

ARTICLE

Classifying Multi-Lingual Reviews Sentiment Analysis in Arabic and English Languages Using the Stochastic Gradient Descent Model

Yasser Alharbi¹ and Sarwar Shah Khan^{2,*}

¹College of Computer Science and Engineering, University of Hail, Hail, 55436, Saudi Arabia

²Department of Computer Science, University of Engineering & Technology, Mardan, 23200, Pakistan

*Corresponding Author: Sarwar Shah Khan. Email: sskhan0092@gmail.com

Received: 26 November 2024; Accepted: 13 January 2025; Published: 26 March 2025

ABSTRACT: Sentiment analysis plays an important role in distilling and clarifying content from movie reviews, aiding the audience in understanding universal views towards the movie. However, the abundance of reviews and the risk of encountering spoilers pose challenges for efficient sentiment analysis, particularly in Arabic content. This study proposed a Stochastic Gradient Descent (SGD) machine learning (ML) model tailored for sentiment analysis in Arabic and English movie reviews. SGD allows for flexible model complexity adjustments, which can adapt well to the Involvement of Arabic language data. This adaptability ensures that the model can capture the nuances and specific local patterns of Arabic text, leading to better performance. Two distinct language datasets were utilized, and extensive pre-processing steps were employed to optimize the datasets for analysis. The proposed SGD model, designed to accommodate the nuances of each language, aims to surpass existing models in terms of accuracy and efficiency. The SGD model achieves an accuracy of 84.89 on the Arabic dataset and 87.44 on the English dataset, making it the top-performing model in terms of accuracy on both datasets. This indicates that the SGD model consistently demonstrates high accuracy levels across Arabic and English datasets. This study helps deepen the understanding of sentiments across various linguistic datasets. Unlike many studies that focus solely on movie reviews, the Arabic dataset utilized here includes hotel reviews, offering a broader perspective.

KEYWORDS: Sentiment analysis; stochastic gradient descent; reviews; English IMDb dataset; Arabic dataset

1 Introduction

Sentiment analysis, an interdisciplinary domain within Natural Language Processing (NLP), goals to systematically recognize, extract, measure, and study affective states and subjective information. In the context of cinematic critique, sentiment analysis is employed to distill the evaluative content of movie reviews, categorizing the expressed sentiment into polarities such as positive, negative, or neutral [1]. This quantitative assessment of qualitative data is instrumental for stakeholders, including filmmakers, marketing personnel, and potential audiences, to discern prevailing attitudes toward a movie's content, themes, and execution [2].

In the current digital era, when individuals seek to invest their time and data usage judiciously, particularly in watching movies, meeting their expectations becomes significant. A film that falls short of these expectations can result in feelings of wasted time and resources. Consequently, many individuals turn to movie review websites to guide their viewing choices. However, with the sheer volume of reviews available online, it becomes exceedingly difficult for viewers to sift through all the available information and gain a



clear, consolidated understanding of whether a movie is worth their time. This overwhelming abundance of reviews necessitates a more efficient method of processing and summarizing sentiments toward movies [3].

An additional concern in the field of movie reviews is the potential inclusion of spoilers. Reviews, while informative, might inadvertently contain spoilers that can significantly diminish the viewing experience for potential audiences [4]. The anticipation and excitement of uncovering a movie's plot naturally contribute to the overall enjoyment of the film. Therefore, spoilers within reviews not only spoil specific plot points but can also ruin the entire viewing experience for an individual. This issue highlights the need for a sentiment analysis tool that can navigate around spoilers, providing audiences with the sentiment of a movie review without exposing them to plot-revealing details.

Despite the global popularity of movies and the universal habit of reviewing them, there is a noticeable scarcity of sophisticated sentiment analysis tools tailored for Arabic content. Arabic, being a rich and complex language with its own set of nuances, poses unique challenges for sentiment analysis. This gap in technology means that Arabic-speaking audiences are underserved in terms of accessing aggregated and spoiler-free sentiments of movie reviews. The development of an advanced sentiment analysis model capable of understanding and processing Arabic movie reviews could significantly enhance the movie selection process for Arabic-speaking audiences, providing them with a reliable, efficient, and spoiler-free way to gauge the worthiness of a movie [5].

This problem statement outlines the critical need for a sentiment analysis solution that addresses not only the challenge of voluminous movie reviews but also the risk of encountering spoilers, specifically catering to the unique linguistic characteristics of Arabic movie reviews.

This article introduces a Stochastic Gradient Descent machine learning model tailored for the task of opinion mining in movie reviews. The focus of the article is to address the challenges posed by dual-language datasets, specifically encompassing movie reviews in both Arabic and English. The aim is to enhance the understanding of sentiments expressed in reviews through the proposed model, offering a solution that accommodates the nuances present in diverse linguistic datasets. Stochastic Gradient Descent (SGD) in the machine learning model provides benefits in the context of movie review sentiment analysis. SGD is well-suited for large datasets, making it efficient for processing vast amounts of movie reviews commonly encountered in sentiment analysis tasks. It processes data in small, random batches, enabling quicker updates to the model parameters and faster convergence. The key contributions of the study are as follows:

- Two datasets, each in distinct languages, namely Arabic and English, were employed for this study. These languages exhibit diverse writing styles and scripts. The simulation was executed using the proposed machine learning model to attain better results. The model was tailored to accommodate the unique characteristics of both languages, ensuring effective performance in opinion mining across the Arabic and English datasets.
- To enhance the quality of datasets for our machine learning algorithms, we incorporated several pre-processing steps. This involved tasks such as Data Cleaning, Label Encoding, Removing Emojis/Emoticon, stop word removal, Tokenization, Stemming, and term frequency-inverse document frequency (TF-IDF). These steps were implemented to address challenges like missing values and special characters. Various techniques in data cleaning were applied to ensure the dataset remained free of errors and inconsistencies. Collectively, these measures were taken to optimize the datasets, improving the overall effectiveness of the machine learning algorithms for subsequent analysis.
- Stochastic Gradient Descent (SGD) machine learning models are proposed to design for two distinct languages to achieve better performance compared to traditional ML models. The intention behind this proposed method is to use SGD capabilities in optimizing model parameters for enhanced results,

tailored specifically to the nuances of each language. The aim is to surpass other existing ML models in terms of accuracy and efficiency within the context of language-specific applications.

2 Literature Review

Sentiment analysis is a significant tool in assisting people in selecting the movies to watch by showing them whether a movie is worth their time. This technique works by analyzing the text in movie reviews to see if the opinions expressed are positive, negative, or neutral. The goal of sentiment analysis is to categorize movie reviews based on the according to the attitudes or feelings they express. In recent years, significant advancement has been made in this research area, and new methodologies have been developed. Opinion mining for Arabic reviews is thoroughly summarized in the literature review shown in [Table 1](#).

The most recent study on sentiment analysis using Arabic reviews has investigated different fields and specially in Arabic movie reviews. Obiedat emphasized the significance of annotated datasets for Arabic sentiment analysis [6]. Al-Mansoori developed a sentiment analysis algorithm with great accuracy that was intended especially for reviews of Arabic films [7]. Alnemer et al. developed a pre-processing technique to improve sentiment analysis results [8] and proposed a novel approach using a ML model, which led to better accuracy [9]. These research demonstrate how sentiment mining for Arabic reviews is gaining popularity and developing.

Al-Moslmi et al. [10] presented the Arabic senti-lexicon, an important tool for opinion mining, and examined its performance across several aspects and classification approaches. The study highlights how important it is becoming to recognize emotions in Arabic texts at the sentence and aspect levels. It also talks about the lack of Arabic annotated datasets and the challenges of processing Arabic. The study emphasises how difficult it is for academics to work in this area due to the restricted availability of such datasets, which only cover a few specialized areas.

A thorough review of aspect-based sentiment analysis in Arabic was offered by Obiedat et al. [6], who emphasized the necessity for further annotated datasets in this field. The paper highlights the challenges associated with working with Arabic natural language processing and explores the growing importance of distinguishing feelings at both the sentence and aspect levels in Arabic texts. One of the main problems raised is the dearth of Arabic-language annotated datasets, which poses challenges for researchers. The study points out that the restricted availability of datasets in certain regions is a major barrier in this discipline, as is the absence of annotated corpora.

Al-Mansoori et al. [7] focus on sentiment analysis in the particular field of Arabic film reviews, developing a system and lexicon that performs exceptionally well. This paper presents the Multi-domain Arabic Sentiment Corpus (MASC), a corpus of 8860 reviews from several categories that show both positive and negative attitudes. Taken as a whole, they highlight the growing interest in and progress made in Arabic sentiment analysis, but they also highlight the need for additional study and resource development. The main findings of the paper are Sentiment analysis is popular and used in various aspects of life, including movie reviews. An Arabic lexicon for movies was created and used in sentiment analysis for Arabic movie reviews. The proposed Movie Rating for Arabic Reviews System (MRARS) showed excellent performance, with accuracy ranging between 94%–100%. The limitations of the study include the lack of sentiment resources in languages other than English, particularly in Arabic and its dialects, the need for further assessment of the quality of the developed language resources, and the study's focus on assessing the quality of the developed resources and integrating different feature sets and classification algorithms, rather than exploring other aspects of sentiment analysis.

Guellil et al. [11] determine sentiment in documents or sentences, with a focus on Arabic and its dialects; to present recent resources and advances in Arabic sentiment analysis, including the construction of sentiment lexicon and corpus; to describe emergent trends related to Arabic sentiment analysis, particularly the use of deep learning techniques. The main findings of the paper are the lack of annotated sentiment resources for Arabic, recent advances in Arabic sentiment analysis, and the emergent trends associated with the use of deep learning techniques.

Brahimi et al. [12] introduce techniques aimed at extracting meaningful opinions and values from online movie reviews to enhance sentiment analysis (SA) in Arabic. First, a method for examining the impact of skip-n-gram and n-gram models on sentiment categorization was presented. Next, investigate a method that employs Part-of-Speech tagging to employ subjective terms like adjectives and nouns. Feature reduction strategies are employed to enhance these approaches and get superior sentiment analysis outcomes. Third, to talk about how to review summaries and conclusions in order to extract important opinions. Lastly, it proposes a combined methodology to improve sentiment classification accuracy and a method of customer opinion analysis by utilizing the customer value model to determine the elements impacting the opinions of the customers.

The SVM Sentiment Analysis for Arabic Students' Course Reviews (SVM-SAA-SCR) algorithm, a comprehensive framework for analyzing student comments, was proposed by Lauati et al. [13]. This approach entails gathering student reviews, preparing the information, and categorizing the sentiments as neutral, negative, or positive using a machine-learning model. Steps like obtaining data, eliminating unnecessary information, tokenizing text, removing common words (stop words), and conducting stemming or lemmatization are all included in the data preprocessing process.

An opinion analysis method was created by Abimanyu et al. [14] to examine Rotten Tomatoes movie reviews. In this article, features were extracted using TF-IDF, significant features were chosen using Information Gain, and classification was done using Logistic Regression (LR).

Table 1: The summery of literature review

S. no.	Ref., year	Techniques	Pre-processing	Dataset	Evaluation measures	Results
1	[8], 2019	Bernoulli Naive Bayes, Decision Tree, Support Vector Machines (SVM), neural network, convolutional neural network	Emoticon Extraction, word normalization, morphological processing, negation processing	Arabic hotel review dataset	Accuracy F-measure	82.60% 79.00%
2	[9], 2020	Ridge, gradient boosting, multi-layer perceptron	Tokenization, normalization, stop words removing, stemming	Arabic hotel review dataset	Accuracy, Precision, Recall, F1-Score, Time	95.06% 94.00% 96.00% 95.00%
3	[15], 2018	Network Hybrid with Naive Bayes Algorithm		Attraction, hotel reviews, restaurant reviews, movie reviews, product reviews	Accuracy	0.16 (S) 93.10%
4	[10], 2017	Naive Bayes, k-Nearest Neighbours, SVM, Logistic Linear Regression, Neural Network	Tokenization, Normalization, Stop Word removal and stemming	Arabic senti-lexicon	Precision, Recall, F1-Score, F1-Macro	

(Continued)

Table 1 (continued)

S. no.	Ref., year	Techniques	Pre-processing	Dataset	Evaluation measures	Results
5	[12], 2019	SVM	POS tagging, review summarization	Arabic movie review dataset	F-measure	76.74%
6	[13], 2023	SVM	Tokenizing, removing stop words, stemming	Arabic course review dataset	Accuracy precision recall	
7	[16], 2023	Logistic regression method, information gain feature selection	Data cleaning, stemming, lemmatization, tokenization	Rotten tomatoes movies reviews	Precision, Recall, F1-Score	75.56% 78.19% 76.50%
8	[17], 2023	Naive bayes, logistic regression, SVM	Data cleaning, stemming	IMDb dataset for movie reviews	Accuracy, Precision, Recall, F1 score.	73.00% 70.00% 68.00% 74.00%
9	[18], 2023	TF-IDF, Chi-Square, KNN	Data cleaning, stemming, case folding	Indonesian-language movie review	Precision, Recall, F1-Score	82.12% 92.35% 86.94%

3 Proposed Methodology

In this paper, a systematic framework is proposed to address the challenges related to opinion in reviews. The framework consists of three main steps: first, preprocessing, where the raw data is prepared for analysis; second, feature extraction, where important information is identified from the preprocessed data; and third, model training, where a machine learning model is used to classify the movie reviews into different categories. Specifically, a Stochastic Gradient Descent (SGD) machine learning model is proposed, designed for analyzing reviews in both Arabic and English datasets.

3.1 Pre-Processing

Pre-processing contributes to cleaning and refining the data, Eliminating noise, irrelevant characters, repeating reviews, and stop words. As a result, sentiment analysis algorithms are not influenced by irrelevant information and the proposed model becomes more accurate as they focus on relevant and meaningful content. Techniques such as text normalization, Removing Emojis, Label Encoding, Stop word removal, lemmatization, Tokenization, and stemming contribute to a consistent representation of words. This consistency is significant for capturing the true sentiment expressed in distinct variations of words.

Pre-processing addresses contractions and abbreviations, assuring that the sentiment indicated in the full form is appropriately captured. This step contributes to a more accurate interpretation of sentiments. Handling with missing values assures that sentiment analysis models can handle instances where user sentiments are not explicitly indicated. Proper handling of missing data provides a more comprehensive sentiment analysis. Removing stop words and irrelevant information contributes to a reduction in the dimensionality of the dataset. This not only speeds up the analysis process but also helps in focusing on the most relevant features for sentiment classification.

In the proposed methodology, we employ pre-processing techniques to eliminate irrelevant content from movie reviews, thereby reducing the features. This ensures that all learning algorithms can effectively utilize the data. One of the key steps in data pre-processing often referred to as a data mining method, involves transforming raw data into a comprehensible format. Real-world data is often inconsistent, fragmented, and lacks precise trends or behaviors, leading to a higher number of inaccuracies. Hence, pre-processing the

data becomes a significant step to address these issues encountered during sentiment analysis. The following seven methods are carried out during the pre-processing to enhance the quality of raw data.

Data Cleaning: Data cleaning involves finding and fixing errors, inconsistencies, and inaccuracies in datasets to enhance the overall quality of the data.

Label Encoding: Label encoding is the method of converting categorical sentiment labels (such as 'positive' or 'negative') into numerical representations for computational analysis in Arabic Language. Example of Label Encoding in Arabic Positive (إيجابي) → 1, Negative (سليبي) → 0 and Neutral (محايد) → 2.

More Detailed Sentiment Labels Very Positive (جداً إيجابي) → 3, Positive (إيجابي) → 2, Neutral (محايد) → 1, Negative (سليبي) → 0, Very Negative (جداً سلبى) → -1. Sentiment for Specific Reviews:

Review: "رائعاً كان الفيلم!" (The movie was wonderful!) → Positive (إيجابي) → 1

Review: "الإطلاق على الفيلم يعجبني لم" (I did not like the movie at all.) → Negative (سليبي) → 0

Review: "متوسطاً الفيلم كان" (The movie was average.) → Neutral (محايد) → 2

Stop words removal: Removing stop words involves eliminating common words that don't significantly contribute to the meaning of the text in the context of text mining. These include words like pronouns, articles, and prepositions, which appear frequently but lack substantial semantic relevance. Examples include words such as 'على' (on), 'هو' (he), and 'من' (of). Since most of these words don't play a crucial role in categorization tasks, they can be discarded without negatively impacting the performance of the classifier. Doing so often enhances the process by reducing noise in the data [9].

Stemming: Stemming involves replacing a word with its root form and is considered a crucial preprocessing step in opinion mining and natural language processing [17]. This method standardizes words by reducing them to their base form. In Arabic, many words share the same root; for instance, the words (تفكير-مفكر-فكرة-تفكر) all stem from the same root, which is (فكر). As a result, these words would be unified under the root (فكر). This method significantly reduces the size of the text corpus, eliminates redundant terms, and ultimately enhances the efficiency of text categorization [9].

Removing Emojis/Emoticon: Emoticon extraction involves identifying textual representations of facial expressions, which serve as valuable emotional indicators in written content. We extracted emoticons and created a dictionary containing 63 different symbols. These emoticons were then organized into nine distinct emotional categories. Ultimately, we replace the emoticons found in a review with the corresponding category label they represent [8]. A subset of this emoticon dictionary can be seen in Fig. 1.

Tokenization: Tokenization in opinion mining is like breaking down sentences into individual words, making it easier for sentiment analysis to understand and interpret the meaning behind each word.

The mentioned pre-processing techniques are commonly beneficial in movie review sentiment analysis. Using a large number of methods can be beneficial if they collectively contribute to solving diverse issues like noise reduction, text normalization, and feature selection. These techniques collectively contribute to preparing the data for sentiment analysis, addressing issues related to noise, irrelevant information, and variations in language expression. Applying them appropriately can enhance the accuracy and meaningful interpretation of sentiments in movie reviews.

Emoticon	Emotional category
;-) ;) *)	Wink
:'-(:'(Cry
:-)) :-) :)	Smiley face
==3 -3	Laugh
:- :@ >:(Angry
: [:[:-<	Sad
>: >:/ :-/	Annoyed
:* ('){'}	Kisses
:\$	Embarrassed

Figure 1: Emoticon dictionary

3.2 Feature Extraction

Feature extraction is essential in movie review sentiment analysis as it identifies key elements or words that carry sentiment, enabling models to focus on relevant aspects. By extracting meaningful features, the analysis can capture nuances, sentiments, and expressions, enhancing the accuracy of sentiment predictions in diverse movie reviews. In essence, it transforms raw text data into informative features, making the sentiment analysis more effective and insightful.

TF-IDF, a technique employed in sentiment analysis, assesses the importance of words in extracted sentences and stands out as one of the most effective methods for identifying key elements in tweets. The term “number of times a certain term is repeated in a document” is also known as the “weight measure for determining the significance of a word in a specific document.” This weight, determined by the frequency of term occurrence in the document, increases when the term appears more than once. The frequency, known as TF, is calculated using Eq. (1) mentioned below.

$$T = \frac{N_s}{TN} \tag{1}$$

In the given equation, the overall count of terms is represented as ‘s,’ and the number of terms in the document is denoted as ‘TN.’ The significance of important terms is determined by IDF, which assigns importance to terms occurring infrequently in the document. This is expressed in Eq. (2).

$$IDF = \log_e \left(\frac{ND}{TD} \right) \tag{2}$$

The total count of documents containing terms is denoted as ‘ND,’ while the total count of documents is represented as ‘TD.’ Ultimately, the weight of the term is computed using Eq. (3).

$$TF - IDF (s, DC) = TF(s, DC) \times IDF(s) \tag{3}$$

In this context, ‘s’ refers to the number of terms, and ‘DC’ represents the document.

3.3 Proposed Machine Learning Model

Stochastic Gradient Descent (SGD) is like the guiding force behind training machines to understand and interpret sentiments from movie reviews. Stochastic Gradient Descent is an optimization algorithm commonly used in machine learning for training models, especially in the context of large datasets. It is an iterative and stochastic (randomized) approach that updates the model parameters based on individual or small batches of training examples.

Let's consider a machine learning model with parameters θ , a loss function $J(\theta)$, and a dataset with m training examples.

Loss Function: The loss function measures the difference between the predicted values $h_{\theta}(x)$ and the actual values (y).

$$J(\theta) = \frac{1}{m} \sum_{i=1}^m L(h_{\theta}(x^{(i)}), y^{(i)}) \quad (4)$$

where L is the loss for a single example.

Gradient Calculation: Compute the gradient of the loss with respect to the parameters θ :

$$\nabla_{\theta} J(\theta) = \frac{1}{m} \sum_{i=1}^m \nabla_{\theta} L(h_{\theta}(x^{(i)}), y^{(i)}) \quad (5)$$

Parameter Update: Update the parameters using the learning rate (α):

$$\theta = \theta - \alpha \nabla_{\theta} J(\theta) \quad (6)$$

The proposed algorithm for SDG model is given in Algorithm 1.

Algorithm 1: Stochastic gradient descent (SGD) machine learning model

// Step 1: Initialize random parameters θ

initialize random θ

// Step 2: Training loop through epochs

For loop epoch do

// Step 3: Shuffle the training data for randomness

shuffle data

// Step 4: Mini-batch training loop

For loop mini-batch do

// Step 5: Compute the gradient of the average loss for the mini-batch

compute gradient: $\nabla_{\theta} J(\theta) = \frac{1}{m} \sum_{i=1}^m \nabla_{\theta} L(h_{\theta}(x^{(i)}), y^{(i)})$ for i in batch using Eq. (3)

// Step 6: Update parameters using the learning rate

update parameters: $\theta = \theta - \alpha \nabla_{\theta} J(\theta)$ using Eq. (4)

End for loop

End for loop

In the pseudo code, α is the learning rate, $h_{\theta}(x)$ is the model prediction, L is the loss function, and ∇_{θ} denotes the gradient with respect to the parameters θ .

Note: The actual implementation might include variations, such as using adaptive learning rates or momentum, but the core idea remains consistent.

Hyperparameters are parameters that are not learned from the data during training but are set manually before starting the training process. They represent higher-level configuration choices for the machine learning algorithm, and their values are typically based on the characteristics of the data being used and the algorithm's ability to learn from that data [19] (Table 2).

Table 2: Hyperparameter descriptions and values

Parameter	Description	Values
Loss	Specifies the loss function used during training.	hinge, log, modified_huber
Penalty	Regularization type to prevent overfitting.	L2, L1, Elasticnet
Alpha	Regularization strength. Higher values apply stronger regularization.	0.0001, 0.001, 0.01
Learning_rate	Controls the step size during updates to parameters.	constant, optimal
Max_Iter	The maximum number of iterations for convergence.	1000, 2000
CV	Number of cross-validation folds to evaluate each parameter set.	5 (5-fold cross-validation)
Grid Search	Systematic process of searching through combinations of parameters.	All combinations of parameters in param_grid

4 Experimentation Setup

The experiment for the proposed SGD method took place on a Windows 10 Pro system equipped with 16 GB of RAM and a 10th-generation Core i5 processor. For the experiment, Jupiter Lab, a web-based IDE for scientific computing, was utilized. The Jupiter Lab IDE was accessed through an Anaconda emulator. Anaconda, a distribution of Python and R tailored for data science, machine learning, and scientific computing, simplifies the setup by providing a user-friendly environment with pre-installed libraries and packages.

4.1 Description of Datasets

This article utilizes datasets containing movie reviews in two distinct languages. The initial dataset comprises, a sample of 100,000 Arabic reviews, while the second dataset involves 50,000 English movie reviews from IMDB.

4.2 Arabic Dataset

This dataset is primarily a combination of various existing datasets, and it involves a sample of precisely 100,000 rows. The compilation includes reviews spanning hotels, books, movies, products, and a select number of airline reviews. Categorized into three classes (Mixed, Negative, and Positive), the ratings are mostly derived from reviewers' assessments, with a rating above 3 classified as positive and below 3 as negative. Notably, the dataset excludes duplicate reviews [20].

The dataset under consideration comprises two key features:

- Text: This field encompasses customer reviews on products in the Arabic language.
- Label: The labels assigned are either 'positive' or 'negative.'

The Problem encountered while working with the Arabic dataset is the presence of diverse dialects. Reviews encompass not only standard Arabic but also various regional variations like Egyptian, Gulf, and Levantine dialects. This introduces a notable challenge in developing a model that can effectively analyze and classify comments, considering the linguistic differences arising from the various dialects. The task involves creating a model robust enough to handle the diversity in language expressions and accurately categorize comments despite the variations in dialects. The snapshot of some examples from Arabic dataset is shown in Fig. 2.

label	text
Positive	..من أجمل ما قرأت.. رواية تستحق القراءة فعلا
Positive	جيد . . لا يوجد غير حمام واحد في الشقة
Positive	روايه مشوقه مشاعر راقية اختيار اسماء متناغم و الاهتمام بتفاصيل ممتع سعدت بقراءتها
Negative	ضعيف جدا. قريب من الحرم لكن يوجد طلعة الفندق صعبة لكبار السن. كل شيء
Negative	اقامه سيئه جددا. تعامل الموظفين في خدمات الغرف والاستقبال مغنيين جدا وقليلي الحياء

Figure 2: Some examples of reviews from the Arabic dataset

As shown in Table 3, the dataset includes frequently occurring words in both Arabic and English, which are crucial for understanding sentiment distribution across languages.

Table 3: Arabic and english words usually in reviews

Positive Arabic words	Positive English words	Negative Arabic words	Negative English words
رائع	Wonderful	فظيع	Terrible
ممتاز	Excellent	سيئ	Bad
رائعة	Fantastic	السمعة سيء	Poor reputation
مذهل	Amazing	للآمال مخيب	Disappointing
ميز	Outstanding	السمعة سيئ	Bad reputation

4.3 IMDB Dataset (English Dataset)

The IMDb 50K review dataset is well-known in the fields of sentiment analysis and natural language processing. It comprises 50,000 movie reviews, segregated into training and test sets, commonly referred to as the “IMDb movie reviews dataset” or simply the “IMDb dataset.” The reviews are categorized as either positive or negative, and the dataset is structured to maintain balance, featuring an equal number of positive and negative reviews [16].

4.4 Performance Matrices

The effectiveness of the proposed Stochastic Gradient Descent (SGD) is assessed using different performance metrics, which are listed below.

Accuracy is determined by counting the instances where the data is correctly classified and dividing it by the total number of data instances.

$$\text{Accuracy} = \frac{\text{TN} + \text{TP}}{\text{TN} + \text{FP} + \text{TP} + \text{FN}} \quad (7)$$

In this calculation, “true negative” (TN), “true positive” (TP), “false positive” (FP), and “false negative” (FN) are used as indicators [16,21].

Precision is a measure of a reliable classifier, aiming for a value close to 1. It reaches 1 when the number of true positives (TP) equals the sum of true positives and false positives (TP + FP), indicating zero false positives (FP) [19,21].

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (8)$$

Recall is computed by considering the correctly identified Positive samples and dividing them by the total number of both Positive and negative samples [16].

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (9)$$

The **F1 score** incorporates both recall and precision in its calculation. It reaches a value of 1 only when both precision and recall are equal to 1 [19].

$$\text{F1 score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (10)$$

4.5 Results and Discussion

In this study, experiments were conducted using movie review datasets in two distinct languages, namely Arabic and English [22,23]. The inclusion of these diverse language datasets serves the purpose of training the model to recognize and adapt to different script styles shown in Tables 4 and 5. Notably, the Arabic script significantly differs from the English alphabet. The comparison of results is conducted separately for the Arabic dataset and the English dataset, allowing for a comprehensive assessment of the model’s performance across different language styles shown in Table 4. This approach ensures the model’s versatility in handling linguistic variations and nuances specific to each language.

Table 4: Performance comparison of ML models on Arabic dataset

Model	Precision (%)	Recall (%)	F1-score (%)	Accuracy (%)
K-Neighbors classifier	56.02	55.18	55.77	55.98
Complement Naive Bayes	81.71	83.70	80.69	81.70
Passive aggressive classifier	77.84	78.82	79.82	78.62
Decision tree	73.27	74.97	76.24	74.67
Gradient boosting classifier	78.20	77.46	75.31	76.22
AdaBoost classifier	76.22	77.22	73.50	75.64
SGD Classifier (Proposed)	83.22	82.82	86.45	84.89

Table 5: Performance comparison of ML models on English movie reviews IMDB dataset

Model	Precision (%)	Recall (%)	F1-score (%)	Accuracy (%)
K-Neighbors classifier	75.48	80.84	78.07	77.28
Complement Naive Bayes	87.25	84.34	85.77	86.00
Passive aggressive classifier	87.33	88.36	87.84	87.76
Decision tree	71.44	72.25	71.84	71.67
Gradient boosting classifier	79.41	87.11	83.08	82.25
AdaBoost classifier	84.47	79.62	81.97	81.42
SGD classifier (Proposed)	89.67	86.35	88.59	87.44

The Arabic dataset utilized in this study is a compilation of various datasets encompassing domains like hotels, books, movies, and products. Despite its diverse nature, the focus remains on training the model using Arabic script. Notably, for the Arabic dataset, a selective set of pre-processing methods was applied, including Data Cleaning, Label Encoding, Removing Emojis/Emotions, Tokenization, and Stemming. On the other hand, the English dataset underwent a different set of pre-processing methods, involving Data Cleaning, Label Encoding, Stop Words Removal, Stemming, and Spelling Correction. These tailored pre-processing approaches acknowledge the linguistic nuances and specific challenges posed by each language, ensuring a more effective training process for the model in both Arabic and English contexts. [Table 4](#) shows the results of the Arabic dataset and [Table 4](#) shows the results of the English dataset using the existing and proposed models.

[Table 4](#) presents a comparison of the performance between the proposed SGD model and an existing models like K-Neighbors Classifier (KNC) [24], Complement Naive Bayes (CNB) [25], Passive Aggressive Classifier (PAC), Decision Tree (DT) [26], Gradient Boosting Classifier (GBC) [27] and AdaBoost Classifier (AC) [28] using various performance metrics such as precision, recall, F1-Score, and accuracy. The Complement Naive Bayes model achieves a higher recall value of 83.70, indicating its ability to identify a larger proportion of relevant instances in the dataset. However, the proposed SGD model surpasses it in precision, F1-Score, and accuracy, with values of 83.22, 86.45, and 84.89, respectively. This means that while the Complement Naive Bayes model is good at capturing relevant instances, the proposed SGD model excels in accurately identifying positive cases, achieving a balance between precision and recall, and overall performance accuracy. SGD allows for flexible model complexity adjustments, which can adapt well to the intricacies of Arabic language data. This adaptability ensures that the model can capture the nuances and specific local patterns of Arabic text, leading to better performance. In [Table 4](#), the bold values represent higher results. [Fig. 3](#) also demonstrates and visually presents how well the proposed model performs when dealing with Arabic data.

[Table 5](#) compares the performance of the proposed SGD model with an existing model using English movie reviews. The Passive Aggressive Classifier achieves a higher recall value of 88.36, indicating its ability to identify a larger proportion of relevant instances. However, the SGD model outperforms in precision, F1-Score, and accuracy, with values of 89.67, 88.59, and 87.44, respectively. This means that while the Passive Aggressive Classifier excels in capturing relevant instances, the SGD model demonstrates better precision, a balanced F1-Score, and overall accuracy. Additionally, [Fig. 4](#) visually illustrates the performance of the proposed model, providing a clear representation of its effectiveness.

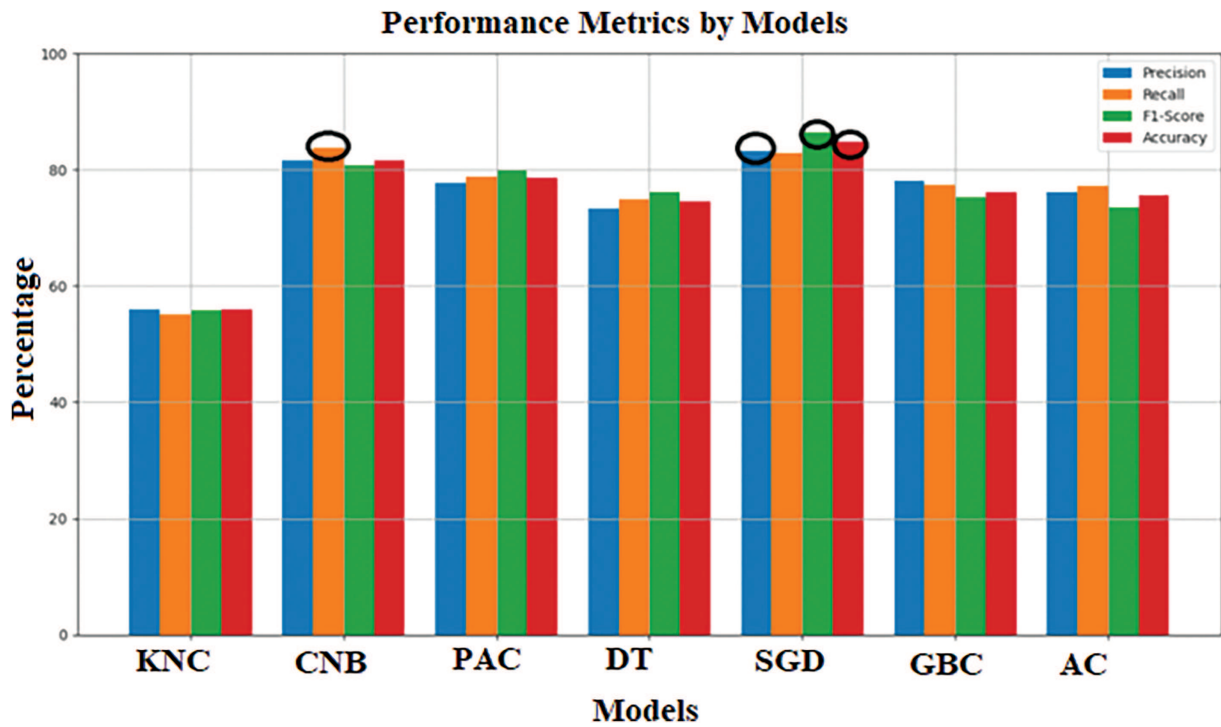


Figure 3: Performance metrics by model evaluation of models on Arabic reviews dataset

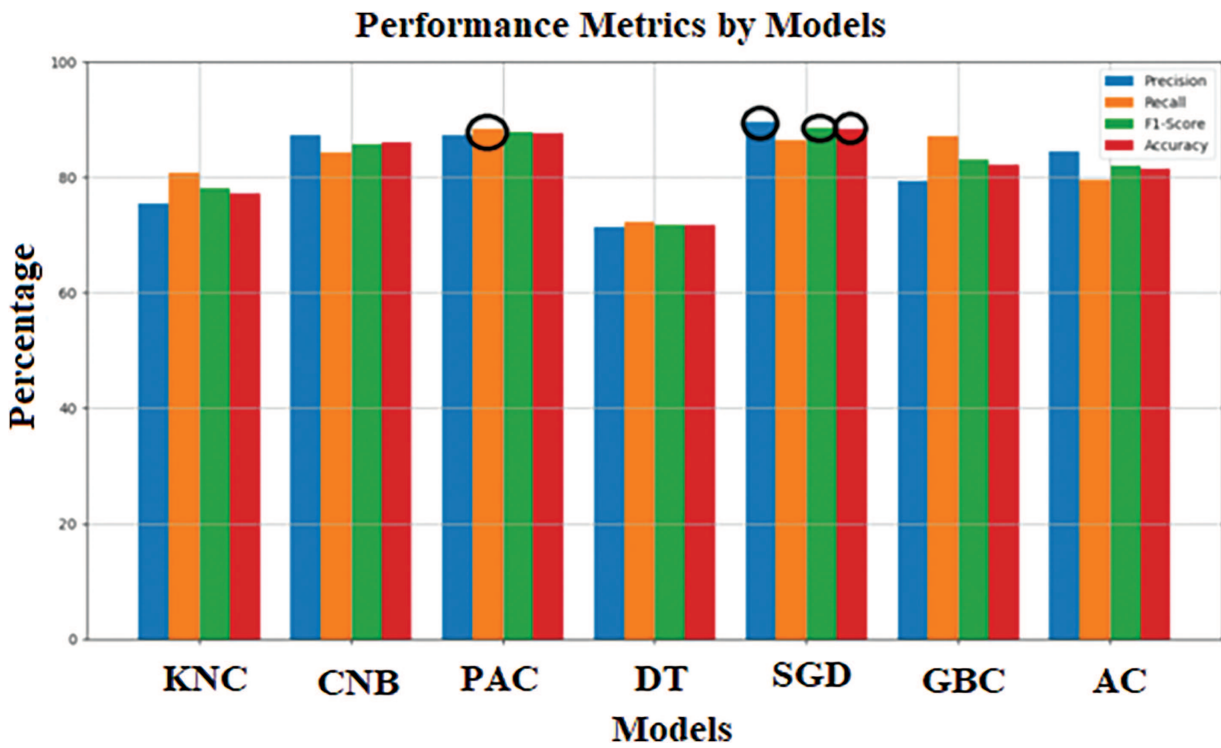


Figure 4: Performance evaluation of models on IMDB movie reviews dataset

5 Time Complexity

The time complexity of various classification models applied to movie reviews in both English and Arabic datasets, notable differences emerge. Conversely, the Decision Tree model showcases significantly higher complexities of 2.09 s for English and 234.44 s for Arabic datasets, suggesting that its computational demands vary drastically depending on the dataset's language. The SGD Classifier presents relatively stable complexities of 3.67 s for English and 2.44 s for Arabic datasets, showcasing its consistency in handling both languages efficiently. Overall, the time complexities of these techniques vary, suggesting the importance of considering computational efficiency when selecting classification models for movie reviews in diverse language datasets showing in Fig. 5.

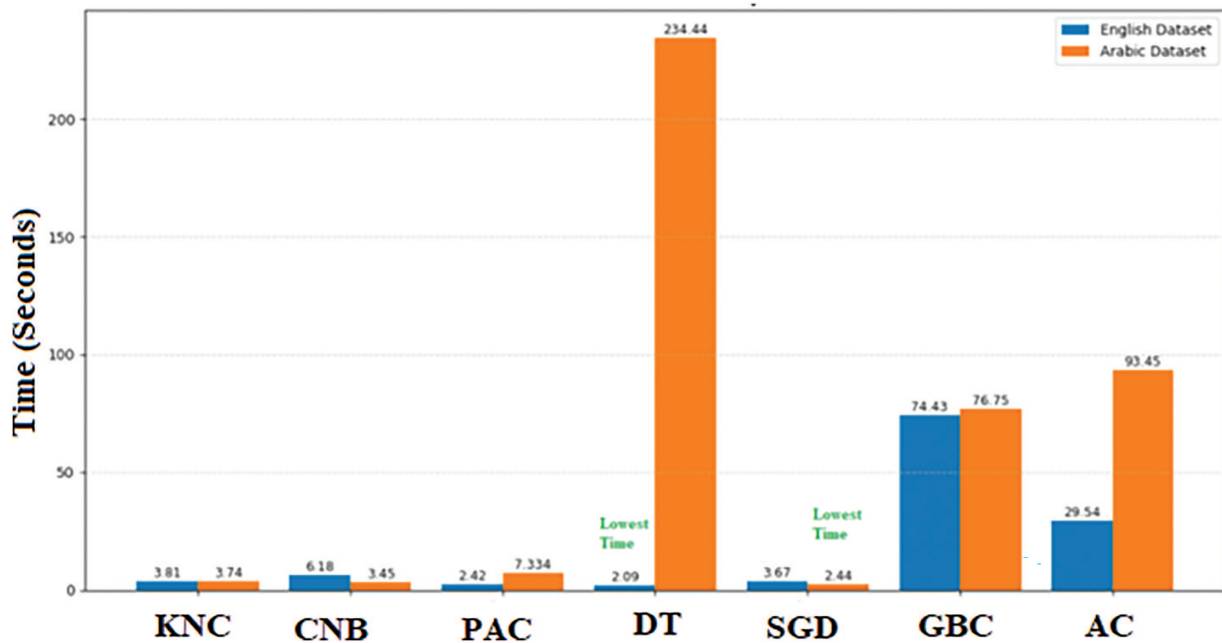


Figure 5: The time complexity of classification models on English and Arabic datasets

6 Conclusion

In conclusion, this study presents a significant advancement in the field of sentiment analysis, particularly focusing on movie reviews in both Arabic and English. By introducing a Stochastic Gradient Descent (SGD) machine learning model, we have addressed the challenges posed by diverse language datasets and the overwhelming number of reviews available online. The model's ability to process large amounts of data efficiently, while also adapting to the unique characteristics of both languages, highlights its potential for accurately classifying sentiments. Moreover, our rigorous pre-processing steps have optimized the datasets, ensuring that the analysis is both reliable and effective. The results demonstrate that the SGD model consistently outperforms existing sentiment analysis tools, achieving impressive accuracy rates across both datasets. This not only enhances the movie selection experience for Arabic-speaking audiences but also paves the way for future research in sentiment analysis within other complex languages. This work contributes to the growing field of Natural Language Processing, offering insights and methodologies that can be applied to various domains beyond movie reviews. Ultimately, our findings support the development of more sophisticated sentiment analysis solutions that are accessible and beneficial to a wider audience. In

the future, we will work with a dataset that includes multiple languages (Mixed Languages) such as English, Urdu, Arabic, and more.

Acknowledgement: None.

Funding Statement: The authors received no specific funding for this study.

Author Contributions: Sarwar Shah Khan: Conceptualization, Methodology, Formal analysis, Investigation, Writing—original draft preparation. Yasser Alharbi: Supervision, Validation, Writing—review and editing. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: Not applicable.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest to report regarding the present study.

References

1. Nagaraj P, Kumar PA, Babu VC, Yedukondalu P, Sai MY, Sudar KM. Movie reviews using sentiment analysis. In: 2023 International Conference on Computer Communication and Informatics (ICCCI); 2023 Jan 23–25; Coimbatore, India: IEEE. p. 1–6. doi:10.1109/ICCCI56745.2023.10128450.
2. Sangeetha B, Sangeetha S, Goutham DT, Vaibhav Ram N. Sentiment analysis on movie reviews: a comparative analysis. In: 2023 International Conference on Intelligent Systems for Communication, IoT and Security (ICISCOIS); 2023 Feb 9–11; Coimbatore, India: IEEE. p. 218–23. doi:10.1109/ICISCOIS56541.2023.10100367.
3. Chatterjee S, Chakrabarti K, Garain A, Schwenker F, Sarkar R. JUMRv1: a sentiment analysis dataset for movie recommendation. *Appl Sci.* 2021;11(20):9381. doi:10.3390/app11209381.
4. Lindo A. Movie spoilers classification over online commentary, using Bi-LSTM model with pre-trained GloVe embeddings [masters thesis]. Dublin, Ireland: National College of Ireland; 2020.
5. Oueslati O, Cambria E, Ben HajHmida M, Ounelli H. A review of sentiment analysis research in Arabic language. *Future Gener Comput Syst.* 2020;112(4):408–30. doi:10.1016/j.future.2020.05.034.
6. Obiedat R, Al-Darras D, Alzaghouel E, Harfoushi O. Arabic aspect-based sentiment analysis: a systematic literature review. *IEEE Access.* 2021;9:152628–45. doi:10.1109/ACCESS.2021.3127140.
7. Al-Mansoori KW, Al-Shalabi R. Sentiment analysis algorithm for Arabic reviews on the movies domain. In: 2022 ASU International Conference in Emerging Technologies for Sustainability and Intelligent Systems (ICETSIS); 2022 Jun 22–23; Manama, Bahrain: IEEE. p. 126–30. doi:10.1109/ICETSIS55481.2022.9888804.
8. Alnemer L, Alammouri B, Alsakran J, El Ariss O. Enhanced classification of sentiment analysis of Arabic reviews. In: *Advances in internet, data and web technologies.* Cham: Springer International Publishing; 2019. p. 210–20. doi:10.1007/978-3-030-12839-5_20.
9. Sayed AA, Elgeldawi E, Zaki AM, Galal AR. Sentiment analysis for Arabic reviews using machine learning classification algorithms. In: 2020 International Conference on Innovative Trends in Communication and Computer Engineering (ITCE); 2020 Feb 8–9; Aswan, Egypt: IEEE. p. 56–63. doi:10.1109/itce48509.2020.9047822.
10. Al-Moslemi T, Albared M, Al-Shabi A, Omar N, Abdullah S. Arabic senti-lexicon: constructing publicly available language resources for Arabic sentiment analysis. *J Inf Sci.* 2018;44(3):345–62. doi:10.1177/0165551516683908.
11. Guellil I, Azouaou F, Mendoza M. Arabic sentiment analysis: studies, resources, and tools. *Soc Netw Anal Min.* 2019;9(1):56. doi:10.1007/s13278-019-0602-x.
12. Brahimi B, Touahria M, Tari A. Improving sentiment analysis in Arabic: a combined approach. *J King Saud Univ Comput Inf Sci.* 2021;33(10):1242–50. doi:10.1016/j.jksuci.2019.07.011.
13. Louati A, Louati H, Kariri E, Alaskar F, Alotaibi A. Sentiment analysis of Arabic course reviews of a Saudi University using support vector machine. *Appl Sci.* 2023;13(23):12539. doi:10.3390/app132312539.

14. Abimanyu AJ, Dwifabri M, Astuti W. Sentiment analysis on movie review from rotten tomatoes using logistic regression and information gain feature selection. *Build Inform Technol Sci (BITS)*. 2023;5(1):162–70. doi:10.47065/bits.v5i1.3595.
15. Al-Batah MS, Mrayyen S, Alzaqebah M. Arabic sentiment classification using MLP network hybrid with naive Bayes algorithm. *J Comput Sci*. 2018;14(8):1104–14. doi:10.3844/jcssp.2018.1104.1114.
16. Danyal MM, Khan SS, Khan M, Ullah S, Mehmood F, Ali I. Proposing sentiment analysis model based on BERT and XLNet for movie reviews. *Multimed Tools Appl*. 2024;83(24):64315–39. doi:10.1007/s11042-024-18156-5.
17. Sharma H, Pangaonkar S, Gunjan R, Rokade P. Sentimental analysis of movie reviews using machine learning. *ITM Web Conf*. 2023;53(1–2):02006. doi:10.1051/itmconf/20235302006.
18. Prayoga I, Purbolaksono MD, Adiwijaya A. Sentiment analysis on Indonesian movie review using KNN method with the implementation of chi-square feature selection. *MIB*. 2023;7(1):369. doi:10.30865/mib.v7i1.5522.
19. Danyal MM, Khan SS, Khan M, Ghaffar MB, Khan B, Arshad M. Sentiment analysis based on performance of linear support vector machine and multinomial Naïve Bayes using movie reviews with baseline techniques. *J Big Data*. 2023;5:1–18. doi:10.32604/jbd.2023.041319.
20. Salloum AM, Almustafa MM. Analysis and classification of customer reviews in Arabic using machine learning and deep learning. *J Data Acquis Process*. 2023;38(4):726. doi:10.5281/zenodo.777803.
21. Danyal MM, Haseeb M, Khan B, Ullah S, Khan SS. Opinion mining on movie reviews based on deep learning models. *J Artif Intell*. 2024;6(1):23–42. doi:10.32604/jai.2023.045617.
22. Danyal MM, Khan SS, Khan M, Ullah S, Ghaffar MB, Khan W. Sentiment analysis of movie reviews based on NB approaches using TF-IDF and count vectorizer. *Soc Netw Anal Min*. 2024;14(1):87. doi:10.1007/s13278-024-01250-9.
23. Khan SS, Alharbi Y. Sentiment analysis of movie review classifications using deep learning approaches. *Int J Adv Appl Sci*. 2024;11(8):146–57. doi:10.21833/ijaas.2024.08.016.
24. Suprayogi S, Sari CA, Rachmawanto EH. Sentiment analyst on twitter using the K-nearest neighbors (KNN) algorithm against covid-19 vaccination. *J Appl Intell Syst*. 2022;7(2):135–45. doi:10.33633/jais.v7i2.6734.
25. Dewi C, Chen RC. Complement naive Bayes classifier for sentiment analysis of internet movie database. In: *Intelligent information and database systems*. Cham: Springer International Publishing; 2022. p. 81–93. doi:10.1007/978-3-031-21743-2_7.
26. Phan HT, Tran VC, Nguyen NT, Hwang D. Decision-making support method based on sentiment analysis of objects and binary decision tree mining. In: *Advances and trends in artificial intelligence. From theory to practice*. Cham: Springer International Publishing; 2019. p. 753–67. doi:10.1007/978-3-030-22999-3_64.
27. Alzamzami F, Hoda M, El Saddik A. Light gradient boosting machine for general sentiment classification on short texts: a comparative evaluation. *IEEE Access*. 2020;8:101840–58. doi:10.1109/ACCESS.2020.2997330.
28. Salman AH, Al-Jawher WAM. Performance comparison of support vector machines, AdaBoost, and random forest for sentiment text analysis and classification. *J Port Sci Res*. 2024;7(3):300–11. doi:10.36371/port.2024.3.8.