



Sentiment analysis of reviews written in the Saudi dialect
to classify customer satisfaction levels in e-commerce.

Prepared By

Areej Haddaj Alqarni

252001729

Submitted in fulfillment of the requirements for the degree Master of
Science in Artificial Intelligence

At

College of Informatics

Midocean University

UAE

Supervisor: Dr. Yasser Abdul Hamid

College of Informatics, Midocean University, Comoros

[2026- Third Trimester]

Abstract

Sentiment analysis is crucial for all businesses and has become an effective method for researching customer satisfaction in e-commerce, given the widespread presence of e-commerce on electronic platforms for purchase decision-making purposes. To address this challenge, in this study, we have developed a sentiment analysis model of Customer Reviews using Arabic, primarily covering the Saudi dialect, to help companies effectively assess the quality of products and services. The second part involves a literature review of Arabic sentiment analysis and six academic papers that have added significant rigor to the theoretical frameworks supporting our study. A dataset of customer reviews in the Saudi dialect corpus was collected and compiled using text preprocessing (cleaning the data and normalizing, filtering, and extracting their features). Then, a sentiment analysis model was developed with different machine learning algorithms, including various classifiers and Aspect-Based Sentiment Analysis (ABSA), that can determine customer satisfaction from positive, negative, and neutral perspectives. In this work, we proposed a model for Arabic text analysis in the Saudi dialect and suggested a robust methodology that allows our language to effectively handle the inclusion of some informal texts. It also offers an annotated dataset available for consideration in future studies. The findings can provide more insight into customer satisfaction trends for Saudi companies in their market, thereby facilitating information-driven decision-making for e-businesses.

Table of Contents

Abstract	1
List of Abbreviations	3
Introduction	4
1.1 Problem Definition	4
1.2 Thesis Objectives	4
1.3 Thesis Contribution	5
1.4 Thesis Organization	خطأ! الإشارة المرجعية غير معروفة.
2 Background	6
2.1 The Role of Sentiment Analysis in E-Commerce	6
2.2 Sentiment analysis in the Arabic language	6
2.3 Machine Learning Approaches	6
2.4 Pre-trained Language Models	6
2.5 Datasets in Arabic Sentiment Analysis	7
.1 Related work	8
2. Literature Review Matrix	9
3 Research Gaps	10
4 Conclusion	10
References	11

List of Abbreviations

- NLP: Natural Language Processing
- SA: Sentiment Analysis
- SVM: Support Vector Machine
- NB: Naive Bayes
- KNN: K-Nearest Neighbors
- DT: Decision Tree
- RF: Random Forest
- ABSA: Aspect-Based Sentiment Analysis
- TF-IDF: Term Frequency – Inverse Document Frequency
- BOW: bag of words
- POS: Part-of-Speech Tagging

Chapter 1

Introduction

1.1 Problem Definition

E-commerce has experienced significant growth in recent years, and dependence on digital platforms is an integral part of customers' decision-making processes. Nowadays, customer reviews and textual evaluations are regarded as the main sources of understanding user experiences with products and services. The quantity of consumer content created by users on e-commerce sites has made the challenge for organizations to deal with the large and continuously growing volume of data generated today an extremely serious one compared to earlier days when data was analyzed by traditional methods or through manual analysis, making data processing no longer possible. This highlights the need to create more sophisticated methods based on artificial intelligence technologies for customer sentiment analysis and the accurate determination of customers' sentiments so that enterprises can modify the quality of products offered, improve the quality of services, and help with future decision-making. Various Arabic studies have tackled the subject of sentiment analysis, but the proposed approaches still need more development in terms of data, accuracy, and the variety of data sources. In addition, there is still a clear need to contribute better natural language processing methods that can process dialectal or linguistics-specific information, analyze textual content, and make the analysis more accurate and effective.

1.2 Thesis Objectives

This thesis presents the following aims:

1. To collect and prepare a dataset of customer reviews written in the Saudi dialect from e-commerce platforms.
2. To apply Natural Language Processing techniques to preprocess, clean, and analyze informal text written in Saudi dialect.
3. To develop a sentiment analysis model, including Aspect-Based Sentiment Analysis (ABSA), capable of classifying customer satisfaction levels (positive, negative, and neutral).
4. To evaluate and compare the performance of various machine learning and deep learning algorithms in sentiment classification.
5. To identify the most influential linguistic features that contribute to determining sentiment in Saudi dialect customer reviews.

1.3 Thesis Contribution

This thesis also makes several scientific contributions, including:

1. A sentiment analysis model tailored for customer reviews, the Saudi dialect (a relatively new area of research).
2. Defining clear processes for processing informal Arabic text in the e-commerce context (e.g., preprocessing, analysis and feature extraction).
3. An annotated dataset, which would be useful to conduct future research on sentiment analysis in Arabic.
4. Making strides in the study of customer satisfaction in the Saudi market, data-driven decision-making for e-commerce platforms.

1.3 Thesis Organization

The thesis is organized into a number of chapters as follows:

1. Chapter 1: Introduction introduces the research topic, objectives, research problem, and study significance.
2. Chapter 2: Background establishes the theory behind the study.
3. Chapter 3: Literature Review: The literature review presents previous studies, and compares them to one another.
4. Chapter 4: Methodology: Methodology describes the research design, data collection methods, preprocessing steps, and the machine learning techniques involved.
5. In Chapter 5: Results we reported the results of experimental results and an evaluation of model performance.
6. Chapter 6: Conclusions and Future Work gives an indication of the major findings and their contributions and shortcomings with recommendations for future research.

Chapter 2

2 Background

2.1 The Role of Sentiment Analysis in E-Commerce

E-commerce is at a period of phenomenal development in the past several years and has become a critical component in consumers' purchasing behavior and is now making a significant contribution towards an overall change in the purchasing practices of buyers. This growth has led organizations to understand customers through the digital channel and by reading opinions and reviews on every digital medium, including every website and social network. Today's market cannot be limited to the traditional feedback methods alone, markets must have smart tools to understand customer digital impressions in order to improve service quality and lead to data-driven decisions and improved levels of customer satisfaction. Thus, the need to develop digital tools in which companies can find out what consumers really think of their products, especially since there is this huge volume of reviews that cannot be done manually, can only give rise to an impression of a good situation. Sentiment analysis provides a good insight of customer beliefs by analyzing the reviews of customers in text with text content processing and showing whether they are positive/negative/neutral.

2.2 Sentiment analysis in the Arabic language

In Arabic, sentiment analysis plays a crucial role, especially in analyzing customer reviews and social media. Language analysts struggle with dialectal variation, morphological complexity, and ambiguity in informal or abbreviated texts. In the face of these challenges, numerous studies have emphasized the importance of constructing corpora and lexicons for various dialects such as Jordanian, Egyptian, and Saudi Arabic in order to create more robust and efficient models of sentiment analysis in Arabic. A blend of traditional and modern algorithms is being used to analyse types of text and to classify sentiment, each with a characteristic value and operation approach. Among the most widely used algorithms in previous studies are the following:

2.3 Machine Learning Approaches

Much work has been done on Arabic sentiment analysis using machine learning techniques, and the studies mainly focus on specific stages starting from text preprocessing to ensure the quality of data before the classification. This part consists of noise removal, text normalization, and dealing with the many language features in the Arabic language. Then, feature extraction techniques were used to make numerical representation of textual data, allowing machine learning algorithms to process this effectively. [1] used six different feature extraction combinations to discover the most robust way to improve model accuracy. Several machine learning classifiers were evaluated, including SVM, NB, RF, and KNN. The main result of this study is that the SVM

model combined with TF-IDF, POS tagging, and bigram features attained the greatest rate of performance among the models evaluated. Additionally, the research conducted by [9] proved that the performance of machine learning algorithms was improved significantly when they convert local dialects texts into Modern Standard Arabic (MSA). Specifically, the SVM classifier achieved the best performance after this conversion, indicating the significance of linguistic normalization in Arabic sentiment analysis. [12] proved that the SVM algorithm shows strong performance very close to the deep learning model CAMELBERT, which is further proof indicating that the SVM algorithm can perform very fast and effectively in Arabic sentiment analysis tasks.

2.4 Pre-trained Language Models

These are artificial intelligence models that were first trained with a large set of texts to learn the linguistic structures, context, and relationships of words. They can later be fine-tuned for particular tasks, such as text classification, sentiment analysis, question answering, or information extraction, without need to train them from scratch: freeing up time and resources and increasing accuracy. There are clear gaps in previous investigations since the amounts of training data used to fit models are small, and many were collected by hand from various platforms. This lack of sufficient data directly affected the models' performance, and it might lead to some inaccuracies for the models' evaluation results due to the limited dataset.

2.5 Datasets in Arabic Sentiment Analysis

In previous studies on sentiment analysis, datasets tend to be sourced from websites like social media or review websites such as Twitter or e-commerce platforms. The latter dataset type includes user-generated texts in Modern Standard Arabic or various Arabic dialects. Texts are typically annotated into positive, negative, neutral, and otherwise, using preprocessing such as the removal of noise, normalization, and stemming to give high quality data and to improve the processing performance. However, compiling high-quality Arabic sentiment data is still difficult due to the range of dialects, the rise of informal language practices, and a shortage of publicly annotated resources. In fact, in the study [1], researchers developed aspect-based sentiment analysis (ABSA) compared to conventional full-text sentiment analysis. For this, a carefully annotated dataset on aspects can be needed which can include factors such as price, quality, service, and so forth, to correctly identify sentiment of the other aspects. Our results showed that ABSA delivers a much more detailed and in-depth analysis than general sentiment analysis, offering actionable and accurate insights into customer opinions and a richer grasp of user experience.

Chapter 3

1. Related work

In study [11], the researchers built a dataset of Arabic tweets in the Jordanian dialect. The data were collected using keywords related to general Jordanian topics. Data was preprocessed: data cleaning, removal of URLs, unwanted characters, repeated words, and hashtags. The tweets were subsequently annotated manually and categorized as positive, negative, or neutral. The dataset was trained with conventional machine learning algorithms like Support Vector Machine (SVM) and Naïve Bayes. TF-IDF representation along with stemming and bigram features in the SVM model led to the best accuracy, F1-score, and recall scores, according to the findings.

Similarly, in [9] in particular, focused on the citizens' opinions towards the traffic conditions in three Moroccan cities. Responses were gathered from social media posts on Twitter and Facebook in the Moroccan dialect. The authors used standard machine learning algorithms, and it was observed that the SVM model performed better than other algorithms after translating the Moroccan dialect into Modern Standard Arabic (MSA) with an accuracy of 92%. These findings also indicated a predominance of negative sentiments, particularly in the city of Rabat.

In 2024, study [1] applied Aspect-Based Sentiment Analysis (ABSA) to analyze customer reviews written in the Saudi dialect using Natural Language Processing techniques and machine learning algorithms. Two separate datasets were extracted from e-commerce platforms such as Amazon and Shein, one for the clothing category and the other for electronics. The data were then annotated by recognizing relevant aspects, such as quality and price in the electronics category, as well as size and color in the clothing category, and all values for sentiment levels were set as 1 (positive), -1 (negative), or 0 (neutral). The results of the evaluation showed that the Support Vector Machine (SVM) method outperformed other models with an accuracy of 86.5%, an F1-score of 85.82%, and a recall of 86.50% when using a combination of TF-IDF, POS, and bigram features.

In study[13] , the sentiment of comments on the YouTube platform to enable content creators to assess viewers satisfaction with their content. This study addressed the challenge resulting from the removal of the dislike count feature on YouTube by relying on user comments as an alternative indicator of audience perception. The data were collected using a Python-based script connected to the YouTube Application Programming Interface (API), through which 4,760 comments were gathered from various video categories. The sentiments were manually labeled by five annotators into two classes: positive and negative. After removing irrelevant comments, a balanced dataset consisting of 4,212 comments was obtained, equally divided into 2,106 positive and 2,106 negative comments. Following text preprocessing, several machine learning classifiers were evaluated. The

results demonstrated that the Naïve Bayes classifier outperformed the other classifiers, achieving an accuracy of 94.62%, a precision of 94.64%, a recall of 94.64%, and an F1-score of 94.62%.

In study [12], The researchers sought to design an opinion-based sentiment system for Arabic language opinions of higher education students in order to enhance and improve the learning process. The paper advanced this SVM-SAA-SCR model which prioritizes text preprocessing prior to classifying opinions under Positive, Negative and Neutral sentiment using SVM (Support Vector Machine). The actual dataset obtained from Prince Sattam bin Abdulaziz University (PSAU) which comprises students perspectives on several educational topics. The performance of the system was assessed by comparing it to the advanced CAMELBERT model. The difference between the two models was small, which confirmed the robustness and effectiveness of the proposed SVM–SAA–SCR algorithm.

2. Literature Review Matrix

This table provides a summary of the most relevant previous studies related to sentiment analysis in the context of e-commerce, highlighting the methodologies used and the key findings of each study.

Study	Algorithm	Dataset	Key Findings
Alrefae, et al., (2024)	SVM, Random Forest (RF), Naïve Bayes (NB), KNN	Amazon, Jarir, Shein, Noon, and Twitter	1. For the electronics dataset, the RF algorithm worked well with TF-IDF, POS tagging, and trigrams, achieving an average accuracy of 86.26% per aspect. 2. For the clothing dataset, using TF-IDF along with POS tagging and bigrams, an average accuracy of 86.51% per aspect was achieved.
Al-Thubaity, et al.,(2018)	Lexicon-based approach	Collect data in twitter	SauDiSenti performed better for neutral tweets and showed promising results despite being smaller than AraSenTi.
Alomari, et al.,(2017)	SVM +Naïve Bayes	collected from the X platform(Twitter).	The results showed that SVM outperformed Naïve Bayes in terms of accuracy and F-score.

Dahbi,et al.,(2019)	Naive Bayes, SVM,KNN, Decision Tree	using a Python script from Twitter and Facebook	SVM algorithm achieved the highest performance after converting the texts from the Moroccan dialect to Modern Standard Arabic (MSA).
Louati, et al., (2023)	SVM	real student opinions from the Department of Computer Science and Information Systems at Prince Sattam University.	The SVM-SAA-SCR model achieved strong performance with an accuracy of 84.7% ,Precision of 86% and Recall 85%.
Musleh, et al.,(2023)	NB, SVM,KNN, DT,RF	YouTube comments	Achieved an accuracy of 94.62%, precision of 94.64%, recall of 94.64%, and F1-score of 94.62%, making it the best-performing classifier among the six evaluated classifiers.

3 Research Gaps

Most previous Arabic sentiment analysis studies have scarce resources and datasets on Arabic language, while limited datasets that do not include all dialects and domains were considered. In addition, the linguistic lexicons and dictionaries are still not well utilized, and have not been integrated with modern techniques, such as Transfer Learning, which limits the model's ability to understand nuanced and figurative expressions and accurately analyze sentiments, delaying the construction of proper tools for Arabic text analysis.

4 Conclusion

The studies highlight the importance of sentiment analysis in business development and in leveraging information extracted from social media texts to enhance and improve business practices. They demonstrate that Arabic text analysis is feasible and effective, whether in Modern Standard Arabic or in various dialects, with the necessity of performing text preprocessing, feature extraction, and considering dialectal differences to improve accuracy. The results also showed the effectiveness of machine learning algorithms, particularly the SVM algorithm, which achieved the best performance and was considered the most efficient scenario among the evaluated classifiers.

References

- [1] Alrefae, R., Alqahmi, R., Alduraibi, M., Almatrafi, S., & Alayed, A. (2024). Enhancing Customer Experience Through Arabic Aspect-Based Sentiment Analysis of Saudi Reviews. *International Journal of Advanced Computer Science & Applications*, 15(7).
- [2] Hicham, N., Karim, S., & Habbat, N. (2023). Enhancing Arabic Sentiment Analysis in E-Commerce Reviews on Social Media Through a Stacked Ensemble Deep Learning Approach. *Mathematical Modelling of Engineering Problems*, 10(3).
- [3] Al-Thubaity, A., Alqahtani, Q., & Aljandal, A. (2018). Sentiment lexicon for sentiment analysis of Saudi dialect tweets. *Procedia computer science*, 142, 301-307.
- [4] Alotaibi, S., Mehmood, R., & Katib, I. (2019, June). Sentiment analysis of Arabic tweets in smart cities: A review of Saudi dialect. In *2019 Fourth International Conference on Fog and Mobile Edge Computing (FMEC)* (pp. 330-335). IEEE.
- [5] Abo, M. E. M., Qazi, A., Saad, A. A. A., Elsaheb, H. A., & Abdelaziz, A. (2024, August). Sentiment Analysis in Saudi Arabic Dialect for HAJJ Season Services Using Twitter Data. In *2024 1st International Conference on Logistics (ICL)* (pp. 1-5). IEEE.
- [6] Alahmary, R. M., Al-Dossari, H. Z., & Emam, A. Z. (2019, January). Sentiment analysis of Saudi dialect using deep learning techniques. In *2019 International conference on electronics, information, and communication (ICEIC)* (pp. 1-6). IEEE.
- [7] Kaseb, A., & Farouk, M. (2022, December). SAIDS: A novel approach for sentiment analysis informed of dialect and sarcasm. In *Proceedings of the Seventh Arabic Natural Language Processing Workshop (WANLP)* (pp. 22-30).
- [8] AlFutamani, A. A., & Al-Baity, H. H. (2021). Emotional Analysis of Arabic Saudi Dialect Tweets Using a Supervised Learning Approach. *Intelligent Automation & Soft Computing*, 29(1).
- [9] Dahbi, M., Saadane, R., & Mbarki, S. (2019, October). Social media sentiment monitoring in smart cities: an application to Moroccan dialects. In *Proceedings of the 4th International Conference on Smart City Applications* (pp. 1-6).
- [10] AlFutamani, A. A., & Al-Baity, H. H. (2021). Emotional Analysis of Arabic Saudi Dialect Tweets Using a Supervised Learning Approach. *Intelligent Automation & Soft Computing*, 29(1).
- [11] Alomari, K. M., ElSherif, H. M., & Shaalan, K. (2017, June). Arabic tweets sentimental analysis using machine learning. In *International conference on industrial, engineering and other applications of applied intelligent systems* (pp. 602-610). Cham: Springer International Publishing.

- [12] Louati, A., Louati, H., Kariri, E., Alaskar, F., & Alotaibi, A. (2023). Sentiment analysis of Arabic course reviews of a Saudi university using support vector machine. *Applied Sciences*, 13(23), 12539.
- [13] Musleh, D. A., Alkhwaja, I., Alkhwaja, A., Alghamdi, M., Abahussain, H., Alfawaz, F., ... & Abdulqader, M. M. (2023). Arabic sentiment analysis of youtube comments: Nlp-based machine learning approaches for content evaluation. *big data and cognitive computing*, 7(3), 127.