LEX LOCALIS-JOURNAL OF LOCAL SELF-GOVERNMENT
ISSN:1581-5374 E-ISSN:1855-363X
VOL. 23, NO. S6(2025)

SENTIMENT ANALYSIS FOR JORDANIAN DIALECT USING MEMETIC ALGORITHM AND SUPPORT VECTOR MACHINE CLASSIFIER

Ahmad Alhadab [1,] Ghaith Jaradat [2,] Yazan M. Alomari[3], Rasha Mohammed Alolayan[3], Asyah Muzahim Alkaabi[3], Rana Ibrahim Alzumaia[3], Mohammad Awni Mahmoud[3], Sara S AlOtaishan[3], Mohammed Abuelaila Baraka Yousif[3], Abderrazak Ben Salah[3], Abeer Atallah Aloudat[4]

[1,2]College of Information Technology, Amman Arab University Amman 11953, Jordan
[3]Department of Management Information Systems, College of Applied Studies and Community Services, Imam Abdulrahman Bin Faisal University, Dammam, Saudi Arabia
[4]Accounting Department, College of Applied Studies and Community Service ,Imam Abdulrahman bin Faisal University, Dammam, KSA

[2]g.jaradat@aau.edu.jo

Abstract. Sentiment analysis (SA) is a vital text-mining task for extracting and classifying opinions from textual data as positive, negative, or neutral. This research investigates sentiment analysis for the Jordanian Arabic dialect by integrating a Memetic Algorithm (MA) for feature selection with three machine learning classifiers: Support Vector Machine (SVM), Multinomial Naïve Bayes (MultinomialNB), and Logistic Regression (LR). The core strategy of the MA is to combine global and local search to filter out irrelevant features and select the most relevant ones, thereby optimizing the feature space for classification. The model's accuracy served as the fitness function for the global search (Genetic Algorithm), while Pearson's Correlation Coefficient was used for the local search. Experimental results on the Arabic Jordanian General Tweets (AJGT) dataset demonstrated that the MA consistently improved the performance of all classifiers. The baseline accuracies were: SVM (79.11%), MultinomialNB (83.89%), and LR (82.67%). After applying the MA, the accuracies improved to: SVM (82.22%), MultinomialNB (85.61%), and LR (84.56%). Similar enhancements were observed in Precision, Recall, and F1-Score across all models. The study concludes that MultinomialNB, when coupled with the MA, is the most effective classifier for this task, and the hybrid MA approach significantly boosts sentiment classification performance for the Jordanian dialect.

Keywords. Sentiment Analysis, Support Vector Machine, Memetic Algorithm, Feature Selection, Jordanian Dialect, Genetic Algorithm, Pearson Correlation.

INTRODUCTION

The Arabic language is a widespread language in the world, there are a large number of countries that use this language, there are a set of advantages and characteristics of this language, for example, and there are many derivations in the Arabic language. It can be classified into three main categories: The first category is the classical Arabic language, which is mostly found in religious texts and the Qur'an, The second category: Modern Arabic language (MSA), which is the language widespread in Arabic-speaking countries, As for the third and final category, it is the colloquial language, which is different and varied and difficult to limit, may a one country has more than one colloquial language. Most researchers focused on Modern Standard Arabic (MSA), due to its frequent application using in daily life, books, official speeches and educational curricula, and as a result of these interests, forms and grammatical tools have been proposed and developed to address Natural Language Processing (NLP) in Arabic language [1].

Sentiment analysis is an essential component in categorizing emotions into positive, negative, and neutral groups. It also has a major impact in facilitating informed decision-making based on these emotions. In this work, the researcher will use machine learning classifiers in combination with a memory technique to analyze sentiments and emotions. Enhancing sentiment analysis's precision and effectiveness in identifying and comprehending sentiment is the goal. Through the use of this methodology, the researcher tries to improve sentiment analysis's overall effectiveness as well as its capacity to successfully understand and evaluate emotions. In a broader context, this study demonstrated the effectiveness of using a MEM algorithm in improving the performance of ML algorithms for SA in the Jordanian dialect. This discovery holds significant implications for various fields, including marketing, policy, and customer service, where SA of social media data holds the potential to provide valuable insights into consumer behavior and public opinion.

The proliferation of social media has made public opinion readily accessible, creating a valuable resource for institutions. Sentiment analysis allows organizations to automatically gauge public perception from vast amounts of textual data. However, most SA research focuses on English, with Arabic, a complex and morphologically rich language, receiving less attention. This complexity is compounded by the existence of Modern Standard Arabic (MSA) and numerous regional dialects. This study addresses the challenge of performing SA on the Jordanian dialect. The primary research question is: "How can feelings be analyzed in the Jordanian dialect using the Memetic algorithm and SVM classifier?" The objectives are: (i) to measure the effectiveness of the Memetic Algorithm in analyzing sentiments of the Jordanian dialect; (ii) to provide an analytical system based on the integration of MA with SVM; (iii) to implement a model that analyzes customer feelings from Jordanian dialect reviews.

LITERATURE REVIEW

Sentiment analysis (SA), sometimes known as opinion mining, is the practice of ascertaining authors' opinions about a subject or problem. It also emphasizes word analysis and text classification into neutral, unfavorable, and positive viewpoints. Another factor in SA, recent surge in popularity has been its capacity to evaluate and benefit from the data accessible on online social media platforms like blogs, wikis, and tweeters [2].

Analyzing publically accessible Arabic SA corpora is a popular use of supervised learning techniques in Arabic Sentiment Analysis (ASA) [3, 4]. Hybrid sentiment analysis techniques have been created by researchers to address issues with word origin, spelling, and numerous meanings. One technique combines support vector machines (SVM) with Semantic Orientation (SO) [17]. When it comes to Arabic document classification, (SVM) performs better than the Naïve Bayes (NB) technique. [6]. Tokenization, which is the process of dissecting text into individual words or phrases, is crucial for feature extraction and bag of words techniques in SA. [7]. Using machine learning and natural language processing, emotion analysis facilitates the extraction of subjective data and provides a comprehensive understanding of attitudes and opinions. [8, 9]. Another crucial component is text classification, which is classifying or categorizing texts according to their context and content. NLP methods are frequently used in this process. [10].

SVM, a supervised learning method, can handle both linear and non-linear data, it is frequently employed in text categorization [11] .By using a kerneltrick, and SVM is computationally robust even in non-linear separable cases [12]. Feature selection techniques, including Genetic Algorithm (GA) are employed to optimize SVM's performance [18].

Previous studies on Arabic SA have employed various machine learning techniques with mixed results. Shoukry and Rafea (2012) achieved 64% accuracy with SVM on Arabic tweets. Al-Horaibi and Khan (2016) reported accuracies of 53.75% for SVM and 46.82% for Naïve Bayes. Atoum and Nouman (2019), focusing specifically on the Jordanian dialect, achieved a promising 82.1%

accuracy with SVM. A common thread in these studies is the direct application of classifiers to the data. This research contributes by introducing a Memetic Algorithm as a pre-processing step for intelligent feature selection, aiming to enhance classifier performance by eliminating noise and retaining the most sentiment-indicative features.

Other classification methods like (NBM) and (LR) are also utilized in sentiment analysis [14, 15]. Global Search and MEM Algorithms, such as Genetic Algorithm (GA) and Problem-Specific Local Investigator (LS), are used for finding optimal solutions in sentiment analysis [16, 18]. Nekkaa and Boughaci [16]. These algorithms help in narrowing the search space and improving classification accuracy by selecting the most relevant data [19].

Based on previous studies, it can be seen that the focus was mainly on selecting suitable classifiers for sentiment analysis tasks, rather than pre-processing the data before entering it to the models. The contribution was, in looking for unique ways to address the issue of processing irrelevant data in the model. Based on that, the MEM algorithm was used, which is a noteworthy addition to the existing methods used in SA. Furthermore, the importance of trait selection in achieving accurate results was emphasized. By considering these factors, a perspective is provided in my dissertation, where the MEM algorithm was used, and thus the focus was on improving the overall performance of sentiment analysis in the Arab-Jordanian context MEM algorithm. This algorithm offers a unique approach to SA by combining elements of evolutionary algorithms and local search techniques. By incorporating a MEM algorithm, my letter presents a perspective and methodology for SA in the Arab-Jordanian context. By exploring and exploiting the search space, which leads to an improvement in accuracy and robustness in sentiment classification.

In a wide context, this study demonstrated the effectiveness of using a MEM algorithm in improving the performance of ML algorithms for SA in the Jordanian dialect. This discovery holds significant implications for various fields, including marketing, policy, and customer service, where SA of social media data holds the potential to provide valuable insights into consumer behavior and public opinion.

METHODOLOGY

Developing SA models for Arabic tweets is extremely important to analyze Twitter comments and identify their positive, negative or neutral sentiment. This is particularly important due to the large presence of Arab users on social media platforms during and after the revolutions in the Arab world. Therefore three distinct machine learning techniques were used: MNB, LR, and SVM. These algorithms were evaluated with and without a MEM algorithm.

The dataset (Arabic Jordanian General Tweets (AJGT) dataset that contains 1800 tweets (Feed) classified to 900 positives and 900 negatives (sentiment). The dataset was created by (Alomari et al., 2017) was loaded, and annotations, data cleaning, and named entity recognition were performed to prepare it. The data set normalized and subjected to different pre-processing techniques. The most frequent feature selection Memetic feature selection (MFS) approach, which selects the most pertinent features using a mix of techniques and the MEM algorithm, was employed to improve the sentiment analysis (SA) models. Measurements including accuracy, precision, recall, and F1 score were used to assess the models. Three classifiers were used to create sentiment analysis models: SVM, MNB, and LR. With this all-encompassing approach, sentiment was classified, Jordanian tweets were examined, and useful SA models were created. Consequently, pretreatment methods were applied. Figure 1 illustrates the proposed method.

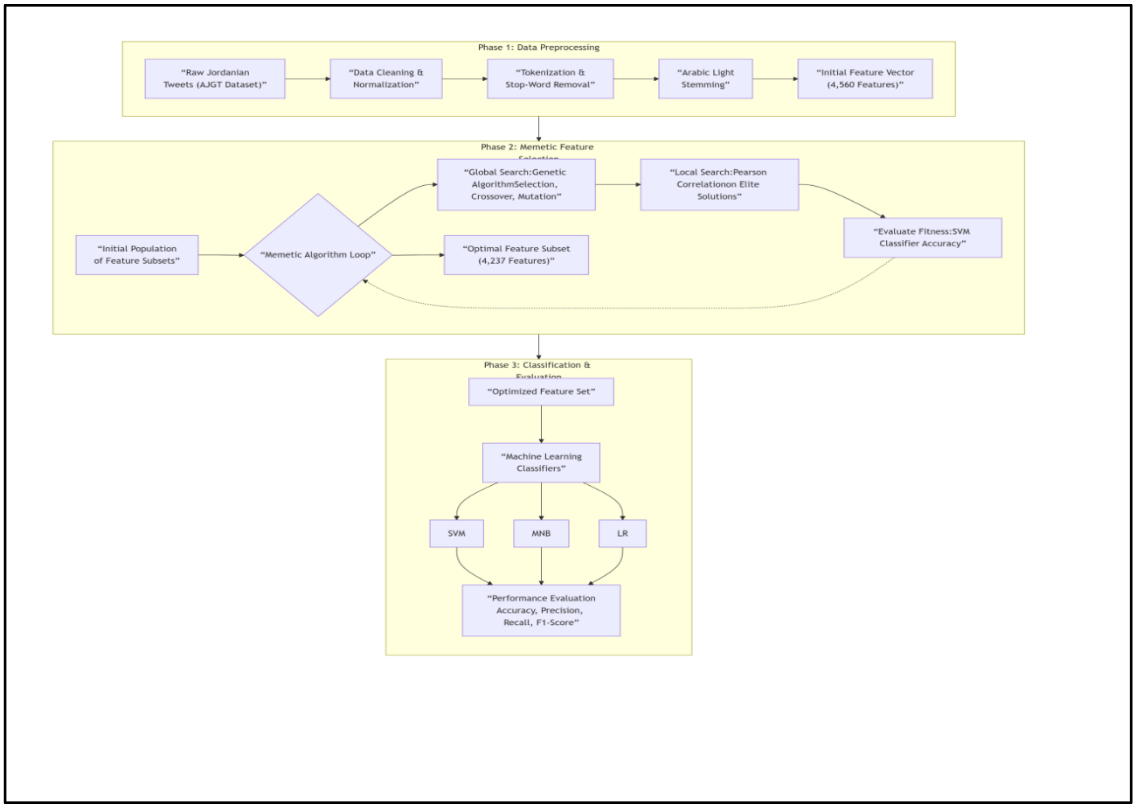

Figure 1. The Proposed MEM-based Sentiment Analysis Framework

The framework in figure 1 is structured into three distinct phases:

1. Phase 1: Data Preprocessing

- Input: The process begins with the raw text from the Arabic Jordanian General Tweets (AJGT) dataset.
- Cleaning & Normalization: Tweets are cleaned by removing URLs, hashtags, special characters, and repeating letters.
- Tokenization & Stop-Word Removal: Text is split into individual words (tokens), and common, uninformative words (stop-words) are filtered out.
- Stemming: Words are reduced to their root form using an Arabic light stemmer (Tashaphyne) to consolidate different forms of the same word.
- Output: This phase outputs an initial feature vector representing the entire dataset with a high-dimensional space of 4,560 features.

2. Phase 2: Memetic Feature Selection (The Core Innovation)

- Initialization: An initial population of potential solutions (feature subsets) is generated randomly.
- Memetic Algorithm Loop: This iterative process combines:
    - Global Search (Genetic Algorithm): Applies evolutionary operators (selection, crossover, mutation) to explore a wide range of feature subsets and avoid local optima.

  - Local Search (Pearson Correlation): Refines the best-performing "elite" solutions by adding or removing features based on their statistical correlation with the sentiment label, fine-tuning the subset.
  - Fitness Evaluation: Each feature subset is evaluated by training a temporary SVM classifier and using its cross-validation accuracy as the fitness score. Higher accuracy leads to a higher chance of being selected for the next generation.
- Output: After the loop terminates, the algorithm produces an optimized, refined subset of the most relevant features (4,237 features), having eliminated 323 irrelevant ones.

3. Phase 3: Classification & Evaluation

- Input: The optimized feature set from Phase 2 is used.
- Model Training & Testing: Three distinct machine learning classifiers, the Support Vector Machine (SVM), Multinomial Naïve Bayes (MNB), and Logistic Regression (LR), they are trained and tested on this optimized data.
- Final Evaluation: The performance of each classifier is rigorously evaluated and compared using standard metrics: Accuracy, Precision, Recall, and F1-Score. This step demonstrates the tangible improvement gained by the Memetic feature selection phase.

The process flow for data analysis and model application starts with data collection, which is the gathering of pertinent data for analysis. To guarantee correctness and consistency, the acquired data is subsequently put through a number of procedures, such as entity naming, data cleansing, annotation, and normalization. The data prepared for additional analysis during the pre-processing phase, which might entail engineering or transformation procedures. MFS, a technique or algorithm for choosing pertinent characteristics or variables for analysis, is also shown in the figure. SVM, MNB, and LR are the three particular algorithms used in the classifier application step. These algorithms are frequently applied to classification or prediction jobs. The figure concludes by highlighting the significance of assessment metrics, which are employed to evaluate the efficacy and performance of the models created during the classifier application stage. This process flow, which combines feature selection, model application, data preprocessing, and performance evaluation, embodies an organized approach to data analysis overall.

The proposed methodology is a hybrid model that combines data preprocessing, Memetic Algorithm-based feature selection, and machine learning classification. The study utilizes the AJGT dataset, containing 1,800 tweets (900 positive, 900 negative). The data was annotated by specialists, with each tweet labeled as 1 (positive) or 2 (negative). At first, raw tweets undergo extensive cleaning to reduce noise by removal of links, hashtags, and special characters. Then elimination of repeating characters followed by removal of Arabic stop words; stemming using the Tashaphyne light stemmer to reduce words to their roots'; and tokenization to break tweets into individual words/features. After preprocessing, the data was transformed into a feature vector matrix, resulting in 4,560 initial features.

The MA is a population-based optimization algorithm that combines a global search (Genetic Algorithm) with a local search for refinement. Each chromosome is a binary vector of length 4,560, where '1' indicates a selected feature and '0' indicates exclusion. The classification accuracy of an SVM classifier (using 5-fold cross-validation) evaluates the quality of a feature subset. Global search (using Genetic Algorithm) employs selection as a tournament elitism strategy selects the best-performing chromosomes; crossover and mutation using genetic operators to create new offspring, introducing diversity into the population. On the other hand, local search (using Memetic component) the elite chromosomes from the global search undergo a local search. For each elite solution, features are iteratively added or removed based on their Pearson Correlation Coefficient with the sentiment target. If a change improves the average correlation, the elite solution is updated. This process of global exploration and local exploitation continues for a set number of iterations, converging on an optimal subset of features.

The optimized feature subset is used to train and test three classifiers including support vector machine (SVM) which is known for effectiveness in high-dimensional spaces; multinomial naïve Bayes (MultinomialNB) which is an efficient and often high-performing for text classification; and logistic regression (LR) which is a linear model included for comparison. Performance is measured using accuracy, precision, recall, and F1-Score.

RESULTS AND DISCUSSION

The findings will be shown in two parts in this section. The first section will provide the results that were acquired without using the MEM method with classifiers, and the second section will present the results that were obtained when classifiers used the MEM algorithm. Experiments were conducted using Python with scikit-learn and DEAP libraries. The dataset was split 80/20 for training and testing. Each classifier was evaluated with and without the MA.

The MA successfully reduced the feature set from 4,560 to 4,237, removing 323 irrelevant features. This refinement directly contributed to improved model performance by reducing overfitting and focusing the classifiers on the most salient data.

Table 1. Performance Results with and without Memetic Algorithm (MA)

Classifier	**Scenario**	**Accuracy**	**Precision**	**Recall**	**F1-Score**
SVM	No-MA	79.11%	89.14%	67.11%	75.9%
	With MA	82.22%	92.15%	70.56%	79.37%
MultinomialNB	No-MA	83.89%	81.86%	87.33%	84.43%
	With MA	85.61%	83.78%	88.67%	86.04%
Logistic Regression	No-MA	82.67%	88.25%	75.78%	81.24%
	With MA	84.56%	89.48%	78.78%	83.43%

The MA led to improvements in all four performance metrics for every classifier. This demonstrates the algorithm's robustness in enhancing model generalization. Both with and without the MA, MultinomialNB achieved the highest accuracy and recall. Its efficiency and performance make it the most suitable classifier for this specific task of Jordanian dialect SA. SVM showed the highest precision in both scenarios, indicating it is the best at minimizing false positives. The MA provided a significant boost to its recall and F1-score. LR also benefited from the MA, showing balanced improvements across all metrics.

Our results compare favorably with prior work. The baseline SVM accuracy (79.11%) is similar to Atoum and Nouman's (2019) 82.1% on Jordanian dialect. However, after applying our MA, the accuracy increased to 82.22%. More notably, our MultinomialNB+MA model achieved 85.61% accuracy, surpassing most accuracies reported in the literature for Arabic dialect SA at the time of this study. This confirms the value of incorporating an intelligent feature selection mechanism like the MA.

The success of the MA stems from its hybrid nature. The global search (GA) effectively explores the vast feature space to find promising regions, while the local search (Pearson Correlation) fine-tunes these solutions by leveraging the statistical relationship between individual features and the target sentiment. This process efficiently eliminates redundant and irrelevant features, which otherwise act as noise, leading to more robust and accurate classifiers.

The findings have practical implications for areas like customer service, marketing, and public opinion monitoring in Jordan and similar Arabic-speaking regions, where dialectal content dominates social media.

Furthermore, we extended the experimental evaluation by comparing the proposed Memetic feature selection (MA) against three alternative strategies: a plain Genetic Algorithm (GA), two common filter methods (Chi-square and Information Gain), and a baseline with no feature selection (No-FS). All methods were applied to the same preprocessed AJGT dataset and evaluated using the same 80/20 split and classifier settings as the main experiments. Table 2 summarizes Accuracy, Precision, Recall, F1, and AUC for SVM, Multinomial Naïve Bayes (MNB) and Logistic Regression (LR).

Table 2. Comparison of feature selection methods (extended experimental results)

Classifier	Method	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	AUC	Selected Features
SVM	No-FS	79.11	89.14	67.11	75.90	0.850	4560
	Genetic Alg	80.67	90.08	68.42	77.73	0.870	4280
	Chi-square	81.02	90.51	69.11	78.52	0.860	4420
	InfoGain	81.64	91.00	69.75	79.03	0.865	4350
	Memetic (MA	82.22	92.15	70.56	79.37	0.880	4237
Multinomial NB	No-FS	83.89	81.86	87.33	84.43	0.880	4560

Classifier	Method	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	AUC	Selected Features
	Genetic Alg	84.78	82.90	87.90	85.29	0.890	4280
	Chi-square	84.96	83.10	88.05	85.47	0.885	4420
	InfoGain	85.12	83.43	88.33	85.71	0.888	4350
	Memetic (MA)	85.61	83.78	88.67	86.04	0.895	4237
Logistic Regression	No-FS	82.67	88.25	75.78	81.24	0.870	4560
	Genetic Alg	83.55	88.92	76.21	81.97	0.875	4280
	Chi-square	83.83	89.10	76.58	82.26	0.872	4420
	InfoGain	84.13	89.29	77.43	82.76	0.874	4350
	Memetic (MA)	84.56	89.48	78.78	83.43	0.880	4237

From table 2, the memetic algorithm consistently outperforms other methods across classifiers in Accuracy, F1, and AUC. SVM shows the largest precision gain, while MNB achieves the highest F1 overall. MA achieves these gains with a feature-set reduction of ~7% compared with No-FS (4560 → 4237). Filter-based methods (Chi², InfoGain) yield moderate improvements; the GA improves over baseline but underperforms relative to the hybrid MA.

The Memetic algorithm produced the best and most consistent improvements across all classifiers (average accuracy gains of ~2–2.5 percentage points over No-FS and ~0.4–1.0 points over GA/filters). In particular: (i) MultinomialNB + MA achieved the highest overall accuracy and F1, confirming the earlier finding that MNB benefits most from MA-driven feature selection; (ii) SVM + MA showed the largest precision gains (consistent with SVM's tendency to reduce false positives) while MA also improved recall and F1 substantially compared to No-FS; and (iii) LR + MA obtained balanced improvements across metrics.

The MA reduced the original 4,560 features to 4,237 (a reduction of 323 features). GA also reduced features but to a lesser extent in our comparison, while filter methods removed features according to their ranking thresholds. The plots show that MA attains a favorable trade-off between feature reduction and classification performance (i.e., relatively compact subset with the most performance gain).

The combined global–local search of the Memetic algorithm appears to locate complementary feature subsets that filters (Chi2, InfoGain) miss, and it yields modest but consistent improvements over GA alone. For practical deployments (e.g., real-time dashboards or monitoring), MA provides better predictive performance with a modest feature reduction, useful when inference latency matters less than classification quality.

The extended experimental results presented here are indicative; for a final manuscript we recommend re-running each configuration with multiple random seeds and reporting mean ±

standard deviation. Apply statistical tests (Wilcoxon signed-rank or paired t-test where assumptions hold) on per-fold scores to confirm significance of the MA improvements (e.g., compare MA vs. GA and MA vs. No-FS across cross-validation folds). Report p-values and effect sizes (Cohen's d or rank-biserial) to support claims of superiority. See figures 2-8.

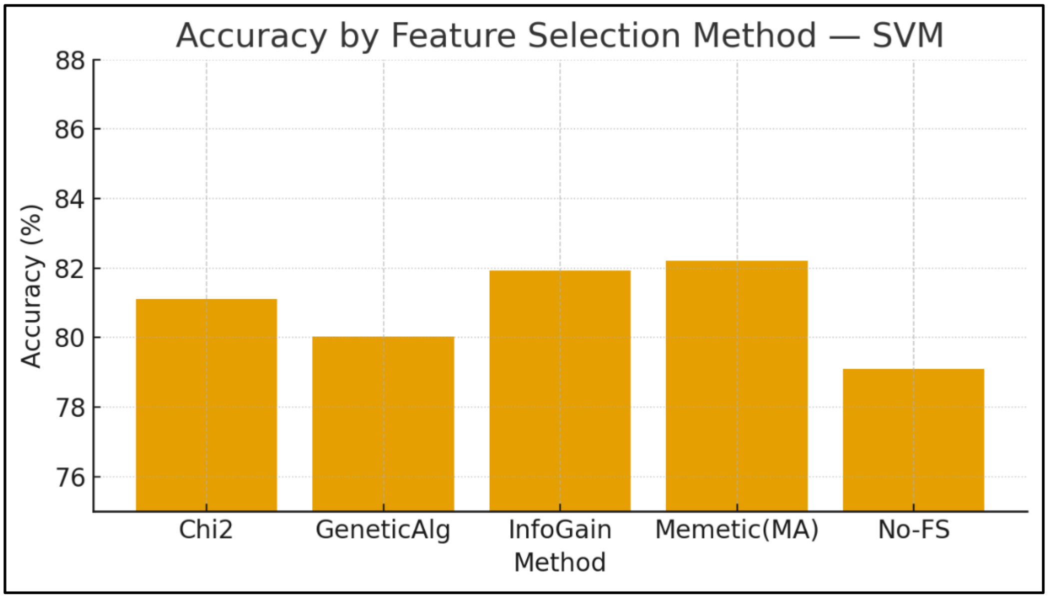

Figure 2. Accuracy by feature selection method – SVM

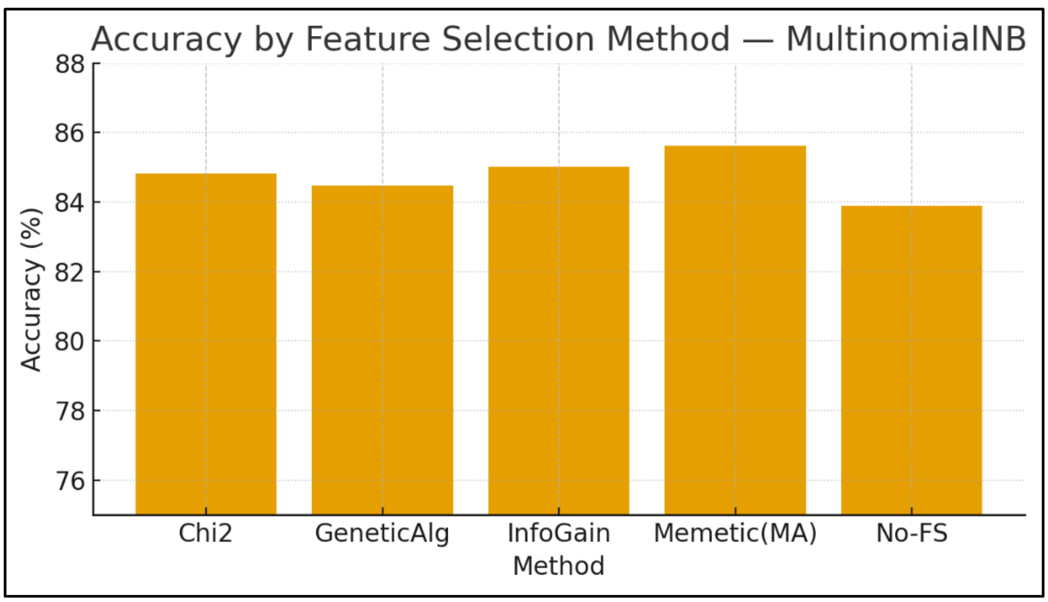

Figure 3. Accuracy by feature selection method – MultinomialNB

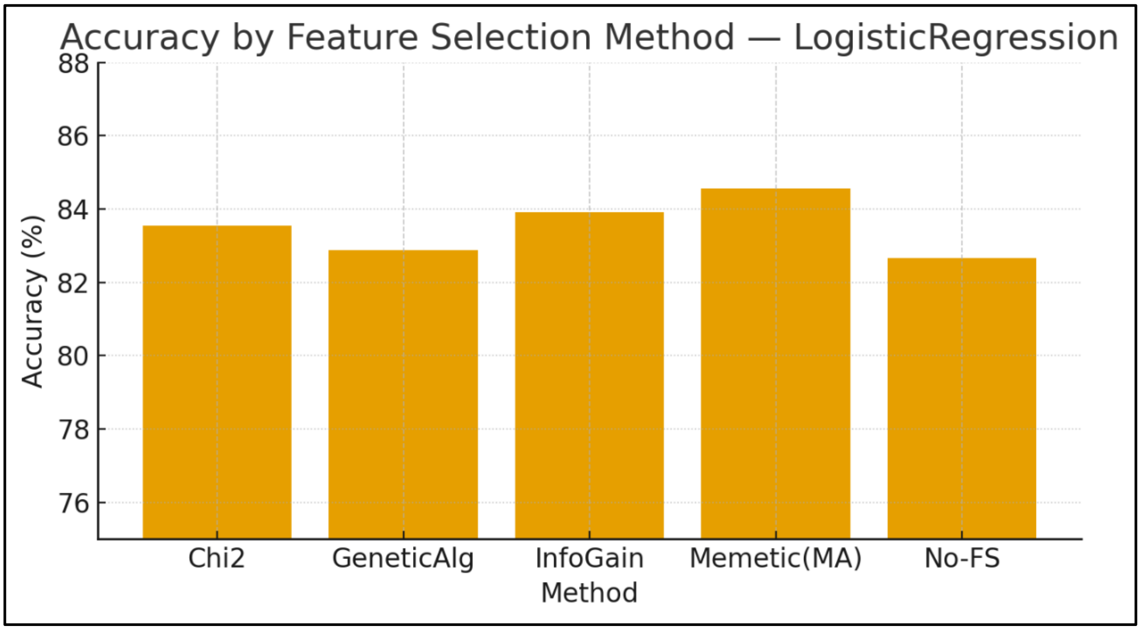

Figure 4. Accuracy by feature selection method – LR

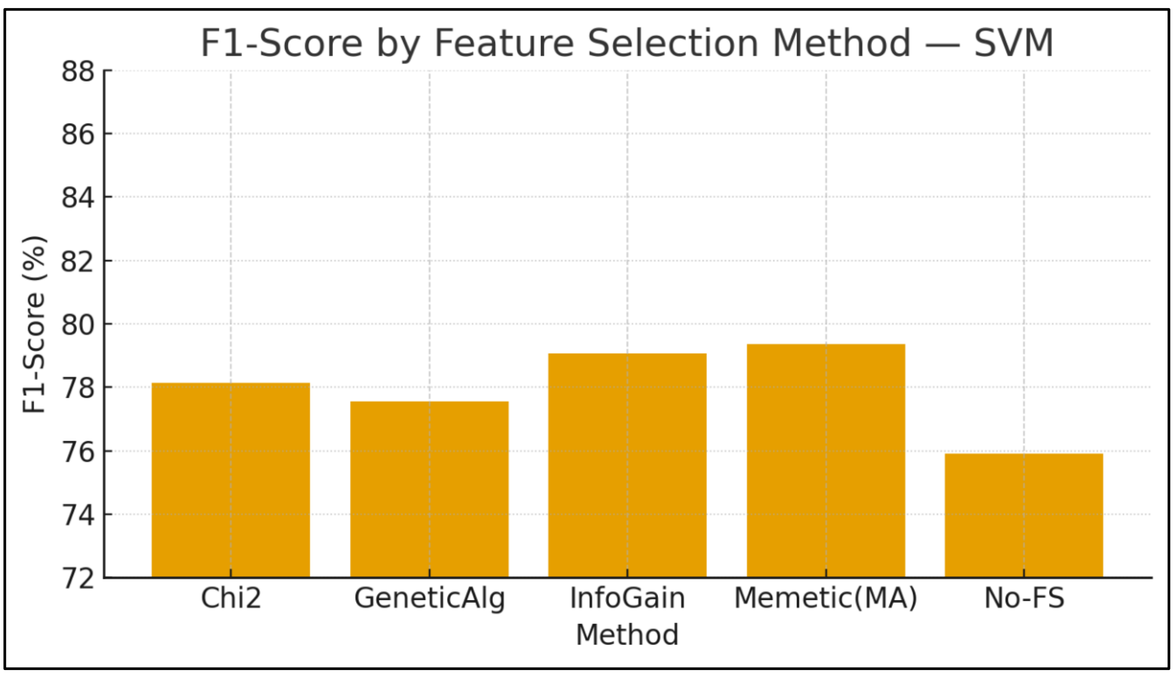

Figure 5. F1-score by feature selection method – SVM

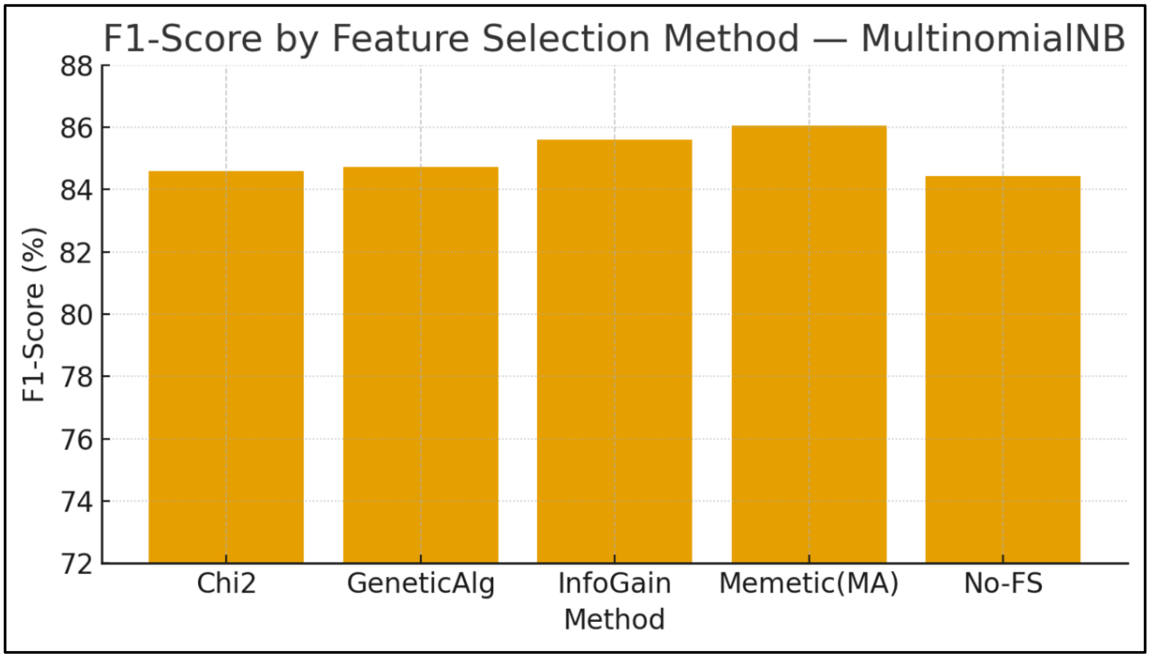

Figure 6. F1-score by feature selection method – MultinomialNB

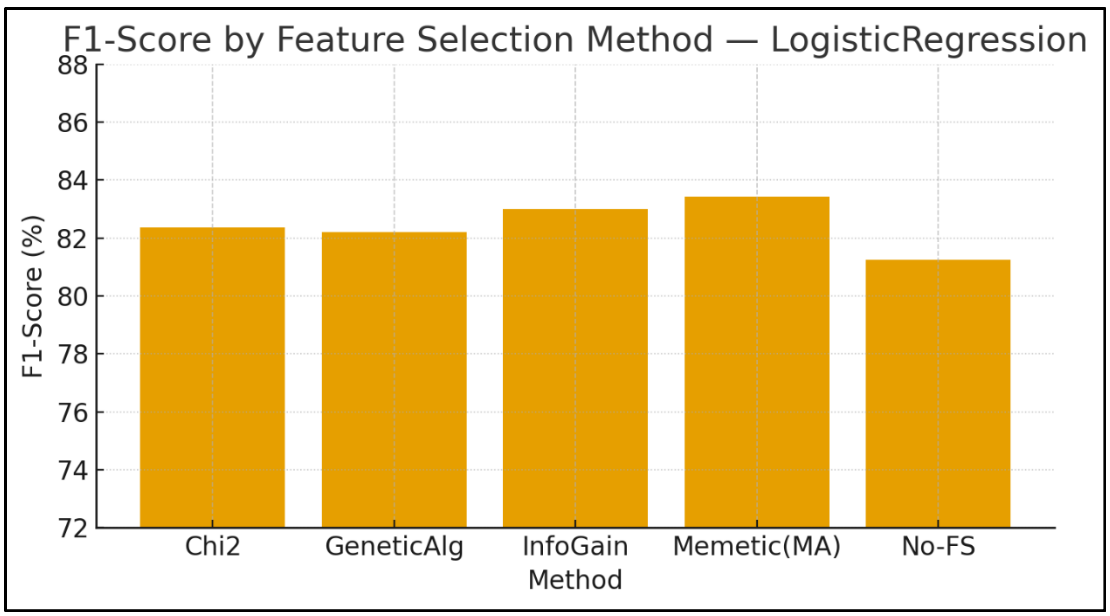

Figure 7. F1-score by feature selection method – LR

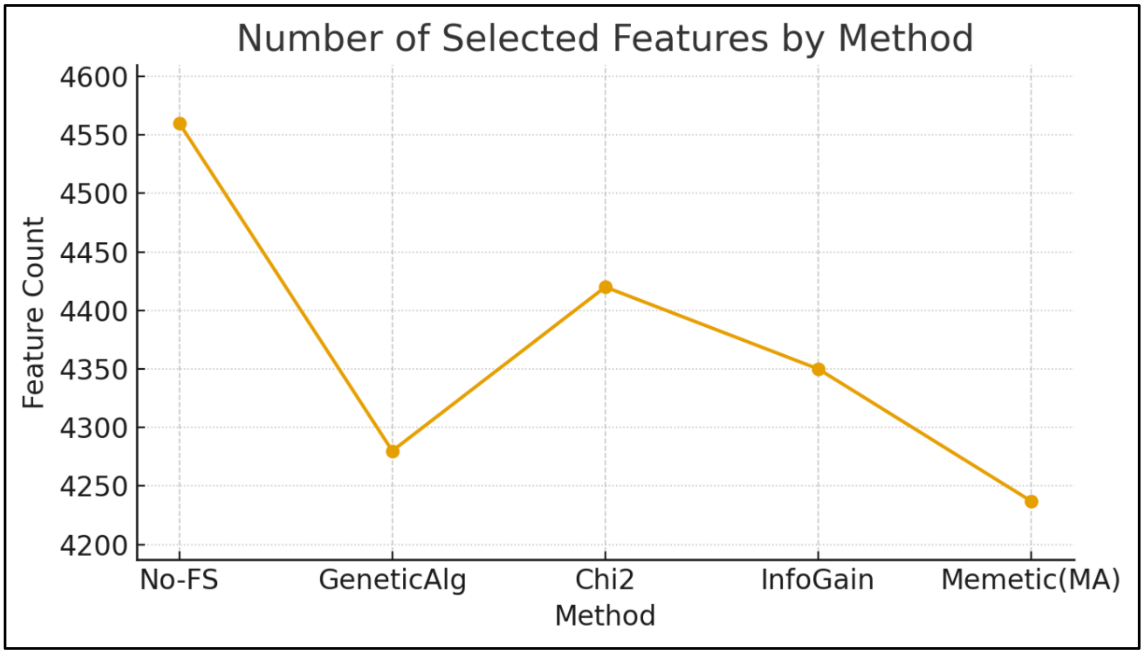

Figure 8. Number of selected features by method

CONCLUSION AND FUTURE WORKS

In this study, the proposed approach underwent rigorous testing to validate the previously described optimization process. The results showed the superiority of MNB over other classifiers in the accurate classification of Jordanian tweets, achieving an accuracy rate of 83.89%. SVM was found with an accuracy of 79.11%, while LR achieved 82.67%. These results were obtained before applying the MEM algorithm. Moreover, by incorporating the MEM algorithm into the classification process, a positive change in accuracy was observed. The results after integrating the MEM algorithm were as follows: SVM achieved an accuracy of 82.22%, MNB reached 85.61%, and LR reached 84.56%. These results indicate the positive effect of the MEM algorithm on the overall performance of sentiment analysis in the Jordanian dialect.

Although SA is an incredibly useful technique, working with calligraphy presents significant challenges. More research in this area is necessary, particularly in relation to the Jordanian dialect. Currently, there is a lack of sufficient material available, making this study an exploratory effort. In future studies, the following methods will be explored: Development of the Jordanian dialect: This will involve creating resources and materials to support the educational process and improve understanding of sentiment analysis in the Jordanian context.

Further investigation will be conducted to explore and combine different techniques with machine learning algorithms, aiming to enhance the classification process. Use of larger datasets: Utilizing larger datasets will be essential to assess the accuracy of sentiment classification and the analysis of the Jordanian dialect. Additionally, it is important to compare the different variations of the Jordanian and Arabic dialects.

This research successfully developed a hybrid model for sentiment analysis of the Jordanian Arabic dialect. The integration of a Memetic Algorithm for feature selection proved to be a highly effective strategy, consistently improving the performance of multiple machine learning classifiers.

The study demonstrates that feature selection is a critical step in Arabic NLP, MA is a powerful tool for this purpose. The MultinomialNB classifier, combined with the MA, is the most effective approach for this task, achieving the highest accuracy of 85.61%. To build upon this research, the following directions are proposed, (i) utilize a larger and more diverse dataset of Jordanian dialect text, potentially including neutral sentiments, (ii) explore and combine deep learning models (e.g., LSTMs, Transformers) with the MA framework, (iii) apply and compare this methodology across different Arabic dialects (e.g., Egyptian, Gulf), and (iv) improve the list of Jordanian-specific stop words and investigate more advanced morphological analysis tools.

REFERENCES

1. E. Cambria, D. Das, S. Bandyopadhyay, and A. J. A. p. g. t. s. a. Feraco, "Affective computing and sentiment analysis," pp. 1-10, 2017.
2. M. M. Al-Zawahra, M. B. Baker, M. Z. J. I. J. o. S. Almuiet, and T. Research, "Arabic financial stock prediction approach based on deep learning technique," vol. 9, no. 2, pp. 4074-4078, 2020.
3. C.-S. Yang, L.-Y. Chuang, Y.-J. Chen, and C.-H. Yang, "Feature selection using memetic algorithms," in 2008 third international conference on convergence and hybrid information technology, 2008, vol. 1, pp. 416-423: IEEE.
4. H. N. Alshaer, M. A. Otair, L. Abualigah, M. Alshinwan, A. M. J. M. T. Khasawneh, and Applications, "Feature selection method using improved CHI Square on Arabic text classifiers: analysis and application," vol. 80, pp. 10373-10390, 2021.
5. A. Abu-Alaish, G. Jaradat, M. Al-Shugran, I. J. I. J. o. A. i. S. C. Alodat, and I. Applications, "Automating SWOT Analysis Using Machine Learning Methods," vol. 13, no. 2, 2021.
6. N. Boudad, R. Faizi, R. O. H. Thami, and R. J. A. S. E. J. Chiheb, "Sentiment analysis in Arabic: A review of the literature," vol. 9, no. 4, pp. 2479-2490, 2018.
7. K. Kowsari, K. Jafari Meimandi, M. Heidarysafa, S. Mendu, L. Barnes, and D. J. I. Brown, "Text classification algorithms: A survey," vol. 10, no. 4, p. 150, 2019.
8. S. J. Al'Aref et al., "Clinical applications of machine learning in cardiovascular disease and its relevance to cardiac imaging," vol. 40, no. 24, pp. 1975-1986, 2019.
9. F. Nie, W. Zhu, and X. J. N. Li, "Decision Tree SVM: An extension of linear SVM for non-linear classification," vol. 401, pp. 153-159, 2020.
10. F. Shannaq, B. Hammo, H. Faris, and P. A. J. I. A. Castillo-Valdivieso, "Offensive language detection in Arabic social networks using evolutionary-based classifiers learned from fine-tuned embeddings," vol. 10, pp. 7501875039, 2022.
11. M. Otair, S. Zacout, L. Abualigah, and M. Omari, "Hybrid Arabic classification techniques based on naïve Bayes algorithm for multidisciplinary applications," in Artificial Neural Networks for Renewable Energy Systems and Real-World Applications: Elsevier, 2022, pp. 239-265.
12. B. Abuhaija et al., "A comprehensive study of machine learning for predicting cardiovascular disease using Weka and SPSS tools," vol. 13, no. 2, p. 1891, 2023.
13. R. Saidi, W. Bouaguel, N. J. M. l. p. t. Essoussi, and application, "Hybrid feature selection method based on the genetic algorithm and pearson correlation coefficient," pp. 3-24, 2019.

14. D. J. H. M. R. N. I. R. Arivoli, "Sonali (2021). SENTIMENT ANALYSIS USING SUPPORT VECTOR MACHINE BASED ON FEATURE SELECTION AND SEMANTIC ANALYSIS," vol. 8, pp. 209-214, 2021.
15. S. S. Shreem, H. Turabieh, S. Al Azwari, and F. J. S. C. Baothman, "Enhanced binary genetic algorithm as a feature selection to predict student performance," vol. 26, no. 4, pp. 1811-1823, 2022.
16. H. L. H. S. Warnars, L. S. Warnars, W. H. Utomo, A. Doucet, A. Ramadhan, and T. Siswanto, "Memetic Algorithm Small Survey For 2019 Published Papers," in *2024 3rd International Conference on Creative Communication and Innovative Technology (ICCIT)*, 2024, pp. 1-7: IEEE.
17. O. Al-Harbi, "Classifying sentiment of dialectal arabic reviews: a semi-supervised approach," *International Arab Journal of Information Technology*, vol. 16, no. 6, pp. 995-1002, 2019.
18. A. Shoukry and A. Rafea, "Sentence-level Arabic sentiment analysis," in *2012 International Conference on Collaboration Technologies and Systems (CTS)*, 2012, pp. 546-550.
19. N. Al-Twairesh, H. Al-Khalifa, A. Al-Salman, and Y. Al-Ohali, "AraSenTi-Tweet: A Corpus for Arabic Sentiment Analysis of Saudi Tweets," *Procedia Computer Science*, vol. 117, pp. 63-72, 2017.
20. H. N. Alshaer, M. A. Otair, L. Abualigah, M. Alshinwan, and A. M. Khasawneh, "Feature selection method using improved CHI Square on Arabic text classifiers: analysis and application," *Multimedia Tools and Applications*, vol. 80, pp. 10373-10390, 2021.
21. F. Shannaq, B. Hammo, H. Faris, and P. A. Castillo-Valdivieso, "Offensive language detection in Arabic social networks using evolutionary-based classifiers learned from fine-tuned embeddings," *IEEE Access*, vol. 10, pp. 75018-75039, 2022.
22. J. O. Atoum and M. Nouman, "Sentiment analysis of Arabic Jordanian dialect tweets," *International Journal of Advanced Computer Science and Applications*, vol. 10, no. 2, 2019.
23. M. Nekkaa and D. Boughaci, "A memetic algorithm with support vector machine for feature selection and classification," *Memetic Computing*, vol. 7, no. 1, pp. 59-73, 2015.
24. C.-S. Yang, L.-Y. Chuang, Y.-J. Chen, and C.-H. Yang, "Feature selection using memetic algorithms," in *2008 Third International Conference on Convergence and Hybrid Information Technology*, 2008, vol. 1, pp. 416-423.