A NEW EVALUATION AND BENCHMARKING FRAMEWORK FOR SENTIMENT ANALYSIS ARABIC DIALECTAL LEXICONS, BASED ON MULTI-CRITERIA DECISION-MAKING

SAMEH MOHAMED SHERIF IBRAHEIM

SULTAN IDRIS EDUCATION UNIVERSITY

2024

A NEW EVALUATION AND BENCHMARKING FRAMEWORK FOR SENTIMENT ANALYSIS ARABIC DIALECTAL LEXICONS, BASED ON MULTI-CRITERIA DECISION-MAKING

SAMEH MOHAMED SHERIF IBRAHEIM

05-4506832 pustaka.upsi.edu.my Perpustakaan Tuanku Bainun Kampus Sultan Abdul Jalil Shah PustakaTBainun ptbupsi

THESIS PRESENTED TO QUALIFY FOR A DOCTOR OF PHILOSOPHY

FACULTY OF COMPUTING AND META-TECHNOLOGY
SULTAN IDRIS EDUCATION UNIVERSITY

2024

05-4506832 pustaka.upsi.edu.my Perpustakaan Tuanku Bainun Kampus Sultan Abdul Jalil Shah PustakaTBainun ptbupsi

UNIVERSITI PENDIDIKAN SULTAN IDRIS
SULTAN IDRIS EDUCATION UNIVERSITY

Please tick (✓)	
Project Paper	
Masters by Research	
Masters by Mix Mode	
Ph.D.	

INSTITUTE OF GRADUATE STUDIES

DECLARATION OF ORIGINAL WORK

This declaration is made on the 18th day of November 2024

i- Student's Declaration:

I Sameh Mohamed Sherif -P20201000019-Faculty of Computing and META Technology (FKMT) hereby declares that the dissertation /thesis for Doctor of Philosophy titled "A New Evaluation and Benchmarking Framework for Sentiment Analysis Arabic Dialectal Lexicons, Based On Multi-Criteria Decision-Making" is my original work. I have not copied from any other students's work or from any sources except where due reference or acknowledgement is made explicitly in the text, nor has any part been written for me by another person.

Sameh Mohamed Sherif

Signature of the student

ii- Supervisor's Declaration:

I Dr. Abdullah Hussein Abdullah Al-Amoodi hereby certifies that the work entitled, "A New Evaluation and Benchmarking Framework for Sentiment Analysis Arabic Dialectal Lexicons, Based On Multi-Criteria Decision-Making" was prepared by the above named student, and was submitted to the Institute of Graduate Studies as a * partial / full fulfillment for the conferment of the requirements for Doctor of Philosophy (By Research), and the aforementioned work, to the best of my knowledge, is the said student's work.

Dr. Abdullah Hussein Abdullah Al-Amoodi
Pensyarah
Fakulti Komputeran dan Meta-Teknologi
Universiti Pendidikan Sultan Idris
35900 Tanjong Malim, Perak

18 – November - 2024

Date

Signature of the Supervisor

UPSI/IPS-3/BO 31
Pind.: 01 m/s:1/1

INSTITUT PENGAJIAN SISWAZAH /
INSTITUTE OF GRADUATE STUDIES

BORANG PENGESAHAN PENYERAHAN TESIS/DISERTASI/LAPORAN KERTAS PROJEK DECLARATION OF THESIS/DISSERTATION/PROJECT PAPER FORM

Tajuk / *Title*: **A New Evaluation and Benchmarking Model for Sentiment Analysis Arabic Dialectal Lexicons, Based On Multi-Criteria Decision-Making**

No. Matrik /*Matric No.*: **P20201000019**

Saya / *I* : **Sameh Mohamed Sherif**

mengaku membenarkan Tesis/Disertasi/Laporan Kertas Projek (Kedoktoran/Sarjana)* ini disimpan di Universiti Pendidikan Sultan Idris (Perpustakaan Tuanku Bainun) dengan syarat-syarat kegunaan seperti berikut:-
acknowledged that Universiti Pendidikan Sultan Idris (Tuanku Bainun Library) reserves the right as follows:-

1. Tesis/Disertasi/Laporan Kertas Projek ini adalah hak milik UPSI.
   The thesis is the property of Universiti Pendidikan Sultan Idris

2. Perpustakaan Tuanku Bainun dibenarkan membuat salinan untuk tujuan rujukan dan penyelidikan.
   Tuanku Bainun Library has the right to make copies for the purpose of reference and research.

3. Perpustakaan dibenarkan membuat salinan Tesis/Disertasi ini sebagai bahan pertukaran antara Institusi Pengajian Tinggi.
   The Library has the right to make copies of the thesis for academic exchange.

4. Sila tandakan (√) bagi pilihan kategori di bawah / Please tick (√) from the categories below:-

☐	**SULIT/*CONFIDENTIAL***	Mengandungi maklumat yang berdarjah keselamatan atau kepentingan Malaysia seperti yang termaktub dalam Akta Rahsia Rasmi 1972. / *Contains confidential information under the Official Secret Act 1972*
☐	**TERHAD/*RES RICTED***	Mengandungi maklumat terhad yang telah ditentukan oleh organisasi/badan di mana penyelidikan ini dijalankan. / *Contains restricted information as specified by the organization where research was done.*
✓	**TIDAK TERHAD / *OPEN ACCESS***	

Sameh Mohamed Sherif

(Tandatangan Pelajar/ Signature)

Tarikh: 18 – November - 2024

Dr. Abdullah Hussein Abdullah Al-Amoodi
Pensyarah
Komputeran dan Meta-Teknologi
Universiti Pendidikan Sultan Idris
35900 Tanjong Malim, Perak

Abdullah Hussein Abdullah Al-Amoodi

(Tandatangan Penyelia / Signature of Supervisor)
& (Nama & Cop Rasmi / Name & Official Stamp)

Catatan: Jika Tesis/Disertasi ini SULIT @ TERHAD, sila lampirkan surat daripada pihak berkuasa/organisasi berkenaan dengan menyatakan sekali sebab dan tempoh laporan ini perlu dikelaskan sebagai SULIT dan TERHAD.

Notes: If the thesis is CONFIDENTAL or RESTRICTED, please attach with the letter from the related authority/organization mentioning the period of confidentiality and reasons for the said confidentiality or restriction.

ACKNOWLEDGEMENT

"In the name of Allah, the Most Gracious and the Most Merciful"

Alhamdulillah, first and foremost, praise be Allah, the Cherisher and Sustainer of the World and to the Prophet Muhammad (Peace and Blessings of Allah Be upon Him) who was sent by Allah to be a great teacher to the mankind.

To date, the last 40 months have been the most hard, intriguing, and gratifying of my life.

I am really glad and grateful to have met so many lovely individuals who have assisted me in finishing my PhD thesis at Sultan Idris Educational University (UPSI) in Malaysia.

Above all, I want to thank Allah for all of his compassion and mercy in allowing me to complete this job. None of this would be possible without Allah's grace and compassion.

Furthermore, many people have been by my side, and their support should be honored.

First and foremost, I would like to thank my supervisors, Dr. Abdullah Huessin Abdullah Al-Amoodi, for his exceptional supervision, teaching, direction, patience, and support in my study, as well as for guiding me through the infinite and wide universe of ideas towards finishing my PhD proposal.

I also want to express my gratitude to my family for making my life so interesting, vibrant, and happy especially to my lovely wife Feeza and my sweet little daughter Farha. Additionally, I want to thank Feeza and Farha from the bottom of my heart for bearing up with my mood swings throughout this stressful time. I love you both, and any success won't mean anything if it's not shared with you both. Furthermore, I was blessed having more magnificent ladies in ly life, including my mother in Egypt, mama Nadia her love, sacrifices, unwavering support and belief in me have been the driving force behind my success. And my mother in Malaysia Mom Bel, which, by means of words, could not convey how grateful I am for all the affection, kindness, sacrifices, or even battels she had gone through for me. I wouldn't be here without all your love, caring and support. In addition to the priceless Doaa from my family in law for my success, which I had seen its results during this path.

Have to mention that, I was so fortunate to be raised by a magnificent Major General Mohamed Sherif who was not only a soulful father, but also an invaluable mentor who instilled in me the correct values, and principles, which molded me into the man I am. Fortunately this blessing continued by Dr. Paoletto Lorenzo the gifted father, who has played the same important role in my life in Malaysia, Dr. Paoletto and

mom Bel were an extension of the parents rule, that I never dreamed I would find in a new land.

I'd also want to thank my two sisters, two nieces, and nephew for their support, courage and doaa, through these years.

And I'll never forget to express my heartfelt gratitude to one of my best friends Fin Madden, who was the first person encouraging me to undertake this PhD journey and to date still supporting me, I'd consider Fin as a pioneer partner in this success as he was the first who has believed in my capability to persevere through this journey.

And with all love, I would like to honor the memory of a noble person, and a loveable family member, Mr. Andrew Beck, who was always a great supporter to our family. While all the gratitude we have, wouldn't be enough to express what he had done for us by himself, and by Linda Beck, who overwhelmed us with boundless love and caring, Linda and Drew gave us a family, which will always remain in our hearts and memories.

ABSTRACT

Sentiment Analysis (SA) field relies on lexicons to analyze people's perceptions. Therefore, evaluating and benchmarking lexicons is crucial for SA development, falling under Multi-Criteria Decision-Making (MCDM). The Fuzzy-Decision-by-Opinion-Score Method (FDOSM) and Fuzzy-Weighted-Zero-Inconsistency (FWZIC) techniques are novel MCDM extensions in a fuzzy-environment that address complexity-problems. Hence, this study aims to use these techniques to evaluate and benchmark dialectal lexicons. The methodology for this research is presented in four phases. The first phase includes developing four main criteria, along with its 28 sub-criteria for lexicon evaluation; the second phase proposes a multidimensional decision-matrix based on evaluation criteria and available lexicons; in the third phase, the multidimensional evaluation and benchmarking results of MCDM are obtained by FDOSM and are achieved by FWZIC; in the fourth phase, the validation was conducted based on the fusion of Systematic Ranking and Sensitivity Analysis to test the weight effect. The findings reveal that variations in weighting can significantly impact sensitivity analysis outcomes, i.e., SUAR (Lex18) ranked (4th) under the original scenario but fell to (5th) in scenario 6. Meanwhile, PADIC (Lex17) ranked (21st) under the original and scenarios 1, 2, 3, and 5, then rose to (20th) ranking position in scenarios 4 and 6. Despite these variations, some lexicons consistently maintained their ranking, i.e., SANAD (Lex21) the (24th), and TSAC (Lex12) the (28th) both maintained their ranking position across all sensitivity analysis interchanged scenarios, indicating their robustness. This emphasizes the reliability of the evaluation process from expert voting to sensitivity analysis of the four main criteria: Labelled-Data, Labelling-Type, Labelling-Techniques, Labelling-Targets, and its derived 28 sub-criteria. The study highlights the significance of addressing issues such as data variation and measuring importance when developing Arabic SA lexicons. The MCDM approach, particularly using FWZIC and FDOSM, proved to be more accurate in handling uncertainty and vagueness compared to other methods.

RANGKA KERJA PENILAIAN DAN PENANDA ARAS BARU UNTUK ANALISIS SENTIMEN LEKSIKON DIALEK BAHASA ARAB, BERDASARKAN PEMBUATAN KEPUTUSAN PELBAGAI KRITERIA

ABSTRAK

Bidang Analisis Sentimen (SA) bergantung pada leksikon untuk menganalisis persepsi orang ramai. Sehubungan dengan itu, aktiviti menilai dan menanda aras leksikon adalah penting untuk pembangunan SA, yang termasuk dalam Keputusan Pelbagai Kriteria (MCDM). Teknik Kaedah Keputusan Fuzzy Berdasarkan Skor Pendapat (FDOSM) dan Pemberat Fuzzy Tanpa Ketidakkonsistenan Sifar (FWZIC) merupakan lanjutan MCDM baharu dalam persekitaran kabur untuk menangani masalah kerumitan dengan lebih berkesan. Oleh itu, kajian bertujuan menggunakan kaedah-kaedah ini untuk menilai dan menanda aras leksikon dialek. Metodologi kajian ini dibentangkan dalam empat fasa. Fasa pertama termasuk membangunkan empat kriteria utama, bersama-sama dengan 28 subkriteria untuk penilaian leksikon. Fasa kedua mencadangkan matriks keputusan pelbagai dimensi berdasarkan kriteria penilaian dan leksikon yang tersedia. Dalam fasa ketiga, penilaian pelbagai dimensi dan hasil penanda aras MCDM diperoleh melalui FDOSM dan dicapai melalui FWZIC. Sementara itu, dalam fasa keempat, pengesahan dilakukan berdasarkan gabungan Pemeringkatan Sistematik dan Analisis Sensitiviti untuk menguji kesan pemberatan. Penemuan menunjukkan bahawa variasi dalam pemberatan mampu memberikan kesan ketara terhadap hasil analisis sensitiviti, iaitu SUAR (Lex18) menduduki tempat (ke-4) di bawah senario asal tetapi jatuh ke kedudukan (ke-5) dalam senario 6. Sementara itu, PADIC (Lex17) menduduki tempat (ke-21) di bawah asal dan senario 1, 2, 3, dan 5, kemudian meningkat ke kedudukan (ke-20) dalam senario 4 dan 6. Meskipun terdapat variasi, beberapa leksikon tetap konsisten dalam pelbagai situasi, iaitu SANAD (Lex21) (ke-24), dan TSAC (Lex12) (ke-28), kedua-duanya mengekalkan kedudukan merentas semua senario pertukaran analisis sensitiviti, membuktikan keteguhan dan keandalan masing-masing dalam pelbagai konteks. Ini menekankan kebolehpercayaan proses penilaian daripada undian pakar kepada analisis sensitiviti bagi empat kriteria utama: Data Berlabel, Jenis Pelabelan, Teknik Pelabelan, Sasaran Pelabelan, dan 28 subkriteria yang diperoleh. Dalam pada itu, kajian ini menekankan kepentingan menangani isu seperti variasi data dan kepentingan ukuran semasa membangunkan leksikon SA bahasa Arab. Pendekatan MCDM terutamanya menerusi penggunaan FWZIC dan FDOSM, terbukti lebih tepat untuk menangani ketidakpastian dan ketidakjelasan berbanding dengan kaedah lain.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABELS

LIST OF EQUATIONS

LIST OF APPENDIX

C2 Expert Questionnaire

LIST OF ABBREVIATIONS

(A)	AUTHORS
(AI)	ARTIFICIAL INTELLIGENCE
(AL)	ALGORITHM
(AlgD)	ALGERIAN DIALECT
(AMT)	AMAZON MECHANICAL TURK
(AOC)	ARABIC ONLINE COMMENTARY
(API)	APPLICATION PROGRAMMING INTERFACE
(ASA)	ARABIC SENTIMENTAL ANALYSIS
(ASO)	ARABIC SENTIMENT ONTOLOGY
(ASTD)	ARABIC SOCIAL SENTIMENT ANALYSIS DATASET
(BAMA)	BUCKWALTER ARABIC MORPHOLOGICAL ANALYZER
(BOW)	BAG OF WORDS
(BTEC)	BASIC TRAVELING EXPRESSION CORPUS
(BTEC)	BASIC TRAVELING EXPRESSION CORPUS
(C)	CROWDSOURCING
(CA)	CLASSICAL ARABIC
(CL)	COMPUTATIONAL LINGUISTICS
(DA)	DIALECTAL ARABIC
(DALG)	DIALECT ALGERIAN
(DL)	DEEP LEARNING
DM	Decision Matrix

(DT)	DECISION TREE
(E)	EXPERTS
(EGY)	EGYPTIAN
(FDOSM)	FUZZY DECISION BY OPINION SCORE
(FWZIC)	FUZZY-WEIGHTED ZERO INCONSISTENCY
(GLF)	GULF
(HPC)	HIGH PERFORMANCE AND CLOUD COMPUTING
(HPC)	HIGH PERFORMANCE AND CLOUD COMPUTING
(IADD)	INTEGRATED DATASET FOR ARABIC DIALECT
(IOT)	INTERNET OF THINGS
(IRQ)	IRAQI
(KNN)	K-NEAREST NEIGHBORS
(LB)	LEXICON BASED
(LEV)	LEVANTINE
(MADA)	MULTI ARABIC DIALECT APPLICATIONS AND RESOURCES
(MAST)	MINI ARABIC SENTIMENT TWEETS DATASE
(MCDM)	MULTI-CRITERIA DECISION-MAKING
(MGH)	MAGHREBI
(ML)	MACHINE LEARNING
(MNB)	MULTINOMIAL NAIVE BAYES
(MRL)	MORPHOLOGICALLY RICH LANGUAGES
(MSA)	MODERN STANDARD ARABIC
(NADiA)	NEWS ARTICLES DATASET IN ARABIC
(NB)	NAIVE BAYES

(Neg)	NEGATIVE
(NilULex)	NILE UNIVERSITY'S ARABIC SENTIMENT LEXICON
(NLP)	NATURAL LANGUAGE PROCESSING
(NLU)	NATURAL LANGUAGE UNDERSTANDING
(OBJ)	OBJECTIVE
(OEA)	OPINION AND EMOTION ANALYSIS
(PL)	PREVIOUS LEXICONS
(Pos)	POSITIVE
(R)	RATING
(SA)	SENTIMENT ANALYSIS
(SANAD)	SINGLE-LABEL ARABIC NEWS ARTICLES DATASET
(SauDiSenti)	SAUDI DIALECT SENTIMENT LEXICON
(SDTC)	SAUDI DIALECT TWITTER CORPUS
(SLR)	SYSTEMATIC LITERATURE REVIEW
(S-NEG)	SUBJECTIVE- NEGATIVE
(S-POS)	SUBJECTIVE-POSITIVE
(SUAR)	SAUDI CORPUS FOR NLP APPLICATIONS AND RESOURCE
(SUBJ)	SUBJECTIVE
(SVM)	SUPPORT VECTOR MACHINES
(TFN)	TERNARY FUZZY NUMBERS
(UN)	UNITED NATIONS
(V)	VOLUNTER

CHAPTER 1

RESEARCH BACKGROUND

1.1 Introduction

This chapter presents a brief overview of the research's background, the problem statement, the research objectives, and questions, in addition to the research scope and significance for this field of study. Section 1.1 is an introduction; Section 1.2 presents the research background. In Section 1.3, the problem statement is defined. In sections 1.4 and Section 1.5, the objectives of the research and the research questions are detailed, respectively. In section 1.6, the research's scope is described. Section 1.7 is the significance of the research. Finally, section 1.8 provides the thesis layout.

1.2 Research Background

In the past decade, the Internet and Social media have become one of the daily life activities essentials (Youssef & El-Beltagy, 2018). Social media applications are considered a famous source for massive amounts of newly streamed data (Najadat, Al-Abdi, & Sayaheen, 2018; Youssef & El-Beltagy, 2018), where people, businesses, and organizations can easily share their information (Abainia, 2019; Younes, Souissi, Achour, & Ferchichi, 2020), thoughts and views (Abdelli, Guerrouf, Tibermacine, & Abdelli, 2019; Alsayat & Elmitwally, 2020; Assiri, Emam, & Al-Dossari, 2016). More than three billion individuals utilize social media to chat, explore, and share their thoughts online (Youssef & El-Beltagy, 2018). As a result of the previous increase in using social media platforms, a tremendous volume of information and user-generated online content have been created (Alnawas & Arici, 2019; Farha & Magdy, 2019; Imane, Kareem, & Faical, 2019) containing generous information (Ombabi, Ouarda, & Alimi, 2020). For example, reviews, tweets, posts, and other forms of social media interactions, such postings have also led to some political events and have shifted public opinion in more than one notable incident, like those in some Arabian countries in 2011. Hence, this necessitates the study of people's ideas and feelings on social media platforms (Zhang, Wang, & Liu, 2018). Due to the growth of data analysis tools in academic literature, it became imperative to implement them, as they are anticipated to play a crucial role in comprehending how people interact, communicate, and exchange data, as well as a dispute with one another (Almanie et al., 2018; Lulu & Elnagar, 2018), Sentiment analysis (SA) is one such instrument, which is part of natural language processing (NLP), SA; commonly known as opinion mining, is a crucial topic in this area, and it has been recognized as an essential method for extracting and analyzing a

text's sentiment and polarity of various online generated contents to study people and public sentiments and emotions (Al-Thubaity, Alqahtani, & Aljandal, 2018; Farha & Magdy, 2019).

SA analyses people's perceptions, attitudes, and emotions relating to broad aspects seen online, such as products, services, business plans, individuals, events, or various themes (Alahmary, Al-Dossari, & Emam, 2019). SA has become essential for decision-makers, business leaders, and everyday consumers (Alotaibi, Mehmood, & Katib, 2019; Assiri, Emam, & Al-Dossari, 2018; Naaima Boudad, Faizi, Thami, & Chiheb, 2018). Reading this description, one may infer that SA is a branch of NLP, and given the long and distinguished history of NLP, it must be the focus of much research. SA emerged in the new millennium thanks to online social media, which allows for evaluating public sentiment on specific issues or products. (Gamal, Alfonse, El-Horbaty, & Salem, 2019; Heikal, Torki, & El-Makky, 2018; Ombabi et al., 2020).

According to Elshakankery and Ahmed (2019), there are various types of SA, and it has been classified into one of three categories: the first is lexicon-based SA, which is dependent on the polarity score of the given text in accordance with the positive and negative values addressed by the word dictionaries. This type also includes so-called topic-based SA, which depends on the given text's topic. The second category of SA is Machine Learning (ML) ML-based, which evaluates texts for polarity, from positive to negative, and other situations. By training ML classifiers with samples of textual emotions, computers learn to identify sentiment without human intervention autonomously. The last kind is hybrid SA, which employs both lexicon-based models and ML models (hybrid). Despite the immense significance of SA, most of its current

study has concentrated on the English language (El-Masri, Altrabsheh, Mansour, & Ramsay, 2017; Elshakankery & Ahmed, 2019; Farha & Magdy, 2019); nevertheless, other languages, notably Arabic, are rapidly becoming a vital portion of the social network community (Alali, Sharef, Murad, Hamdan, & Husin, 2019; Alayba, Palade, England, & Iqbal, 2018; El-Taher, Hammouda, & Abdel-Mageid, 2016).

Recent years have witnessed an increase in multinational analysts focusing on the Arab world (Almanie et al., 2018; Alwakid, Osman, & Hughes-Roberts, 2017) since it represents a significant participant in the global economy and international politics (Abdelli et al., 2019). Arabic is the official language of 22 countries, spoken by more than 400 million people. It is ranked as the fourth most widely used language on the Internet (N Boudad, Faizi, Oulad Haj Thami, & Chiheb, 2017), which makes it famous and notable for some of its related notable events, such as the Arab Spring (Alsayat & Elmitwally, 2020), let alone the growing of social media content for Arabic texts on the Internet. Arabic Language SA requires difficult preprocessing since Arabic as a language is classified in three main varieties (Farghaly & Shaalan, 2009; Habash, 2010; Harrat, Meftouh, & Smaili, 2019): (1) Classical Arabic (CA) which is a form of Arabic language used in literary texts and the Quran (Sharaf & Atwell, 2012),(2) Modern Standard Arabic (MSA) which is a form of Arabic language used in everyday life (N Boudad et al., 2017), and the (3) Dialectal Arabic (DA) which is spoken as street or slang languages which various across the regions and countries. The availability of several dialects, such as (1) Egyptian (Al-Azani & El-Alfy, 2019; Imène Guellil & Azouaou, 2016; Imane et al., 2019), (2) Gulf (Al-Azani & El-Alfy, 2019; Assiri et al., 2018), (3) Iraqi (Al-Azani & El-Alfy, 2019; Assiri et al., 2018; Imène Guellil & Azouaou, 2016), (4) Maghrebi (Al-Azani & El-Alfy, 2019; Alnawas & Arici, 2019),

and numerous more, further contributes to the complexity of Arabic SA (Habash, 2010; Sadat, Mallek, Boudabous, Sellami, & Farzindar, 2014). In addition to these complexities, Arabic speakers on social media often utilize a non-standard Romanization known as *'Arabizi'* (Bies et al., 2014; Darwish, 2013), described as an encoding method that employs the Latin alphabet plus Arabic numerals in place of Arabic letters.

Therefore, developing NLP and SA technologies capable of adequately processing the vast volume of (DA) content on the Internet is crucial. However, before the construction of tools, survey research initiatives on the issue merit examination. To present this language's characteristics and classify the works handling it across various academic contexts. Various and many works were presented as follows.

In the work of Imane Guellil, Saâdane, Azouaou, Gueni, and Nouvel (2021), the authors presented diverse surveys for Arabic SA were offered based on the academic literature; for instance, they examined the literature on Arabic SA studies published between 2015 and 2018 that included all dialects. Another review article by Abo, Raj, and Qazi (2019) examined various characteristics of Arabic SA, such as modern standards and dialects of Arabic languages, in addition to various ML processes and a few well-known algorithms. The following evaluation by Alotaibi et al. (2019) examined resources on (SA) in the Arabic and Saudi dialects by analyzing important tools and approaches, as well as their accuracy in Arabic SA. The authors in Al-Ayyoub, Khamaiseh, Jararweh, and Al-Kabi (2019) evaluated the literature on Arabic sentimental analysis (ASA) and classified the studies according to the addressed issue, the considered data scope, the employed technique, and the used resources. Following

the work of Oueslati, Cambria, HajHmida, and Ounelli (2020), a qualitative evaluation of the most significant research works in this context of Arabic was conducted, and their merits and weaknesses were discussed, specifically, machine translation or transfer learning to convert English resources to Arabic, as well as methodologies originating straight from Arabic. Another study by Imane Guellil, Azouaou, and Mendoza (2019) examined the literature until 2019 to identify the most current resources and advancements in Arabic (SA) regarding resources and tools. The following review by Alsayat and Elmitwally (2020) was conducted in 2019 about the challenges and applications of ASA. The authors of Younes et al. (2020) were concerned with assessing the literature in terms of Maghrebi Arabic dialects (MAD) language resources produced by the different work conducted on MAD language processing.

It is clear from observing a sample of the academic literature that SA studies concerning the Arabic Language are gaining significant momentum, this is apparent given that the Arabic Language is ranked fifth as the most spoken and used language among the top 100 languages around the world (Alotaibi et al., 2019; Mohammed & Kora, 2019; Soliman, Eissa, & El-Beltagy, 2017) with usage by more than 400 million speakers (Alahmary et al., 2019; Imane Guellil et al., 2019; Tartir & Abdul-Nabi, 2017), and it is the official language of 27 countries (Mohamed Elhag Mohamed Abo et al., 2019; Almouzini & Alageel, 2019; Imane et al., 2019). Arabic is a vibrant inflectional language considered one of the richest in terms of morphology and dialect. (Albukhitan, Alnazer, & Helmy, 2020; Alnawas & Arici, 2019; Alwakid et al., 2017), it has some unique characteristics that distinguish it from other languages (Alsayat & Elmitwally, 2020), and these characteristics result in some challenges when trying to study it in the

SA context. These characteristics, along with their challenging aspects and their relation to SA and the problem of this research, are detailed in the following problem statement section.

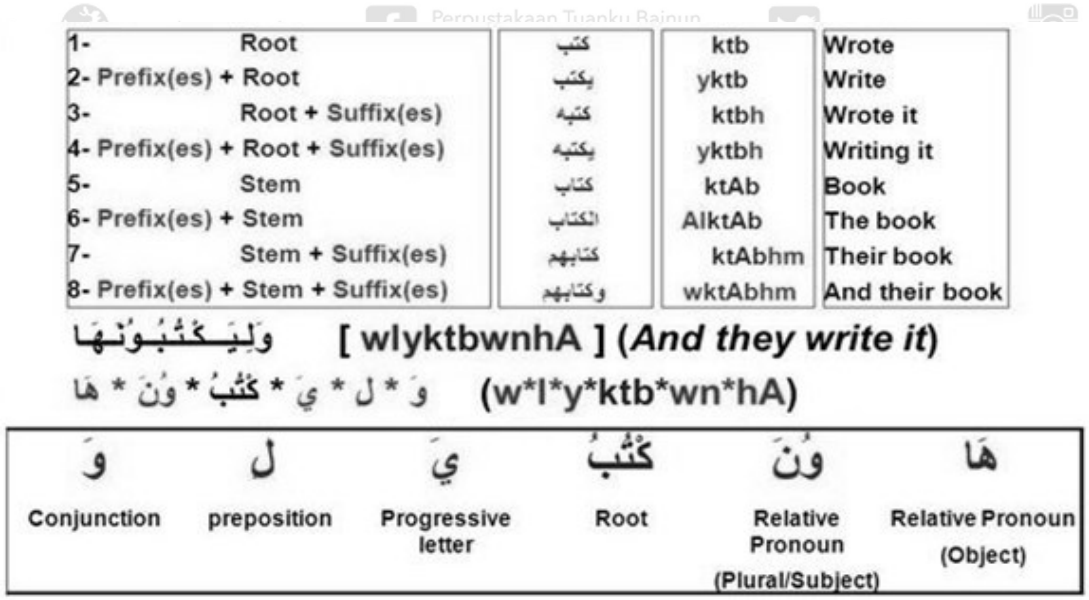
1.3 Problem Statement

While exploring the literature to analyze the (SA) research in terms of the Arabic language, several significant challenges were highlighted in many works; they are deemed significant and play an essential role in SA research in the Arabic language. One of the major challenges is linked to the Arabic Language in terms of its nature and how it has been processed and analyzed in previous SA research, as it is widely recognized that the Arabic language has many characteristics (Al-Twairesh, Al-Khalifa, Al-Salman, & Al-Ohali, 2017; Younes et al., 2020), dialects (Mohamed Elhag Mohamed Abo et al., 2019; Al-Azani & El-Alfy, 2019; Alawami, 2016), vocabularies, grammar, syntaxes, and all other linguistic features (Abainia, 2019; Al-Twairesh et al., 2018; Naaima Boudad et al., 2018), even how it is being used in everyday spoken communications (Lulu & Elnagar, 2018) is challenging since it presents different meanings of the same word (Alahmary et al., 2019) when used in different contexts. It is noteworthy that in Arabic, a single word can express several polarity classes for different expressions (Ombabi et al., 2020) and that simply by looking at where the word is placed in the sentence (Alahmary et al., 2019; Ombabi et al., 2020) and the punctuation of the word could have different meanings, which may cause ambiguity in the meanings (Alotaibi et al., 2019).

Aside from the grammatical challenges of the Arabic language, dialectal challenges are another story. In dialectical Arabic, words sometimes emerge from roots of foreign languages (El-Taher et al., 2016; Soumeur, Mokdadi, Guessoum, & Daoud, 2018; Younes et al., 2020), adding new suffixes, prefixes, words, and expressions and changing the morphology of the language (El-Taher et al., 2016; Younes et al., 2020), adding new words over time the meanings of words (Elshakankery & Ahmed, 2019) or changing topics of interests (Al-Thubaity, Alqahtani, et al., 2018). Figure 1.1presents samples of the suffixes and prefixes in Arabic language.

Figure 1.1

Suffixes and Prefixes Samples



Moreover, aside from the Arabic language nature challenges, another primary class of challenges is also significant regarding how the language has been analyzed. Several research works have encountered so-called technical challenges, where in these

challenges come the resources faced to study the Arabic language, which is ironic given the significant number of Arabic speakers (Alahmary et al., 2019; Imane Guellil et al., 2021; Tartir & Abdul-Nabi, 2017). Extensive studies focused on the English language, while very little focused on the Arabic language (Altowayan & Elnagar, 2017; Elshakankery & Ahmed, 2019; Gamal et al., 2019), especially in terms of SA work on the Arabic language and its dialects (El-Masri et al., 2017; Mohammed & Kora, 2019; Rahab, Zitouni, & Djoudi, 2019); this resulted in a slow growth in SA of the Arabic language (Mohammed & Kora, 2019). The fact that SA for Arabic is considered in the very beginning stages (Al-Kabi et al., 2016); for instance, the accomplished works fundamentally focused on the MSA (Alahmary et al., 2019), while few works mentioned dialects (Imane Guellil et al., 2019). For SA research in Arabic needs data (Abdellaoui & Zrigui, 2018) and standard datasets and lexicons to be used with ML classifiers training and testing purposes (Al-Kabi et al., 2016; Elnagar, Lulu, & Einea, 2018). However, the limited availability (Abdul-Mageed, 2019; El-Taher et al., 2016; Elnagar et al., 2018), with many restrictions in terms of covered dialects, size, and domain, is worthy of investigation (Abdul-Mageed, 2019), especially in comparison with other languages like English (Abdul-Mageed, 2019). Therefore, it is challenging to conduct SA Arabic research given the limited available resources.

Arabic datasets known as Arabic language lexicons have been created in some research work to be used in SA research; however, they suffer several issues like the long annotation process (Al-Kabi et al., 2016), and because many of the currently existing lexicons are questionable due to the absence of orthographic basis for dialects annotated (El-Taher et al., 2016), being built based on the judgment with conflicting agreement (Al-Kabi et al., 2016). In addition, many professional Arabic datasets are

unavailable for research usage (Elnagar et al., 2018), and the available datasets are significantly small compared to large-scale datasets available in the English language (Alsayat & Elmitwally, 2020; Elnagar et al., 2018). It has been demonstrated that the need for more available corpora for ASA is an intricate challenge that Arabic SA is facing (Al-Twairesh et al., 2017). With the absence of helpful large-corpus (Fsih, Boujelbane, & Belguith, 2018) and the unavailability of datasets for research purposes (Elnagar et al., 2018), it is therefore significantly essential to encourage development efforts for such language for large-scale datasets (Elnagar et al., 2018), which can be primarily utilized in SA and NLP works. Therefore, more corpora are required to support the sentiment model's comparative evaluations through Arabic dialect variations (Baly et al., 2017). To establish such corpora and lexicons, it is essential to understand how the current lexicons are built and evaluated, and based on such information, more development efforts can be distinguished to make research in ASA more apparent than before. For example, some currently available Arabic lexicons contain thousands of annotated tweets, which are labeled using normal users. On the other hand, some smaller available lexicons are annotated by language experts. In such a conflicting environment, which of these lexicons can be considered better standards for future development efforts: the one with more amateur-annotated data or the one with less experienced individuals and less data? This creates a conflicting evaluation environment when two measures/criteria are considered, which are the size of the lexicon and who annotated it. Not all criteria has similar importance levels, and their importance varies from one person to another, let alone their variety, and conflict between their values amongst lexicons involved. Each demonstrate ceerrtain characteristics; some of the criteria correspond to benefit analysis, while others demonstrate cost analysis. For instance, the time consumption and cost criterion both

display an inverse proportion, by which the increase in value leads to a decrease in performance indicators such as those measured in terms of either time consumption or cost. In the case of data volume or resource allocation, the benefit criteria reflect a positive correlation, wherein an increase in value leads to a rise in performance. This is shown by the fact that there is a direct link between the two variables. Furthermore, some criteria have (yes/no) values, either by indicating the presence or absence of certain elements. In point of fact, increasing the number of factors and variables to take into account will make the process of lexicon selection a more challenging endeavor. In the context of Arabic lexicons, these factors provide significant obstacles for any possible SA study or lexicon creation.

Based on the above issues, a complex multi-criteria decision-making (MCDM) problem appears in the evaluation and bench-marking of Arabic language lexicons. If these issues are addressed using MCDM, this will create an evaluation and bench-marking decision support platform to support Arabic language lexicon development and, at the same time, assist data scientists and researchers in analyzing more Arabic-related sentiments. Therefore, in this research, an evaluation and bench-marking approach is proposed for Arabic Language SA Lexicons. While achieving this goal, many aspects are considered, like available data characteristics, available lexicons for evaluation, evaluation criteria from literature, their significance, and many more.

Overall, the above issues are outlined as below:

- From the systematic literature review we found that, the most important challenge is how to make the labeling or the annotation of the data, not how to collect the data

- No previous work targeted benchmarking platforms or platforms for decision support.
- When building, developing lexicons, or choosing which criteria is more effective or important than others, several factors need to be taken in consideration such as diglossia, Grammar and Structures, Code-switching, and many other Language.
- To establish lexicons it is essential to understand how the current lexicons are built and evaluated, accordingly more development efforts can be achieved.
- Hence, Multi-criteria Decision-Making (MCDM) approach is essential to bridge the gap and solve the complexity problem by addressing, evaluating and benchmarking platform that supports Sentiment Analysis Arabic dialectal lexicons.

1.4 Research Objectives

This research aimed to develop an evaluation and bench-marking MCDM methodology for Arabic SA lexicons. The main objectives of this research are as follows:

RO1 - To specify the existing methods and techniques for developing Sentiment Analysis Arabic dialectal lexicons through SLR settings.

- **RO2** - To identify and weigh the most significant evaluation criteria utilized in Sentiment Analysis Arabic dialectal lexicons development.
- **RO3** - To propose a multi-criteria Decision Matrix (DM) for Sentiment Analysis Arabic dialectal lexicons evaluation based on the intersection between weighted criteria and alternatives.

- **RO4** - To develop evaluation methodology for the proposed Sentiment Analysis Arabic dialectal lexicons (DM) based on the intersection between weighted criteria and alternatives.
- **RO5** - To test and validate the proposed evaluation methodology for Sentiment Analysis Arabic dialectal lexicons based on systematic ranking and sensitivity analysis.

1.5 Research Questions

The main research questions are as follows:

- **RQ1** - What are the adopted methods and techniques utilized for Sentiment Analysis Arabic dialectal lexicons development? And how they are used for the evaluation of lexicons?
- **RQ2** - Which evaluation criteria is most significant in the evaluation process of the Sentiment Analysis Arabic dialectal lexicons? And what are the levels of importance for this criteria in decision-making context for lexicons development and selection tasks?
- **RQ3** - What are the integrated decision platforms for Sentiment Analysis Arabic dialectal lexicons selection based on weighted criteria?
- **RQ4** - What are the suitable techniques for selecting Sentiment Analysis Arabic dialectal lexicons, based on integrated platforms and weighted criteria?

  RQ5 - To what extent are the results of the proposed methodology for the evaluation of Sentiment Analysis Arabic dialectal lexicons are valid?

Table 1.1 demonstrates the difference in the current state of knowledge regarding what was known from the previous researches before achieving this research objectives, along with the added value that this research contributes after the achievement of its objectives.

Table1.1

What Was Known Before And After Achieving The Research Objectives

Obj	Before Achievement	After Achievement	Scientific Research Added Value
1	• No sufficient knowledge about adopted methods and techniques utilized for Arabic-Based Dialectal Lexicons development. •No sufficient knowledge about how such methods and techniques have been used for the assignment and selection in Arabic-Based Dialectal Lexicons development.	• All Adopted and existing methods utilized in the literature for Arabic-Based Dialectal Lexicons development are identified. •The working mechanism and issues of All existing methods for Arabic-Based Dialectal Lexicons development are studied and analysed	• Detailed knowledge and analysis about Arabic-Based Dialectal Lexicons, and their current state of art, limitations, and potentials for improvements
2	•No sufficient knowledge and understanding about what the most significant factors/criteria utilized in the Arabic-Based Dialectal Lexicons development process. •No sufficient knowledge and understanding about levels of importance for these factors/criteria in the context Arabic-Based Dialectal Lexicons development process.	•All decision making most significant factors/criteria for Arabic-Based Dialectal Lexicons have been identified from literature. •All weights and levels of importance for each criterion in its specific Case Study across both Arabic-Based Dialectal Lexicons have been independently determined	•Different Set of decision criteria for Arabic-Based Dialectal Lexicons were highlighted, and their weights to be used in the decision-making context in each case study of this research has been measured and determined using proposed MCDM solution.
3	•No sufficient knowledge about decision support platforms utilized in Arabic-Based Dialectal Lexicons based on weighted critical Lexical criteria.	•Most decision platforms utilized in the literature in the context of Arabic-Based Dialectal Lexicons based are limited.	• Independent decision matrix for Arabic-Based Dialectal Lexicons was proposed where criteria are weighted and alternatives for various case considered scenarios.
4	•No sufficient knowledge about current selection techniques to be used for Arabic-Based Dialectal Lexicons techniques and their reliance on weighted criteria in the selection.	•Current selection techniques of Arabic-Based Dialectal Lexicons doesn't consider weights and importance levels of criteria in the decision- making context for the selection tasks.	•Detailed analysis on how weighted criteria can be integrated with MCDM while considering various criteria importance levels.
5	•No sufficient knowledge on the validity of proposed MCDM approach for Arabic-Based Dialectal Lexicons assignment tasks.	•The validity of the proposed MCDM approach for Arabic-Based Dialectal Lexicons have been determined using sensitivity analysis and systematic ranking.	•The proposed MCDM approach using integrated Arabic-Based Dialectal Lexicons is valid.

1.6 Research Scope

Regarding the scope of this research, a few points need to be considered.

- This research is aimed at ASA research, and other languages are not considered.
- As part of this research scope, the measures considered for the creation of the decision matrix evaluation and bench-marking platform will be carried with the publicly available Arabic lexicons and any restricted lexicons which are not accessible will not be considered.
- The main output expected from this study is an evaluation and bench-marking decision support platform completed using MCDM integrated methodology for the process of Arabic lexicons evaluation/selection and providing development recommendations. The proposed MCDM approach will be validated using two methods, using various scenarios to demonstrate its effectiveness and validity.

1.7 Research Significance

The findings of this dissertation will render the benefits of different areas related to Society and Research. As for the area of Society, it will contribute towards identifying a better understanding of human feelings and emotions through SA, followed by benefits for better business decision-making strategies. As for the other part of the research, it will discover how SA lexicons evaluation can assist in enlarging development efforts for opinion data on social networks like Twitter, Facebook,

Instagram, etc., and the wide range of applications that can benefit from harvesting the public opinion, and the enormous amounts of opinionated data that are available online. Recommendations and benefits in this line of research vary, and demostrating them in one section cannot cover all their areas, and notions. Therefore, and for more topics and research significances in this area of science, section 2.4.2 Motivation was designed.

1.8 Thesis Layout

This thesis consists of six chapters. Chapter One provides a background about the area of Arabic Language SA. After that, a brief about the current background of research and current gaps concluded by the state of the problem with regards to evaluation and bench-marking of Arabic Language SA Lexicons, research objectives, questions, scope, and research significance, the rest of the thesis is organized as the following:

Chapter Two: In Chapter Two, an in-depth investigation was conducted for the SA research in Arabic Language. This includes defining the terms (Queries) used for investigating the current literature. An (SLR) systematic literature review protocol is adapted to review and analyze the literature to construct a taxonomy where we distribute our result articles to map out this area of science and extract essential elements like challenges, which later allow us to draw our gaps and research problems.

Chapter Three: In this chapter, the research methodology and the research follow-up are designed and reported. In addition to that, the main experiments to achieve the

research objectives are designed. This includes experiments for creating the requirements for the evaluation and bench-marking methodology.

Chapter Foure: Provides an illustration of the benchmarking criteria, including both the phase involving participating experts and the follow-up evaluation process. The evaluation procedure include all of the main criteria, which will be elaborated upon extensively. The aforementioned criteria were operationalized into a questionnaire format, whereby each point to be assessed was assigned five degrees of relevance.

Chapter Five: Presents a comprehensive analysis of the assessment procedure used to evaluate and benchmark the results of Multi-Criteria Decision Making (MCDM). The discussions presented herein are predicated upon the discoveries facilitated by the utilisation of the fuzzy-weighted zero inconsistency (FWZIC) and fuzzy decision by opinion score method (FDOSM) approaches, correspondingly. The evaluation technique played a crucial role in empirical research, as it was necessary to establish the reliability and validity of the results obtained by the proposed approaches.

Chapter Six: This chapter presents a comprehensive review of the key findings derived from the previous chapters, contributions, limitations, and future research prospects for more research that emerges from this study.

1.9 Operational Definitions

Below in Table 1.2 a detailed explanations of the technical terms aused during this PhD thesis.

Table 1.2

Operational Definitions Explanations

Term	Definition	Chapter
ALGORITHM	Refers to a sequence of instructions used to perform calculations and data processing.	Chapter 01
ALTERNATIVES	Refers to lexicons which are to be used in the evaluation and benchmarking process.	Chapter 03
AMAZON MECHANICAL TURK	Refers to service to manually annotate the dataset using the Boto API.	Chapter 02
ANNOTATATION	Refers to a process of labeling of emotion, opinion, or sentiment inside a body of text	Chapter 01
AUTHORS	Refers to persons who have been included in the list of authors for a certain paper, piece of study, article, manuscript, or publication because they made significant contribution to its creation and participated or fully performed the annotation process.	Chapter 01
BENCHMARKING	Refers to the act of measuring the quality of an entity by comparing it with another one of an accepted standard.	Chapter 01
CORPORA	Refers to dataset, consisting of natively digitalized language resources in a machine-readable format,, either annotated or unannotated.	Chapter 01
CORPUS	Refers to dataset, consisting of natively digitalized language resources in a machine-readable format,, either annotated or unannotated.	Chapter 01
CRITERIA	Refers to a principle for judging, and evaluating lexicons different properties and attributes	Chapter 01
CROWDSOURCING	Refers to a method of enrolling a "crowd" or group of individuals to gain required expertise or services about a given instance," such as manual annotation.	Chapter 02
DATASET	Refers to a structured collection of lexical data organized and stored together for analysis or processing.	Chapter 01
DECISION TREE	Refers to a decision support hierarchical model that uses a tree-like model of decisions and their possible consequences,	Chapter 06
DEEP LEARNING	Refers to a subset of machine learning methods that uses multilayered neural networks, called deep neural networks, to simulate the complex decision-making power of the human brain.	Chapter 02
DIALECT	Refers to regional speech patterns that is distinguished from other varieties of the same language by features of phonology.	Chapter 01
EMOTION ANALYSIS	Refers to classifying text into different types of categories anger, disgust, fear, happiness, sadness, and surprise.	Chapter 02
EXPERTS	Refers to humans who have expertise in linguistics fields linked to sentiment analysis and data annotation, and who have been enlisted to do a particular activity as part of a scientific experiment or study.	Chapter 01
GLOBAL WEIGHT	Refers to values assigned to criteria which represent their overall importance in the decision-making process.	Chapter 04
GULF	Refers to Saudi Arabia, Kuwait, Oman, Qatar, Bahrain, United Arab Emirates, Yemen and Iraq	Chapter 01
K-NEAREST NEIGHBORS	Refers ton non-parametric algorithm which uses supervised learning to make classifications or predictions about the grouping of an individual data point.	Chapter 06
LABELING	Refers to a process of identifying raw data and adding one or more meaningful and informative labels to provide context so that a machine learning model can learn from it	Chapter 01
LEVANTINE	Refers to Syria, Jordan, Lebanon, and Palestine	Chapter 01

Term	Definition	Chapter
LEXICONS	Refers to a collection of words or phrases along with additional information about them, such as their meanings, pronunciations, parts of speech, and grammatical	Chapter 02
LOCAL WEIGHT	Local weight refers to the relative importance assigned to each criterion at the local level within a decision-making process. These weights are specific to a particular decision context and are used to prioritize criteria based on their significance in the decision-making process.	Chapter 04
MULTIDIMENSIONAL DECISION MATRIX	Refers to a qualitative technique used to rank the multi-dimensional options of an option set	Chapter 03
NAIVE BAYES	Refers to a linear "probabilistic classifiers" which assumes that the features are conditionally independent, given the target class.	Chapter 02
NATURAL LANGUAGE PROCESSING	Refers to a branch of artificial intelligence that enables computers to comprehend, generate, and manipulate human language.	Chapter 01
NEGATIVE	Refers to expressions that don't like to appear in positive sentences, and can be used only in certain environments, negative environments in particular.	Chapter 02
NEUTRAL	In the sentiment level neutral usually means no opinion	Chapter 02
OPINION ANALYSIS	Refers to classifying a piece of text into positive versus negative classes.	Chapter 02
POSITIVE	Refers to expressions that may normally only be used in a semantically or syntactically positive or affirmative context	Chapter 02
RATING SYSTEM	Refers to any application, portal, or system procedures, data control tools that allow the implementation of classification rules, rating plans, and rating values.	Chapter 02
SENTIMENT ANALYSIS	Refers to a method of extracting and analyzing people's thoughts , attitudes, and emotions relating things such as products, services, businesses, individuals, events, or in any variety of themes.	Chapter 01
SOCIAL MEDIA PLATFORMS	Refers to a type of online social media service which people use to build social networks or social relationships with other people who share similar personal or career content, interests, activities, backgrounds or real-life connections.	Chapter 01
SUPPORT VECTOR MACHINE	Refers to a type of supervised learning algorithm used in machine learning to solve classification and regression tasks.	Chapter 01
SYSTEMATIC LITERATURE REVIEW	Refers to the value when a criteria is placed on a weighing scale	Chapter 01
TERNARY FUZZY NUMBERS	Set of three digits numbers	Chapter 03
VOLUNTEERS	Refers to an ordinary members of the general public who participated in the annotation	Chapter 02
WEIGHT	Refers to determining the value representing the relative importance of an object according to decision-makers.	Chapter 02