Machine Learning and AraBERT Models for Arabic Online Reviews Sentiment Analysis

Rahaf A Alsuhemi, Soha M Zarbah

Department of Information System, King Abdulaziz University

CPIS 604 Emerging Technologies in Business

Dr. Ghada Amoudi

May 21, 2022

Machine Learning and AraBERT Models for Arabic Online Reviews Sentiment Analysis

Romanian Journal of Information Technology and Automatic Control

Rahaf Alsuhaimi[1], Soha Zarbah[2]

[12] Institution King Abdulaziz University

[1]email ralsuhaimi0008@stu.kau.edu.sa

[2]email szarbah@stu.kau.edu.sa

Abstract Despite the fact that Arabic is the sixth most spoken language in the world, it is a less studied syntax compared to English, where most sentiment analysis and data sets focus on English. This is because the subtleties of the Arabic language and the different Arabic dialects that exist in each Arab country. However, most studies on Arabic sentiment analysis use a dataset specialized in a particular domain, which results in the resulting model being difficult to apply or having low accuracy when tested with data from other domains. Therefore, in this work, different datasets from different domains are processed and then combined into one dataset so that the results can be compared between the performances of all cases in the dataset. Logistic regression, random forest, support vector machine, and Naïve Bayes classifiers and the Arabert model are applied to determine the best performance. The results show that the two best models that perform better are logistic regression and support vector machine. Moreover, the Naïve Bayes classifier achieved the highest accuracy when the datasets were combined. It is also noted that the size of the dataset, the number of classes, balancing, and the combination of different fields affect the performance.

Keywords: Arabic Sentiment Analysis (ASA). Deep Learning (DL). Machine Learning (ML). Natural Language Processing (NLP) Sentiment Analysis (SA).

1. Introduction

Recently, natural language processing (NLP) has become an important research topic. Its popularity has attracted the interest of researchers in the fields of data, web and text mining, and information extraction (Setyanto et al., 2022). One of the most common tasks in NLP is the identification and classification sentiment (Folorunsho, 2020). Sentiment Analysis (SA) has been the subject of numerous research due to the enormous range of possible applications. It is used to determine whether a text is positive, negative, or neutral. In the SA system, NLP and machine learning (ML) are combined to provide scaled sentiment scores for the entities, subjects, and categories within a phrase. ML in SA enhances and automates the limited text analysis capabilities that SA relies on (Contributor, 2021).

Since SA works on written expressions of different languages, Arabic is one of the most popular web languages and is one of the top four languages with over 226 million users(Alharbi et al., 2021). Although Arabic is one of the most widely spoken languages in the world, it is hardly considered in sentiment analysis (Alharbi et al., 2021). Furthermore, the analysis of Arabic texts poses an additional challenge, since Arabic has a variety of morphological features as well as several dialects in addition to its standard form.

However, there are a set of studies which focus on sentiment analysis of Arabic. But it is noticeable in most of these studies that they use a data set specialized in a particular field, which produces small accuracy when tested on data from different fields. The majority of previous research has concentrated on tweet and book review analysis whereas there are many different areas to apply the SA. Therefore, in this paper, sentiment analysis is performed using various datasets from different domains, including products, movies, and restaurants, as well as combinations of these datasets. In addition, various ML classifiers are used in conjunction with the Arabert model, these models are validated against a number of metrics. The research questions that this paper attempts to answer are depicted below.

1.1 Research questions:

1- What are the characteristics of the Arabic movies, products, and restaurants reviews datasets?

2- How models perform when the datasets are separated and combined?

3- What are the most effective classifiers for dealing with Arabic content?

The rest of the paper proceeds as follows: literature review, the methodology, the implemented algorithms, the results, discussion are presented before concluding of the paper and future work.

2. Literature Review

2.1 ML and Sentiment Analysis

Hence sentiment analysis through reviews and opinions is really important. This survey research (Jain & Pamula, 2020) aimed to perform a review of a number of previous studies (68 papers) on the topic of sentiment analysis in the context of tourism and hotels. Despite the fact that the study was focused on this topic, the result was a framework that describes how to do sentiment analysis using ML. These stages also extended beyond data preparation to include anything from topic selection to model adjustment for increased accuracy.

Another study on the issue of sentiment analysis to anticipate global disease outbreaks and epidemics was undertaken between 2010 and 2017 (Singh et al., 2018). The studies surveyed were either discussing the application of ML in that discipline, sentiment analysis categorization and prediction, or epidemic and health-care forecasting. The findings support the use of sentiment analysis for future prediction in a variety of areas, not only health care.

2.2 Arabic Sentiment Analysis

Since there is a lack of many resources in the field of Arabic sentiment analysis, this work (ElSahar & El-Beltagy, 2015) addressed this problem by providing a large dataset in various domains (products, movies, hotel, restaurant). The SVM model provided the best performance, and the best feature models were the lexicon-based feature combined with other features such as TFI-IDF and number.

The authors of (Nabil et al., 2015) present the largest Arabic sentiment analysis dataset with over 63,000 book reviews rated 1 to 5 stars. The dataset is then used to investigate two tasks: sentiment polarity classification and rating classification. To evaluate this dataset, the authors applied eight classifiers with OVA technique. They revealed that the SVM and logistic regression classifiers have high accuracy. Eventually, the authors provided a seed sentiment lexicon derived from the dataset that can easily extend to other datasets or domains.

A deep learning-based sentiment analysis model was proposed (Alharbi et al., 2021) to predict the polarity of personal views and sentiments. The datasets used were product, restaurant, and hotel datasets from(ElSahar & El-Beltagy, 2015), as well as book reviews and tweets. Results showed that the DeepASA model outperformed all tested datasets with accuracy ranging from 81.11% to 94.32%. In addition, the model lowered the comparable classification error rate by up to 26%.

3. Methodology

3.1 Dataset

This work will employ sentiment analysis for Arabic reviews of movies, products, and restaurants. The datasets were used from (ElSahar & El-Beltagy, 2015), but it's unbalanced and not well cleaned. A sample of some texts is illustrated in Figure 1. All datasets are labeled and have two columns: text and polarity. The text contains the reviews, and polarity includes three classes: 1 as positive, -1 as negative, and 0 as a mixed review as shown in Figure 2. The movie dataset has 1524 reviews, whereas the restaurant dataset has 10970 reviews, and the product dataset has 4272 reviews.

Buona scelta gluten-free \nIl ristorante "Il Grissino" è un onesto ristorante la famiglia, senza troppe pretese e, coerentemente, con costi contenuti; caratteristiche che ne fanno un buon locale d'appoggio se si vuole mangiare fuori per comodità ma non si sta cercando l'evento memorabile. Tutte le portate sono più che discrete, il servizio cortese e ragionevolmente rapido. Fra i luoghi che fanno pizza senza glutine a Milano, le loro proposte (basate su farina DS) sono da contarsi fra le migliori. Raccomandabile per una cena tranquilla senza impegno.\n

\nمطعم هندي لذيذ بميلان وقريب من الدومو انصح فيه وبشده ،، حطوا اسم المطعم بخرائط القوقل ويطلع لكم العنوان وقريب حيل من السنتر بس مكانه داخل شوي\n مطعم هندي بميلان

هناك صعوبة في ايجاد المطعم لكنه يستحق العناء فالمالك كان ودودا جدا والاستقبال والخدمة ممتازة\nتقريبا يبعد اقل من كيلومتر والمسافه تقطعها ماشيا\nالمطعم قريب من الدومو\n مطعم هندي ممتاز بأكله واستقباله وضيافته \nانصح به بشدة\nوالأكل لذيذ وخفيف وسعره معقول

\nمطعم هندي مأكولاته لذيذه وهو قريب من الدومو وانصح بزيارته لديه وجبات متنوعه وخدمته جيده وسريع في تحضير الوجبه المطلوبه بس حاول تخبره بتقليل الفلفل حتي لاتولع\n رائع

\nانا سافرت إيطاليا و أكلت بيتزا و كانت حلوى اوى والمطعم ممتاز و الخدمة حلو آوى و المكان متميز و المعاملة راقية\n مطعم ممتاز

\n♥♥♥ قريبا\nو سأعود له\nولن انسى مذاقه ابدا!!\n♥♥الاكل لذيذ جدا و اطيب بيتزا و سيزر سلاد تذوقتها في هذا المطعم\n ♥♥♥♥ مطعم رائع

\nالاكل لذيذ انصح به ريش غنم و بلي اربيانا ومعاملتهم رائعه انصح بتجربه الاكل فيه بس مشكله مافيه بيتزا مع انها اكلتهم المفضله\n مطعم رائع يستحق التجربه

Fossil CH2688 For Women (Analog, Dress Watch) إيجابيات و سلبيات -

صغيرة و مش بتقفل كويس و بتقع كتير شكلها كويس بس حجمها صغير و وقوعها المتكرر ممكن يخليها تضيع في اي حتة بس شكلها شغال مش وحشة المفروض تتحط كساعة يدالي مش حريمي بسبب حجمها الصغير ده

.... الساعة جميلة جدا وخاماتها ممتازة وقوية وعملية جدا وسعرها رائع جدا!!!!!!!!

اكثر من رائعة والبائع محترم وملتزم , وتغليف السلعة كان ممتاز .

Figure 1 Sample of Some Reviews in the Dataset

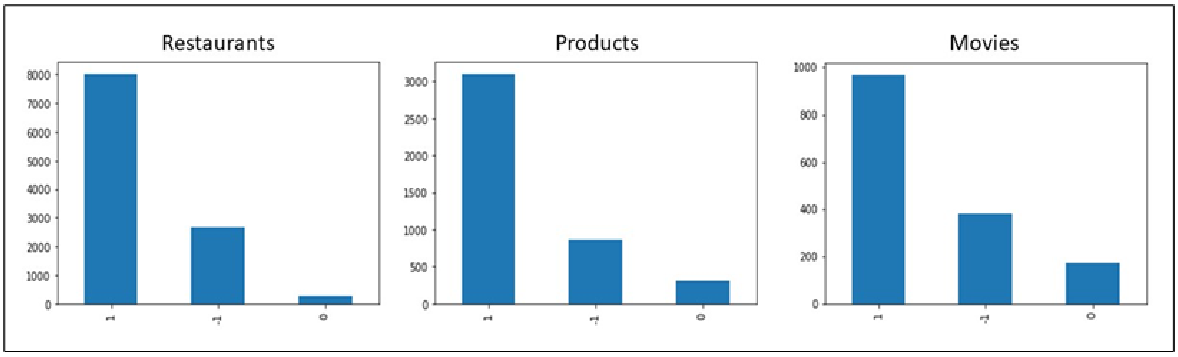

Figure 2 Datasets Classes and Size

3.2 Preprocessing

The product, restaurant, and movie datasets were cleaned using Text & Numbers Normalization, Removing Hashtags, Punctuations, Diacritics, English Words, Null cells, repeated characters, and stop words. The movie dataset did not need much cleaning. But additional cleaning was applied to restaurant data to remove emojis and other languages. Furthermore, class 0 was dropped because it has a weak effect, and the volume of data is minimal compared to the other classes The following sources have been used to import basic data preparation and cleaning fundamentals (ANLP_Workshop/TextMiningWorkshop, 2022; Arabic Text Preprocessing, 2019/2022).

3.3 Resampling

The issue of data imbalance arises as a consequence of its escalation and growth over the years (Mohammed et al., 2020). Because the data will be viewed as balanced, the result and accuracy will be biased to the side that contains the majority of the data (Iqbal et al., 2019). As a result, there is a growing interest in solving this challenge and finding new ways to deal with this sort of data (Mohammed et al., 2020). One of these ways was resampling, which have a variety of techniques (Mohammed et al., 2020), including the following techniques, which were clarified in Figure 3:

Random Over-Sampling (ROS) The minority group is enlarged by random repeating of the data in this category using this approach (Iqbal et al., 2019; Taghizadeh-Mehrjardi et al., 2020). The categories will be balanced in this manner, but the problem of overfitting may arise, affecting the model's accuracy (Iqbal et al., 2019).

Random Under-Sampling (RUS) This strategy involves deleting data from the majority category at random until the minority category is balanced (Iqbal et al., 2019; Taghizadeh-Mehrjardi et al., 2020). This should balance the categories, but some critical data may be removed, thus increasing the variance of machine learning algorithms (Taghizadeh-Mehrjardi et al., 2020).

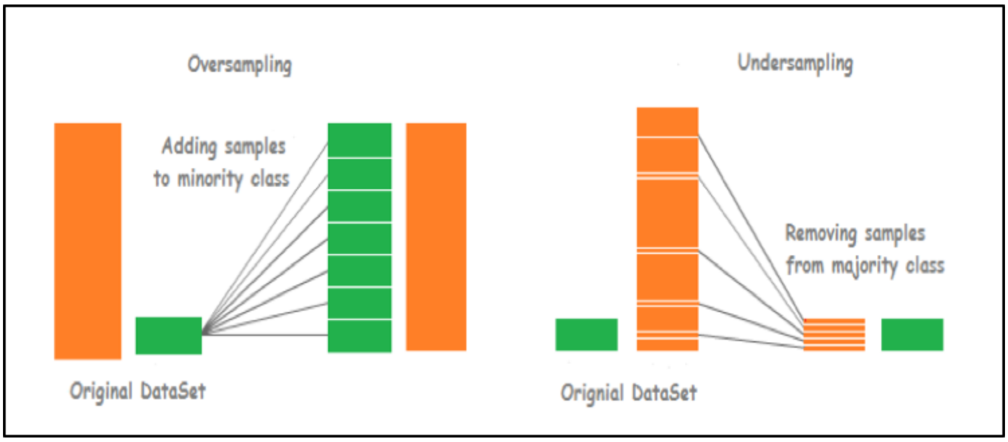

Figure 3 ROS and RUS Illustration. This figure adopted from (Mohammed et al., 2020)

4. Algorithm Implemented

For modeling using ML and DL in sentiment analysis, the data is divided into features (Text) and label (Polarity). Next, splitting the data into training and testing sets, and different classifiers will be implemented which are: Logistic Regression, Random Forest Classifier, Multinomial NB, and SVC. To get better results, the Pipeline class in Scikit-Learn which combines vectorization using TfidfVectorizer(), transformation, Gridsearch with Gridsearch() and classification is used. Except using Gridsearch() with naive Bayes because it has almost no hyperparameters to tune. Also, a pretrained AraBert for classification problems will be used. Each of the classifiers used will be briefly described here:

4.1 Logistic Regression

Logistic regression is a supervised machine learning technique for solving classification problems. The model's aim is to learn and estimate a mapping function $f(X_i) = Y$ from the input variables (X_1, X_2, and X_n) to the output variable (Y). It is referred to as supervised because the model predictions are evaluated iteratively against the output variables and corrected until appropriate performance is achieved (Kolamanvitha, 2021).

4.2 Random Forest Classifier

Random Forest is a classification technique used for sentiment identification and emotion analysis in large datasets. Random Forest is a classification method that combines tree building with training on sample data (Saifullah et al., 2021).

4.3 Multinomial NB

The Multinomial Naive Bayes classification algorithm is a good starting point for sentiment analysis tasks. The basic idea behind the NB technique is to use the joint probabilities of words and classes to determine the probabilities of the classes assigned to the texts (Smetanin, 2018).

4.4 SVC

It is an effective classifier-building approach. Its goal is to build a decision boundary between two classes known as the "hyperplane," which allows labels to be predicted from one or more feature vectors. It's oriented such that it's as far away as feasible from each of the classes' nearest data points, or "support vectors" (HUANG et al., 2018).

4.5 TF-IDF Vectorization

It is a statistical approach for determining the relevance of a word in a document or a file set category. The basic idea is that if a word or phrase shows repeatedly in one article but infrequently in others, it is assumed that the word or phrase has strong class distinguishing capacity and may be classified. The number of times a specific word appears in the file is referred to as word frequency (TF). The inverse file frequency (IDF) is a measure of a word's overall significance (Liu et al., 2018).

4.6 Gridsearch

It is a method of parameter tuning that builds and evaluates a model for each combination of algorithm parameters supplied in a grid in a systematic manner. The following are the crucial parameters to remember when utilizing Grid Search CV (Ranjan et al., 2019):

4.6.1 Estimator

This parameter receives the classifier that will be taught. We used a pipeline in our code that included TF-IDF vectorization and one of our earlier classification techniques.

4.6.2 Parameter grid

It is a Python dictionary having keys for parameter names and values for parameter settings. To provide the highest accuracy, all combinations of these parameters are examined.

4.6.3 Cross Validation

CV is a method for resampling available data in order to test machine learning models. The key benefit of this method is that it produces less bias or optimistic findings than a standard train-test split. It works by shuffling the dataset randomly at first. The dataset is then divided into k groups. The test data comes from each group, while the training data comes from the other groups. The assessment ratings for each group are saved and then summed up to assess the model's performance.

We will also use the Precision, Recall, F-score, and Accuracy metrics to evaluate the performance of the model. The evaluation will be performed separately for each dataset in different cases (balanced & unbalanced) and (with & without class 0). Then, a combination of the three datasets into one dataset is used to demonstrate the performance in different cases.

4.7 AraBERT with Fast-BERT library

AraBERT is a pre-trained Arabic language model built on Google's BERT structure. This model is a pretrained model for Arabic Reviews and it is suitable for reviews analysis (sentiment analysis). Fast-Bert is a DL library that allows developers and data scientists to train and implement natural language processing models based on BERT and XLNet.

Using Python, we used aubmindlab Bert base Arabert pretrained model form (AraBERTv2 / AraGPT2 / AraELECTRA, 2022). This model creates a data bunch object that contains training, validation, and test csv files and converts them to internal BERT representations. The model was then trained on the data sets with a learner object that held everything.

5. Results

5.1 Preprocessing Results:

The dataset is now being cleaned and preprocessed using the previously mentioned methods. The Figure 4 below displays samples of the cleaned dataset after applying the preprocessing phase.

Figure 4 Samples of Cleaned Dataset

5.2 Resampling Results:

In this paper, we adopted the RUS method, where we wanted to obtain high accuracy while avoiding overfitting. The total data in the restaurant data set was 10609 and then it became 5306 after using the RUS technique as shown in histogram in Figure 5.

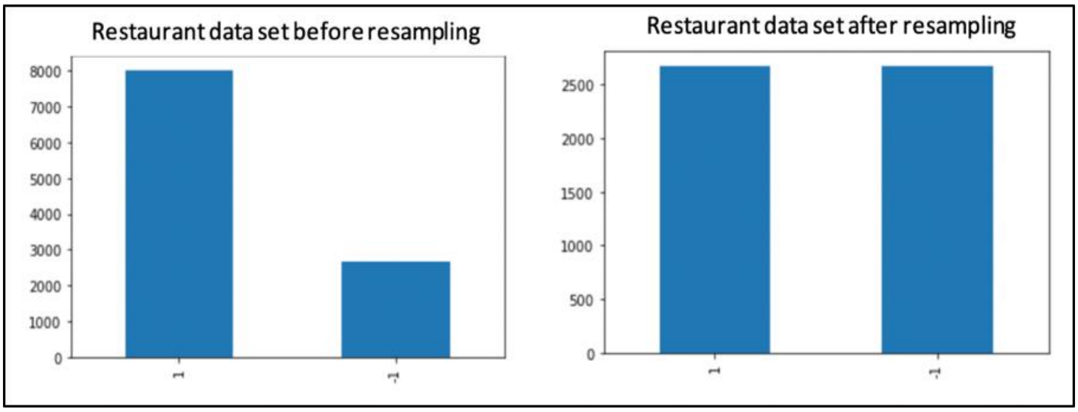

Figure 5 Restaurant Dataset Before and After Resampling

The following Figure 6 shows the change that took place in the product data set after resampling, where its total was 3689 and became 1612.

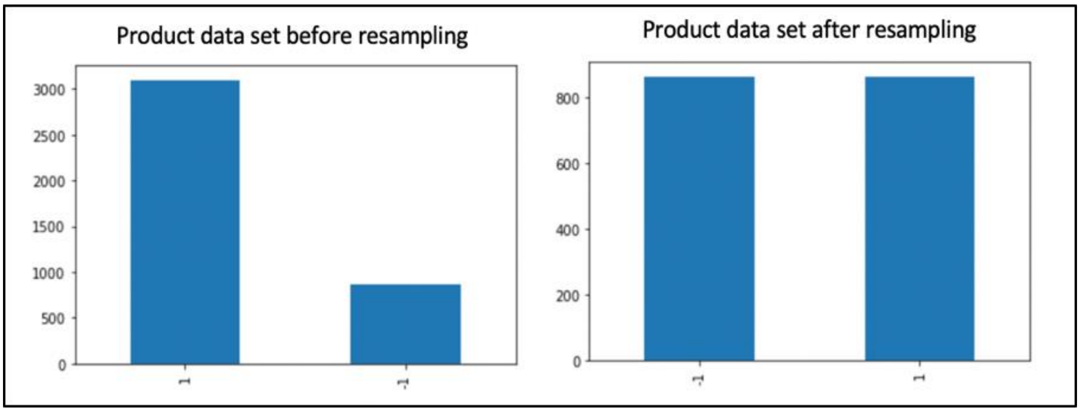

Figure 6 Product Dataset Before and After Resampling

As for the movie data set, before the resampling the total of data was 1353 and then it became 768 as shown in Figure 7.

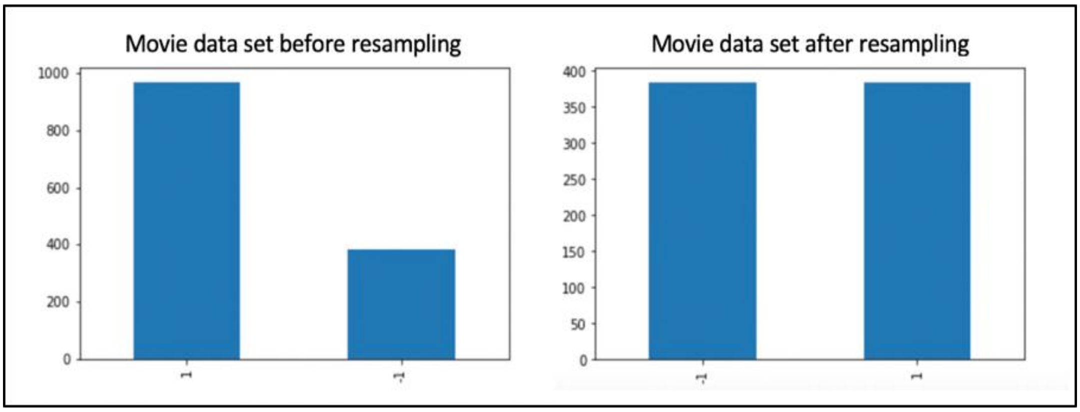

Figure 7 Movie Dataset Before and After Resampling

Moreover, we obtained comprehensive statistics for all datasets in different cases after applying preprocessing and under sampling. Table I summarizes the properties and statistics of the datasets.

Table I The Properties and Statistics of the Datasets

Dataset	**Original Unbalanced dataset**	**Unbalanced data with 3 classes**	**Unbalanced data with 2 classes**	**Balanced data with 2 classes**
Product Reviews	4272	3974	3689	1612
Restaurant Reviews	10970	10871	10609	5306
Movies Reviews	1524	1524	1353	768

5.3 Modeling Results:

5.3.1 Imbalanced with 3 classes

Table II illustrates the results of the initial modeling of the imbalanced data for the three data sets separately. Almost both logistic regression and SVC were the highest metrics for all datasets.

Table II The Results of the Imbalanced Dataset with Three Classes

	Restaurants				Products				Movies			
	Accuracy	Precision	Recall	F1	Accuracy	Precision	Recall	F1	Accuracy	Precision	Recall	F1
Logistic Regression	**86%**	85%	86%	84%	**81%**	78%	81%	77%	73%	65%	73%	68%
Random Forest Classifier	83%	81%	83%	80%	77%	73%	77%	72%	70%	67%	70%	61%
SVC	**86%**	85%	86%	85%	80%	77%	80%	76%	**74%**	65%	74%	69%
Multinomial NB	73%	78%	73%	62%	72%	73%	72%	61%	65%	42%	65%	51%

3.5.2 Imbalanced with 2 classes

As can be seen in Table III, all classifiers for Imbalanced gave higher metrics with 2 classes than with three classes.

Table III The Results of the Imbalanced Dataset with Two Classes

	Restaurants				Products				Movies			
	Accuracy	Precision	Recall	F1	Accuracy	Precision	Recall	F1	Accuracy	Precision	Recall	F1
Logistic Regression	**88%**	87%	88%	87%	85%	84%	85%	84%	**80%**	82%	80%	78%
Random Forest Classifier	85%	85%	85%	83%	**87%**	86%	87%	85%	71%	77%	71%	62%
SVC	**88%**	88%	88%	88%	**87%**	86%	87%	86%	79%	81%	79%	77%
Multinomia l NB	77%	82%	77%	68%	83%	83%	83%	76%	67%	45%	67%	54%

3.5.3 Balanced with 2 classes

Table IV shows the metrics after removing class 0 and balancing the data sets. Compared to the previous table, we see a decrease in some metrics and an increase in others.

Table IV The Results of Balanced Dataset

	Restaurants				Products				Movies			
	Accuracy	**Precision**	**Recall**	**F1**	**Accuracy**	**Precision**	**Recall**	**F1**	**Accuracy**	**Precision**	**Recall**	**F1**
Logistic Regression	84%	84%	84%	84%	79%	79%	79%	79%	**80%**	80%	80%	80%
Random Forest Classifier	**85%**	85%	85%	85%	76%	76%	76%	76%	79%	79%	79%	79%
SVC	**85%**	85%	85%	85%	**78%**	78%	78%	78%	78%	78%	77%	77%
Multinomial NB	**85%**	85%	85%	85%	**78%**	78%	78%	78%	76%	78%	76%	76%

3.5.4 Datasets Combination

After combining the three datasets from different domains and got a size of 7686, we can say that all the fitted classifiers provided excellent accuracy values between 82% and 84%. The models were trained only on the balanced dataset with two classes. Table V summarized the accuracy of all implemented models for each dataset and all datasets together.

Table V The Results of the Combined Balanced Dataset

	Restaurants	**Products**	**Movies**	**All datasets**
Logistic Regression	84%	79%	**80%**	83%
Random Forest Classifier	**85%**	76%	79%	82%
Multinomial NB	**85%**	78%	76%	**84%**
SVC	**85%**	78%	78%	83%
AraBert	**85%**	**80%**	70%	82%

6. Discussion

In the preceding part, is was demonstrated that the findings were highly accurate, ranging from 74% to 88% in various data sets and algorithms. This might be because the methods we used to clean and prepare the data was superior to that employed in the study (ElSahar & El-Beltagy, 2015). As shown in the Table VI, the variance in accuracy across data sets while employing TF-IDF Vectorization with neutral classification is compared.

Table VI Comparison Between Accuracy in Our Paper and (ElSahar & El-Beltagy, 2015)

Datasets	**(ElSahar & El-Beltagy, 2015)**		**Our Paper**	
	3 Classes	**2 Classes**	**3 Classes**	**2 Classes**
Restaurants	51%	76%	86%	88%
Products	47%	72%	81%	87%
Movies	34%	55%	74%	80%

It's worth mentioning that, as we mentioned before, the data was imbalanced, so we resample it and then presented it on the algorithms, and the results were also a little bit higher than traditional paper's (ElSahar & El-Beltagy, 2015) overall accuracy, as seen in the Table VII.

Table VII Comparison Between Overall Accuracy

Datasets	(**ElSahar & El-Beltagy, 2015**)	**Our Paper**
	2 Classes (imbalanced)	**2 Classes (balanced)**
Restaurants	84%	85%
Products	76%	78%
Movies	74%	80%

The contribution of our article, which is machine learning for sentiment analysis on a data set from many areas (i.e. gathering data in a single set), as well as the use of one of the deep learning models AraBERT with the Fast-BERT library, has helped to improve the accuracy result. The aggregate data accuracy varied from 82 to 84 %, with Multinomial NB having the best accuracy.

The constancy of height in accuracy for the combined data and the restaurant data set is also noteworthy. This might be due to the vast amount of data in it, which causes the system to get saturated with various sentiments and therefore appropriately assess them. The contrary is obvious in the movie data, where the accuracy fluctuated owing to a lack of data and its significant variation before to balancing, and then the high accuracy after balancing. As a result, the algorithms were able to better understand and evaluate the data as shown in following chart Figure 8.

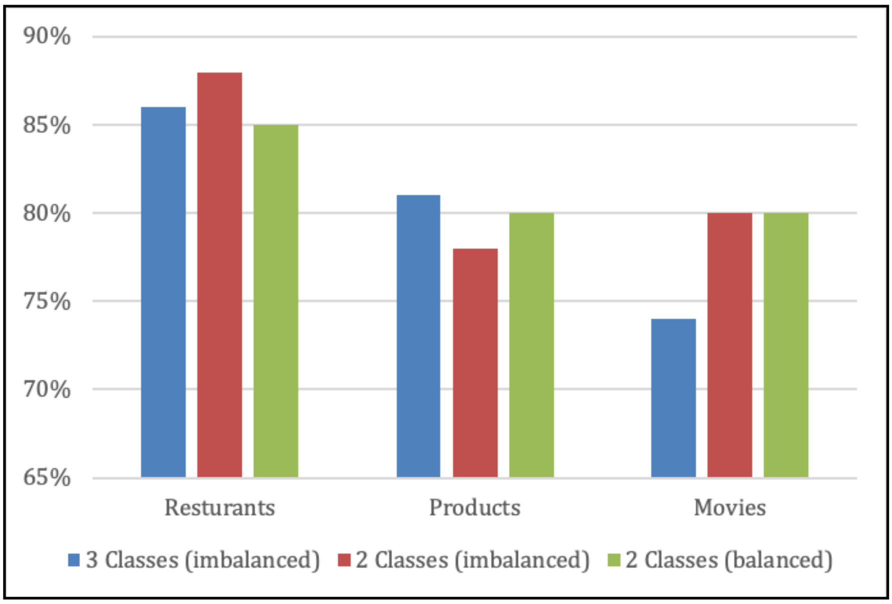

Figure 8 Accuracy Varies Depending on the Amount of Data Among Datasets.

Finally, the performance of the algorithms in sentiment analysis varied from one data set to another in our investigation, and we did not have a model that stood out from the others, despite the fact that the SVC model emerged in the publication (Nabil et al., 2015).

7. Conclusion and Future Work

In this paper, we used sentiment analysis to analyze three datasets that contain evaluations and opinions regarding (restaurants, products, and movies). These datasets were imported from a prior research in order to attempt to get high accuracy results and to experiment with different data preparation and machine learning model training techniques. The resampling of the data to balance them was one of the things that was employed as a new addition to the previous research in preparing the data. The data sets were also combined into one set, and different algorithms were tested to see how well the models performed when examined on data from different sources. To supplement studies and to experiment alternative techniques to explore and analyze sentiments, the AraBERT deep learning model with Fast-BERT library was employed. The results were more accurate than the initial study, and the models performed well with the combined dataset.

One of the difficulties that limited improved outcomes was the lack of Arab evaluations and comments, as well as the imbalance in data sets, which led to a further drop in data after the balancing procedure. The more data, obviously, the better the results and accuracy, as demonstrated by the restaurant data set, which produced the best results in every situation.

One of the most important things that we think it will lead to clear improvements and better results is to provide additional datasets to the Arab side so research can be expanded and improved. We hope to build machine learning algorithms that specialize in the Arabic language and its grammar in the future. One thing we saw was that instead of removing and discarding emojis totally, we might improve the accuracy of the models by processing and replacing them with the meanings they represent.

This research examines and analyzes attitudes from three separate and merged, balanced and unbalanced data sets. However, today's research of user and customer opinions may not be restricted to direct evaluations; instead, data from social media may be more useful. All of the aforementioned insights and limitations serve as a stepping stone for further research.

REFERENCE

1. Alharbi, A., Kalkatawi, M., & Taileb, M. (2021). Arabic Sentiment Analysis Using Deep Learning and Ensemble Methods. *Arabian Journal for Science and Engineering*, *46*(9), 8913–8923. https://doi.org/10.1007/s13369-021-05475-0
2. *ANLP_Workshop/TextMiningWorkshop*. (2022). GitHub. https://github.com/ghadamoudi/ANLP_Workshop
3. *AraBERTv2 / AraGPT2 / AraELECTRA*. (2022). [Python]. AUB MIND. https://github.com/aub-mind/arabert/blob/d5977729968f921b7463037e878142b9121bcc2e/examples/old/AraBERT_with_fast_bert.ipynb
4. *Arabic text preprocessing*. (2022). [Jupyter Notebook]. saobou. https://github.com/saobou/arabic-text-preprocessing/blob/9a801912fa1c43ab59b2430ba5003e6965dfd866/Preprocess.ipynb (Original work published 2019)
5. Contributor, T. (2021). *What is sentiment analysis (opinion mining)? Definition from SearchBusinessAnalytics*. SearchBusinessAnalytics. https://www.techtarget.com/searchbusinessanalytics/definition/opinion-mining-sentiment-mining
6. ElSahar, H., & El-Beltagy, S. R. (2015). Building Large Arabic Multi-domain Resources for Sentiment Analysis. In A. Gelbukh (Ed.), *Computational Linguistics and Intelligent Text Processing* (pp. 23–34). Springer International Publishing. https://doi.org/10.1007/978-3-319-18117-2_2
7. Folorunsho, D. (2020, October 13). *Arabic Sentiment Analysis*. Medium. https://towardsdatascience.com/arabic-sentiment-analysis-5e21b77fb5ea
8. HUANG, S., CAI, N., PACHECO, P. P., NARRANDES, S., WANG, Y., & XU, W. (2018). Applications of Support Vector Machine (SVM) Learning in Cancer Genomics. *Cancer Genomics & Proteomics*, *15*(1). https://doi.org/10.21873/cgp.20063
9. Iqbal, A., Aftab, S., & Matloob, F. (2019). *Performance Analysis of Resampling Techniques on Class Imbalance Issue in Software Defect Prediction*. 10.
10. Jain, P. K., & Pamula, R. (2020). *A systematic literature review on machine learning applications for consumer sentiment analysis using online reviews* (arXiv:2008.10282). arXiv. http://arxiv.org/abs/2008.10282
11. Kolamanvitha. (2021, May 9). Twitter sentiment analysis using Logistic Regression. *Nerd For Tech*. https://medium.com/nerd-for-tech/twitter-sentiment-analysis-using-logistic-regression-ff9944982c67
12. Liu, C., Sheng, Y., Wei, Z., & Yang, Y.-Q. (2018). Research of Text Classification Based on Improved TF-IDF Algorithm. *2018 IEEE International Conference of Intelligent Robotic and Control Engineering (IRCE)*, 218–222. https://doi.org/10.1109/IRCE.2018.8492945
13. Mohammed, R., Rawashdeh, J., & Abdullah, M. (2020). *Machine Learning with Oversampling and Undersampling Techniques: Overview Study and Experimental Results* (p. 248). https://doi.org/10.1109/ICICS49469.2020.239556
14. Nabil, M., Aly, M., & Atiya, A. (2015). LABR: A Large Scale Arabic Sentiment Analysis Benchmark. *ArXiv:1411.6718 [Cs]*. http://arxiv.org/abs/1411.6718
15. Ranjan, G. S. K., Verma, A., & Sudha, R. (2019). *K-Nearest Neighbors and Grid Search CV Based Real Time Fault Monitoring System for Industries* (p. 5). https://doi.org/10.1109/I2CT45611.2019.9033691

16. Saifullah, S., Fauziyah, Y., & Aribowo, A. S. (2021). Comparison of machine learning for sentiment analysis in detecting anxiety based on social media data. *Jurnal Informatika*, *15*(1), 45. https://doi.org/10.26555/jifo.v15i1.a20111

17. Setyanto, A., Laksito, A., Alarfaj, F., Alreshoodi, M., Kusrini, Oyong, I., Hayaty, M., Alomair, A., Almusallam, N., & Kurniasari, L. (2022). Arabic Language Opinion Mining Based on Long Short-Term Memory (LSTM). *Applied Sciences*, *12*(9), 4140. https://doi.org/10.3390/app12094140

18. Singh, R., Singh, R., & Bhatia, A. (2018). *Sentiment analysis using machine learning techniques to predict outbreaks and epidemics*.

19. Smetanin, S. (2018, September 1). *Sentiment Analysis of Tweets using Multinomial Naive Bayes*. Medium. https://towardsdatascience.com/sentiment-analysis-of-tweets-using-multinomial-naive-bayes-1009ed24276b

20. Taghizadeh-Mehrjardi, R., Schmidt, K., Eftekhari, K., Behrens, T., Jamshidi, M., Davatgar, N., Toomanian, N., & Scholten, T. (2020). Synthetic resampling strategies and machine learning for digital soil mapping in Iran. *European Journal of Soil Science*, *71*(3), 352–368. https://doi.org/10.1111/ejss.12893