

**USING WEB SCRAPING IN SOCIAL MEDIA TO DETERMINE
MARKET TRENDS WITH PRODUCT FEATURE-BASED
SENTIMENT ANALYSIS**

Thamarasi Nanayakkara

199480A

Thesis/Dissertation submitted in partial fulfillment of the requirements for the Degree Master of
Science

Department of Computational Mathematics

University of Moratuwa

Sri Lanka

July 2022

Declaration

I declare that this is my own work and this thesis/dissertation does not incorporate without acknowledgement any material previously submitted for a Degree or Diploma in any other University or institute of higher learning and to the best of my knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgement is made in the text. Also, I hereby grant to University of Moratuwa the non-exclusive right to reproduce and distribute my thesis/dissertation, in whole or in part in print, electronic or other medium. I retain the right to use this content in whole or part in future works (such as articles or books).

Name of Student: Thamarasi Nanayakkara

Signature of Student:

Date: 25.07.2022

The above candidate has carried out research for the Masters dissertation under my supervision.

Name of Supervisor(s): Dr. Sagara Sumathipala

Signature of Supervisor(s):

Date: 25.07.2022

Acknowledgement

I would like to express my honest appreciation and thankfulness to my supervisor, Dr. Sagara Sumathipala for all the support, assistance and guidance throughout all the phases of this research project. He not only encouraged my efforts, but also helped me in every step of my research work. This project would not have not been possible and completed without his guidance, support and advice. I would also like to thank all the lecturers in MSc in AI Batch 12, University of Moratuwa for their support and guidance. I am also very grateful to my parents and all friends who assisted me in my project.

Abstract

Customer product reviews are openly available online and they are now widely used for deciding quality of product or service and to determine market trends and influence decision making of users. Due to the availability of a massive number of customer reviews on the web, summarizing them requires a fast classification system. Compared to supervised and unsupervised machine learning techniques for binary classification of reviews, fuzzy logic can provide a simple and comparatively faster way to model the fuzziness existing between the sentiment polarities classes due to the uncertainty present in most of the natural languages. But the fuzzy logic techniques are not much considered in this domain. This thesis proposes a model which measures product market value by using sentiment analysis conducted on the reviews of online products which are collected from a well known ecommerce website “Amazon”. Fuzzy logic approach is used in calculating the final product market demand.

Hence, in this paper we propose a fine grained classification of customer reviews into weak positive, average positive, strong positive, weak negative, average negative and strong negative classes using a fuzzy logic model based on the most popularly known sentiment based lexicon SentiWordNet. By creating rules and relationships between fuzzy membership functions and linguistic variables, we can analyze the customer opinions towards online products. This proposed model provides the most reasonable sentiment analysis because we try to reduce all the problems from the related past researches. The outcomes can allow the business organization to understand their customer’s sentiments and improve customer loyalty and customer retention techniques in order to increase customer values and profits result. Fine grained classification accuracy approximately in the range of 74% to 77% has been obtained by experiments conducted on datasets of electronic products containing reviews of smart phones, TV and laptops.

Keyword- Sentiment analysis, Fine grained classification, Fuzzy Logic, SentiWordNet, Online reviews

Table of Contents

Declaration	i
Acknowledgement	ii
Abstract.....	iii
Table of Contents	iv
List of Figures.....	v
List of Tables	v
List of Abbreviations	v
CHAPTER 1	1
Introduction	1
1.1 Prolegomena	1
1.2 Aims and Objectives.....	2
1.3 Background and Motivation	3
1.4 Problem definition	4
1.5 Fuzzy based Sentiment Analysis Solution.....	5
1.6 Resource Requirements	6
1.7 Structure of the Thesis.....	6
1.8 Summary.....	8
CHAPTER 2	9
2.1 Introduction	9
2.2 Problem definition	12
2.3 Summary.....	12
CHAPTER 3	13
3.1 Introduction	13
3.2 Hypothesis	13
3.3 Input.....	13
3.4 Output	13
3.5 Process	13
3.6 Users	14
3.7 Features.....	14
3.8 Summary.....	14
CHAPTER 4	15
4.1 Introduction	15
4.2 Summary.....	27
CHAPTER 5	28
5.1 Introduction	28

5.2 Summary.....	33
CHAPTER 6.....	34
6.1 Conclusion and Further Work	34
6.2 Summary.....	34
References	35

List of Figures

Figure 4. 1 High Level Diagram of the proposed Methodology.....	18
Figure 4. 2 Extraction of Crisp Input from SentiWordNet	19
Figure 4. 3 Membership Function for Adjective with Six Linguistic Terms.....	22
Figure 4. 4 Membership Function for Verb with six Linguistic Terms	23
Figure 4. 5 Membership Function for Adverb with Three Linguistic Terms	23
Figure 4. 6 Membership Function for Output with Six Linguistic Terms	26

List of Tables

Table 4. 1 Extracted phrase patterns and example opinionated phrases	16
Table 4. 2 Linguistic Variables and Fuzzy Sets	20
Table 4. 3 Sample If – then Rules	24
Table 5. 1 Corpus Statistics.....	29
Table 5. 2 Results of sample opinion phrases	31
Table 5. 3 Range for the six classes	31
Table 5. 4 Evaluation results for fine grained classification.....	32

List of Abbreviations

Abbreviation	Description
POS	Part of Speech
NLP	Natural Language Processing