Sentiment Analysis for Egyptian Socioeconomic Topics

Omar Reda-211001265

Yousef Akram-211000014

**Presented to Dr:** Ghada Khoriba

**Eng**/Amira Tarek

**Table of Contents**

[Introduction 3](#_Toc187163856)

[Methodology 4](#_Toc187163857)

[Dataset Description 4](#_Toc187163858)

[Data Preprocessing 4](#_Toc187163859)

[Feature Extraction 4](#_Toc187163860)

[Model Selection and Traning 5](#_Toc187163861)

[Evaluation Metrics 5](#_Toc187163862)

[Results 5](#_Toc187163863)

[Conclusion 6](#_Toc187163864)

**Abstract**

The rapid growth of social media platforms has made them a crucial source of public opinion on various socioeconomic issues. This research focuses on analyzing public sentiment expressed in Arabic tweets related to Egyptian socioeconomic topics using machine learning techniques. The study employs a robust framework for sentiment classification, utilizing a dataset of tweets annotated with positive and negative sentiments. Machine learning models, including Logistic Regression, Support Vector Machines (SVM), Naïve Bayes, and Recurrent Neural Networks (RNN), were implemented and evaluated. The SVM model achieved the highest accuracy of 87%, showcasing its effectiveness in handling high-dimensional text data. This work highlights the importance of understanding public sentiment for informed policymaking and improved communication strategies, offering valuable insights into societal trends and public concerns.

# Introduction

The surge in social media usage has transformed platforms like Twitter into vital arenas for public discourse, offering insights into opinions on critical issues such as economic reforms, government policies, and social initiatives. Sentiment analysis, a subset of natural language processing (NLP), has emerged as a powerful tool to analyze these opinions, enabling organizations and policymakers to assess public sentiment and align their strategies accordingly.

Arabic sentiment analysis presents unique challenges due to the linguistic diversity and complexity of Arabic dialects, particularly Egyptian Arabic. Unlike Modern Standard Arabic (MSA), Egyptian Arabic incorporates colloquialisms, regional expressions, and informal structures, making computational analysis challenging. Despite these difficulties, analyzing sentiment in Egyptian Arabic is crucial, as it reflects the perspectives of one of the most widely spoken dialects in the Arab world.

The primary objective of this research is to develop a sentiment analysis framework tailored for Arabic tweets, particularly those written in Egyptian Arabic. By focusing on socioeconomic topics, this study provides actionable insights into public opinion, offering a foundation for effective communication and decision-making.

# Methodology

This study utilizes machine learning techniques to analyze and classify sentiment in Arabic tweets. The workflow consists of data preprocessing, feature extraction, model training, and evaluation to ensure reliable sentiment classification.

## Dataset Description

Tweets were collected using the Twitter API, focusing on hashtags and keywords related to Egyptian socioeconomic issues. The dataset includes tweets labeled as positive or negative, covering topics such as economic reforms, environmental policies, and social initiatives.

## Data Preprocessing

* **Text cleaning:** Removal of special characters, URLs, and emojis to ensure the text is in a clean format for analysis.
* **Tokenization and stopword removal:** Breaking down the text into individual tokens and removing non-informative words using Arabic-specific stopword lists
* **Stemming:** Application of Arabic text stemming techniques to normalize word forms.

## Feature Extraction

* **TF-IDF Vectorization:** The Term Frequency-Inverse Document Frequency method was applied to convert text data into numerical features.

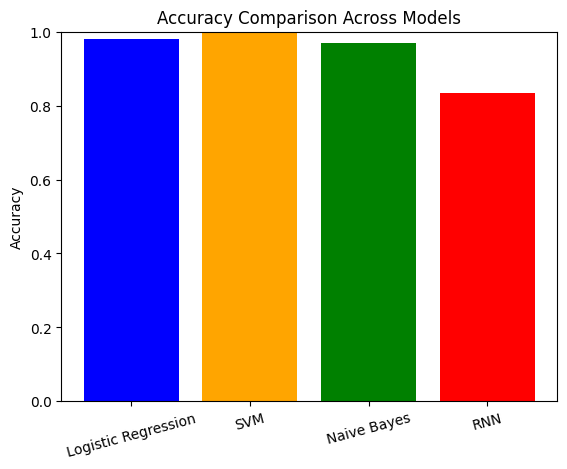
## Model Selection and Traning

Logistic Regression: A simple yet effective linear classifier for binary sentiment classification. Showing accuracy of 98%

SVM: Demonstrated the best performance with an accuracy of 99%,

Naïve Bayes: Showed efficiency with accuracy 97%

RNN: Captured sequential patterns in text, achieving an accuracy of 89%.



# Results

The results of the sentiment analysis revealed significant trends in public opinion:

The best-performing model was the Support Vector Machine. Positive sentiment dominated topics related to technological advancements and environmental initiatives, whereas economic reforms exhibited a more mixed sentiment profile. A word cloud analysis highlighted frequently used terms, providing qualitative insights into public discourse.

# Conclusion

This sentiment analysis project successfully demonstrated the utility of machine learning techniques in classifying Arabic text data. By analyzing tweets on Egyptian socioeconomic topics, the study provided valuable insights into public opinion trends. The findings underscore the importance of NLP in understanding societal issues and offer a robust framework for further research in sentiment analysis for Arabic text.

Future work could focus on expanding the dataset to include other forms of social media, incorporating deep learning models for enhanced accuracy, and applying the methodology to other regional contexts to assess cross-cultural sentiment trends.