

# Sentiment Analysis for Government E-Services: A CRISP-DM Machine Learning Approach

Saeed AlRashdi

Department of Government Enablment  
(GovDigital)

AbuDhabi, U.A.E.

Saeed.AlRashdi@dge.gov.ae

<https://doi.org/10.63962/UVVI9200>

## II. METHODOLOGY

**Abstract**—The increasing adoption of digital platforms by government organizations necessitates a deeper understanding of user sentiment to enhance public services. This paper presents an AI-driven sentiment analysis framework focused on Arabic-language Twitter data related to Abu Dhabi's TAMM e-services platform. Utilizing the CRISP-DM methodology, this study collected and preprocessed over 3,000 Arabic tweets, applied machine learning algorithms including Logistic Regression, Naïve Bayes, SVM, Random Forest, and MLP, and evaluated their performance in classifying sentiment. The results reveal that Multinomial Naïve Bayes achieved the best balance of accuracy (79.4%) and training efficiency (0.68s). Comparative evaluations using Cohen's Kappa and Matthews Correlation Coefficient validate inter-model agreement. This work demonstrates the feasibility and value of leveraging Arabic sentiment analysis for real-time feedback, aligning with national strategies such as the UAE Happiness Index [1], [2].

**Keywords**— *Sentiment Analysis, Arabic NLP, CRISP-DM, TAMM, Machine Learning, Public Services, UAE Happiness Index*

## I. INTRODUCTION

As governments embrace digital transformation, e-services like TAMM in Abu Dhabi aim to streamline service delivery and improve citizen satisfaction. Understanding public sentiment through social media provides actionable insights that support policymaking and service enhancement. Sentiment analysis, particularly for Arabic text, poses unique challenges due to linguistic complexity and dialectal variation [3], [4].

Previous research has explored various machine learning methods for sentiment analysis [3], with specific focus on Arabic sentiment modeling [4]. However, limited studies have applied these techniques to the domain of government e-services, especially in the Arabic-speaking context. This study bridges that gap by applying a structured data mining methodology (CRISP-DM) and comparative machine learning evaluation to Arabic tweets related to TAMM.

This study followed the CRISP-DM methodology [5], a structured process that includes six major phases, applied specifically to sentiment analysis:

- **Business Understanding:** The primary objective was to extract and analyze public sentiment about the TAMM government e-service platform in Abu Dhabi to help enhance decision-making and improve user satisfaction.
- **Data Understanding:** We collected Arabic-language tweets related to TAMM using Tweet Flash [6]. This data was derived from both labeled sentiment datasets and live user-generated content.
- **Data Preparation:** Preprocessing was a critical step to ensure data quality and accuracy. The steps included:
  - Removing non-Arabic words, emojis, URLs, and repeated characters.
  - Normalizing Arabic text by unifying variations in spelling (e.g., replacing  $\overset{\sim}{\text{w}}$  with  $\text{w}$ ).
  - Tokenizing and applying TF-IDF vectorization to transform the raw text into numerical feature representations [7], [8].
  - Merging normalized labeled data (Arabic Sentiment Corpus) with the preprocessed TAMM tweets to form a unified dataset.

The full flow of handling the data is illustrated in **Figure 1**, showcasing the end-to-end pipeline from tweet collection to model training and evaluation

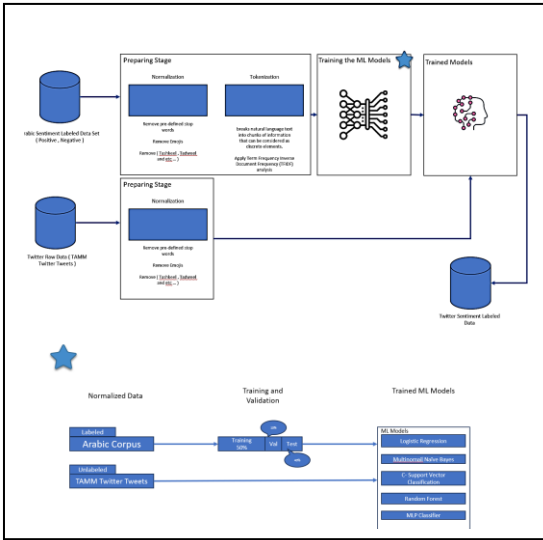


Figure 1: Workflow for Arabic Sentiment Analysis (from preprocessing to model evaluation)

accuracy, training time, Cohen’s Kappa, and MCC [9], [10].

### III. RESULTS

The performance of each model is summarized in Table 1. Figures 2 and 3 illustrate comparative accuracy and training time.

TABLE 1: Model Performance Comparison

Algorithm	Accuracy (%)	Training Time (s)	Notes
Logistic Regression	78.9	7.28	Balanced accuracy with fast training
Multinomial Naïve Bayes	<b>79.4</b>	<b>0.68</b>	Best balance between accuracy and efficiency
SVM (C-SVC)	79.1	3694.44	High test score but slowest training
Random Forest	78.9	966.59	High test score; potential overfitting
MLPClassifier	77.4	866.90	Complex model, risk of overfitting

Agreement metrics:

- **Cohen’s Kappa:** Strong agreement between SVM and Logistic Regression (0.80).
- **MCC:** Validated model consistency across predictions.

These results indicate that while complex models such as Random Forest and MLP offer high performance, their training cost and overfitting potential make them less ideal for real-time applications. Multinomial Naïve Bayes, on the other hand, combines strong accuracy with minimal computational overhead, making it a suitable candidate for integration into live systems.

### IV. DISCUSSION

Arabic sentiment analysis requires rigorous preprocessing and model tuning [4], [11]. The Multinomial Naïve Bayes model proved most efficient, showing both high accuracy and the fastest training time—making it ideal for government applications where responsiveness is key.

A qualitative complement was added through sentiment word clouds. These visualizations (Figure 4) reveal key user themes. Positive sentiments included words like "سهل" (easy), "ممتاز" (excellent), and "سريع" (fast), indicating satisfaction with usability. Negative terms like "تأخير" (delay), "صعب" (difficult), and "غير واضح" (unclear) reveal areas for service refinement.

```

svm_counts = df['svm'].value_counts()
mlp_counts = df['mlp'].value_counts()
logistic_regression_counts = df['logistic regression'].value_counts()
naive_bayes_counts = df['naive bayes'].value_counts()
random_forest_counts = df['random forest'].value_counts()

print("SVM Counts:\n",svm_counts)
print("MLP Counts:\n",mlp_counts)
print("Logistic Regression Counts:\n",logistic_regression_counts)
print("Naive Bayes Counts:\n",naive_bayes_counts)
print("Random Forest Counts:\n",random_forest_counts)

SVM Counts:
pos    462
neg    305
Name: svm, dtype: int64
MLP Counts:
pos    451
neg    396
Name: mlp, dtype: int64
Logistic Regression Counts:
pos    433
neg    424
Name: logistic regression, dtype: int64
Naive Bayes Counts:
pos    467
neg    388
Name: naive bayes, dtype: int64
Random Forest Counts:
pos    547
neg    388
Name: random forest, dtype: int64

```

Fig 2 : Training Time for Each Model

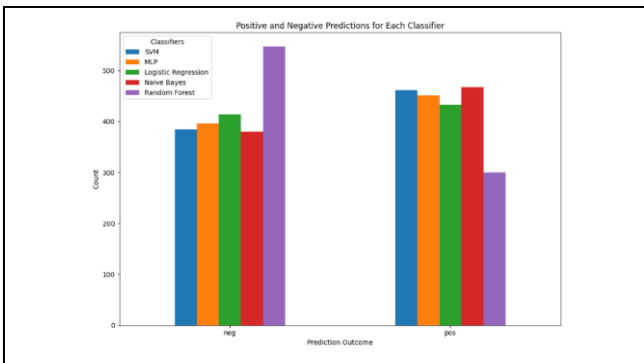


Fig 3: Accuracy of Sentiment Models

- **Modeling:** Five classifiers were selected to evaluate their suitability for real-time sentiment analysis in Arabic: Logistic Regression, Multinomial Naïve Bayes, SVM (C-SVC), Random Forest, and MLP. Naïve Bayes was selected for its simplicity and performance with sparse data; SVM and MLP for their ability to capture nonlinear patterns; and Random Forest for robustness against overfitting.
- **Evaluation:** The dataset was split into 50% training, 10% validation, and 40% testing. Metrics included



Fig 4: Word Clouds for Positive and Negative Sentiment Classification

These visuals provide intuitive support for identifying service strengths and weak points. Importantly, they align directly with real-world citizen experiences, informing data-driven improvements in TAMM.

## V. CONCLUSION

This study presents a CRISP-DM-based framework to apply sentiment analysis to Arabic-language Twitter data regarding TAMM e-services. It highlights the Multinomial Naïve Bayes model as a practical, efficient classifier for deployment in real-time feedback systems.

The findings have direct implications for service development teams: user feedback can be rapidly interpreted and categorized to inform dynamic improvements, aligning with goals like the UAE Happiness Index.

Limitations: The analysis focused on Modern Standard Arabic and did not account for dialectal diversity. Also, social media users may not represent the full demographic range of TAMM users.

Fig 3: Accuracy of Sentiment Models

Future Work: Research can expand into dialect detection, apply deep learning models like AraBERT, and integrate with real-time dashboards used by government support teams.

## REFERENCES

- [1] Abu Dhabi Government, "TAMM – Abu Dhabi Government Services", 2018. [Online]. Available: <https://www.tamm.abudhabi/>
- [2] The Official Portal of the UAE Government, "Happiness and National Agenda", 2023. [Online]. Available: <https://u.ae/en/about-the-uae/the-uae-government/government-of-future/happiness>
- [3] [M. Wankhade, A. C. S. Rao, and C. Kulkarni, "A survey on sentiment analysis methods, applications, and challenges," *Artif. Intell. Rev.*, vol. 55, no. 7, pp. 5731–5780, 2022.
- [4] S. Mohammad, "Sentiment Analysis in Arabic," 2021. [Online]. Available: <https://saifmohammad.com/WebPages/ArabicSA.html>
- [5] C. Schröer, F. Kruse, and J. M. Gómez, "A systematic literature review on applying CRISP-DM process model," *Procedia Comput. Sci.*, vol. 181, pp. 526–534, 2021.
- [6] Shane, "Tweet Flash", Apify. [Online]. Available: <https://apify.com/shanes/tweet-flash>
- [7] M. I. Alfarizi, L. Syafaah, and M. Lestandy, "Emotional Text Classification Using TF-IDF And LSTM," *JUITA*, vol. 10, no. 2, pp. 225–232, 2022.
- [8] H. T. Sueno, B. D. Gerardo, and R. P. Medina, "Multi-class document classification using SVM based on improved Naïve Bayes vectorization," *Int. J. Adv. Trends Comput. Sci. Eng.*, vol. 9, no. 3, 2020.
- [9] G. Rau and Y. S. Shih, "Evaluation of Cohen's kappa and other measures of inter-rater agreement," *J. Engl. Acad. Purp.*, vol. 53, 2021.
- [10] J. Yao and M. Shepperd, "Assessing software defection prediction performance: Why using the Matthews correlation coefficient matters," in *Proc. Eval. Assess. Softw. Eng. Conf. (EASE)*, 2020.
- [11] S. H. Janjua et al., "Multi-level aspect based sentiment classification of Twitter data using hybrid approach in deep learning," *PeerJ Comput. Sci.*, vol. 7, p. e433, 2021