



Faculty of Cognitive Sciences and Human Development

**SENTIMENT ANALYSIS: ANALYZING ONLINE FOOD REVIEWS  
FOR BEST DISHES**

**Fadhila Binti Hassan**

BF  
311  
F144  
2019

**Bachelor of Science with Honours  
(Cognitive Science)  
2019**



1000283518

UNIVERSITI MALAYSIA SARAWAK

Grade: A-

Please tick one

Final Year Project Report

Masters

PhD

DECLARATION OF ORIGINAL WORK

This declaration is made on the 17 day of MAY year 2019.

**Student's Declaration:**

I, FADHILA BINTI HASSAN, 58587, FACULTY OF COGNITIVE SCIENCES AND HUMAN DEVELOPMENT, hereby declare that the work entitled, SENTIMENT ANALYSIS: ANALYZING ONLINE FOOD REVIEWS FOR BEST DISHES is my original work. I have not copied from any other students' work or from any other sources with the exception where due reference or acknowledgement is made explicitly in the text, nor has any part of the work been written for me by another person.

11 JUNE 2019

  
FADHILA BINTI HASSAN (58587)

**Supervisor's Declaration:**

I, MOHAMAD HARDYMAN BIN BARAWI, hereby certify that the work entitled, SENTIMENT ANALYSIS: ANALYZING ONLINE FOOD REVIEWS FOR BEST DISHES was prepared by the aforementioned or above mentioned student, and was submitted to the "FACULTY" as a \*partial/full fulfillment for the conferment of BACHELOR OF SCIENCE WITH HONOURS (COGNITIVE SCIENCE), and the aforementioned work, to the best of my knowledge, is the said student's work

Received for examination by:   
(MOHAMAD HARDYMAN BIN BARAWI)

Date: 11/6/19

I declare this Project/Thesis is classified as (Please tick (√)):

- CONFIDENTIAL** (Contains confidential information under the Official Secret Act 1972)\*  
 **RESTRICTED** (Contains restricted information as specified by the organisation where research was done)\*  
 **OPEN ACCESS**

I declare this Project/Thesis is to be submitted to the Centre for Academic Information Services (CAIS) and uploaded into UNIMAS Institutional Repository (UNIMAS IR) (Please tick (√)):

- YES**  
 **NO**

### Validation of Project/Thesis

I hereby duly affirmed with free consent and willingness declared that this said Project/Thesis shall be placed officially in the Centre for Academic Information Services with the abide interest and rights as follows:

- This Project/Thesis is the sole legal property of Universiti Malaysia Sarawak (UNIMAS).
- The Centre for Academic Information Services has the lawful right to make copies of the Project/Thesis for academic and research purposes only and not for other purposes.
- The Centre for Academic Information Services has the lawful right to digitize the content to be uploaded into Local Content Database.
- The Centre for Academic Information Services has the lawful right to make copies of the Project/Thesis if required for use by other parties for academic purposes or by other Higher Learning Institutes.
- No dispute or any claim shall arise from the student himself / herself neither a third party on this Project/Thesis once it becomes the sole property of UNIMAS.
- This Project/Thesis or any material, data and information related to it shall not be distributed, published or disclosed to any party by the student himself/herself without first obtaining approval from UNIMAS.

Student's signature: \_\_\_\_\_

Date: 11/6/19

Supervisor's signature: \_\_\_\_\_

Date: 11/6/19

Current Address:

64, Kpg Balan Hilir,  
96300 Dalat, Sarawak.

Notes: \* If the Project/Thesis is **CONFIDENTIAL** or **RESTRICTED**, please attach together as annexure a letter from the organisation with the date of restriction indicated, and the reasons for the confidentiality and restriction.

**SENTIMENT ANALYSIS: ANALYZING ONLINE FOOD REVIEWS FOR BEST  
DISHES**

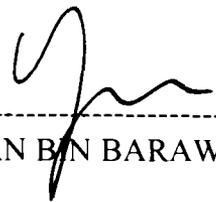
**FADHILA BINTI HASSAN**

This project is submitted  
in partial fulfilment of the requirements for a  
Bachelor of Science with Honours  
(Cognitive Science)

Faculty of Cognitive Sciences and Human Development  
UNIVERSITI MALAYSIA SARAWAK  
(2019)

The project entitled 'Sentiment Analysis: Analyzing Online Food Reviews for Best Dishes' was prepared by Fadhila Binti Hassan and submitted to the Faculty of Cognitive Sciences and Human Development in partial fulfillment of the requirements for a Bachelor of Science with Honours (Cognitive Science)

Received for examination by:



-----  
(MOHAMAD HARDYMAN BIN BARAWI)

Date:

-----  
21/6/19

Grade

A -

## **ACKNOWLEDGMENTS**

In the name of Allah, the Most Gracious and the Most Merciful. Alhamdulillah, praise to Allah the Almighty for the strengths and His blessing in completing this thesis. I would like to express my profound and sincere gratitude to my supervisor, Mohamad Hardyman Bin Barawi, for his supervision and constant support. His extensive knowledge has been of great value for me. His invaluable help of constructive comments and suggestions which were of critical importance have contributed throughout this process.

I would like to dedicate this work to my beloved parents, Hassan Bin Mut and Hadiyah Binti Arip, and also to my sisters who always believed in me. Thank you for your unconditional love and encouragement.

Special thanks to my friends for their advice and their continuous support, without your moral support and love, I could not have made it until the end.

**TABLE OF CONTENTS**

LIST OF TABLES.....	v
LIST OF FIGURES .....	vi
ABSTRACT.....	viii
CHAPTER ONE INTRODUCTION.....	1
CHAPTER TWO LITERATURE REVIEW .....	5
CHAPTER THREE METHODOLOGY .....	9
CHAPTER FOUR RESULTS .....	13
CHAPTER FIVE DISCUSSION.....	25
REFERENCES .....	28

## LIST OF TABLES

Table 1 Restaurants' Information .....	13
--	----

## LIST OF FIGURES

Figure 1 Sentiment Analysis Model .....	9
Figure 2 Python source code for data pre-processing.....	14
Figure 3 Example of pre-processing result for Aroma Cafe.....	14
Figure 4 Example of Python source code to generate bigrams.....	14
Figure 5 Example of bigrams result for Aroma Cafe .....	15
Figure 6 Source code to generate word cloud.....	15
Figure 7 Aroma Café word cloud .....	15
Figure 8 Bella Italia word cloud .....	16
Figure 9 Burger o Myy word cloud .....	16
Figure 10 Ceylonese word cloud .....	16
Figure 11 Dapur Makwa word cloud .....	16
Figure 12 Feast and Furious word cloud.....	17
Figure 13 Fork and Knife word cloud.....	17
Figure 14 Kings Curry word cloud .....	17
Figure 15 Kopi o Corner word cloud.....	17
Figure 16 Lepau word cloud.....	17
Figure 17 Lok Lok word cloud .....	17
Figure 18 Mom’s Laksa word cloud.....	18
Figure 19 New Atmosferah word cloud.....	18
Figure 20 RJ Ayam Bakar word cloud .....	18
Figure 21 Rumah Asap Dayak Samarahan word cloud.....	18
Figure 22 Rumah Asap Tabuan Dayak word cloud.....	18
Figure 23 Sharing Downtown word cloud.....	19
Figure 24 The Container word cloud.....	19

Figure 25 Tok Janggut word cloud .....	19
Figure 26 Example Python source code of VADER for sentiment classification .....	20
Figure 27 Sample Output for VADER sentiment classification .....	20
Figure 28 Example Python source code to generate pie chart .....	20
Figure 29 Sentiment classification pie chart of Tok Janggut restaurant .....	21
Figure 30 Sentiment classification for each restaurant .....	22
Figure 31 Sentiment classification for each restaurant .....	23
Figure 32 Sentiment classification for each restaurant .....	24

## **ABSTRACT**

The Internet is a massive repository of natural language. People share their personal opinions and experiences which are subjective in nature. Every day, millions of online informal reviews are generated by non-experts on various sites, about places, products and services. It is time consuming for most customers to read, comprehend and make decisions based on all of these restaurant reviews. Sentiment analysis allows analysing of the existing reviews and extract meaningful element. In this paper, data on restaurant reviews were analysed and from the bigrams method, the top dishes of each restaurants were extracted. Next, customer's opinions were analysed to understand their sentiment towards a particular restaurant and classified into positive, negative and neutral opinions by a lexicon known as Valence Aware Dictionary and sEntiment Reasoner (VADER). Hence, a summary of positive and negative reviews was generated. The results were visualized in form of word cloud which highlight the restaurant's top dishes while the positive and negative review counts of each restaurant were represented in pie chart.

*Keywords:* restaurant reviews, sentiment analysis, opinion mining, VADER

# **CHAPTER ONE**

## **INTRODUCTION**

### **Introduction**

This chapter outlines the background of study, the problem statement, the purpose and objectives of the study. It also discusses the significance of study and scope of study in this chapter.

### **Background of Study**

Over the last decade, the Internet is the most commonly used and effective medium of communication (Lagrosen, 2005). The amount of information available online increased along with the internet use. Now, internet users experienced a great opportunity for information search from the internet as they are able to share and acquire information easily due to its lower search costs and efficient access to the related information (Moon, 2004).

In recent years, an increasing number of various sites introducing online platform offers variety of choices in different categories of product such as restaurants, hotels, movies and so on for online users (Choi & Chihyung, 2011). Moreover, users can access multiple sites in the field of foodservice business such as TripAdvisor, Yelp, Zomato and others (Holleschovsky & Constantinides, 2016). User reviews are one of major source of information for online users to choose where to go or what to eat among various options (Kim, Maslowska & Malthouse, 2017). According to Chen, Chen and Wang (2017), user review is an unstructured textual information. Furthermore, every restaurant consumer able to write their online reviews in the available online platform. The reviews are open-ended comments where restaurant consumers wrote their personal dining experience, giving information and evaluating the product as well (Anjum & Dev, 2016). Consequently, restaurant consumers can make their personal reviews increasingly accessible to virtual communities or individual

who relies on the information as a factor facilitating their purchasing decisions (Dellarocas, 2003).

With the undeniable growth and popularity of online platforms for user opinion towards products, online product reviews influence consumer purchase decisions (Zhao, Wang, Guo & Law, 2015). Online reviews also provide useful information about the product and services and do recommendation (Zhang, Law, Ye & Li, 2010). In Holleschovsky and Constantinides (2016) findings, online reviews influence users in their purchasing behavior. This is because when potential consumers have a little knowledge of the product, they tend to use available resources such as guidance from others' opinions and recommendations (Tolon & Ozkan, 2015). As reported by a survey by marketing firm Comscore, nearly 81% of Internet users have performed online research on a product at least once and, among them, up to 87% reported that reviews have a significant influence on their purchases (Cataldi, Ballatore, Tiddi, & Aufaure, 2013). The consumers involved in this survey revealed to be willing to pay for products that received very positive reviews. Additionally, potential restaurant consumers usually went through previous consumers' reviews to seek for information ahead of time (Parikh, Behnke, Vorvoreanu, Almanza & Nelson, 2014). It is also noted by Parikh et al. (2014), user reviews and recommendations help consumers to make restaurant selections beforehand.

Advanced in text mining and sentiment analysis allows analysing of the existing consumers' reviews and also extracting meaningful elements (Dong, Mahony, McCarthy & Smyth, 2015). A study conducted by Anjum and Dev (2016) shows that the reviews provides extra information about the best dishes available at particular restaurants and states the positive or negative sentiment of dishes popularity. Sentiment analysis, also known as opinion mining, is the process of determining a text unit is positive or negative (Yu, Zhou, Zhang & Cao, 2017). It was divided in three levels: document level, sentence level and

feature level. The sentiment analysis at sentence level classifies sentence into a positive, negative or neutral sentiment for a product (Patil & Yalagi, 2016). This level is used for reviews that consists of one sentence and provided by the user (Patil & Yalagi, 2016).

## **Problem Statement**

Nowadays, rise of user-generated content (UGC) leads to information overload towards users (Monanhal, Prabhu & Yuan, 2016). The need to solve the problems regarding these unstructured opinionated data contributes to the importance of sentiment analysis field. In addition, restaurant reviews posted via online provide rich information in terms of food, service quality, restaurant atmosphere, and price (Yang, Hlee, Lee & Koo, 2017). The reviews are a large quantity of data, which leads difficulty in searching for useful information.

There are two significant problems that need to be solved. People usually focused more on the existing reviews to narrow down their personal selection of restaurants in terms of food. However, consumers encounter an enormous number of alternative recommendations of food and places and it will be time consuming to sift through many reviews in order to get recommendation based on their specific interest. The problem arises because of the overwhelming number of reviews making it is almost impossible for potential consumers to go through all reviews and attain the information they are looking for. Next, sentiment analysis would not be able to create the local food names word dictionary automatically. Therefore, the word dictionary needs to be created manually.

In response to the problem mentioned previously, user reviews were analyzed and extract meaningful information, such as food names, to know the popular dishes at the restaurant.

## **Purpose**

The purpose of this study is to analyze the user reviews and provide potential consumers with the top dishes at the restaurant.

## **Objective of Study**

- To perform sentiment analysis towards user reviews to know sentiment polarity of their reviews
- To extract food names of restaurants from the user reviews

## **Significance of Study**

The findings of this study were redounded to the benefit of potential restaurant customers to know the right restaurants that provide dishes matched with their tastes. Therefore, customers able to experience opportunity to try new dishes which they have never knew before but are among the best dishes offered by the restaurant. From the restaurant owner view, they able to know their popular dishes among previous consumers.

## **Scope of study**

In this paper, the scope of study was confined to obtaining and analyzing datasets of restaurants in Kuching and Kota Samarahan. Extract information on food only, does not extract information about other aspects reviewed by customers such as services and physical environment.

## **CHAPTER TWO**

### **LITERATURE REVIEW**

#### **Introduction**

In this chapter, past researches related to this study are presented in detail.

#### **User Generated Content (UGC)**

Nowadays, everyone can post their opinions online and also access and perceive others' opinions. Consequently, an enormous amount of data rich of opinions generated from various sources including reviews and microblogs (Baruah, 2012). In other word, these data are also known as user-generated content (UGC), which is “any positive or negative statements made by potential, actual, or former customers about a product or company, which is made available to a multitude of people and institutions via the Internet” (Henning–Thurau et al., 2004, p. 39). Online UGC also often considered more trustworthy and credible than those granted by companies (Burgess, Sellitto, Cox & Bultjens, 2009). Besides its importance as the information source to aid consumers in their decision-making, UGC in the form of online also helps consumers reconsider their purchase decisions and influence their buying behaviour (Schmunk, Höpken, Fuchs & Lexhagen, 2013).

#### **Sentiment Analysis**

Sentiment analysis or also known as opinion mining, analyzes the text written in a natural language on a topic and categories it as positive, negative or neutral based on the emotions and opinions asserted in it (Chinsha & Joseph, 2015). Sentiment analysis is intended for textual data processing, the overall polarity of text is obtained in accordance to positive or negative words and phrases expressed in the text data (Pang & Lee, 2008). Furthermore, as

Pang and Lee (2008) described in their research report, semantic orientation (SO) in the particular document determined automatically by performing sentiment analysis. Semantic orientation also (SO) marks the polarity whether it is positive, negative or neutral and determining the strength of words, phrases or sentences (Turney, 2002).

Sentiment analysis conducted at three different levels: aspect-level, document-level, and sentence-level. In document level, role of sentiment analysis is to classify the entire document that reviews about an object indicate an overall negative or positive opinion (Turney, 2002). Previous researchers focus on automatically separate positive texts from negative texts. Thus, Liu (2012) states that lack of thorough analysis is the disadvantage of document level sentiment analysis. Next, sentiment analysis in sentence level is to determine whether the sentence expressed a positive, negative or neutral opinion (Liu, 2012). Liu (2012) put it up that there is no significant difference between sentence and document level as sentences are part of the documents. Furthermore, an aspect level sentiment analysis was conducted to extract the features from the sentiment express in each aspect and the given target entities (Pang & Lee, 2008).

### **Existing Systems**

Early works by researchers were seen constructing sentiment analysis systems by utilizing lexical resources like WordNet. An experiment on Twitter sentiment analysis and their research outcomes shows that the application of lexicon resources was more helpful compare to the part-of-speech features in the microblogging domain in order to classify the polarity of emotions or opinions (Kouloumpis, Wilson & Moore, 2011).

Apart from that, Vu, Li, Law and Zhang (2017) presents a method processing restaurant reviews and apply text mining techniques in analysing the dining behaviours of tourists and their dining preferences. The SentiStrength sentiment estimation tool has been

utilized in their research (Vu et al., 2017). They analyse the food items and restaurant features based on the sentiment labels of the sentences. A food item or restaurant feature is classified as negative sentiment if it contains keyword in a sentence with negative sentiment, and vice versa (Vu et al., 2017).

In Sasikala and Sheela (2018), they proposed a method to mine food reviews based on score together with text analysing packages. The outcome of the proposed system was a good result using the ratings scale (Sasikala & Sheela, 2018). Thus, the drawback of this system is better outcomes was only generated when it applied towards open sentiments such as scores or ratings (Sasikala & Sheela, 2018).

Other existing research have conducted review summarizations and sentiment analysis. In (Hu & Liu, 2004), they focus on summarizing each customer reviews about a product by mining the features of the product on which the customers have expressed their opinions and classify the sentiment as positive or negative. In Pang and Lee (2008) findings, they implement machine learning techniques to categories documents by overall sentiments.

Lastly, Anjum and Dev (2016) proposed a system to assist and find out good restaurants and its dishes that seek by the potential consumer who relies on past consumers' reviews. The method is to obtain reviews from the websites. After that, meaningful information from the sentences were extracted by utilizing Natural Language Processing techniques then they assigned a 5-point scale rating.

### **Valence Aware Dictionary and Sentiment Reasoner (VADER)**

The VADER sentiment analysis tool is a tool which make use of specially developed lexicon to classify the sentiment based on the sentiment intensity (Hutto & Gilbert, 2014). Sentiment lexicon is a list of words and phrases that indicates positive or negative polarity information. Much of sentiment analysis relies heavily on sentiment lexicon as it stores

sentiment information about the smallest possible linguistic unit (Hutto & Gilbert, 2014). Implementation of a reliable lexicon will help enhance the performance of sentiment analysis system. Apart from that, VADER is a rule-based model with rich lexical features. It is applicable on social media style text and achieves generalizable results (Hutto & Gilbert, 2014). Hutto and Gilbert (2014) states that VADER Sentiment Lexicon is a human-validated inclusive list of gold-standard sentiment words that provides both polarity and intensity. VADER incorporates common dictionary words, including information on emoticons(for example,“:-)”) represents a “smiley face” and generally conveys positive sentiment), acronyms(“LOL”, “LMAO” etc.) and slang(“nah”,“meh” etc.) (Hutto & Gilbert, 2014).

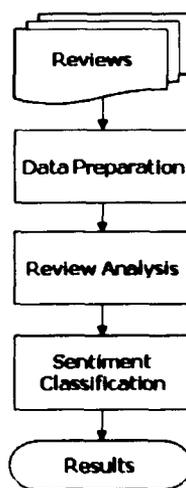
## CHAPTER THREE

### METHODOLOGY

#### Introduction

This chapter discussed about the details of the methodology that is being used for this research work to be done.

#### Reviews



*Figure 1.* Sentiment Analysis Model

Firstly, data from restaurant review websites were scraped using a web scraper tool such as Agency, a chrome browser extension. To use Agency extension, we were required to have an Agency account. It can be easily created by completing the sign-up forms in Agency website. Then a username and password were provided for login purpose. Upon completing this process, any data from website can be scraped easily.

In this study, the reviews were taken from Google Reviews. The selected restaurants were chosen based on a quite high number of comments from customers. In addition, restaurant review website usually consists of restaurant name, review text, username, star rating, and date visited. From the dataset, review text are comments posted by reviewers after visiting restaurants. It provides useful information about a consumer's preferences, such as

their consumed dishes from the particular restaurants. Data from review text about the restaurants was crucial in the analysis. However, they are usually made in an unstructured format and unable to be directly analysed.

### **Data preparation**

Python's Natural Language Toolkit (NLTK) was utilized to conduct the pre-processing of reviews raw data. Next, each review text was normalized into lower case along with the removal of stopwords, punctuations, non-alphabet characters and numbers. After that, the sentences undergo a text tokenization algorithm, in which the sentences were split into words, phrases or symbols known as "tokens." (Vu et al., 2017). Moreover, the data preparation steps performed necessary data pre-processing and cleaning on the dataset for the subsequent analysis. Some commonly used pre-processing steps include removing non-textual contents and tags and removing vague information about the reviews that are not required for sentiment analysis, such as star rating, review dates and reviewers' names.

### **Review Analysis**

In review analysis, the remaining tokens processed from previous step were inputted into the n-gram method. Bigrams function from NLTK was used to generate bigrams from the data. Common forms of n-gram are: unigram, bigram and trigram. Bigram refers to n-gram size of 2 when a pair of word are considered. For instance, " 'The best', 'best laksa', 'laksa in', 'in town' ". For this study, bigrams technique was chosen in order to obtain the food names from reviews and count the number of occurrences of each pair. The benefit of using bigram instead of using a unigram (single word) is due to dependencies between some words and also the importance of individual phrases (Yousefpour, Ibrahim, Hamed & Hajmohammadi, 2014). In this paper, some of the terms (for food names) were constructed of

more than one word. By considering bigrams, the process able to embed some context from the combination of a few words. Then, from the bigrams result, the most frequently occurring bigrams were extracted from the food reviews and listed as the top dishes.

### **Sentiment Classification**

This process step is to classify reviews with respect to its sentiment classes. The idea is that the polarity of reviews determined which are positive, neutral and negative reviews. Using NLTK's VADER, the sentiment score (positive, neutral and negative, compound) was generated for each review text input. VADER is unique as the intensity of sentiment were taken into account. For example, VADER scores "okay" moderately positive and "great" extremely positive. It also able to identify and score lexical features which commonly appears in informal online text such as punctuations, capitalizations, degree modifiers, conjunctions(i.e. use of conjunctions such as "but"), emoticons, slang, and acronym (Hutto & Gilbert, 2014).

VADER generates four sentiment metrics from these word ratings such as positive, negative, neutral and compound. The compound score is obtained by summing all of the valence scores for every single word in the lexicon, adjusted based on the rules, then normalized in between scale of -1 (most extreme negative) and +1 (most extreme positive).

The default value of the compound score are:

1. **positive sentiment:** compound score  $\geq 0.05$
2. **neutral sentiment:** (compound score  $> -0.05$ ) and (compound score  $< 0.05$ )
3. **negative sentiment:** compound score  $\leq -0.05$

Therefore, the review text with compound score of  $\geq 0.05$  are considered as positive reviews while review text with compound score of  $\leq -0.05$  are considered as negative reviews.

## Result

The results obtained from the bigrams method were visualized in form of word cloud. Word cloud also known as Tag cloud, is a visual representation which composed of words from the dataset where the size of the word indicates its frequency or importance(Vu, 2018). Python's modules such as Matplotlib and WordCloud were used to generate word cloud. The most frequent words occurring for each restaurant which appears larger than other words were determined as the best dishes for the restaurant. Next, for the sentiment classification result, the percentage of positive, neutral and negative reviews were represented in the form of pie chart. For this purpose, Matplotlib was used as it supports pie charts using the pie() function and meaningful pie charts were generated.

## CHAPTER FOUR

### RESULTS

#### Introduction

This chapter comprises the analysis, presentation and interpretation of the findings resulting from this study.

#### Reviews Data

A total of 2695 reviews from 19 restaurants were scraped by using Agenty chrome extension.

Table 1. Restaurant information

Restaurants' name	Review counts(lines)
Aroma Café	49
Bella Italia	103
Burger o Myy	74
Ceylonese	175
Dapur Makwa	50
Feast and Furious	187
Fork and Knife	85
Kings Curry	192
Kopi o Corner	340
Lepau	285
Lok Lok	94
Moms Laksa	213
New Atmosferah	184
RJ Ayam Bakar	104
Rumah Asap Dayak Samarahan	95
Rumah Asap Tabuan Dayak	184
Sharing Downtown	98
The Container	84
Tok Janggut	99
<b>Total</b>	<b>2695</b>