

# **Mining social sentiments for demand analysis in Footwear industry**

A.M.A.N.P.W.M.R.P.D.B. Athurupane

198742X

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

**July 2022**

# Declaration

We declare that this thesis is our 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 Information derived from the published or unpublished work of others has been acknowledged in the text, and a list of references is given.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

A.M.A.N.P.W.M.R.P.D.B. Athurupane  
.....

Signature of Student

***UOM Verified Signature***  
.....

Date: 30/07/2022

Supervised by

Name of Supervisor

Dr. Saminda Premaratne  
.....

Signature of Supervisor

***UOM Verified Signature***  
.....

Date: 30/07/2022

# **Acknowledgment**

This dissertation would have been impossible without the tremendous support and motivation of those who assisted me throughout this journey. First and foremost, my heartfelt gratitude goes to my supervisor and Programme coordinator, Dr. Saminda Premaratne, who took responsibility for supervising me without hesitation. I am eternally grateful for him, for always being patient, flexible and empathetic. Also, I am thankful for his expertise, encouragement, moral support and guidance which he provided with generosity to facilitate my task.

Next, my heartfelt gratitude is extended to my friend Ms. S. N. Rambukkanage for the immeasurable support provided to me through these two years of period.

It is with great pleasure I thank the University of Moratuwa, Sri Lanka, for all its efforts and facilities that it has taken to contribute towards this postgraduate Programme. Specially, I would like to thank my colleagues for their support to complete this research in many ways.

Finally, all my thanks and respect go to my family and friends for all the unceasing love and support they provided me.

## Abstract

The public tends to express their thoughts about particular goods and or services through popular social media networks such as Twitter and Facebook, while the firms also use social media to communicate with their consumers. As a result, this beneficial information can be used to make marketing and business decisions. However, due to the vast, noisy, and dynamic nature of these information, capturing the true public opinion has become a key challenge. Sentiment analysis is one of the methods that is employed to extract positive or negative attitudes from this social media information and thus, it has drawn the attention of the scholars during last decade. Different scholars have used a variety of techniques and methods to capture accurate results. Data mining is one of the recent approaches adopted by them to obtain better results from sentiment data. Moreover, some scholars have extended their studies to gain more insights from different topics. Such studies conducted to predict future results, analyze trends, detect anomalies etc. It can be beneficial for massive industries like footwear to understand their market through these approaches and streamline their product and service catalogs to meet the needs of their customers. This research aims to analyze previous studies conducted on this area, identify their contribution, challenges and limitations, and build a new comprehensive demand prediction model for footwear industry using data mining techniques.

Key words : *Sentiment Analysis, Data Mining, Demand prediction, Footwear Industry, Polarity detection, Lexicon, CrowdTangle, Twitter API, Postman, RapidMiner*

# Table of Contents

	Page
<b>Declaration.....</b>	<b>i</b>
<b>Acknowledgment.....</b>	<b>ii</b>
<b>Abstract.....</b>	<b>iii</b>
<b>Table of Contents .....</b>	<b>iv</b>
<b>List of Figures.....</b>	<b>vii</b>
<b>List of Tables .....</b>	<b>vii</b>
<b>1. Introduction .....</b>	<b>1</b>
1.1 Background of the study .....	1
1.2 Problem Statement .....	3
1.3 Aim and Objectives.....	4
1.3.1 Aim .....	4
1.3.2 Objectives .....	4
1.4 Proposed solution .....	4
1.5 Summary .....	5
<b>2. Literature Review .....</b>	<b>6</b>
2.1 Introduction .....	6
2.2 Data mining in online social media.....	6
2.3 Social sentiment analysis .....	7
2.4 Demand analysis using social media sentiments.....	8
2.5 Summary .....	9
<b>3. Technology adopted .....</b>	<b>10</b>
3.1 Introduction .....	10
3.2 Tools and software used.....	10
3.2.1 RapidMiner Studio 9.10.....	10
3.2.2 CrowdTangle.....	10
3.2.3 Upscene Advanced Data Generator 4.....	11
3.2.4 Twitter API .....	12
3.2.5 Postman API .....	12
3.2.6 Microsoft Excel.....	13
3.3 Data mining algorithms and other techniques used .....	13
3.3.1 Naïve Bayes classification .....	13
3.3.2 K-NN Algorithm.....	14
3.3.3 Random forest.....	14

3.3.4	Deep learning .....	15
3.3.5	Rule induction .....	15
3.3.6	Decision tree .....	15
3.3.7	SentiWordNet .....	16
3.4	Summary .....	16
<b>4.</b>	<b>Approach for mining social sentiments for demand analysis in Footwear industry.....</b>	<b>17</b>
4.1	Introduction .....	17
4.2	Inputs to the model.....	17
4.3	Output from the model .....	17
4.4	Process in brief.....	17
4.5	Platform selection.....	18
4.6	Data collection.....	18
4.6.1	Collection of Facebook post reactions data .....	19
4.6.2	Collection of Twitter post data (Tweets) .....	20
4.7	Summary .....	23
<b>5.</b>	<b>Analysis and Design.....</b>	<b>24</b>
5.1	Introduction .....	24
5.2	Analysis of the study .....	24
5.3	Design of the study.....	25
5.3.1	Attribute selection.....	25
5.3.2	Text mining.....	26
5.3.3	Sentiment Analysis .....	26
5.3.4	Prediction models.....	30
5.4	Final design of the study .....	30
5.5	Summary .....	31
<b>6.</b>	<b>Implementation of the model.....</b>	<b>32</b>
6.1	Introduction .....	32
6.2	Data preprocessing and Text mining.....	32
6.2.1	Facebook post reaction data.....	32
6.2.2	Twitter post data (Tweets) .....	35
6.3	Sentiment analysis.....	35
6.3.1	Sentiment analysis in Facebook post reaction .....	35
6.3.2	Sentiment analysis in Tweets.....	35
6.4	Process of creating aggregated dataset.....	37

6.5	Sentiment prediction .....	38
6.6	Identifying trend patterns and predict demand.....	40
6.7	Summary .....	40
<b>7.</b>	<b>Evaluation .....</b>	<b>41</b>
7.1	Introduction .....	41
7.2	Evaluation of the Sentiment predictions .....	41
7.2.1	Decision Tree algorithm .....	41
7.2.2	Random Forest algorithm .....	42
7.2.3	Naïve Bayes algorithm.....	42
7.2.4	Deep learning algorithm .....	42
7.2.5	Rule induction algorithm .....	43
7.3	Evaluation of identifying trend patterns and predict demand .....	45
7.4	Summary .....	46
<b>8.</b>	<b>Conclusion and Future work.....</b>	<b>47</b>
8.1	Conclusion of the study.....	47
8.2	Limitations of the study.....	48
8.3	Future work .....	49
<b>9.</b>	<b>Reference.....</b>	<b>50</b>
	<b>Appendix A .....</b>	<b>52</b>
	<b>Appendix B .....</b>	<b>53</b>
	<b>Appendix C.....</b>	<b>54</b>

## List of Figures

	<b>Page</b>
Figure 3-1: CrowdTangle Dashboard .....	11
Figure 3-2: Upscene Advanced Data Generator .....	11
Figure 3-3: Postman dashboard .....	12
Figure 4-1: Raw data collected from CrowdTangle .....	22
Figure 4-2: Raw data collected from Twitter.....	22
Figure 5-1: Proposed design of the model .....	25
Figure 5-2: Actual design of the model .....	31
Figure 6-1: Process of extracting attribute values from FB data .....	33
Figure 6-2: RapidMiner Operators used to analyse sentiment in Tweets.....	36
Figure 6-3: Parameter setting in Extract Sentiment operator.....	36
Figure 6-4: Example of the Final Dataset .....	38
Figure 6-5: Process of Predicting Sentiments.....	39
Figure 6-6: Field configuration of Pivot table .....	40
Figure 7-1: Sentiment trend for Flip Flops .....	45
Figure 7-2: Sentiment trend for Sandals .....	45

## List of Tables

	<b>Page</b>
Table 4-1: Additional Attributes and values.....	19
Table 4-2: Filters used in CrowdTangle .....	20
Table 4-3: Query parameters used in Postman API.....	21
Table 5-1: Sentiment score of Facebook reactions (Source: [21]) .....	29
Table 6-1: Tasks performed by RapidMiner Operator in attribute value extraction ...	34
Table 6-2: Tasks performed by RapidMiner operators in sentiment prediction.....	39
Table 7-1: Accuracy percentages of Decision Tree algorithm .....	41
Table 7-2: Accuracy percentages of Random Forest algorithm .....	42
Table 7-3: Accuracy percentages of Naive Bayes algorithm.....	42
Table 7-4: Accuracy percentages of Deep Learning algorithm.....	42
Table 7-5: Accuracy percentages of Rule Induction algorithm.....	43
Table 7-6: Suitable ML algorithms for Predict Sentiments.....	43