentiment of many attributes. Which can be useful when a person is talking about an overall experience but has different experiences at different parts. For example, if we were to look at a university course review, each module on the course would appeal differently to each person. One student could really enjoy one module but had a bad overall experience, with this type of analysis the data scientist can pin point the sentiment for each module. This would be effective to find one person’s sentiment over time for a certain subject.

FIND REAL EXAMPLES

Advantages

Disadvantages

Conclusion

### 2.2.4 - Comparative sentiment analysis;

Comparative sentiment analysis looks at the sentences which are comparing a product/service to a similar product/service. This would be great if comparing 2 similar products and seeing how they compare to one another. An example of this would to look at 2 products on a website and seeing how model 1 is better than model 2. If being an iteration of the same product they could easily compare them both to one another with advantages and disadvantages.

When looking at comparative analysis, there is a very interesting paper which helped to improve the accuracy of comparative analysis

Comparing a product to another & Looks for words like, (More, less, lighter) – Look for an existing example of comparative analysis!

FIND REAL EXAMPLES

Advantages

Disadvantages

Conclusion

### 2.2.5 - Sentiment lexicon acquisition

Lexicon based sentiment analysis is the most crucial resource (Feldman 2013), this is due to the use of dictionaries which can be hand coded and unique for a specific use case. Alternatively, the dictionaries can be crowd sourced, such as Bag of Words which uses a dictionary of positive and negative words which are all matched up against a score. This is done by following a calculation of:

$$\sum{positive\_matches} - \sum{negative\_matches}$$

Which was pointed out by LyonEye (2016), the scores are then normalised to the form of 1 to 5. There are also other dictionaries the user can choose from such as WordNet, which is described as a ‘Large lexical database of English nouns, verbs, and objectives’ (WordNet 2019).

FIND REAL EXAMPLES

Advantages

Disadvantages

Conclusion

Whilst this method of sentiment analysis is considered the most crucial, it can also prove difficult for when the context starts to get more complex which as stated by Ding, et al. (2018) “This approach allows the system to handle opinion words that are context dependent, which cause major difficulties for existing algorithms”.

## 2.3 – Methodologies

2.3.1 – Sentiment Analysis Model

For the methodology for sentiment analysis, the methodology would follow as shown:

Reviews > Data Preparation > Reviews Analysis > Sentiment Classification > Results

(Create a graph of this)

## 2.4 – Current Software (IBM & Google)

Sentiment Analysis has become so popular that even the big names in the tech industry have provided their own data analysis tools. In this part of the report we will explore these different types of software, which makes it a lot easier for people to use sentiment analysis and data analysis tools within the work place.

Note: Whilst most of these tools require a premium subscription to use, a lot of them provide a free trial and some of them are provided by the University of South Wales. Which If I use any premium features to show in this report, will be displayed with a \*.

The first sentiment analysis tool this report will explore is SAS which claimed to be the “Analytics Leader” (SAS 2019).

SAS - VISUAL TEXT ANALYTICS

https://www.sas.com/en\_us/software/visual-text-analytics.html

SPSS – Sentiment Analysis

<https://www.ibm.com/analytics/spss-statistics-software>

Google Cloud Natural Language

<https://cloud.google.com/natural-language/>

Watson Tone Analyzer

<https://www.ibm.com/watson/services/tone-analyzer/>

Amazon Comprehend

<https://aws.amazon.com/comprehend/>

# Chapter 3 – Design

## 3.1 – Use outputs

# Chapter 4 – Implementation (Prototype/Deliverable)

## 4.1 – Based on design

# Chapter 5 – Evaluation of prototype/deliverable

# Chapter 6 - Conclusions

# LSEPI – (Put in appendix in final)

## Legal

Software licenses – The aim of this project is to use open source tools which will give anyone the ability to follow along with this project.

* R - An open source front end for the programming language R, which is great for creating and manipulating scripts and data frames.
* R-Studio -
* Git - is an open source version-control system for keeping track of changes in code.
* GitHub is a website for developers to upload their code externally, which was built on Git and allows for collaboration. (Microsoft)
* Visual Studio Code - An open source text editor for developers.

## Social

Anonymity of user input – With the use of web scrapping for this project, this could be a difficult issue for some websites. For the types of websites this project will focus on, have a developer version which allows users to download and analyse their data sets. This has been confirmed in the Terms Of Service (TOS) and usually requires the user to create an account and tell the company why/what you’ll be doing with the data.

## Ethical

To lay out the rules for ethics, we would have to consider of how personal it can be from person to person. Whilst it can be interchangeable usually the work place would follow some general ethical concerns.

* To treat people fairly
* To respect the autonomy of individuals
* To act with integrity
* To seek the best results

## Professional

Five characteristics of a professional!?

A professional should be someone who has the following:

* A duty of care
* A responsibility for consequences

This also falls under the BCS Code of Conduct which is the body of British Computer Society, who sets out the professional standards required as part of their members.

Within the BSC Code Of Conduct they state that all members, no matter their rank or length of being a member will provide public interest.

(Look at BSC Code Of Conduct)

## Issues

# 

# References

Behdenna S., Barigou F., Belalem G. (2016) Sentiment Analysis at Document Level. In: Unal A., Nayak M., Mishra D., Singh D., Joshi A. (eds) Smart Trends in Information Technology and Computer Communications. SmartCom 2016. Communications in Computer and Information Science, vol 628. Springer, Singapore.

BCS (2019) *‘BCS, THE CHARTERED INSTITUTE FOR IT CODE OF CONDUCT FOR BCS MEMBERS’.* Available at: <https://cdn.bcs.org/bcs-org-media/2211/bcs-code-of-conduct.pdf> (Accessed 18/11/19).

Bibi, M. (2017) ‘*Sentiment Analysis at Document Level*’. Available at: <https://www.researchgate.net/publication/320729882_Sentiment_Analysis_at_Document_Level> (Accessed 30/10/2019)

Ding, X., Liu, B., & Yu, P, S. (2008) ‘A Holistic Lexicon-Based Approach to Opinion Mining’. Proceedings of the 2008 International Conference on Web Search and Data Mining., Pages 231-240, Palo Alto, California, USA. February 11 - 12, 2008.

Feldman, R. (2013) ‘Techniques and Applications for Sentiment Analysis’. *Communications of the ACM*, vol. 56, no. 4.

Jindal, N. & Liu, B. (2008) ‘*Opinion Spam and Analysis*’. Available at: <https://www.researchgate.net/publication/200044297_Opinion_Spam_and_Analysis> (Accessed 16/11/19)

Lexalytics (2019) *‘Sentiment Analysis Explained’* Available at: <https://www.lexalytics.com/technology/sentiment-analysis> (Accessed 18/11/19)

Lima, A, C, E, S., & de Castro, L, N. (2012) ‘Automatic Sentiment Analysis of Twitter Messages’. Fourth International Conference on Computational Aspects of Social Networks (CASoN). November 2012.

LyonEye (2016) ‘*Lexicon-based Bag of Words Sentiment Analysis’*. Available at: <https://smartcity.readthedocs.io/en/latest/BOW/> (Accessed 17/11/19)

Nair, V. (2017) ‘*The rise of big data*’. Available at: <https://www.bcs.org/content-hub/the-rise-of-big-data/> (Accessed 28/10/2019)

Pang, B. & Lee, L. (2008) ‘Opinion mining and sentiment analysis’. *Foundations and Trends in Information Retrieval*, 2(1-2), pp. 1-135.

Sas (2019)*‘Analytics Software & Solutions’.* Available at: <https://www.sas.com/en_gb/home.html> (Accessed 17/11/19)

Silge, J. & Robinson, D. (2017) *Text Mining with R*. 1st ed. O'Reilly Media.

Silge, J. ‘*tidytext’*. Available at: <https://github.com/juliasilge/tidytext> (Accessed 19/11/19)

Taboada, M. et al. (2011) ‘Lexicon-Based Methods for Sentiment Analysis’. *Computational Linguistics*, 37(2), pp. 267-307.

<http://ataspinar.com/2016/01/21/sentiment-analysis-with-bag-of-words/>

https://towardsdatascience.com/supervised-vs-unsupervised-learning-14f68e32ea8d