Journal of Autonomous Intelligence (2024) Volume 7 Issue 3
doi: 10.32629/jai.v7i3.940

ORIGINAL RESEARCH ARTICLE

Sentiment analysis for Arabic call center notes using machine learning techniques

Abdullah Alsokkar[1], Mohammed Otair[2], Hamza Essam Alfar[2], Ahmad Yacoub Nasereddin[3], Khaled Aldiabat[4], Laith Abualigah[5,6,7,8,9,10,*]

[1] *Applied Science Private University, Amman 11931, Jordan*

[2] *Khawarizmi University Technical College, Amman 25669, Jordan*

[3] *Faculty of Business Administration, Middle East University, Amman 11831, Jordan*

[4] *Department of Management Information Systems, Ajloun National University, Ajloun 26810, Jordan*

[5] *Computer Science Department, Al al-Bayt University, Mafraq 25113, Jordan*

[6] *Department of Electrical and Computer Engineering, Lebanese American University, Byblos 13-5053, Lebanon*

[7] *Hourani Center for Applied Scientific Research, Al-Ahliyya Amman University, Amman 19328, Jordan*

[8] *Applied Science Research Center, Applied Science Private University, Amman 11931, Jordan*

[9] *School of Computer Sciences, Universiti Sains Malaysia, Pulau Pinang 11800, Malaysia*

[10] *School of Engineering and Technology, Sunway University Malaysia, Petaling Jaya 27500, Malaysia*

*** Corresponding author:** Laith Abualigah, aligah.2020@gmail.com

ABSTRACT

Call centers handle thousands of incoming calls daily, encompassing a diverse array of categories including product inquiries, complaints, and more. Within these conversations, customers articulate their opinions and interests in the products and services offered. Effectively categorizing and analyzing these calls holds immense importance for organizations, offering a window into their strengths, weaknesses, and gauging customer satisfaction and needs. This paper introduces an innovative approach to extract customer sentiments through an advanced sentiment analysis technique. Leveraging two distinct yet synergistic algorithms—Support Vector Machine (SVM) and Neural Networks (NNs)—on the Kaggle machine-learning platform, our method discerns the polarity of each note, classifying them as positive, negative, or neutral. To enhance the quality of our analysis, we employed Natural Language Processing (NLP) and a range of preprocessing tools, including tokenization. The dataset comprises three thousand notes from various telecommunication companies, authored during real call center interactions. These notes form the basis of a specialized corpus, notable for being composed in the Jordanian dialect. Rigorous training and testing procedures were conducted using this corpus. The results are notable: our proposed algorithms displayed strong performance metrics. SVM yielded a commendable accuracy rate of 66%, while NNs excelled, boasting an impressive accuracy rate of 99.21%. These achievements are substantiated by comprehensive confusion matrices. In conclusion, our research provides a novel and robust framework for customer sentiment analysis in call centers, underpinned by the fusion of SVM and NNs. This technique promises valuable insights into customer feedback, facilitating informed decision-making for businesses seeking to enhance their services and products.

Keywords: sentiment analysis; support vector machine; bidirectional long short term memory; natural language processing; Jordanian dialect; customer experience

1. Introduction

Recently, companies have become highly interested in customer opinions by following their reviews, comments, and calls. Customers have become expressing their views through many available channels such as social media platforms, websites, phone calls, and others[1]. Therefore, researchers have become very interested in the field of data mining and text classification. There is a substantial need to extract valuable

ARTICLE INFO

Received: 7 July 2023
Accepted: 21 November 2023
Available online: 9 January 2024

information in an easy, automated, and effective manner that can be relied upon, especially when we have an enormous volume of data[2]. Sentiment analysis is needed to analyze customer's opinions and reviews to extract information for decision-makers in the companies. This information will provide robust functionality for competitive analysis, marketing analysis, and detection of unfavorable rumors for risk management[3].

Sentiment analysis or opinion mining determines the expressions, evaluations, judgments, and sentiment polarity of people and their attitude and behavior about topics such as products and services. In another way, sentiment analysis in the field of determining the polarity of specific text and classifies this text as: positive, negative, or neutral[4]. The purpose of sentiment analysis is to extract polarity, feelings, behavior, and feedback of people, which is difficult and needs a lot of research and attention. In other words, Sentiment analysis or opinion mining is the field of how to determine the expressions, evaluations, judgments, and sentiment polarity of people and their attitude and behavior about topics such as products and services.

Detecting opinions polarity helps in analyzing competition and marketing. Machine learning techniques are way better than traditional ways such as surveys and questionnaires. Detecting opinions using sentiment analysis became the focus of researchers' concern around the world; it will have a great interest in the future[5]. One of the most used channels that expressed customer's opinions is call center phone calls; therefore, thousands of calls are received by call centers. These calls are divided into several categories, including questions about products or services, complaints, etc. In these calls, customers are expressing their opinions and their interests in the products and services.

Companies need to categorize these calls; the results of this categorization will make their decision easier. This categorization is considered an essential indicator for decision-makers to know the company's strengths and weaknesses, besides see the level of customer satisfaction and their needs. The collected dataset made up of call center notes provides a vast amount of raw data that can be processed to generate valuable and structured information. This process needs to apply automated techniques such as natural language processing (NLP) techniques and pre-processing tools to clean the dataset and make it ready to use by machine learning techniques, classify the notes, and extract the polarity from each one of them. Sentiment analysis and its tools and applications have attained significant attention and popularity in the last few years. Thus, they are needed in all fields and domains. Several sentiment analysis applications focused on English, while only a few research focused on the Arabic language. In this research, the focus was on Arabic Jordanian sentiment analysis.

Call centers receive thousands of calls every day, making an enormous volume of data that needs to be processed. This process is strenuous to achieve manually. Therefore, sentiment analysis techniques are required to analyze such notes and detecting their polarity as: positive, negative, and neutral. This analysis is necessary for companies to help decision-makers know the strengths and weaknesses of products and services and facilitate and develop their services. This paper focuses on call center notes in Jordan, written as received, in Arabic Jordanian dialect. Each dialect has its unique words and idioms. In the Jordanian dialect, some numerous words and expressions need more focus from researchers. After further investigation, it was figured that the Arabic language and Arabic dialects lack in research, especially in the Sentiment Analysis (SA) field[6].

English language and other dialects take the more significant share of research. Thus, this paper brings attention to the needed information to construct a focused model of idioms and expression in Jordanian dialect using machine learning techniques, such as Support Vector Machine (SVM) and Bidirectional Long Short Term Memory (BiLSTM). This paper focuses on sentiment analysis in the Arabic language by applying sentiment analysis techniques on phone call notes to determine their polarity and classify them. This process plays a significant role in helping decision-makers know the Pros and cons of their products and services; it gives them a considerable indicator to improve and develop their products and services. This paper uses machine learning techniques and NLP tools to classify and detect the polarity of the Arabic Jordanian dataset to help researchers find an Arabic Jordanian dialect corpus, which contains a set of Arabic Jordanian words and idioms that trained using machine learning techniques. The main contributions of this paper are presented as follows:

- Detecting customer's opinions using sentiment analysis to create an indicator to decision-makers in order to enhance the process of customer services.
- Evaluating and comparing classification techniques, in addition to compare the classification process with and without applying preprocessing techniques.
- Providing an Arabic Jordanian corpus available online.

The motivation behind the work is to address the challenge faced by call centers, where they receive a large number of calls daily, each falling into different categories such as product inquiries, complaints, and more. During these calls, customers express their opinions and interests in the products and services offered by the company. It's crucial for companies to categorize and analyze these calls to gain insights into their strengths, weaknesses, customer satisfaction levels, and needs.

The rest of this paper is organized as follows. Section 2 present the related works. Section 3 presents the proposed sentiment analysis technique. Section 4 presents the results and discussion. The conclusion and future works are given in section 5.

2. Related works

In this section, the related works are presented to show the most relevant research in this area.

Majumder et al.[7] proposed a hybrid technique extract the features from the text, in addition to text classification, the used algorithms in this research were BiLSTM and CNN, they focused on predicting and classifying the text from Chinese news. They applied pre-processing tools such as stop word removal, tokenization, and punctuation removal, comparing to our research, the pre-processing tools didn't involve stop word removal, and punctuation removal. They used CNN to feature extraction, which is not included in this paper as well, the dataset in their research concluded 74,000 notes, they classify it into 14 different categories according to the news field, their technique achieved accuracy rate of 91.56% for BiLSTM.

Researchers such as Abandah et al.[8] classified the Indian's people sentiment about Covid-19 pandemic and lockdown in India using sentiment analysis techniques, they have collected their data from Twitter for over a period of three months of lockdown, they used twitter API to collect the dataset, then they applied natural language processing tools such as removing unneeded characters, which is used in this paper, in addition to using Panda library in order to cleaning the dataset. The used algorithm to classify the dataset was SVM, their technique achieved accuracy of 91.5%.

Kumar et al.[9] studied the Arabic poetry and the 16 Arabic meters and the importance and need to know the poem's meter and digitize it to chant the poem are studied in the right way. They proposed a solution for Arabic poetry discretization and classification using machine learning. They developed an Arabic poetry comprehensive dataset by adding prose to it. Their dataset divided to make the process of classification easier, they used recurrent neural network RNN algorithm, the proposed algorithm discovered the whole classes and got excellent results with accuracy (97.27%), it worth to mention the verses with diacritics get better and more accurate results than those without diacritics as they found.

Zhan et al.[10] used sentiment analysis to know and conclude the influence of users' feelings based on gender and age. They collected the dataset from Facebook by users' answers about their favorites Books in addition to their genders and ages. In this study, the researchers present a careful analysis through comprehensive investigations on the influence of users' expression, then they reviewed and compared various machine learning techniques. They applied pre-processing tools such as tokenization, transmuting smiley faces into proper sentiment, stop words removal, and converting the data into lowercase. Then they used Support Vector Machine, Naïve Bayes, and Maximum Entropy to apply Bag of words in Feature extraction. In addition to applying Word2Vec using Neural Network.

Samuels and Mcgonical[11] developed a sentiment analysis system to explore and detect customer feedback and reviews to help companies know the strengths and weaknesses of products and services and reserve their reputation in the market, using sentiment analysis and natural language processing techniques. They used the (http://forum.lendacademy.com) website to collect their data set, which consists of more than 4000 reviews; they concluded that the negative posts get more reviews. Hence, decision-makers should pay attention to it. Shaukat et al.[12] used sentiment analysis and opinion mining techniques to display consumers' reviews and their feedback about the products. They collected their data from the Amazon website, more than five hundred reviews were collected, and their reviews were about products such as electronics, laptops, flash memories, and mobile phones. They applied polarity classification and part of speech techniques with detailed explanations.

Rane and Kshatriya[13] applied sentiment analysis techniques to classifying movie reviews into (positive, negative, and neutral) reviews using an artificial neural network (ANN). They collected their dataset from the IMDB website and Stanford University. Their methodology included a sentiment lexicon, a bag of words, part of speech, and semantic relationship. They focused on sentiment analysis problems and challenges; in conclusion, their approach achieved accurate results. Lakshmi Devi et al.[14] used sentiment analysis to build their approach, which concerns audio to classify services and products. They believed that their approach would help decision-makers know their products' level and make the right decisions to enhance the company's performance. Yin et al.[15] proposed a system to classify movie reviews using sentiment analysis techniques. They used Decision Tree and Naïve Bayes algorithms in their approach reviews divided into positive, negative, and neutral. They used training and testing datasets to build the classifiers. Their system got accurate results in reviews ranking.

Arulmurugan et al.[16] mentioned coronavirus pandemic in their study, how this pandemic affected the entire world, and how governments took policies and means to prevent the spread of this disease. They mentioned the importance of detecting the feelings of people and their rational health to decision-makers. Social networks became the place where people are expressing their feelings. The authors in this study proposed a sentiment analysis system to detect posts about Covid-19. They collected about thirteen million tweets about coronavirus and classified them into positive and negative using a lexicon-based sentiment analysis tool called Valence Aware Dictionary and Sentiment Reasoner (VADER).

Some researchers using sentiment analysis to classify the whole document as one piece to get better precision[17] illustrated the SA approach to extract the polarity of a specific subject in the document as positive or negative. The polarity of a particular subject is different from comprehensive documents, such as when we say: "this network is fast, but it's expensive", this sentence has two sentiments: This network is fast (which is positive sentiment), but it is expensive, (which is negative sentiment). So they tried to extract the view of each one. Opinion mining has become an essential domain for researchers. Most researchers conducted in the English language. On the contrary, few researchers deal with the Arabic language. Nowadays, people express their opinions freely on social media. One of the most popular platforms for Arab social media users is Twitter; they use it frequently to give their views on various topics.

Kumar and Garg[18] detected sentiment analysis and sarcastic text in their research, using neural network classifier. Their dataset contains about one thousand notes; their proposed method exceeds the other techniques. They noted that combined ways better than separate sarcasm and sentiment classifiers. Also, to obtain good results, they concluded that detecting sentiment analysis and detecting sarcastic text are similar and relevant. Mammadli et al.[19] build multimedia sentiment analysis systems to discover the polarity of tweets, and they focused on images, texts, typography, and infographic. Senti-Bank used to score image sentiment, and scoring (RCNN), Regions with convolution neural network done using Senti-Strength. The hybrid methodology used to sentiment the text, which consists of both machine learning and lexicon techniques, got more than 90% accuracy, which is an outstanding performance recognized for the casual multimedia tweet data used to estimate the recommended method.

Cunha et al.[20] studied the importance of sentiment analysis and how it became the focus of scientists' attention in recent years. A few studies in this research area in the Azerbaijani language collected about 30 thousand essays from different news websites. Then, they annotated each one of them into positive and negative depending on their sentiment. After that, they applied pre-processing tools to facilitate the classification process and make the text more structured. They then applied a bag of words techniques like frequency model and TF-IDF. The classifiers used in their approach were Support Vector Machine, Naïve Bayes, and Random Forest. They concluded that SVM had achieved the best results.

YouTube became one of the most popular and used websites in the whole world. Millions of people use it every day; these people express their feelings and give their feedback by leaving their comments on those videos. Farisi et al.[21] proposed a sentiment analysis system to classify these comments into positive, negative, and neutral words. They used ANN in their proposed system. They used comments of 2 videos in their dataset to test their design; their proposed approach outperforms the manual classification against expectations with an accuracy of about 80%. Rahab et al.[22] proposed a sentiment analysis system to help hotels administrations to disclose and follow-up customer's feedback about their experiments in hotels, this feedback considered very important and adequate to the hotel status in the market so that the hotel's administration can know the strengths and weaknesses of their services, and other customers take it in account to decide when they want to use one of these hotels. The researchers in this study collected about five thousand reviews about hotels, labeling step done manually. They applied the Naïve Bayes algorithm to classify these reviews into negative and positive; pre-processing tools are used in addition to featuring selection and extraction. The study got accurate and good results. Noor et al.[23] proposed a method to detecting and classifying text through sentiment analysis technologies. They focused on the Arabic language, and in particular Algerian dialect. They mentioned the increasing importance of seeing public opinion to know people's directions and feelings and how these opinions become existing in social networks and websites. They collected their dataset from newspaper articles written in the Arabic language and located in Algeria, OCA and SANA corpus utilized to manage the dataset. The dataset labeling was done manually by people who are using Algerian language as their mother language. They used KNN, SVM, and Naïve Bayes algorithms. They got accurate and good results. They mentioned that KNN over performed the rest of the used algorithms.

Zaw and Tandayya[24] mentioned the considerable expansion that is taking place in the world of online shopping, and importance of websites that provide these services, and how it facilitates the process of shopping in addition to saving efforts and time, hence it saves money as well. They mentioned the importance of detecting customers' feedback about the products they buy from such websites because customers used to express their input through websites and other customers' interest in these reviews to benefit from others' experience. Therefore, this research proposed a system to explore and follow-up customer's feedback about online products. They classified it into positive, negative, or neutral classes using a support vector machine algorithm. They collected more than 20 thousand reviews, their dataset from different online shopping websites. Their dataset was in Urdu Roman. The labeling step was done manually.

To help the organizations develop making their decisions and enhance their customers' satisfaction, Appel et al.[25] used a rule-based algorithm named sentiment analysis contrast rule-based to discover the reviews of customers by classifying them automatically. They proposed an algorithm to extract sentiment data automatically, they collected their dataset from Amazon, and they compared their system with the Senti Strength tool. The results showed that their system is better than Senti Strength, and it is reliable in consumer relief and satisfaction detection. Ahmad et al.[26] developed crossbred system for sentiment analysis, using linguistic, Variables, Grammatical Rules, and sentiment dictionary or lexicon. Their datasets consist of tweets and reviews about movies. Their approach achieved better and more accurate results than machine learning classifiers based on supervised algorithms such as Maximum Entropy and Naïve Bayes after comparing the dataset with their proposed system.

The importance of sentiment analysis evolving, different reasons making this field of study one of the most used in the research area, such as the massive volume of available online data that need to reveal, discover, follow-up, and analyzed. This process could be utterly impossible if we talk about applying it manually. Therefore, sentiment analysis is fundamental since it enables us to detect people's feelings and feedback from their text. Chao[27] sought to examine and review the efficiency of the SVM algorithm through classifying tweets. The classification process aims to discover and extract each tweet's direction, which is divided into (negative, positive, or neutral). They collected two different datasets from Twitter, about 10 thousand tweets. The tweets were already annotated. They used the WEKA tool to achieve the classification process and test the algorithm execution, and then they compared the results of their dataset using (F-Measure, Recall, and Precision). The results showed efficiency and accurate results of using machine learning algorithms in sentiment analysis.

In this paper, the focus was on the Arabic language and the Jordanian dialect, which needs more research due to the lack of previous studies. Through the mentioned previous studies, the focus in this paper was on the best classification algorithms and using preprocessing tools since they are enhancing the accuracy of the classification process. The proposed technique in this paper contributes to developing the Arabic Jordanian dialect corpus. It will be uploaded online to help other researchers in sentiment analysis in Arabic Jordanian dialect because most previous studies focused on English. This paper concluded that BiLSTM and SVM are considered the best algorithms in text classification and BiLSTM achieved better results than SVM in text classification. This draws the researchers' attention to the BiLSTM algorithm's focus and uses it in their future research.

3. The proposed method

This paper's proposed technique is to classify the dataset by applying two existing machine learning algorithms: Bidirectional Long Short Term Memory (BiLSTM) and Support Vector Machine (SVM), besides, to compare the results of each one of them. The development of the classification process was positive, negative, and neutral. These results can help the company's focus on the issues related to products and services to avoid them in the future, besides to know the positive notes and improve them. There are four steps in the proposed technique of this paper, the first one is to prepare the data, the second one is data annotation to use in the machine learning algorithms, the third one is the preprocessing step, and the last one is text classification by applying SVM algorithm and Neural Networks. **Figure 1** shows the model of this method.

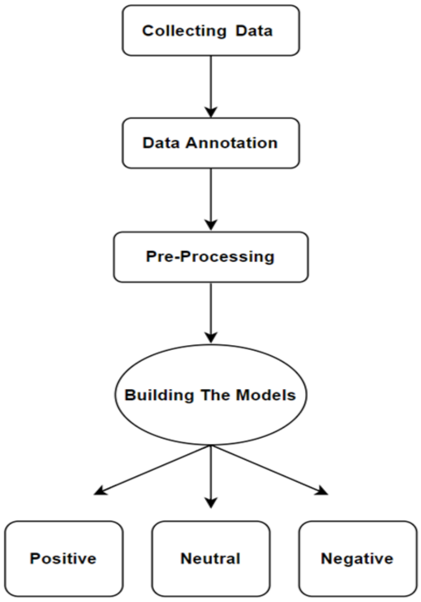

Figure 1. The proposed Model.

Both models were implemented using Python 3 because python allows developers to write reliable machine learning systems efficiently and quickly; different packages were used for each model. For data preparation, NumPy, pandas, and transformers were used. SciKit learn used to implement, train, and evaluate the SVM. For the RNN (BiLSTM), Keras from Tensorflow was used.

It is worth mentioning that BiLSTM is a sequence processing model that consists of two LSTMs: one taking the input in a forward direction and the other in a backward direction. BiLSTM effectively increases the amount of information available to the network, improving the algorithm's content. SVM is an algorithm that determines the best decision boundary between vectors that belong to a given group and vectors that do not belong to it. It can be applied to any vectors which encode any data. The proposed technique of this paper was done using Kaggle. It is an online community of data scientists and machine learning practitioners from Google. It is a diversified data-science platform that provides opportunities for students, researchers, and companies from all over the world to solve their problems related to machine learning and predictive analytics. It provides training, education, and solving real-life problems for companies through cash prizes or job opportunities for its members.

Data scientists can get priceless experience from Kaggle, and it prepares them to understand what goes into finding suitable solutions for big data. It provides the ability to write machine learning codes, run them, and share them, in addition to hosting thousands of available datasets that could be used to train and test machine learning techniques. Therefore, Kaggle was used to upload the dataset and to apply every step in this paper using Python 3.

3.1. Data preparation

In this paper, 3000 notes have been collected by the author from different telecommunication companies. The dataset was from letters that call center employees wrote; these notes were written during the customer's calls with the call center. Call center employees summarize the calls that they receive as notes. It is worth mentioning that these notes are written in the Arabic Jordanian dialect specifically. Thousands of calls are received daily by the call center. This paper used a sample of these notes and applied the proposed technique to detect their polarity. The dataset has been divided into three categories based on their contradictions: positive, negative, and neutral.

3.2. Data annotation

As mentioned before, the dataset has been divided into three different categories as positive, negative, and neutral. The dataset has been annotated by experts who are qualified to annotate each note with its suitable category. Each of the three categories has been linked with a number to facilitate the classification process as 0 for neutral, 1 for positive, and 2 for negative. The annotation process was repeated another time by another group to ensure that each note is connected with the correct category. **Table 1** shows some examples of the collected dataset after the annotation process from the three different categories:

Table 1. Dataset example.

Note	Polarity
اشتريت حزم و النت مش شغال	2
في عندي حزمة مش عارف كيف الغيها	2
في رسائل بضل توصلني و انا ما اشتركت فيها و بدي الغيها	2
ضفت حزمة و خلصت و انا ما لحقت استخدمها	2
بحاول اجدد اشتراكي و مش زابط معي	2
اشتريت بطاقة و مش راضية تشحن	2
بحاول احكي مكالمة عم بضل يفصل	2
في حزمة مش عارف الغيها بصير تلغيلي إياها من عندك	0
بدي إعدادات الانترنت	0
بدي ألغي خدمة صوتك عكيفك	0
سرعة الانترنت صارت منيحة بتلاع العلي	1
العروض الجديد مشجعة	1
المعرض حلولي مشكلتي	1
التطبيق سهل علينا كثير	1

3.3. Preprocessing

Pre-processing is particularly an essential step in sentiment analysis to make the text more appropriate and coordinated to make the text representation easier. In other words, text preprocessing is the process of cleaning the data and make it ready to use in machine learning techniques[28]. Text preprocessing includes several methods such as stemming and stop word removal etc. In this paper, the dataset consists of Arabic notes collected during calls by call center staff. These notes are written in a hurry and in an abbreviated way; therefore, it will probably be like raw data that has additions such as symbols, unneeded numbers, etc. Thus, to enhance the process of the classification and prediction, all non-Arabic characters have been removed from the dataset, except for the question marks, because it sometimes indicates that the note could be neutral since it's used with questions. **Table 2** show some examples of the dataset before and after applying preprocessing by removing unneeded characters:

Table 2. Pre-processing examples.

Before pre-processing	After pre-processing
!!!في عندي حزمة مش عارف كيف الغيها	في عندي حزمة مش عارف كيف الغيها
شكاوي// بحاول اجدد اشتراكي و مش زابط معي	بحاول اجدد اشتراكي و مش زابط معي
!اذا شحنت بطاقة -20- دينار اكم بتعطيني صلاحية	اذا شحنت بطاقة دينار اكم بتعطيني صلاحية

Another essential step in preprocessing is tokenization. Tokenization is converting a sentence to tokens represented by numbers. It is a crucial step in preprocessing. The used tokenizer in this paper was the ArabicBERT tokenizer proposed by Afifi et al.[29].

3.4. Text classification

In this step, text classification algorithms were applied, using the SVM algorithm and Neural Network (BiLSTM), in addition to comparing the results of each one of them after training and testing each model. According to several studies, SVM and Neural Networks (BiLSTM) are considered the best algorithms in text classification. They showed accurate results by Rahat et al.[30]. Therefore, these algorithms were used in this paper.

- Support Vector Machine (SVM)

Support vector machines is a supervised machine learning algorithm, which works with classification problem, and regression. SVM should train from labelled data to classify new data, SVM works on the training and testing principle with classification problems. Based on previous experiences, SVM obtained accurate results, and it is considered as one of the best algorithms in machine learning[31]. SVM algorithm determines the better decision boundary between vectors that belong to a category and vectors that do not belong to it. It can be applied to any kind of vectors which encode any type of data.

SVM begins the work by looking at the dividing line that divides data into different categories, this line named a hyper plane, which is presented as follows:

$$w.x + b = 0 : x_i \in R \quad (1)$$

The following equation divides the data into different categories:

$$g(x) = sign(w.x + b) \quad (2)$$

SVM use this rule to divide the hyper plane using the data point to deduce the maximum margin x_i. As shown in **Figure 2**, the data is far from the hyper plane, so this will make the right decision in classifying the data.

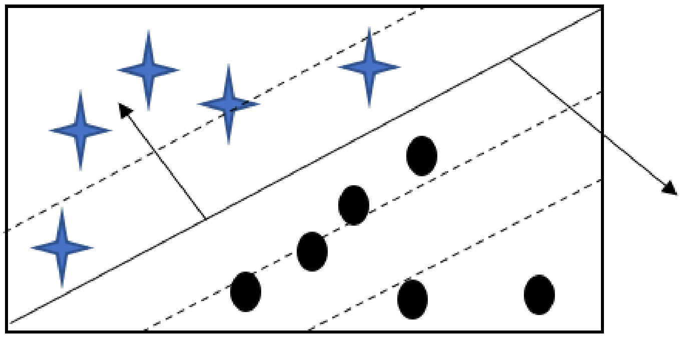

Figure 2. The hyperplane in SVM.

- Neural Networks (BiLSTM):

Recurrent neural network (RNN) is a type of neural network algorithm that is useful in designing sequence data. Obtained from forward network, Recurrent Neural Network exhibits attitude like human brain work. In other words: RNN predict results in sequential data, while other algorithms cannot do. Bidirectional recurrent neural network proposed by Galdi and Tagliaferri[32], in order to find a solution to issues that need utilization in the next and previous context. The bi-directional recurrent neural network layer contains two networks, each layer is unidirectional ($x1, 2, ..., xN$). The forward direction is through one sequence to train the front network, and the back direction is through another sequence to train the back network ($xN, N-1, ..., x1$), bidirectional layer produces forward networks and back networks as a group of outputs.

Bidirectional Long-Short Term Memory Networks (BiLSTM) is the type of recurrent neural network (RNN) which means the signal spreads backward and forward in time. In this paper, the proposed technique applying these two algorithms to classifying customers' notes is written as text. The meaning of text classification is to train a set of data. The machine can essentially learn the boundaries that split the group of data. Therefore, new data inputted to the model can be classified based on where the point exists.

4. Experiments and results

This paper aims to extract the polarity of notes that call center employees to write as a summary of customers' calls with the call center. To classify these notes, Recurrent Neural Networks (RNN), Bidirectional Long Short Term Memory (BiLSTM), and Support Vector Machine (SVM) algorithms have been applied. It is worth mentioning that the classifier's result was measured by accuracy by dividing the correctly predicted notes by the total number of the notes. Accuracy has the following definition as taken from studies by Abuaddous et al.[33], Al Sokkar and Law[34], and Almajali et al.[35]:

$$Accuracy = \frac{Number\ of\ correct\ predictions}{Total\ number\ of\ predictions} \tag{3}$$

The notes that were collected from call center notes contain different categories. Each one of the notes has been classified manually. Each note has been matched with its suitable category as positive, negative, and neutral by expert call center employees. This manual classification of the notes is based on their experience in this field. The dataset was divided into training, validation, and testing as 80%, 10%, and 10%. The largest portion of the dataset was used for training to train the classifier as much as possible. The reason for splitting the dataset is to train the classifier using a part of the data. Then the classifier is tested using the remaining data, which the classifier has never prepared; because of invalidation and testing steps, the classifier needs to be tested by new data that has never been tried before. The dataset consists of different numbers of words for each note, as two words for the shortest note and 19 words for the longest note. The number of characters was ten characters for the most concise note and 108 characters for the longest note. **Figure 3** shows the count distribution of sentences words used in the dataset. **Figure 4** shows the count distribution of sentences characters used in dataset.

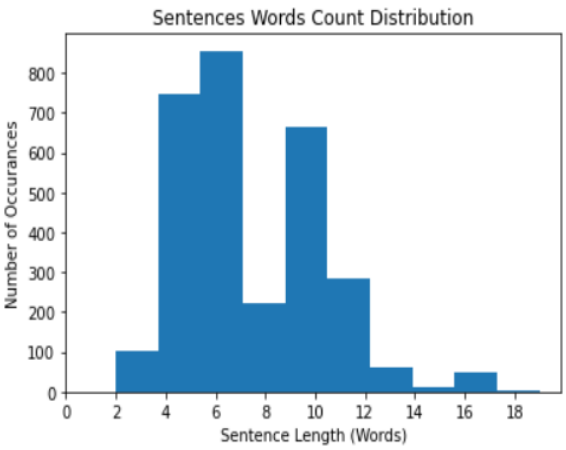

Figure 3. Sentences words count distribution.

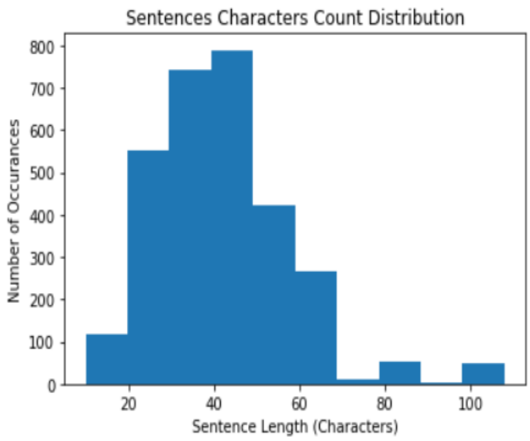

Figure 4. Sentences characters count distribution.

4.1. Accuracy results

Many experiments were conducted in order to obtain the best results on both algorithms, so that the proposed models were trained several times with applying pre-processing tools and without, in addition to use the dataset with numbers, and without in order to get the best accuracy of results. The results of training the two models show that the RNN (BiLSTM) algorithm outperformed the SVM algorithm, with 99.12% accuracy for RNN (BiLSTM), and 66.00% for SVM. The reason of this difference between the accuracy of the models is that SVM needs bigger size of dataset to train, while BiLSTM can train with such size of dataset. According to several previous studies, SVM achieved better accuracy results with big size of dataset comparing to the used dataset in this paper.

It is worth mentioning that applying pre-processing techniques contributed to enhancing the results. Accuracy results without applying pre-processing techniques were as follows: 97.60% for BiLSTM and 60.67% for SVM. Additional experiment was test by evaluating the proposed technique without removing the numbers from the dataset, the accuracy results are as follows: BiLSTM: (testing: 99.67%, validation: 100%, training: 99.33%), SVM: (testing: 68.10%, validation: 65.66%, training: 63.47%). Accuracy results for each model after applying pre-processing techniques described in details as follows: BiLSTM Accuracy: **Figures 5–7** show the accuracy results, loss, and the time taken to complete the evaluation process of the BiLSTM model for testing, validation, and training, respectively, after train the model and evaluated in Kaggle:

```
In [14]: model.evaluate(test_texts, test_labels)

10/10 [==============================] - 0s 26ms/step - loss: 0.0048 - accuracy: 1.0000

Out[14]: [0.004833649378269911, 1.0]
```

Figure 5. BiLSTM test accuracy.

```
In [15]: model.evaluate(val_texts, val_labels)

10/10 [==============================] - 0s 24ms/step - loss: 0.0021 - accuracy: 1.0000

Out[15]: [0.0020635202527046204, 1.0]
```

Figure 6. BiLSTM validation accuracy.

```
In [16]: model.evaluate(train_texts, train_labels)

75/75 [==============================] - 2s 29ms/step - loss: 0.0282 - accuracy: 0.9912

Out[16]: [0.028246574103832245, 0.9912499785423279]
```

Figure 7. BiLSTM training accuracy.

SVM Accuracy: Figures 8–10 show the accuracy results, loss, and the time taken to complete the evaluation process of the SVM model for testing, validation, and training, respectively, after train the model and evaluated in Kaggle.

```
In [17]: print("svm test accuracy:")
         svm.score(test_texts, test_labels) #Evaluate on test split

svm test accuracy:

Out[17]: 0.7441860465116279
```

Figure 8. SVM test accuracy.

```
In [16]: print("svm validation accuracy:")
         svm.score(val_texts, val_labels) #Evaluate on validation split

svm validation accuracy:

Out[16]: 0.72
```

Figure 9. SVM validation accuracy.

```
In [15]: print("svm train accuracy:")
         svm.score(train_texts, train_labels) #Evaluate on train split

svm train accuracy:

Out[15]: 0.66
```

Figure 10. SVM train accuracy.

Table 3 shows the result of SVM and RNN classifiers measured by accuracy. The results of the two used models show that the accuracy of validation and testing were better than the effects of training. In this case, the reason for this result is that the dataset size is relatively small. Hence the size of testing and validation of the dataset is small as well, and the dataset with numbers enhanced the accuracy. **Figure 11** shows the accuracy of the trained models.

Table 3. The accuracy results of the classifiers before and after pre-processing.

	Before pre-processing			**After pre-processing**			**After pre-processing with numbers**		
Model	Train	Validation	Test	Train	Validation	Test	Train	Validation	Test
BiLSTM	97.60%	91.35%	59.05%	99.12%	100%	100%	99.67%	100%	99.33%
SVM	60.67%	70.19%	39.04%	66.00%	72.00%	74.41%	68.10%	65.66%	63.47%

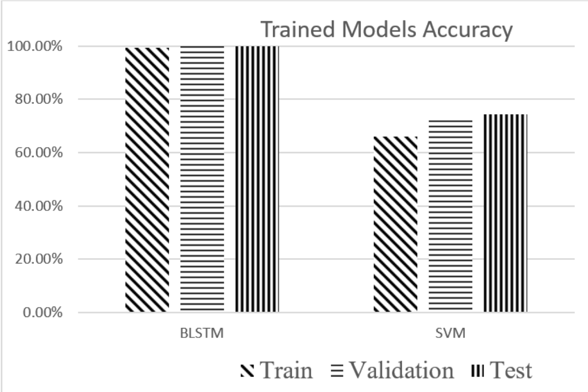

Figure 11. The accuracy of trained model.

4.2. Confusion matrix

Additional measures used to evaluate the models show the true positive, true negative, false positive, and false negative for each algorithm using a confusion matrix. BiLSTM Confusion Matrix for neutral expected 179 true notes and one note was incorrect, and all positive and negative notes were expected truly (8 notes for positive, and 113 for negative). **Table 4** shows the confusion matrix of the BiLSTM model.

Table 4. Confusion matrix of BiLSTM.

		0	1	2
Expected classes	0	179	0	1
	1	0	8	0
	2	0	0	113
	Predicted classes			

SVM Confusion Matrix for neutral expected 150 true notes and 46 notes were false: (7 as positive, 39 as negative), and expected one-note truly for positive. Five notes were false (3 as positive and two negatives). For negative notes, 71 notes were expected truly, 26 were false as neutral. **Table 5** shows the confusion matrix of SVM model.

Table 5. Confusion Matrix of SVM.

		0	1	2
Expected classes	0	150	7	39
	1	3	1	2
	2	26	0	71
	Predicted classes			

Here are some potential limitations that the author may need to consider and discuss in their research:

Generalization to Other Dialects or Languages: The proposed sentiment analysis technique was applied to notes written in the Jordanian dialect. One limitation could be the generalization of this approach to other dialects or languages, as the linguistic characteristics may differ significantly, affecting the model's performance.

Data Quality and Quantity: The success of machine learning algorithms often relies on the quality and quantity of the training data. If the dataset used in this research is limited in size or contains noisy or biased data, the model's performance may not generalize well to a broader context.

5. Conclusion and future works

In this paper, the sentiment analysis technique was used to classify the customers' calls with the call center written by call center employees from different telecommunication companies. These calls contain the customer's opinions about the products and services provided by the company. It is essential to analyze customer's behavior through their views during these calls. The dataset contains three different classes: positive, negative, and neutral. In this paper, the collected dataset was classified to detect each note's polarity; the dataset contains about 3000 notes. These notes were compiled from accurate data that were telecommunication companies used to summarize the customers' calls with the call center. Natural language processing tools were applied, such as removing unneeded characters. The proposed technique applied the classification process using two algorithms: BiLSTM, and SVM, to detect the polarity of each note. The results show that the training model got better accuracy results with BiLSTM compared with SVM. Several experiments were conducted in this paper, including applying the classification algorithms for the dataset without using preprocessing tools to remove unneeded characters. The results showed that applying preprocessing tools obtain much better accuracy results. Besides, the expansion of the dataset from 1000 notes to 3000 notes received more accurate results. Hence, the more data, the more the accuracy. In conclusion, it is worth mentioning that using preprocessing tools is a crucial step in the machine learning process. Another finding is that the outputs of this classification are considered a vital indicator to decision-makers. It will help them focus on the most critical weaknesses through negative notes to avoid them in the future and improve and develop strengths through the positive notes. It is worth mentioning that the difference between the two models' accuracy is that SVM needs a more extensive dataset to train, while BiLSTM can train with such size of the dataset. In this paper, after classifying call center notes using machine learning techniques, the results showed that the classification process was efficient, accurate, and fast. It contributes to the operation of text classification. In future work, there are several suggestions for improving this technique. The first one is to expand the dataset to get better and more accurate results. The second suggestion is to integrate this technique with other solutions, such as a dictation system, that converts customer's calls directly into the text to facilitate the process of collecting the data, since the current way of many companies is to write the notes manually during the calls as notes, this traditional way requires time and effort, in addition to that there is a possibility for human errors. The dictation system will make the process of collecting the notes faster and smoother. The last suggestion is to use additional preprocessing tools that may get better results in the classification process and classify the data into different categories based on its content or field.

Author contributions

Conceptualization, AA, MO, HEA, AYN, KA and LA; methodology, HEA; software, HEA; validation, HEA; formal analysis, HEA; investigation, HEA; resources, HEA; data curation, HEA; writing—original draft preparation, AA, MO, HEA, AYN, KA and LA; writing—review and editing, AA, MO, HEA, AYN, KA and LA; visualization, HEA; supervision, HEA; project administration, HEA. All authors have read and agreed to the published version of the manuscript.

Informed consent

Informed consent was obtained from all individual participants included in the study.

Data availability statements

Data is available from the authors upon reasonable request.

Conflict of interest

The authors declare no conflict of interest.

References

1. Prabaningtyas NI, Surjandari I, Laoh E. Mining Customers Opinion on Services and Applications of Mobile Payment Companies in Indonesia Using Sentiment Analysis Approach. 2019 16th International Conference on Service Systems and Service Management (ICSSSM); July 2019. doi: 10.1109/icsssm.2019.8887643
2. Hartama D, Perdana Windarto A, Wanto A. The Application of Data Mining in Determining Patterns of Interest of High School Graduates. Journal of Physics: Conference Series. 2019, 1339(1): 012042. doi: 10.1088/1742-6596/1339/1/012042
3. Sadr H, Pedram MM, Teshnehlab M. A Robust Sentiment Analysis Method Based on Sequential Combination of Convolutional and Recursive Neural Networks. Neural Processing Letters. 2019, 50(3): 2745-2761. doi: 10.1007/s11063-019-10049-1
4. Yadav A, Vishwakarma DK. Sentiment analysis using deep learning architectures: a review. Artificial Intelligence Review. 2019, 53(6): 4335-4385. doi: 10.1007/s10462-019-09794-
5. Gamal D, Alfonse M, El-Horbaty ESM, Salem ABM. Twitter Benchmark Dataset for Arabic Sentiment Analysis. International Journal of Modern Education and Computer Science. 2019, 11(1): 33-38. doi: 10.5815/ijmecs.2019.01.04
6. Wang M, Cai Q, Wang L, et al. Chinese news text classification based on attention-based CNN-BiLSTM. In: Liu Z, Udupa JK, Sang N, Wang Y (editors). MIPPR 2019: Pattern Recognition and Computer Vision. 2020. doi: 10.1117/12.253813
7. Majumder S, Aich A, Das S. Sentiment Analysis of People During Lockdown Period of COVID-19 Using SVM and Logistic Regression Analysis. SSRN Electronic Journal. 2021. doi: 10.2139/ssrn.3801039
8. Abandah GA, Khedher MZ, Abdel-Majeed MR, et al. Classifying and diacritizing Arabic poems using deep recurrent neural networks. Journal of King Saud University - Computer and Information Sciences. 2022, 34(6): 3775-3788. doi: 10.1016/j.jksuci.2020.12.002
9. Kumar S, Gahalawat M, Roy PP, et al. Exploring Impact of Age and Gender on Sentiment Analysis Using Machine Learning. Electronics. 2020, 9(2): 374. doi: 10.3390/electronics9020374
10. Zhan G, Wang M, Zhan M. Public Opinion Detection in an Online Lending Forum: Sentiment Analysis and Data Visualization. 2020 IEEE 5th International Conference on Cloud Computing and Big Data Analytics (ICCCBDA); April 2020. doi: 10.1109/icccbda49378.2020.9095690
11. Samuels A, Mcgonical J. Sentiment Analysis on Customer Responses. arXiv. 2020, arXiv:2007.02237. doi: 10.48550/arXiv.2007.02237
12. Shaukat Z, Zulfiqar AA, Xiao C, et al. Sentiment analysis on IMDB using lexicon and neural networks. SN Applied Sciences. 2020, 2(2). doi: 10.1007/s42452-019-1926-x
13. Rane AL, Kshatriya AR. Audio Opinion Mining and Sentiment Analysis of Customer Product or Services Reviews. ICDSMLA 2019. 2020, 282-293. doi: 10.1007/978-981-15-1420-3_30
14. Lakshmi Devi B, Varaswathi Bai V, Ramasubbareddy S, et al. Sentiment Analysis on Movie Reviews. Advances in Intelligent Systems and Computing. 2020, 321-328. doi: 10.1007/978-981-15-0135-7_31
15. Yin H, Yang S, Li J. Detecting Topic and Sentiment Dynamics Due to COVID-19 Pandemic Using Social Media. Lecture Notes in Computer Science. 2020, 610-623. doi: 10.1007/978-3-030-65390-3_46
16. Arulmurugan R, Sabarmathi KR, Anandakumar H. RETRACTED ARTICLE: Classification of sentence level sentiment analysis using cloud machine learning techniques. Cluster Computing. 2017, 22(S1): 1199-1209. doi: 10.1007/s10586-017-1200-1
17. Majumder N, Poria S, Peng H, et al. Sentiment and Sarcasm Classification With Multitask Learning. IEEE Intelligent Systems. 2019, 34(3): 38-43. doi: 10.1109/mis.2019.2904691
18. Kumar A, Garg G. Sentiment analysis of multimodal twitter data. Multimedia Tools and Applications. 2019, 78(17): 24103-24119. doi: 10.1007/s11042-019-7390-1
19. Mammadli S, Huseynov S, et al. Sentiment Polarity Detection in Azerbaijani Social News Articles. Proceedings - Natural Language Processing in a Deep Learning World; 22 October 2019. doi: 10.26615/978-954-452-056-4_082
20. Cunha AAL, Costa MC, Pacheco MAC. Sentiment Analysis of YouTube Video Comments Using Deep Neural Networks. Lecture Notes in Computer Science. 2019, 561-570. doi: 10.1007/978-3-030-20912-4_51
21. Farisi AA, Sibaroni Y, Faraby SA. Sentiment analysis on hotel reviews using Multinomial Naïve Bayes classifier. Journal of Physics: Conference Series. 2019, 1192: 012024. doi: 10.1088/1742-6596/1192/1/012024
22. Rahab H, Zitouni A, Djoudi M. SANA: Sentiment analysis on newspapers comments in Algeria. Journal of King Saud University-Computer and Information Sciences. 2021, 33(7): 899-907. doi: 10.1016/j.jksuci.2019.04.012
23. Noor F, Bakhtyar M, Baber J. Sentiment Analysis in E-commerce Using SVM on Roman Urdu Text. Emerging Technologies in Computing. 2019, 213-222. doi: 10.1007/978-3-030-23943-5_16
24. Zaw M, Tandayya P. Multi-level Sentiment Information Extraction Using the CRbSA Algorithm. 2018 15th International Joint Conference on Computer Science and Software Engineering (JCSSE); July 2018. doi: 10.1109/jcsse.2018.8457328
25. Appel O, Chiclana F, Carter J, et al. Successes and challenges in developing a hybrid approach to sentiment analysis. Applied Intelligence. 2017. doi: 10.1007/s10489-017-0966-4
26. Ahmad M, Aftab S, Ali I. Sentiment Analysis of Tweets using SVM. International Journal of Computer

Applications. 2017, 177(5): 25-29. doi: 10.5120/ijca2017915758
27. Chao Z. RETRACTED: Research on English translation long text filtering based on LSTM semantic relevance. Microprocessors and Microsystems. 2021, 80: 103574. doi: 10.1016/j.micpro.2020.103574
28. Safaya A, Abdullatif M, Yuret D. KUISAIL at SemEval-2020 Task 12: BERT-CNN for Offensive Speech Identification in Social Media. Proceedings of the Fourteenth Workshop on Semantic Evaluation. 2020. doi: 10.18653/v1/2020.semeval-1.27
29. Afifi S, GholamHosseini H, Sinha R. Dynamic hardware system for cascade SVM classification of melanoma. Neural Computing and Applications. 2018, 32(6): 1777-1788. doi: 10.1007/s00521-018-3656-1
30. Rahat AM, Kahir A, Masum AKM. Comparison of Naive Bayes and SVM Algorithm based on Sentiment Analysis Using Review Dataset. 2019 8th International Conference System Modeling and Advancement in Research Trends (SMART); November 2019. doi: 10.1109/smart46866.2019.9117512
31. Schuster M, Paliwal KK. Bidirectional recurrent neural networks. IEEE Transactions on Signal Processing. 1997, 45(11): 2673-2681. doi: 10.1109/78.650093
32. Galdi P, Tagliaferri R. Data Mining: Accuracy and Error Measures for Classification and Prediction. Encyclopedia of Bioinformatics and Computational Biology. 2019, 431-436. doi: 10.1016/b978-0-12-809633-8.20474-3
33. Abuaddous HY, Al Sokkar AA, Abualodous BI. The Impact of Knowledge Management on Organizational Performance. International Journal of Advanced Computer Science and Applications. 2018, 9(4). doi: 10.14569/ijacsa.2018.090432
34. Al Sokkar A, Law E. Validating an episodic UX model on online shopping decision making. Proceedings of the 5th ACM SIGCHI symposium on Engineering interactive computing systems; 24 June 2013. doi: 10.1145/2494603.2480301
35. Almajali DA, Omar F, Alsokkar A, et al. Enterprise resource planning success in Jordan from the perspective of IT-Business strategic alignment. Cogent Social Sciences. 2022, 8(1). doi: 10.1080/23311886.2022.2062095