

Midocean University
Master's Student, Department of Artificial Intelligence

Project Proposal

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

Submitted by

Areej Haddaj ALqarni

Supervisor

Dr. Yasser Abdul Hamid

1. Project Overview

This project aims to develop a sentiment analysis system that makes use of the Saudi dialect in classifying customer satisfaction levels in the e-commerce sector.

The importance of this research is very much related to the rapid growth of online business activities within Saudi Arabia, leading to a huge volume of customer reviews and feedback reflecting user experiences.

Since most of the customers express their opinions using the Saudi dialect, the traditional sentiment analysis systems poorly capture the actual context and meaning of such texts written in Modern Standard Arabic.

This project aims at filling this gap by developing a model that will be able to comprehend the linguistic and cultural traits of the Saudi dialect and, after analysis, present conclusions on the level of customer satisfaction, whether positive, neutral, or negative.

The outcome that can be expected is improved service quality, building customer trust, and facilitating data-driven decisions by business houses for better customer relationship management.

2. Project Goal

- To develop a model capable of analyzing sentiments written in the Saudi dialect.
- To classify customer satisfaction levels as positive, negative, or neutral.
- To build a system that understands the Saudi dialect, which contributes to a more accurate interpretation of customer reviews.
- To enable e-commerce platforms to make data-driven decisions that enhance service quality and improve customer experience.

3. Approach

- **Data Collection:** Customer review data will be collected from e-commerce platforms such as Amazon.sa and Noon, with a focus on customer comments written in the Saudi dialect. The collected dataset will include key fields such as the product name, rating, review date, and review text, and it will be stored in CSV format for further processing and analysis.

- **Data Processing:** includes several steps to prepare the text for analysis. First, the text is cleaned by removing unnecessary symbols, non-Arabic characters, and repetitive letters. Next, a specialized Saudi dialect lexicon is created or existing linguistic resources are utilized to interpret local vocabulary. Then, the tokenization process is performed using Arabic NLP tools such as Farasa, CAMEL Tools, or the AraBERT Tokenizer. Stop words that do not contribute significant meaning to the context are removed, while sentiment-relevant words are retained. Finally, the processed text is converted into numerical representations (encoding) using BERT-based models such as SaudiBERT, AraBERT, or MARBERT, in preparation for training and analysis.
- **Model Building:** in this stage, the proper algorithm or model will be chosen for learning patterns in the labeled sentiment data. Machine learning algorithms commonly include SVM, Naive Bayes, and Random Forest. Deep learning models include LSTM, GRU, CNN, and transformer-based models like BERT and AraBERT. Sometimes, hybrid models with several combined algorithms or combinations including both machine learning and deep learning are put into place, aiming to increase accuracy. The model then performs sentiment classification after being trained, where each review is categorized into a sentiment class like Positive, Negative, or Neutral.
- **Testing and Evaluation:** In this stage, the model will be tested on a separate dataset that was not used during training, to ensure its accuracy and reliability in handling sentiment classification tasks.

[illegible]

5. Expected Results

- Compiling a dataset containing customer reviews in the Saudi dialect from e-commerce websites.
- Training a model capable of classifying reviews based on sentiment analysis in Arabic, especially the Saudi dialect into positive, negative, or neutral.

6. Resources Needed

Python , pandas, Natural language (NLP) tools: Farasa , AraBERT , jupyter Notebook .

7. References

[1]. Alahmary,R,Aldosari,H,Emam,A.(2019 January).“Sentiment Analysis of Saudi Dialect Using Deep Learning Techniques”. International Conference on Electronics, Information and Communications (ICEIC).