

**Feature Comparison of Yelp Restaurants Based on  
Sentiment Analysis**

**W.N.N De Silva**

**169307b**

**Faculty of Information Technology**

**University of Moratuwa**

**September 2020**

**Feature Comparison of Yelp Restaurants Based on  
Sentiment Analysis**

**W.N.N De Silva**

**169307b**

**Dissertation submitted to the Faculty of Information Technology, University of  
Moratuwa, Sri Lanka for the partial fulfillment of the requirements of the  
Degree of Master of Science in Information Technology.**

**September 2020**

# Declaration

I declare that this thesis is my own work and has not been submitted in any form for another degree or diploma at any university or other institution of tertiary education. Information derived from the published or unpublished work of others has been acknowledged in the text and a list of references is given.

Name of Student

Signature of Student

Date:

Supervised by

Name of Supervisor

Signature of Supervisor

Date:

## **Acknowledgement**

Foremost, I would like to express my sincere gratitude to my advisor Senior Lecturer, Mr. Saminda Premaratne for the continuous support of my MSc study and research, for his motivation, enthusiasm, and immense knowledge. His guidance helped me in all the time of research and writing of this thesis.

Last but not the least, I would like to thank my family, my parents for supporting me spiritually throughout my life and to all of my university colleagues who helped me in this venture.

# Abstract

Today the internet has given a huge space for people to explore new experiences. It has made lives more comfortable. Global information can be retrieved within seconds over any topic. Advance of such internet technologies made people to share own experiences over various topics. The extensive use of social media, forums, blogs or various e-commerce platforms have made this easier. People give opinions on various products or services based on their experiences. These reviews affect marketing strategies of online businesses. Also, comments are important for the staff or the owners as well. Positivity or the negativity of these comments is important. But thousands of reviews are collected per a day. In order to get an idea about the reviews one has to spend many hours reading them. This is an impossible task. Hence there should be an easier way to explore huge number of reviews with in seconds. This research is based on implementing a system to facilitate this task. The domain of the business is restaurants. People often search for good restaurants. Most of them are used to explore various customer reviews to find the best one. The solution will be a hybrid sentiment-based system which aids customers to find good restaurants in a particular city. Customer review datasets of various restaurants in a particular city are collected from Yelp.com. The reviews are analyzed based on popular restaurant features such as price, quality of food, ambience or service. The main functionality of the system is to deliver a sentimental comparison between various restaurants over their features. Several NLP tasks with machine learning techniques such as multilabel Naïve Bayes model, SVM model, deep learning convolution neural networks and word embeddings are used in this research.

# Table of Contents

Declaration.....	i
Acknowledgement .....	ii
Abstract.....	iii
List of Figures.....	vii
List of Tables .....	vii
Chapter 1 .....	1
Introduction to Systems based on Natural Language Processing Techniques.....	1
1.1 Introduction .....	1
1.1.1 Background .....	1
1.2 Aim.....	2
1.3 Objectives.....	2
1.4 Proposed Solution.....	3
Chapter 2 .....	5
Literature Review on Natural Language Processing Systems .....	5
2.1 Introduction .....	5
2.2 Related work based on Natural Language Processing .....	5
Chapter 3 .....	12
Automation of Restaurant Feature Comparison Based on Customer Experience .....	12
3.1 Introduction .....	12
3.2 Natural Language Processing Techniques.....	13
3.3 Sentiment Analysis .....	13
3.4 Feature/Aspect based sentiment analysis.....	15
Chapter 4 .....	16
Automated system for Feature Comparison of Yelp Restaurants .....	16
4.1 Introduction .....	16
4.2 Dataset preparing and preprocessing.....	17
4.2.1 Lowercase .....	20
4.2.2 Stop words and Noise Removal .....	20
4.2.3 Tokenization.....	20
4.3 Machine Learning Approach for Review Classification .....	20
4.3.1 Naïve Bayes Classification .....	21
4.3.2 Naïve Bayes Model.....	21
4.3.3 Support Vector Machines .....	23

4.3.4 Deep Learning Methods .....	23
4.4 Rule based Approach for Aspect Term Extraction and Polarity Detection	24
4.4.1 Dependency Parser .....	25
4.5 Aspect Categorization .....	25
4.5.1 Word Embedding Techniques .....	26
Chapter 5 .....	27
Analysis and Design of Restaurant Reviews Mining .....	27
System .....	27
5.1 Introduction .....	27
5.2 Text/Review Classification .....	29
5.3 Extracting terms of Yelp reviews and generating dictionary with polarity for each term.....	29
5.4 Assigning aspect terms in to aspect categories and detecting the overall sentiment (positive/negativity) towards aspect categories.....	29
5.5 Finding most positive or most negative aspect terms of a category.....	30
5.6 Visualizing Aspect categories of multiple restaurants for decision making. .....	30
Chapter 6 .....	31
Implementation of Restaurant Feature Comparison System .....	31
6.1 Introduction .....	31
6.2 System Development .....	31
6.3 Dataset preparing and preprocessing.....	32
6.3.1 Lowercase .....	33
6.3.2 Stop words and Noise Removal .....	33
6.3.3 Tokenization.....	33
6.4 Text/Review Classification using machine learning models.....	33
6.4.1 Naïve Bayes Classification method .....	34
6.4.2 Text Classification using Support Vector Machines .....	34
6.4.3 Deep Learning Method .....	35
6.5 Extracting terms of review sentences and generating dictionary with polarities for each term.....	35
6.6 Assigning aspect terms in to aspect categories and detecting the overall sentiment (positive/negativity) .....	36
6.6.1 Finding most positive or most negative aspect terms of a category .....	36
6.6.2 Visualizing Aspect categories of multiple restaurants for decision making .....	36

6.7 Pseudocode for Extracting aspect terms and polarity detection.....	37
6.8 Pseudocode for Assignment of aspect terms in to aspect categories.....	37
Chapter 7 .....	39
Results and Evaluations .....	39
7.1 Introduction .....	39
7.2 Measuring performance of classification models. ....	39
7.2.1 Finding the most common and most positive aspects of a particular restaurant .....	41
7.3 Comparison for Restaurants found in Yelp.com .....	42
Chapter 8 .....	44
Conclusion .....	44
References.....	45



## **List of Figures**

<b>Figure 3.1 - Sentiment Analysis Techniques Summary</b>	<b>14</b>
<b>Figure 4.1 - Yelp Review Details Dataset</b>	<b>19</b>
<b>Figure 5.1 - Top Level Design of the system</b>	<b>28</b>
<b>Figure 6.1 - Final Customer Reviews Dataset for multiple Restaurants</b>	<b>32</b>
<b>Figure 7.1 - Loss and accuracy of Convolution Neural Network</b>	<b>40</b>
<b>Figure 7.2 - Most common and positive Aspects</b>	<b>41</b>
<b>Figure 7.3 - Visualization of Restaurant Comparison</b>	<b>43</b>

## **List of Tables**

<b>Table 4.1 - SemEval-2014 Dataset</b>	<b>17</b>
<b>Table 4.2 - Yelp Business Details Dataset</b>	<b>18</b>
<b>Table 4.3 - Yelp Customer Review Details Dataset</b>	<b>19</b>
<b>Table 7.1 - Classification Report of Naïve Bayes model</b>	<b>39</b>
<b>Table 7.2 - Classification Report of SVM model</b>	<b>40</b>
<b>Table 7.3 - Values of Positivity/Negativity for restaurant features</b>	<b>42</b>