

Article

Evaluating User Satisfaction Using Deep-Learning-Based Sentiment Analysis for Social Media Data in Saudi Arabia's Telecommunication Sector

Majed A. Alshamari

Department of Information Systems, College of Computer Sciences and Information Technology, King Faisal University, Hofuf 31983, Saudi Arabia; smajed@kfu.edu.sa

Abstract: Social media has become common as a means to convey opinions and express the extent of satisfaction and dissatisfaction with a service or product. In the Kingdom of Saudi Arabia specifically, most social media users share positive and negative opinions about a service or product, especially regarding communication services, which is one of the most important services for citizens who use it to communicate with the world. This research aimed to analyse and measure user satisfaction with the services provided by the Saudi Telecom Company (STC), Mobily, and Zain. This type of sentiment analysis is an important measure and is used to make important business decisions to succeed in increasing customer loyalty and satisfaction. In this study, the authors developed advanced methods based on deep learning (DL) to analyse and reveal the percentage of customer satisfaction using the publicly available dataset AraCust. Several DL models have been utilised in this study, including long short-term memory (LSTM), gated recurrent unit (GRU), and BiLSTM, on the AraCust dataset. The LSTM model achieved the highest performance in text classification, demonstrating a 98.04% training accuracy and a 97.03% test score. The study addressed the biggest challenge that telecommunications companies face: that the company's services influence customers' decisions due to their dissatisfaction with the provided services.



Citation: Alshamari, M.A. Evaluating User Satisfaction Using Deep-Learning-Based Sentiment Analysis for Social Media Data in Saudi Arabia's Telecommunication Sector. *Computers* **2023**, *12*, 170. <https://doi.org/10.3390/computers12090170>

Academic Editors: Phivos Mylonas, Katia Lida Kermanidis and Manolis Maragoudakis

Received: 26 June 2023

Revised: 15 August 2023

Accepted: 23 August 2023

Published: 26 August 2023



Copyright: © 2023 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: user satisfaction; human–computer interaction; deep learning algorithms; sentiment analysis; social media; artificial intelligence; telecommunication

1. Introduction

Measuring users' satisfaction is a critical part of assessing successful interaction between humans and technologies. The telecommunications industry has emerged as a prominent sector in developed nations. The escalation of competition has been propelled by the proliferation of operators and advancements in technology [1]. Enterprises are implementing diverse tactics to sustain themselves in this highly competitive marketplace. According to the extant literature [2], three principal strategies have been proposed to augment revenue generation: (1) procuring a new clientele, (2) upselling to the extant clientele, and (3) prolonging the retention duration of the clientele. Upon analysing these strategies while considering their respective return on investment (RoI), it has been determined that the third strategy yields the greatest financial benefit [2]. This discovery corroborates the idea that maintaining an existing customer is more cost effective than obtaining a new one [3] and is also regarded as a less complicated tactic than the upselling technique [4]. To execute the third strategy, corporations must address the potential occurrence of customer churn, which describes the phenomenon of customers transitioning from one provider to another [5].

The pursuit of customer satisfaction is a key driver for telecommunication companies in the face of intense global competition. Numerous studies have established a positive correlation between customer satisfaction and both customer loyalty and customer churn [6–8].

The phenomenon of customer churn is characterised in the telecommunications industry as the act of customers switching from one telecommunications service provider to another [9]. According to recent research, the expense associated with acquiring a new customer surpasses that of retaining an existing customer [10]. Currently, corporations exhibit a heightened level of interest in retaining their clientele. As demonstrated in the literature review, a multitude of investigations have been carried out in various sectors pertaining to customer relationship management (CRM) to handle customer retention and develop a proficient framework for forecasting customer churn.

Customer involvement is a crucial aspect of the operations of diverse small, medium, and large enterprises (SMEs). The success or failure of businesses or industries can be influenced by various factors, such as customer relations, loyalty, trust, support, feedback, opinions, surveys, and other forms of commentary, either independently or in conjunction with one another. Comprehending the requirements and comfort levels of customers holds great importance in both commercial and individual-based sectors, particularly in terms of customer satisfaction during or after service consumption. Researchers have used a range of techniques to obtain and anticipate customer feedback, such as social media platforms, electronic questionnaires, telephone calls, email correspondence, online mobile applications, and websites [8]. Through the incorporation of customer feedback, comments, advice, suggestions, recommendations, and viewpoints, it is feasible to augment and broaden the calibre and volume of services [11,12].

Telecommunications is a vital global industry that has the potential to make a significant impact across various sectors, including business, defence, investment, production, and individual domains. The provision of a swift, dependable, protected, and precise service has the potential to enhance the calibre of service offered by the relevant communication enterprises. Thus, the anticipation of customer feedback holds significant importance for the progress of the nation. Several techniques from statistics, computer science, and theoretical and mathematical fields have been suggested and simulated to precisely forecast customer satisfaction. This is carried out to enhance service quality in accordance with the requirements and expectations of customers [13,14].

In Saudi Arabia, customer feedback holds significant importance for both government and private sector companies. The establishment of diverse departments to oversee and address customer grievances across multiple industries through varied approaches and mediums has fostered a highly competitive telecommunications market [15]. The telecommunications sector in Saudi Arabia is currently experiencing notable changes in various aspects, including technological innovations, service provision, a competitive landscape, and the extension of telecommunications services to non-traditional domains. The services mentioned above include managed infrastructure, data/colocation centres, and cloud services. Some of the most prominent telecommunication firms operating in Saudi Arabia are the STC, Integrated Telecom Company (ITC), Saudi Mobily Company (Etisalat), Zain, Virgin, and Go Telecom [16]. Saudi Arabia is one of the most densely populated nations in the Gulf Cooperation Council (GCC) region, with a demographic composition that is predominantly youthful. Emerging nations exhibit a strong inclination towards the adoption and application of cutting-edge technology in various domains, such as education, research, commerce, manufacturing, and more. The uncertain business situations of the future have been amplified by the high-speed 5G network and the COVID-19 pandemic, as stated in [17]. In 2021, the 5G awards were bestowed upon STC, Mobily, and Zain. According to [18], approximately 11 million individuals use the Twitter platform via both smartphones and computers. The user base is observed to be expanding at a swift pace owing to the growing population and interest [18].

Our study centred on the assessment and examination of the efficacy of a collection of deep learning (DL) models in predicting customer attrition in the telecommunications industry. Various algorithms, including long short-term memory (LSTM), gated recurrent unit (GRU), BiLSTM, and convolutional neural networks (CNN) with LSTM (CNN-LSTM), were used to develop methods for data preparation, feature engineering, and feature selection.

This research contributes to the domain of customer satisfaction analysis by using Arabic tweets about Saudi telecommunications companies. It demonstrates the ability of several models using DL, including LSTM, GRU, BiLSTM, and CNN-LSTM, to predict customer satisfaction. The significance of social media as a platform where customers may express their positive and negative experiences with telecommunications services and products was further confirmed. The study's findings have real-world relevance for Saudi Arabia's telecommunications sector because they shed light on customer satisfaction and reveal opportunities for service enhancement. This information can inform business decisions, reduce customer churn due to dissatisfaction, and enhance customer service and loyalty.

The present study is organised as follows: The literature review thoroughly examines the pertinent research within the discipline. The methodology section comprehensively describes the dataset and the model architectures utilised. The section dedicated to experimental results comprehensively examines the obtained findings and their subsequent analysis. Finally, the study concludes by engaging in a comprehensive discussion and providing a conclusive summary.

2. Background of Study

Various methodologies have been used to forecast customer attrition in telecommunications firms. The majority of these methodologies employ machine learning (ML) and data mining techniques. The predominant body of literature has centred on the implementation of a singular data-mining technique for knowledge extraction, while alternative studies have prioritised the evaluation of multiple approaches for the purpose of churn prediction.

In their study, Brandusoiu et al. [19] introduced a sophisticated data-mining approach to predict churn among prepaid customers. This approach involved the use of a dataset containing call details for 3333 customers, which included 21 distinct features. The dependent churn parameter in this dataset was binary, with values of either 'Yes' or 'No'. The features encompass details pertaining to the quantity of incoming and outgoing messages as well as voicemail for individual customers. The PCA algorithm was used by the author to perform dimensionality reduction on the data. The research used three discrete ML algorithms—specifically neural networks, a support vector machine (SVM), and Bayes networks—to predict the churn factor. The author evaluated the algorithms' performance using the area under the curve (AUC) as a metric. The present study involved the computation of the area under the receiver operating characteristic curve (AUC-ROC) for three distinct ML models: Bayes networks, neural networks, and SVM. The AUC values acquired were 99.10%, 99.55%, and 99.70%, respectively. The current study used a restricted dataset that was free from any instances of missing data. He et al. [20] proposed a model that employed the neural network algorithm to tackle the problem of customer churn in a large telecommunications company in China that had a customer base of around 5.23 million. The metric used to assess the precision of predictions was the general accuracy rate, which yielded a score of 91.1%. Idris [21] addressed the issue of churn in the telecommunications industry by presenting a methodology that employed genetic programming alongside AdaBoost. The efficacy of the model was evaluated using two established datasets: one from Orange Telecom and the other from cell2cell. The cell2cell dataset achieved an accuracy rate of 89%, while the other dataset achieved a rate of 63%. Huang et al. [22] investigated customer churn within the context of the big data platform. The aim of the researchers was to exhibit noteworthy enhancement in churn prediction by leveraging big data, which is dependent on the magnitude, diversity, and speed of the data. The handling of data derived from the Operation Support and Business Support divisions of the largest telecommunications corporation in China required the implementation of a big data platform to enable the requisite manipulations. The use of the random forest algorithm was evaluated using the AUC metric.

A rudimentary set theory-based churn prediction model was proposed by Makhtar et al. [23] for the telecommunications sector. The rough set classification technique outper-

formed the linear regression, decision tree, and voted perception neural network methods, as indicated in the aforementioned research. The problem of skewed data sets in churn prediction has been the subject of several studies. When the number of churned client classes falls below the number of active customer classes, this phenomenon occurs. In their research, Amin et al. [24] compared six alternative oversampling strategies in the context of telecommunication churn prediction. The results showed that genetic algorithm-based rules-generation oversampling algorithms outperformed the other oversampling techniques evaluated.

Burez and Van den Poel [25] investigated the issue of imbalanced datasets in churn prediction models. They conducted a comparative analysis of the efficacy of random sampling, advanced undersampling, gradient-boosting models, and weighted random forests. The model was evaluated using metrics such as AUC and Lift. The findings indicate that the undersampling technique exhibited superior performance compared to the other techniques tested. Individuals who use social media platforms, including Twitter, Facebook, and Instagram, tend to provide commentary and evaluations regarding a company's offering because these platforms provide a means for users to express their opinions and exchange ideas concerning products [9]. The process of sentiment analysis, also referred to as feedback mining, involves the use of natural language processing (NLP), statistical analysis, and ML to extract and classify feedback from textual inputs based on criteria such as subjectivity and polarity recognition [6]. Individuals who use social media platforms, including but not limited to Twitter, Facebook, and Instagram, have been observed to provide commentary and evaluations regarding a company's offerings because these platforms provide an avenue for individuals to express their viewpoints and exchange their perspectives on products [9]. The process of sentiment analysis, also referred to as feedback mining, involves the use of NLP, statistical analysis, and ML to extract and categorise feedback from textual inputs based on factors such as subjectivity and polarity recognition [6]. Furthermore, Pavaloaia and colleagues provided a concise definition of sentiment analysis as a social media tool that entails evaluating the presence of positive and negative keywords in text messages linked to a social media post [10].

Recognition of the need for sentiment analysis is increasing [9,25]. This is attributed to the growing demand for the estimation and organisation of unstructured data from social media. The task of text mining is challenging because it involves the identification of topical words across various subjects. To effectively categorise these words into either positive or negative polarity, it is imperative to conduct sentiment analysis. Additionally, selecting appropriate sentiment signals for real-time analysis is crucial in this process [26,27]. The increasing prevalence of textual content sharing on social media has led to an increase in the use of text-mining and sentiment analysis techniques [28–30].

The study conducted by [31] involved an analysis of consumer sentiment expressed in a Jordanian dialect across the Facebook pages of multiple telecommunication businesses in Jordan, as well as on Twitter. The four fundamental classifiers used for the manual categorisation of all the gathered and processed attitudes are the SVM, K-nearest neighbour (k-nn), naïve Bayesnaïve (), and decision tree (DT). The present study used its results to exhibit the superiority of SVM over three other widely used sentiment classifiers. In [27], the researchers aimed to ascertain the sentiment of user-generated content but were constrained to classifying comments instead of the actual posts.

Furthermore, [32] employed Twitter as a medium for conducting sentiment analysis by scrutinising tweets in the English language originating from diverse businesses in Saudi Arabia. The researchers used K-nearest neighbour and naïve Bayes algorithms to classify attitudes into three categories: positive, negative, and neutral. These classifications were based on their daily and monthly trend observations. Furthermore, the K-nearest neighbour algorithm, an ML methodology, was employed to examine user sentiment in the present investigation. Nonetheless, the exclusion of Arabic opinions may have resulted in a less comprehensive dataset.

The study used a sentiment analysis of Facebook posts as a means of assessing the efficacy of social media posts in supporting effective self-marketing strategies on social media platforms. A reference for this study is available. Furthermore, according to a study conducted by [33], the implementation of sentiment analysis results in a rise in negative sentiment among followers during phases of reduced user-generated activity. This is noteworthy, as sentiment analysis consistently yields supplementary insights beyond those derived from solely analysing comments, likes, and shares of articles. Research has demonstrated that a single published article has the potential to generate a substantial number of comments, which can be subjected to sentiment analysis using an ML-based approach.

The researchers [34–36] used a range of deep learning approaches to establish the correlation between several organizations and their clients, drawing from feedback, quality assessments, comments, and surveys conducted across many domains. In the field of natural language processing (NLP), these three approaches have garnered significant interest because of their exceptional accuracy in text categorization analysis. These methods have shown to be indispensable in many sectors, including commercial and consumer interactions, as well as in predicting societal implications on future trends. The user has provided a numerical sequence.

3. Materials and Methods

Measuring people’s opinions about a product or service is significant for assessing the percentage of customer satisfaction with the service provided by the provider. At present, social media has become a means of expressing opinions and, so, these data were collected and analysed to contribute to decision making that led to an increase in the percentage of customer satisfaction and loyalty. To analyse a large number of these opinions, this study presents DL models that analysed the negative and positive feelings expressed through people’s opinions about telecommunication services in Saudi Arabia. Figure 1 displays the framework of the proposed system. Our research method began with pre-processing the data to clean them up and remove irrelevant data, and then applying several DL models, such as convolutional neural networks, LSTM and BiLSTM, GRU, and CNN-LSTM, to compare the classification results. This methodology is applicable to many languages, and only the pre-processing process differed from one language to another.

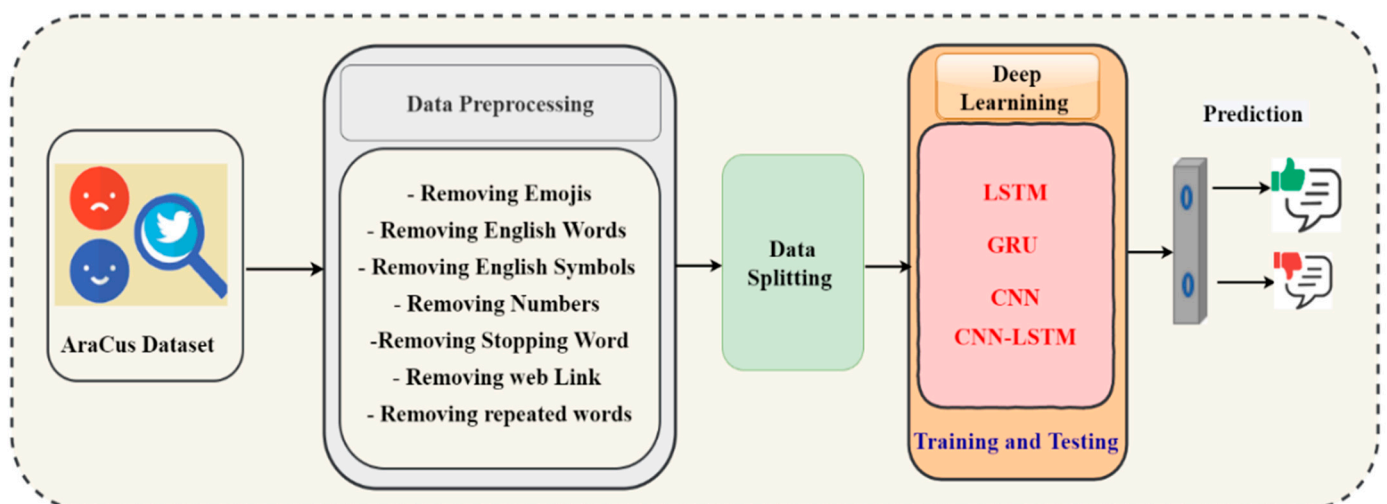


Figure 1. Framework of the proposed methodology.

3.1. Dataset

This study used the publicly available dataset AraCust [37], which was collected via Twitter. The data collection period was from January to June 2017. It included data such as negative and positive opinions about the services of three telecommunication services

in Saudi Arabia—STC, Zain, and Mobily—as shown in Figure 2. It consisted of around 20,000 tweets. Table 1 details the number of reviews for each company: STC had 7590, Mobily had 6450, and Zain had 5950.

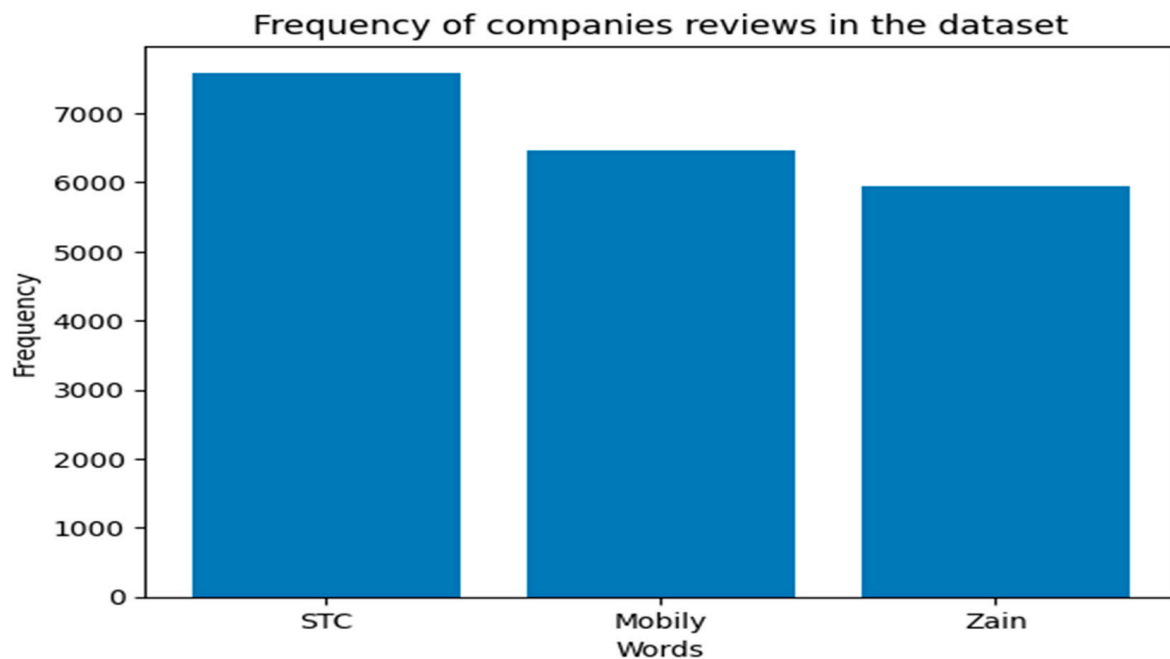


Figure 2. Samples from the AraCust dataset.

Table 1. Samples from the AraCust dataset.

| Company | Number of Tweets |
|---------|------------------|
| STC | 7590 |
| Mobily | 6450 |
| Zain | 5950 |
| Total | 20,000 |

3.2. Data Exploration

Data exploration aims to analyse customer opinions from the AraCust dataset to better understand the sentiment frequencies represented for each service provider. This study was primarily interested in counting the number of positive and negative instances in the dataset. The authors could better analyse and interpret the data by studying these sentiment frequencies.

3.2.1. Represented Sentiment Analysis of Customer Satisfaction

Table 2 provides information about the sentiment frequencies for three telecommunication providers: STC, Mobily, and Zain. Each provider had two sentiment categories, ‘negative’ and ‘positive’, which indicated neither satisfaction nor dissatisfaction. The numbers in the table represent the frequencies or counts of each sentiment category for each provider. For STC, there were 5065 occurrences of negative sentiments and 2524 of positive sentiments (Figure 3a). Similarly, for Mobily, there were 4530 occurrences of negative sentiments and 1930 occurrences of positive sentiments (Figure 3b). Lastly, for Zain, there were 3972 occurrences of negative sentiment and 1978 occurrences of positive sentiment (Figure 3c).

Table 2. Sentiment frequencies for STC, Mobily, and Zain.

| Provider | Negative | Positive |
|----------|----------|----------|
| STC | 5065 | 2524 |
| Mobily | 4530 | 1930 |
| Zain | 3972 | 1978 |

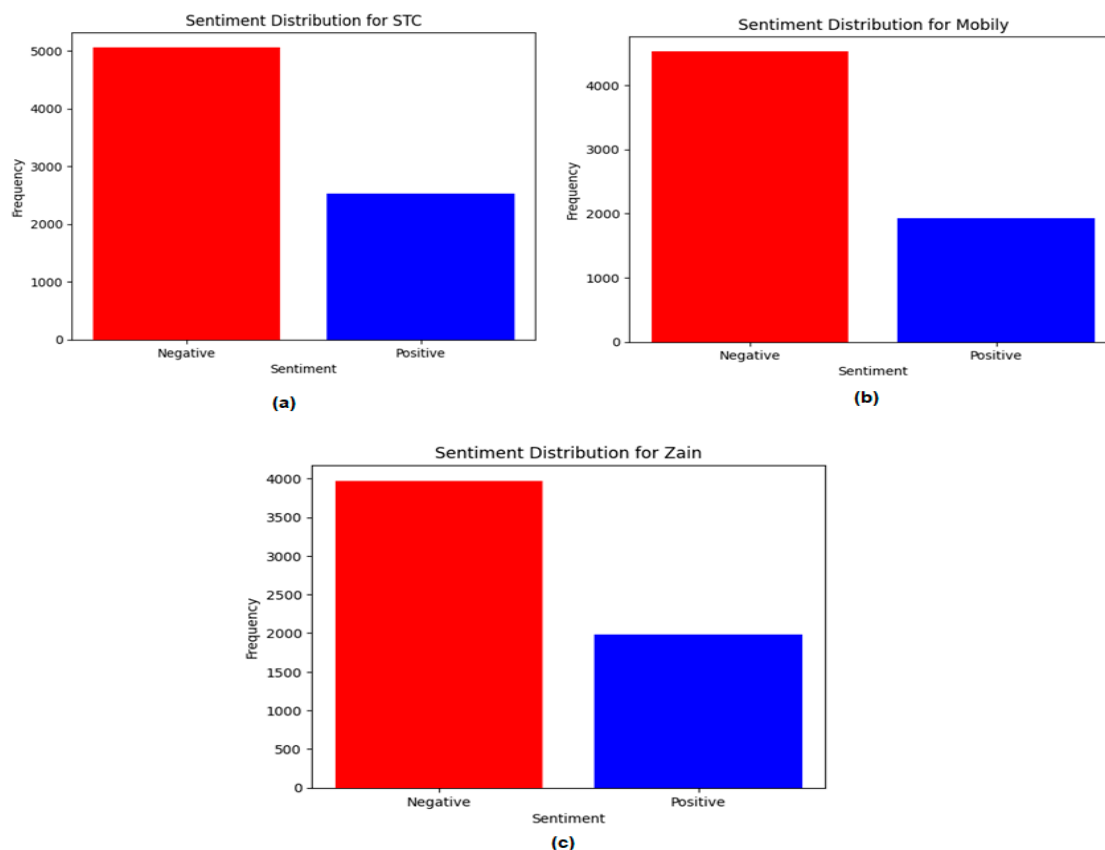
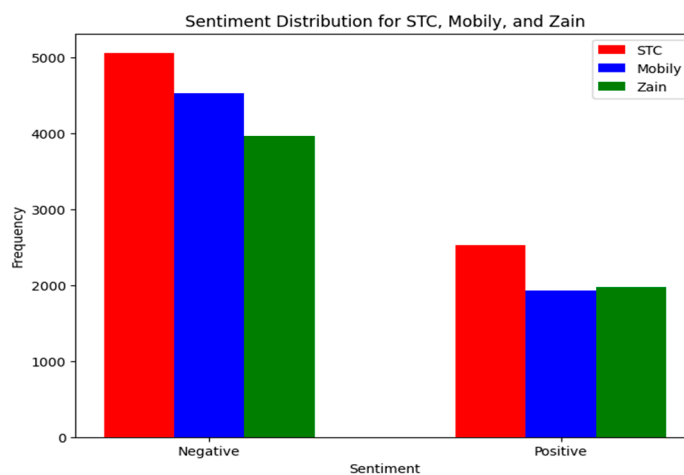
**Figure 3.** Positive and negative sentiments for (a) STC, (b) Mobily, and (c) Zain.

Table 2 and Figure 4 permit a quick comparison of sentiment frequencies between the three telecommunication providers, providing insights into the overall sentiments associated with each company.

**Figure 4.** Comparison of the positive and negative sentiments for each company.

3.2.2. Arabic Bigrams

Bigrams can be used for Arabic linguistic structures. They demonstrate Arabic text structure and meaning and capture the relationship between adjacent words [38]. Bigrams are essential to text analysis and NLP. Frequency, co-occurrence, and context are ways to analyse Arabic bigrams. Language modelling, sentiment analysis, and machine translation involve identifying positive or negative sentiment patterns and capturing language-specific dependencies and collocations. Bigrams are needed to improve Arabic machine translation, sentiment analysis, and linguistic models. Table 3 shows the Arabic bigram of the AraCust dataset.

Table 3. Summary of the Arabic bigrams.

| Bigram (Arabic) | Frequency | Bigram (English Translation) |
|---------------------|-----------|------------------------------|
| ('عليكم', 'السلام') | 972 | ('Peace', 'be upon you') |
| ('من', 'أكثر') | 386 | ('More', 'than') |
| ('علي', 'اتصل') | 316 | ('Call', 'me') |
| ('مشكله', 'عندي') | 289 | ('I have', 'a problem') |
| ('بيانات', 'شريحه') | 258 | ('Data', 'SIM card') |
| ('الله', 'حسي') | 232 | ('Allah is', 'sufficient') |
| ('العملاء', 'خدمه') | 225 | ('Customer', 'service') |
| ('لكم', 'شكرا') | 217 | ('Thank you', 'all') |
| ('المشكلة', 'حل') | 189 | ('Problem', 'solution') |
| ('سمحت', 'لو') | 179 | ('If', 'you please') |

3.3. Data Preprocessing

In this subsection, we detail the preprocessing steps that were taken to prepare the dataset to be passed into DL models for training and testing.

3.3.1. Data Cleaning

Raw data-processing is crucial when working with any kind of data [39]. It is also vital when using a sentiment extraction method because the extracted text needs high-quality data for producing accurate simulation results, such as emojis, English words, English symbols, and mobile numbers. Table 4 shows the basic NLP steps applied in this dataset, with some random examples.

Table 4. Summary of the data preprocessing process.

| Operation | Original | Cleaned | Translate to English |
|---|---|--|--|
| Removing emojis | @STC1100 شكرًا 🙏 | شكرا | Thank you |
| Removing English symbols and removing English words | @STC_KSA جميل | جميل | Beautiful |
| Removing mobile No/ID/numbers | @Mobily1100 الله لا يوفقكم | الله لا يوفقكم | May God not grant you success |
| Remove tashkeel | جداً | جدا | Very |
| Removing emojis, English words, English symbols | @Mobily ماشاء الله تبارك الله جودة شبكة ممتازة 🙏🙏 | ماشاء الله تبارك الله جوده شبكه ممتازة | MashaAllah, may Allah bless you Excellent network quality. |
| Remove hashtags and web links | بالتوفيق MySTC تطبيق https://t.co/I0Tv083lYr (accessed on 22 May 2023) | بالتوفيق تطبيق | Good luck with the application. |

3.3.2. Data Tokenisation

Tokenisation is essential to pre-processing text data because it converts unstructured text into ML models [40]. Tokenisation breaks text into smaller tokens for better analysis. The tokeniser creates a vocabulary or word index for each text token. Common tokens have lower vocabulary indices. Once vocabulary is established, tokenisers encode text sequences into sequences of word indices for ML algorithms. Tokenisers allow developers to use DL models for sentiment analysis, text classification, and language generation. Tokenising words across texts makes finding patterns and insights in text data easier.

Various preprocessing strategies were applied to the AraCust dataset that reduced the Arabic text to its essential linguistic nature. This improved dataset can be used for model training and evaluation in the fields of Arabic sentiment analysis, machine translation, named-entity identification, and more.

3.3.3. Label Encoder

Label encoder instances are needed to assign numeric values to categorical class labels [38]. The encoder was applied to the dataset's target column to encode class labels. The negative was repeated as 0, while the positive was repeated as 1. These initial steps prepare data for ML model training.

3.4. Data Splitting

The data were tokenised and then split 80:20 between the training and testing sets. Here, 80% of the data were used in the training phase and 20% were reserved for testing. This was carried out to see how well the models performed on data they had never seen before and to ensure that the models could generalise beyond the samples used for training. The researcher could evaluate the efficacy of a model on a test set and then decide if it was suitable for use in real-world scenarios.

3.5. Deep Learning Algorithms

DL is a subfield of ML that concentrates on developing artificial neural networks to make predictions or decisions by processing huge amounts of data [41]. The NLP field has garnered considerable interest due to its focus on the interaction between computational systems and human language [42]. The AraCust initiative is a scholarly endeavour that integrates DL and NLP techniques to construct an Arabic sentiment analysis framework. The objective is to develop models using extensive Arabic textual data to effectively categorise sentiment, thereby furnishing significant perspectives for commercial enterprises, social media scrutiny, and customer response evaluation.

3.5.1. Recurrent Neural Networks

Recurrent neural networks (RNNs), a type of DL model, are used in NLP applications because they can process sequential data. They use a 'hidden state' to store information from previous time steps and influence future predictions. RNNs can process inputs of various lengths, making them suitable for sentiment analysis, machine translation, text generation, and named-entity recognition [43]. Advanced RNN architectures avoid the vanishing gradient problem. These architectures can store and update data longer.

3.5.2. Long Short-Term Memory Network

LSTM architecture is a subtype of RNNs developed to address the challenges of capturing long-term dependencies in sequential data [44]. Due to their constructed memory cells and gating mechanisms, LSTMs can selectively remember and update data at different time steps. An LSTM cell comprises four basic elements: the cell state, input gate, forget gate, and output gate. The LSTM architecture's vanishing gradients allow it to store and recall long-term dependencies. Language modelling, machine translation, sentiment analysis, and speech recognition are some of the NLP tasks that have significantly benefited from using LSTMs.

The mechanical equations for the LSTM are as follows:

Input gate:

$$i_t = \sigma(W_{ix_t} + U_{ih_{t-1}} + b_i) \quad (1)$$

Forget gate:

$$f_t = \sigma(W_{fx_t} + U_{fh_{t-1}} + b_f) \quad (2)$$

Memory cell:

$$c_t = f_t * c_{t-1} + i_t * \tanh(W_{cx_t} + U_{ch_{t-1}} + b_c) \quad (3)$$

Output gate:

$$o_t = \sigma(W_{ox_t} + U_{oh_{t-1}} + b_o) \quad (4)$$

Hidden state:

$$h_t = o_t * \tanh(c_t), \quad (5)$$

where x_t is the current input, h_t is the hidden state, c_t is the memory cell state, i_t is the input, f_t is the forget state, o_t is the output gate, and W , U and b are the network weights and biases. Sigma and tanh represent the sigmoid and hyperbolic tangent activation functions, respectively.

In this study, an LSTM model was developed for binary text classification tasks. The model's architecture consisted of the input, hidden, and output layers, as shown in Figure 5. First, the model's embedding layer was built. The embedding layer converted text into numerical form for the neural network. The tokeniser's index-word dictionary's mappings between word indices and words determined vocabulary size. The embedding layer's 64 dimensions enabled dense vector representations of vocabulary words. Then, three LSTM layers were added, the first layer with 512 and the second with 128 units, both returned sequences = True, which returned all previous outputs from these LSTM layers. The third layer was 64 units LSTM. Finally, a dense layer of two neurons completed the model. In this dense layer, all neurons communicated with each other and with the layer below. This layer used the sigmoid function for binary classification because it worked well. Each class's likelihood was compressed to 0 and 1, representing the negative and positive; for more details of the model parameters, see Table 5.

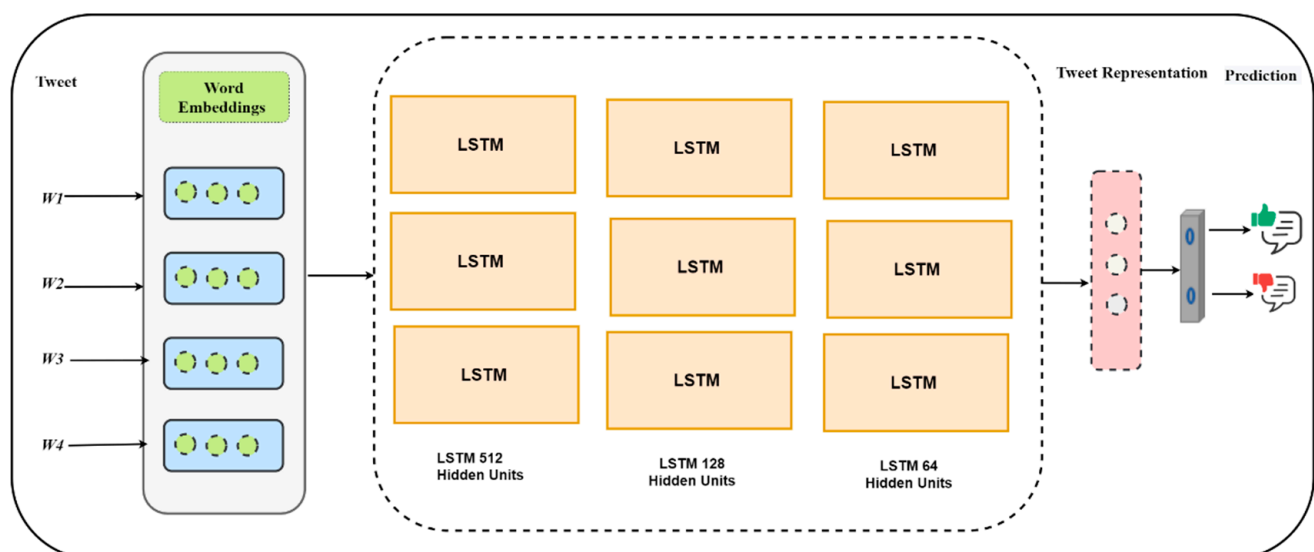


Figure 5. LSTM model.

Table 5. LSTM model parameters.

| Parameter | Value |
|-----------------|--|
| Embedding layer | Embedding(len(tokenizer.index_word) + 1, 64) |
| layer 1 | LSTM (512, return_sequences = True) |
| layer 2 | LSTM (128, return_sequences = True) |
| layer 3 | LSTM (64) |
| Dense layer | Dense (2, activation = 'sigmoid') |
| Optimiser | Adam |
| Loss function | Binary cross entropy |
| Batch size | 256 |
| Epochs | 20 |
| Callbacks | Early stop with 5 patience |

3.5.3. Gated Recurrent Networks

The GRU network is a type of recurrent neural network architecture for sequential data-processing tasks, such as NLP and speech recognition. This improves on the original RNN's flaws [45]. GRU nodes act as update gates and reset gates in a GRU network. The update gate decides how much information from the previous time step to carry over into the current time step, while the reset gate decides how much to forget when computing the current hidden state. When modelling long-term dependencies, GRU networks excel because they have fewer parameters and process longer sequences more efficiently. In machine translation, speech recognition, sentiment analysis, and language modelling, GRUs have performed well.

In this study, the GRU model was used for binary text classification tasks. The model's architecture consisted of the input, hidden, and output layers, as shown in Figure 6. First, the embedding layer's 64 dimensions enabled dense vector representations of vocabulary words. Then, three GRU layers were added to the first layer with 512 and the second 128 units, both returned sequences = True, which returned all previous outputs from these GRU layers. The third layer was 64 units of GRU. Finally, a dense layer of two neurons completed the model. In this dense layer, all neurons communicated with each other and with the layer below. This layer used the sigmoid function for binary classification because it worked well. Each class's likelihood was compressed to 0 and 1, representing the negative and positive; for more details of the model parameters, see Table 6.

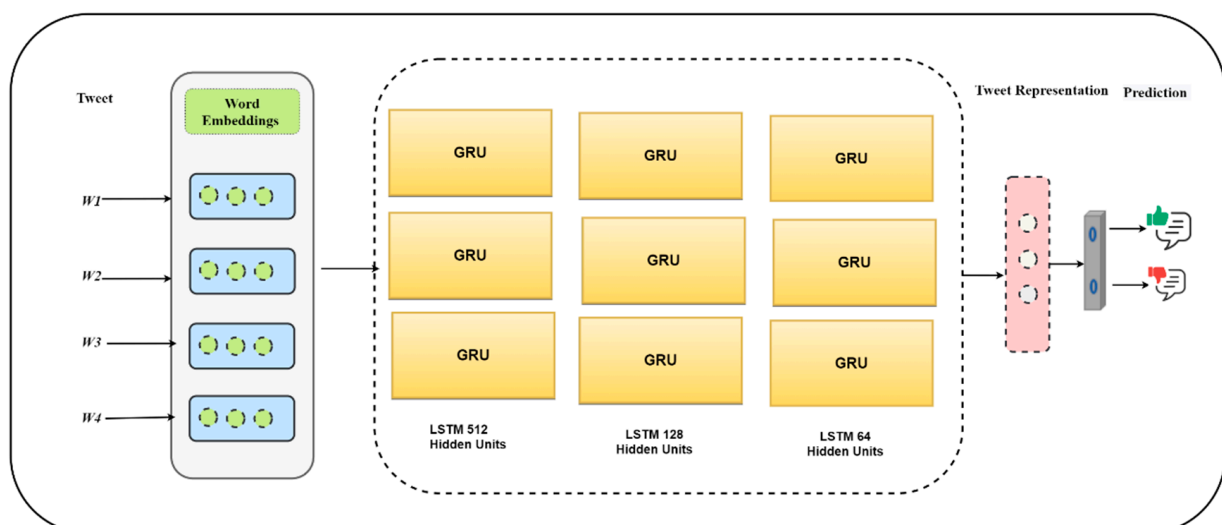
**Figure 6.** GRU model.

Table 6. GRU model parameters.

| Parameter | Value |
|-----------------|--|
| Embedding layer | Embedding(len(tokenizer.index_word) + 1, 64) |
| GRU layer 1 | GRU (512, return_sequences = True) |
| GRU layer 2 | GRU (128, return_sequences = True) |
| GRU layer 3 | GRU (64) |
| Dense layer | Dense (2, activation = 'sigmoid') |
| Optimiser | Adam |
| Loss function | Binary cross entropy |
| Batch size | 256 |
| Epochs | 20 |
| Callbacks | Early stop with 5 patience |

3.5.4. Bidirectional LSTM Networks

Bidirectional LSTM networks can consider past and future data. This method excels at sequential data processing with context from both directions [46]. First, it engages in 'encoding' the input sequence into numerical representations. A forward LSTM layer processes the encoded input sequence step by step, encoding the past context. Backward LSTM Pass: A backward LSTM layer reverses the encoded input sequence. For more details of the model parameters, see Table 7.

Table 7. BiLSTM model parameters.

| Parameter | Value |
|-----------------------|--|
| Embedding layer | Embedding(len(tokenizer.index_word) + 1, 64) |
| Bidirectional layer 1 | LSTM (128, return_sequences = True) |
| Bidirectional layer 3 | LSTM (32) |
| Dense layer | Dense (2, activation = 'sigmoid') |
| Optimiser | RMSprop |
| Loss function | Binary crossentropy |
| Batch size | 256 |
| Epochs | 20 |
| Callbacks | Early stop with 5 patience |

Like the forward LSTM, the internal memory state is updated based on the input and memory states at each time step. Bidirectional LSTM networks combine forward and backward LSTM layers to obtain a more complete picture of the input sequence at a given time. Routing the output layer through fully connected layers generates the final output. Despite their higher training and inference costs, bidirectional LSTM networks are often a good choice for sequence modelling due to their improved performance and ability to capture bidirectional dependencies.

The BiLSTM model has several key components for text analysis, as shown in Figure 7. An embedding layer converts each word or token into a dense vector in continuous space. Len(tokeniser.index_word) + 1 vocabulary and 64 embedding dimensions were used in this model's embedding layer. Each word in the input text becomes a 64-dimensional semantic vector. Next is a 128-unit LSTM bidirectional layer. LSTM and RNN are optimised for sequential data processing by long-term dependencies and a memory state. Bidirectional layers analyse input sequences. Return sequences = True ensures that the layer returns the hidden state output for each time step in the sequence without disrupting

the data's natural order. After that, a 32-unit LSTM bidirectional layer uses a second. This layer gathers contextual information from the input sequence's forward and backward motions. Fully connected, the final dense layer has two units. Since the model predicts two classes—satisfactory and unsatisfactory—this layer completes the classification. Each class's probability estimates are compressed to a single value between 0 and 1 by the dense layer's 'sigmoid' activation function.

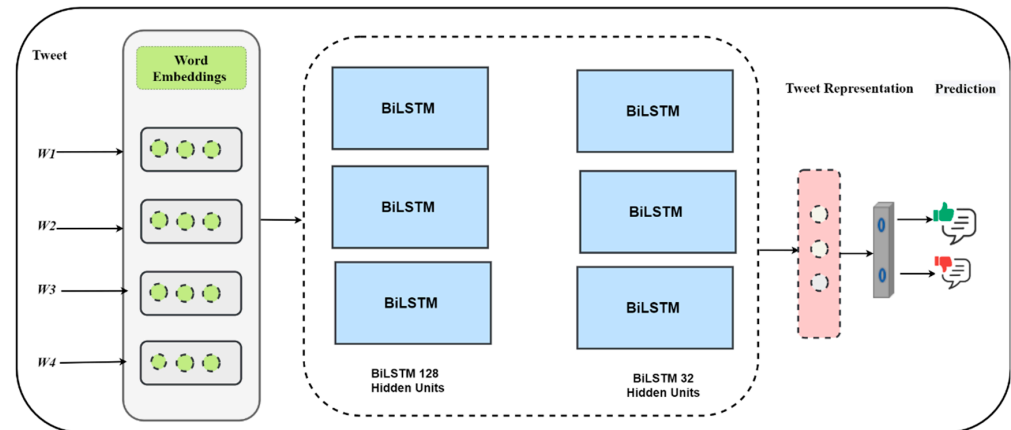


Figure 7. BiLSTM model.

3.5.5. Convolutional Neural Networks

CNNs, a type of DL model, process and analyse grid-like data, such as images and sequences [47]. They were developed for computer vision but are now used in NLP [13]. CNNs learn hierarchical representations of input data automatically. Convolution, pooling, and fully connected neurons achieve this. CNNs have revolutionised many fields, including computer vision, by performing well in image classification, object detection, and segmentation.

CNNs can process text and audio signals, making them useful for NLP. One-dimensional CNNs use sequence data such as word embeddings or character-level encodings. Parsed sentences and documents are converted into neural network-readable numerical representations. One-dimensional filters or kernels scan input data in the convolutional layer to apply convolutional operations across sequential dimensions. The pooling layer reduces the convolutional layer's feature maps, and the fully connected layers continue processing and learning relevant feature combinations.

One-dimensional CNNs are useful for text classification, sentiment analysis, named-entity recognition, and speech recognition. To improve NLP performance, 1D CNNs have been modified architecturally. These changes improve the modelling of one-dimensional sequential data and feature extraction for NLP.

3.5.6. CNN-LSTM Network

The CNN-LSTM model uses the LSTM's temporal dependencies and the CNN's spatial features because the CNN feeds the LSTM [14]. CNN-LSTM models combine spatial feature extraction power and sequential modelling precision. CNNs and LSTMs process sequential data to extract features. It can be used for video analysis, sentiment, and text classification, requiring spatial and temporal data. The CNN extracts spatial features from input data, while the LSTM handles sequential or temporal dependencies.

Each component of the model architecture has several crucial parameters. The model can process up to the tokeniser's index–word dictionary length plus one distinct word or token. Each word or token in the input has a dense vector representation with 64 dimensions that captures its semantic meaning. The Conv1D layer's 128 filters detect patterns in the input data and serve as feature detectors. The Conv1D layer extracts features from 5-word or token blocks with a kernel size of 5. Conv1D uses ReLU, a nonlinear activation function, to better capture intricate data patterns. The LSTM layer's 64 units, which determine

the dimensionality of the hidden state and the number of memory cells, are essential for capturing complex temporal dependencies. Two binary classification units in the dense layer generate output probabilities. The model's architecture is made up of these parameters, which affect data processing and learning. Table 8 shows the parameters of the CNN-LSTM model. The architecture of the CNN-LSTM model for analysis sentiment of customer satisfaction from social media is presented in Figure 8.

Table 8. CNN-LSTM Model parameters.

| Parameter | Value |
|---------------------|--|
| Vocabulary size | $\text{len}(\text{tokenizer.index_word}) + 1$ |
| Embedding dimension | 64 |
| Conv1D filters | 128 |
| Conv1D kernel size | 5 |
| Conv1D activation | ReLU |
| LSTM units | 64 |
| Dense units | 2 |
| Dense activation | Sigmoid |
| Optimiser | Adam |
| Loss function | Binary Cross Entropy |
| Metrics | Accuracy |
| Batch size | 256 |
| Number of epochs | 20 |
| Callbacks | Early stop with 5 patience |

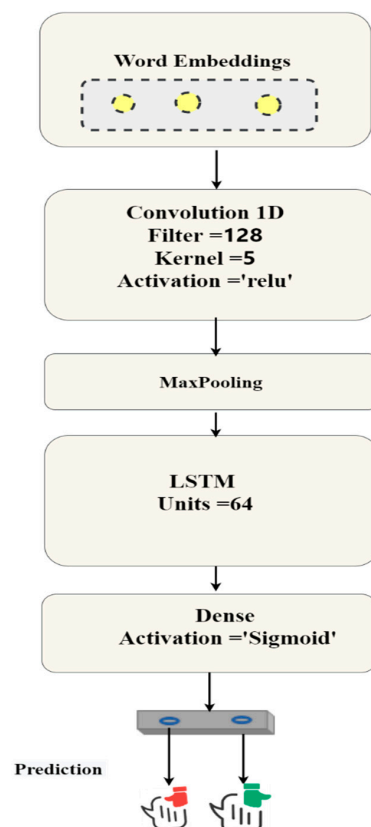


Figure 8. CNN-LSTM Model.

4. Experimental Results

In this section, the experimental setup and results are presented, in which different models were assessed based on standard evaluation metrics.

4.1. Environmental Setup

The experiments used a laptop with an NVIDIA graphics processing unit (GPU; RTX model with 8 GB of VRAM) and 16 GB of RAM. The DL libraries Keras [48] and TensorFlow [49] were used to create and train neural network models. Because it can parallelise and efficiently process the massive matrix operations needed to train neural networks, the GPU speeds up DL computations. Keras is a simple neural network library that hides low-level operations. TensorFlow is a popular open-source DL framework with a more granular programming interface for manipulating and tailoring neural network models. This environment is ideal for training and experimenting with DL models for computationally intensive ML tasks, such as image classification and NLP.

4.2. Evaluation Metrics

Several factors determine a DL model's overall effectiveness. The following metrics are used to evaluate DL models:

Accuracy: The percentage of correctly tagged data points was compared to the total to determine accuracy. In classification tasks, this statistic provides a complete model performance evaluation. Accuracy can be calculated by Equation (6).

$$Accuracy = \frac{TP + TN}{FP + FN + TP + TN} \times 100 \quad (6)$$

Confusion Matrix: The confusion matrix analyses model predictions. True positive (TP), true negative (TN), false positive (FP), and false negative (FN) are shown above. In addition to the F1 score, the confusion matrix can measure precision, recall, sensitivity, and specificity.

Precision: Precision, or positive predictive value, measures how well a model predicts the future. Precision was calculated by dividing the correct results by positive and negative findings. This metric shows the percentage of confirmed positive instances predicted correctly. Precision can be calculated by Equation (7).

$$Precision = \frac{TP}{TP + FP} \times 100 \quad (7)$$

Sensitivity: Sensitivity (recall) or TP rate (TPR) measures a model's positive event detection accuracy. Sensitivity was calculated by dividing the number of correct diagnoses by the number of incorrect diagnoses. This metric shows the percentage of positive achievements. Sensitivity can be calculated using Equation (8).

$$Sensitivity = \frac{TPs}{TPs + FPs} \quad (8)$$

Specificity: Specificity is a model's ability to identify outliers. The actual negative rate can be calculated using Equation (9). This represented the percentage of negative cases that were accurately detected.

$$Specificity = \frac{TNs}{TNs + FNs} \quad (9)$$

F1 score: The F1 score balances precision and recall to assess a model's performance. The F1 score was calculated by dividing the harmonic mean of precision and recall accuracy

ratings by their sum, as in Equation (10). When recall and precision are equal or datasets are unbalanced, the F1 score is useful.

$$F1 - score = 2 * \frac{precision \times Sensitivity}{precision + Sensitivity} \times 100 \quad (10)$$

Receiver Performance Curve (AUC–ROC): A model’s AUC–ROC can be used to evaluate its ability to distinguish between positive and negative examples at different classification levels. This method graphs the TPR and *FP* rate (FPR) and calculates the AUC. Higher AUC–ROC values indicate better model discrimination.

These indicators can evaluate the categorisation model. The above measures help determine how well the system distinguishes true positives and negatives, classifies events accurately, and balances precision and recall. These indicators allow us to evaluate our DL model’s performance and make informed decisions about its use and future improvements.

5. Results

This section presents the results of various DL models, namely BiLSTM, CNN-LSTM, GRU, and LSTM, for sentiment analysis of Arabic customer satisfaction. Several evaluation metrics, such as accuracy, precision, and the F1 score, were used to assess the quality of these models. Table 9 shows the results of the DL models. The training accuracy for BiLSTM was 97.84%, while the test accuracy was 96.40%. With a sensitivity of 91.67% and a specificity of 98.58 percent, it showed a healthy middle ground. The overall classification ability was measured by an AUC score of 96.44% and an F1 score of 94.14%, which considered both precision and recall. CNN-LSTM scored 96.82% on the accuracy test, which was slightly higher than BiLSTM’s score of 96.80%. Its specificity remained high, at 98.58%, while its sensitivity increased to 93.1%. In spite of a slight drop in AUC (96.17%), the F1 score improved to 94.86%. The test results showed that GRU, similar to CNN-LSTM, had a sensitivity of 93.02% and a specificity of 98.58%. However, it improved upon the previous version’s AUC score of 96.57% and F1 score of 94.86%.

Table 9. Results of deep learning (DL).

| Model Name | Training Accuracy (%) | Test Accuracy (%) | Sensitivity (%) | Specificity (%) | F1 Score (%) | AUC Score (%) |
|------------|-----------------------|-------------------|-----------------|-----------------|--------------|---------------|
| BiLSTM | 97.84 | 96.40 | 91.67 | 98.58 | 94.14 | 96.44 |
| CNN-LSTM | 97.82 | 96.82 | 93.02 | 98.58 | 94.86 | 96.17 |
| GRU | 98.07 | 96.82 | 93.02 | 98.58 | 94.86 | 96.57 |
| LSTM | 98.04 | 97.03 | 93.34 | 98.72 | 95.19 | 96.35 |

When compared to other models, LSTM achieved the best results. Its test accuracy was 97.03%, which was nearly as high as its 98.04% training accuracy. LSTM also had the highest sensitivity (93.34%) and specificity (98.72%) of all the models, indicating that it was the best at making the right positive and negative identifications. It performed admirably across the board, with an F1 score of 95.19% and an AUC of 96.35%. Figure 9 shows a comparison of the performance of the models.

The models’ performance on the task was very high. However, LSTM excelled above all other models in terms of accuracy, sensitivity, specificity, F1 score, and AUC.

The LSTM model trained for 20 epochs and early stopping at 8 epochs. The performance of the model in training accuracy was 98.04%, and the testing accuracy was 97.03%, as shown in Figure 10a,b. The model achieved a sensitivity of 93.34%, a specificity of 98.72%, and an F1 score of 95.19%. Additionally, the model achieved an AUC of 96.35%.

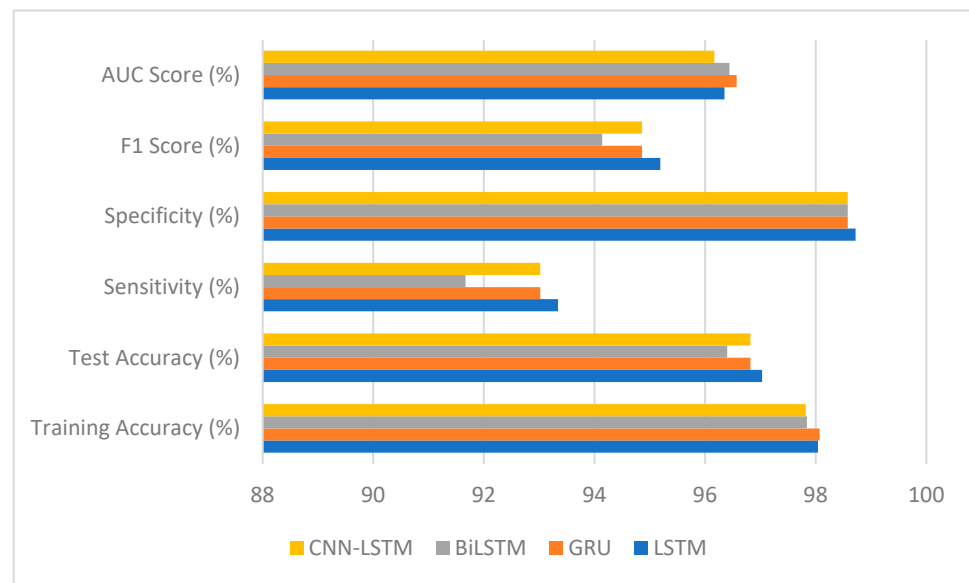


Figure 9. Comparison of performance models.

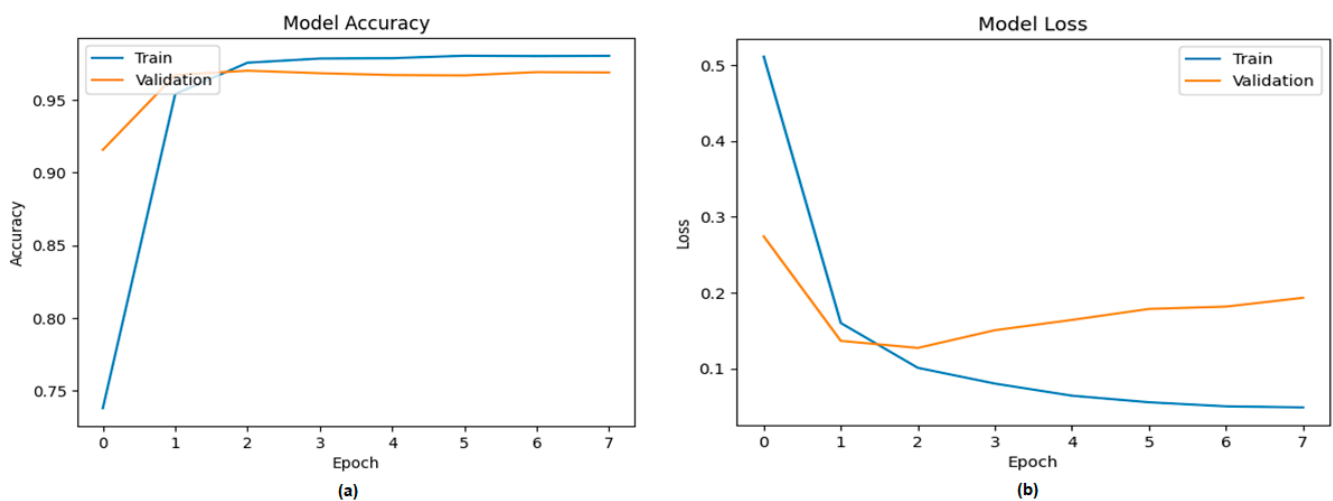


Figure 10. Training plots and testing accuracy and loss for LSTM models: (a) accuracy (b) loss.

The GRU model trained for 20 epochs and early stopping at 7 epochs. The performance of the model in training accuracy was 98.07%, and the testing accuracy was 96.82%, as shown in Figure 11a,b. The model achieved a sensitivity of 93.2%, a specificity of 98.58%, and an F1 score of 94.86%. Additionally, the model achieved an AUC of 96.57%.

The BiLSTM model was trained for 20 epochs and early stopping at 12 epochs. The performance of the model in training accuracy was 97.84%, and the testing accuracy was 96.40%, as shown in Figure 12a,b. The model achieved a sensitivity of 91.67%, a specificity of 98.58%, and an F1 score of 94.14%. Additionally, the model achieved an AUC of 96.44%.

The BiLSTM model was trained for 20 epochs and early stopping at 12 epochs. The performance of the model in training accuracy was 97.82%, and the testing accuracy was 96.82%, as shown in Figure 12a,b. The model achieved a sensitivity of 93.02%, a specificity of 98.58%, and an F1 score of 94.86%. Additionally, the model achieved an AUC of 96.17%.

This study's customer satisfaction level findings help improve services and retain regular clients. This research detailed the models' sensitivity, specificity, and positive and negative predictive values, as described in Figure 13. With only 35 FPs and 84 FNs, LSTM achieved 2704 true positives and a 1177 TN accuracy ratio. Among the 2700 positive results and 1173 negative results that GRU found, there were only 39 FPs and 88 FNs. Exactly the

same numbers of true positives (2700), FPs (39), FNs (88), and true negatives (1177) were generated by both BiLSTM and CNN-LSTM.

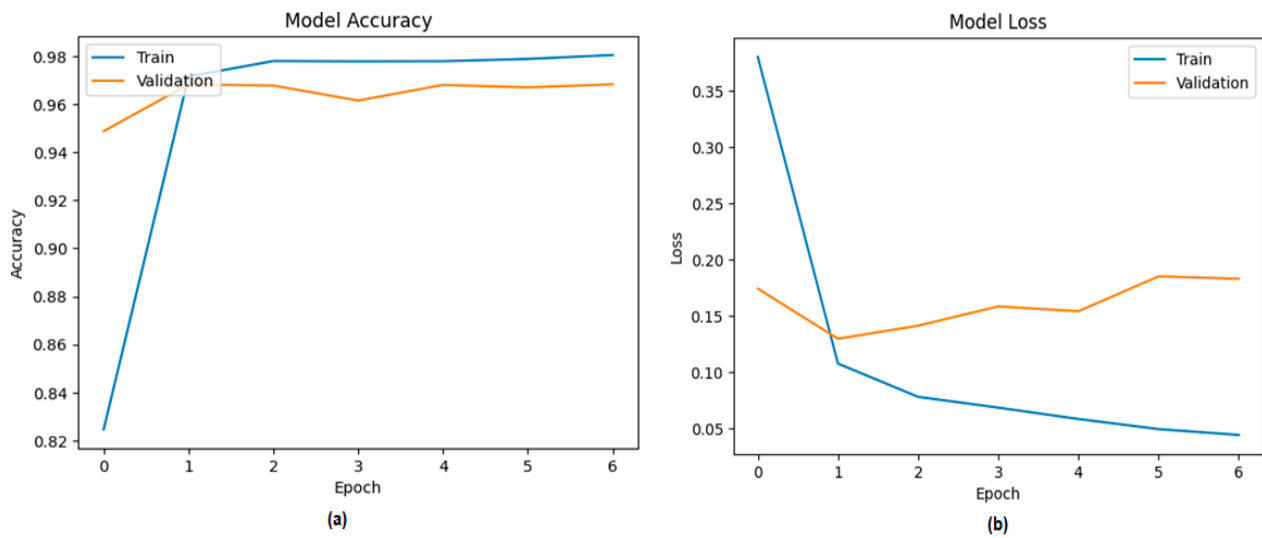


Figure 11. Training plots and testing accuracy and loss for GRU models: (a) accuracy (b) loss.

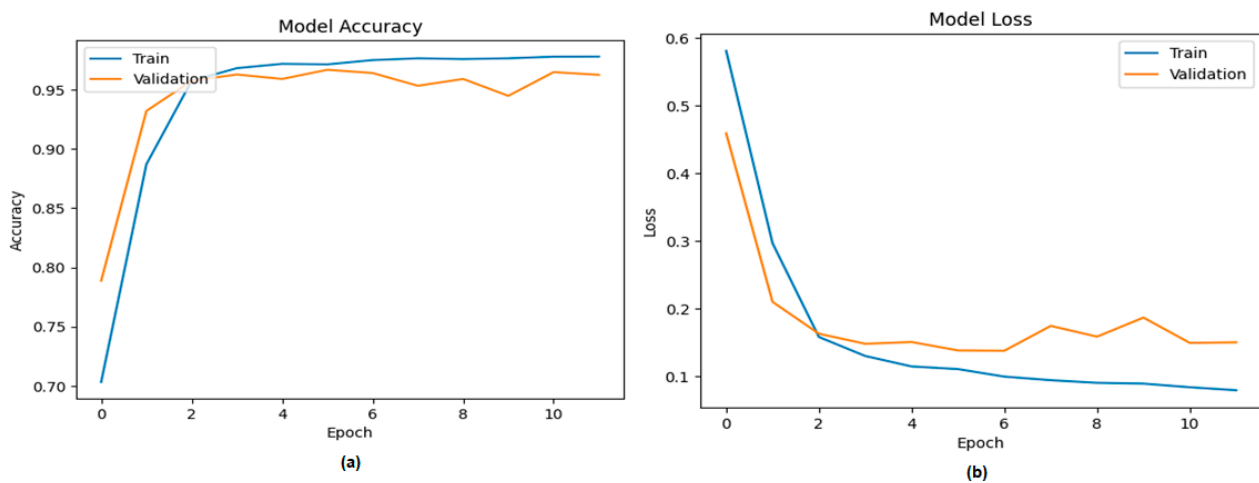


Figure 12. Training plots and testing accuracy and loss for BiLSTM models: (a) accuracy (b) loss.

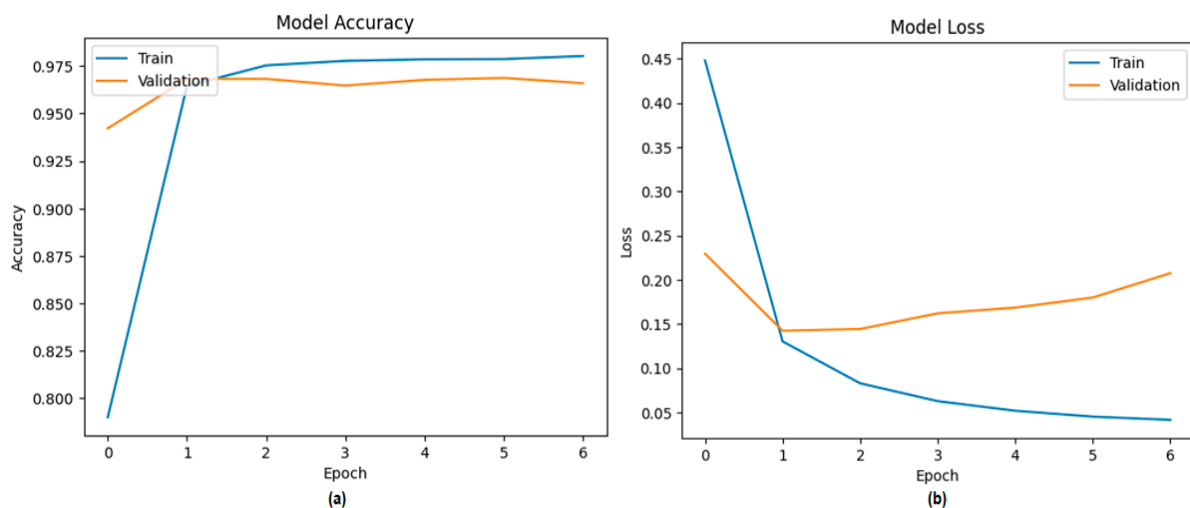


Figure 13. Training plots and testing accuracy and loss for CNN-LSTM models: (a) accuracy (b) loss.

Although there were some differences between the models in terms of the proportions of correct predictions, incorrect predictions and FNs, all of them performed a respectable job. LSTM had the highest proportion of correct positive and negative identifications, demonstrating its superior ability to detect customer satisfaction. The confusion metrics of the deep learning models is presented in Figure 14. Figure 15 shows a comparison of the confusion metrics of DL models.

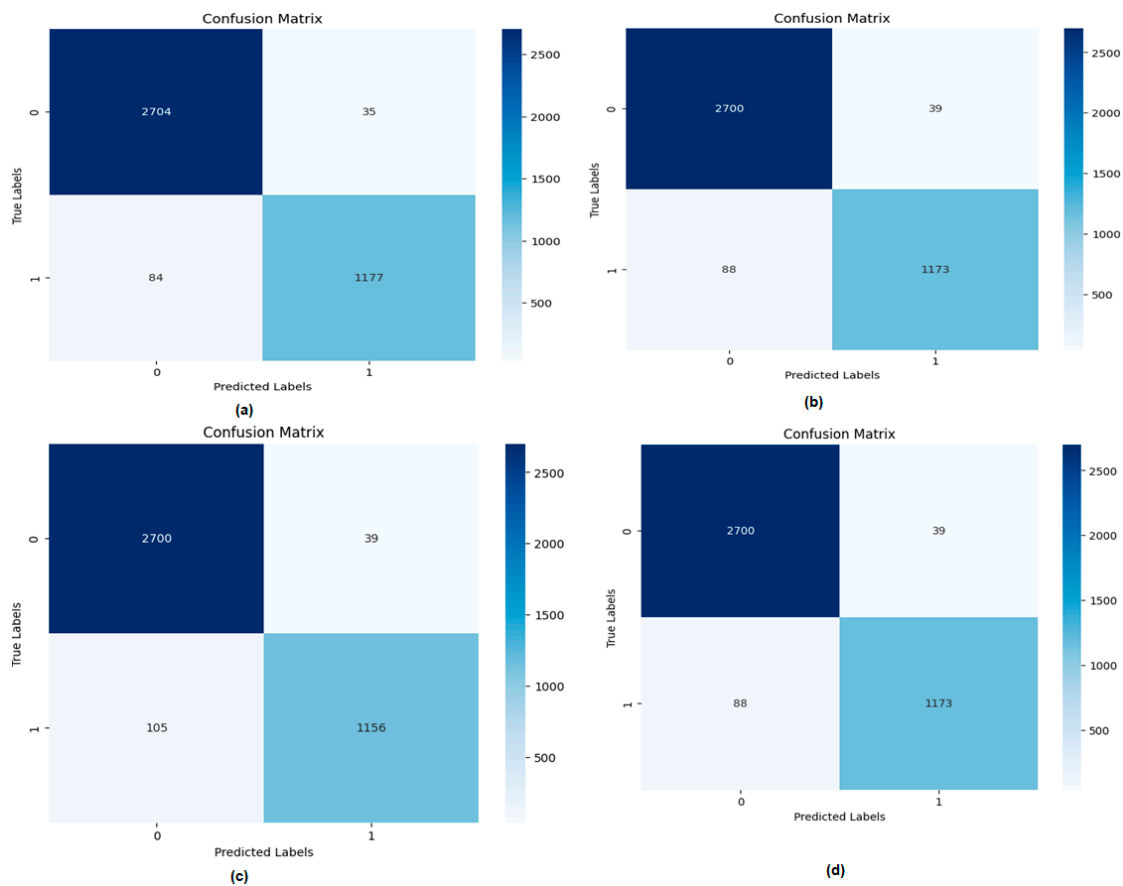


Figure 14. Confusion matrix plots of the study models: (a) LSTM (b) GRU (c) BiLSTM (d) CNN-LSTM.

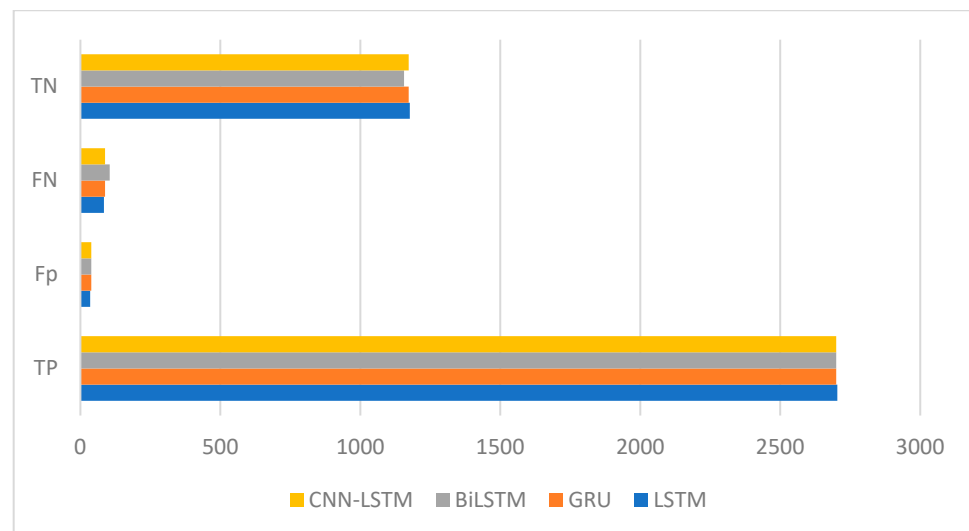


Figure 15. Comparison of model performance based on positive and negative factors.

6. Discussion

The phenomenon of customer churn represents a significant challenge and a top priority for major corporations. Owing to its significant impact on corporate revenues, particularly within the telecommunications industry, companies are actively pursuing strategies to forecast potential customer churn. Hence, identifying the determinants that contribute to customer attrition is crucial to implementing appropriate measures aimed at mitigating this phenomenon. Our study's primary objective was to create a churn prediction model that can aid telecommunication operators in identifying customers who are at a higher risk of churning.

This paper used Arabic tweets from Saudi telecommunications companies. The new restrictions on Twitter prevent data collection from tweets using the Python scripiter. The restrictions, which were put in place in January 2023, limit the number of tweets a single user or application can collect in a given period. This makes it more difficult to collect large datasets of tweets, which is often necessary for data mining and other research purposes. This study compared four models for predicting customer satisfaction. Models such as LSTM, GRU, BiLSTM, and CNN-LSTM were tested. The research confirmed the significance of customers' use of social media to share their experiences, both good and bad, with a company's services or products. Figure 16 shows the ROC of the deep learning models. The problem was solved by creating and training DL methods on the open-source AraCust dataset. The LSTM model stood out because it had the highest training and test accuracy for text classification: 98.04% and 97.03%, respectively.

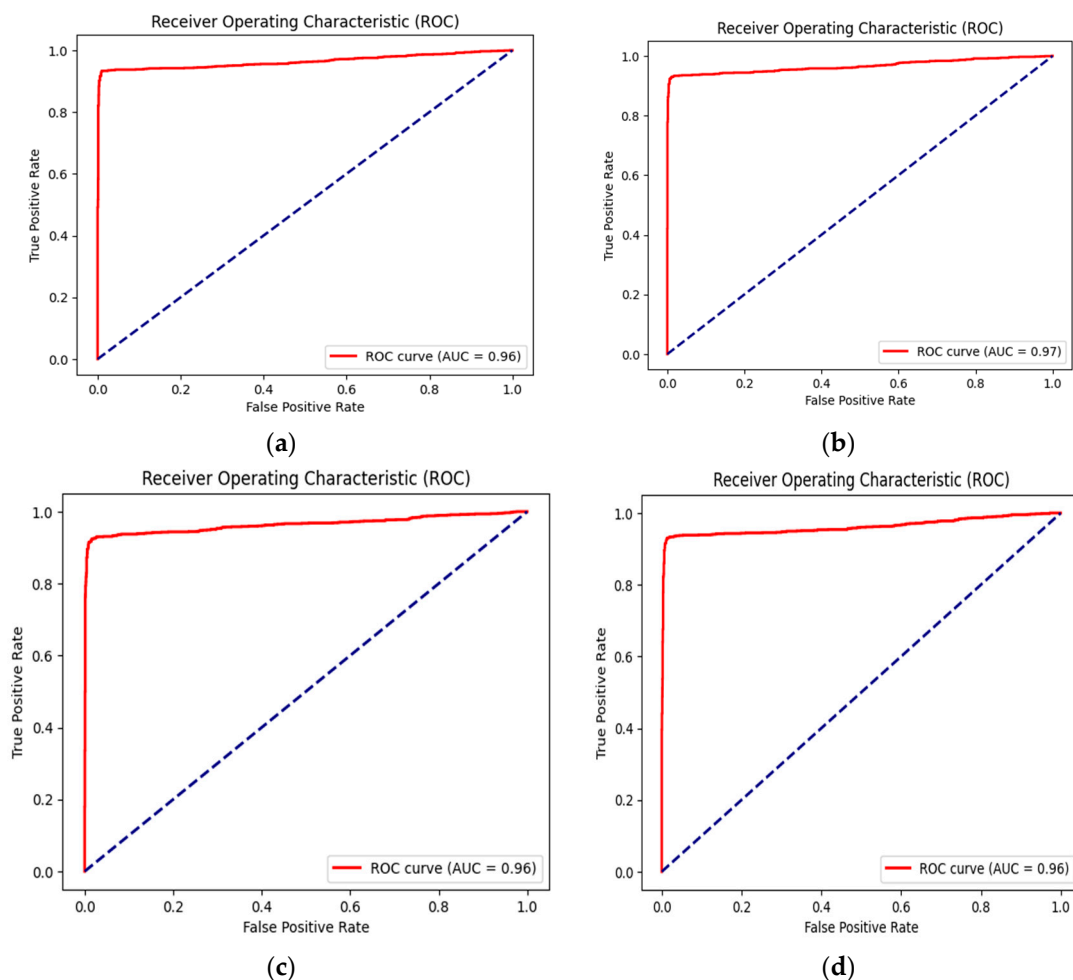


Figure 16. Training plots and testing accuracy and loss of the study models: (a) LSTM (b) GRU (c) BiLSTM (d) CNN-LSTM.

The comparison results of proposed deep learning models and existing models for sentiment analysis for Arabic customer satisfaction on the racist dataset are presented in Table 10. This is related to the telecommunication sectors of Saudi Arabia. Almuqren et al. [49] proposed two models: Bi-GRU and LSTM. The BiGRU model achieved an accuracy of 95.16%, while the LSTM model achieved 94.66% accuracy. Aftan and Shah [50] proposed three other models: RNN, CNN, and AraBERT. The AraBERT model achieved 94.33% accuracy, the RNN model achieved an accuracy of 91.35%, and the CNN model achieved 88.34% accuracy. Almuqren et al. [46] proposed a SentiChurn model and obtained an accuracy of 95.8%. In this study, we proposed several DL models; the best accuracy result achieved by an LSTM model was 97.03%, and it also achieved the highest accuracy among the existing studies.

Table 10. Comparison of various proposed DL models and existing research papers.

| Study | Model | Accuracy |
|-----------------------------|-------------------|---------------------------|
| Almuqren et al. (2021) [49] | Bi-GRU and LSTM | 95.16%, 94.66% |
| Almuqren et al. (2021) [46] | SentiChurn | 95.8% |
| Aftan and Shah (2023) [50] | RNN, CNN, AraBERT | 91.35%, 88.34% and 94.33% |
| Our model | LSTM | 97.03% |

7. Conclusions

The significance of conducting research in the telecommunications industry lies in its potential to enhance the interaction between users and technologies and, therefore, to improve companies' profitability. It is widely acknowledged that the ability to forecast customer churn is a critical revenue stream for telecommunications enterprises. Therefore, the objective of this study was to construct a predictive system for customer churn in Saudi Arabian telecommunications companies. This study used DL and sentiment analysis to make important decisions about increasing customer loyalty and satisfaction. This research can help the telecommunications industry better serve its customers and address their concerns as social media continues to shape public opinion. This study used sentiment analysis to assess customer satisfaction with STC, Mobily, and Zain services, and to inform business decisions. The study confirmed social media's value as a platform for consumers to share their positive and negative experiences with a company's products or services. Communication is vital to Saudi life and, so, online discussions are inevitable. In this study, sophisticated DL models were trained on the available online dataset AraCust, which was collected from Arabic tweets. The proposed models in this study were LSTM, GRU, BiLSTM, and CNN-LSTM. The LSTM model had the highest training (98.04%) and test accuracy (97.03%) in text classification. The model's superior sensitivity to identifying customer satisfaction showed its potential to help telecommunications providers reduce customer churn caused by dissatisfaction with their offerings. The researcher aimed to enhance their existing research model by incorporating sophisticated DL techniques, such as transform models and time series models, to enhance its precision.

This research paper provided a substantial contribution to the domain of customer satisfaction analysis in the Arabic language. It is a crucial area of investigation, given the population of Arabic speakers in the world. The study effectively showed the ability of different deep learning models to accurately predict customer satisfaction through analysing Arabic tweets. This study highlighted the importance of social media platforms as valuable mediums through which customers can share their experiences, which helps business owners improve service quality and maintain customer loyalty.

Funding: This research was funded by the Deputyship for Research and Innovation, Ministry of Education in Saudi Arabia, grant number 523.

Data Availability Statement: The datasets are available at the following link: DOI: <https://doi.org/10.7717/peerj-cs.510/supp-2>, accessed on 15 May 2023.

Conflicts of Interest: The author declares no conflict of interest.

References

1. Gerpott, T.J.; Rams, W.; Schindler, A. Customer retention, loyalty, and satisfaction in the German mobile cellular telecommunications market. *Telecommun. Policy* **2001**, *25*, 249–269. [\[CrossRef\]](#)
2. Wei, C.P.; Chiu, I.T. Turning telecommunications call details to churn prediction: A data mining approach. *Expert Syst. Appl.* **2002**, *23*, 103–112. [\[CrossRef\]](#)
3. Qureshii, S.A.; Rehman, A.S.; Qamar, A.M.; Kamal, A.; Rehman, A. Telecommunication subscribers' churn prediction model using machine learning. In Proceedings of the Eighth International Conference on Digital Information Management, Islamabad, Pakistan, 10–12 September 2013; pp. 131–136.
4. Ascarza, E.; Iyengar, R.; Schleicher, M. The perils of proactive churn prevention using plan recommendations: Evidence from a field experiment. *J. Mark. Res.* **2016**, *53*, 46–60. [\[CrossRef\]](#)
5. Adwan, O.; Faris, H.; Jaradat, K.; Harfoushi, O.; Ghatasheh, N. Predicting customer churn in telecom industry using multilayer perceptron neural networks: Modelling and analysis. *Life Sci. J.* **2014**, *11*, 75–81.
6. Afful-Dadzie, E.; Nabareseh, S.; Oplatková, Z.K.; Klímek, P. Enterprise competitive analysis and consumer sentiments on social media: Insights from telecommunication companies. In Proceedings of the 3rd International Conference on Data Management Technologies and Applications (DATA 2014), Vienna, Austria, 29–31 August 2014; pp. 22–32.
7. Aghakhani, N.; Asllani, A. A Text-mining approach to evaluate the importance of information systems research themes. *Commun. IIMA* **2020**, *18*, 3. Available online: https://scholarworks.lib.csusb.edu/ciima/vol18/iss1/3/?utm_source=scholarworks.lib.csusb.edu%2Fciima%2Fvol18%2Fiss1%2F3&utm_medium=PDF&utm_campaign=PDFCoverPages (accessed on 2 June 2023). [\[CrossRef\]](#)
8. Alalwan, A.A.; Rana, N.P.; Dwivedi, Y.K.; Algharabat, R. Social media in marketing: A review and analysis of the existing literature. *Telemat. Inform.* **2017**, *34*, 1177–1190. [\[CrossRef\]](#)
9. El Rahman, S.A.; Alotaibi, F.A.; Alshehri, W.A. Sentiment analysis of Twitter data. In Proceedings of the 2019 International Conference on Computer and Information Sciences (ICCIS 2019), Aljoud, Saudi Arabia, 3–4 April 2019.
10. Pavaloia, V.D.; Teodor, E.M.; Fotache, D.; Danileț, M. Opinion mining on social media data: Sentiment analysis of user preferences. *Sustainability* **2019**, *11*, 4459. [\[CrossRef\]](#)
11. Aldhyani, T.H.H.; Alsubari, S.N.; Alshebami, A.S.; Alkahtani, H.; Ahmed, Z.A.T. Detecting and Analyzing Suicidal Ideation on Social Media Using Deep Learning and Machine Learning Models. *Int. J. Environ. Res. Public Health* **2022**, *19*, 12635. [\[CrossRef\]](#)
12. Susanti, C.E. The effect of product quality and service quality towards customer satisfaction and customer loyalty in traditional restaurants in East Java. In Proceedings of the International Conference on Managing the Asian Century, Singapore, 11–13 July 2013; Springer: Singapore, 2013; pp. 383–393.
13. Abiodun, R. Development of mathematical models for predicting customers satisfaction in the banking system with a queuing model using regression method. *Am. J. Oper. Manag. Inf. Syst.* **2017**, *2*, 86–91.
14. Mugion, R.G.; Musella, F. Customer satisfaction and statistical techniques for the implementation of benchmarking in the public sector. *Total Qual. Manag. Bus. Excell.* **2013**, *24*, 619–640. [\[CrossRef\]](#)
15. Al-Ghamdi, S.M.; Sohail, M.S.; Al-Khaldi, A. Measuring consumer satisfaction with consumer protection agencies: Some insights from Saudi Arabia. *J. Consum. Mark.* **2007**, *24*, 71–79. [\[CrossRef\]](#)
16. The Communication and Information Technology Commission. Annual Report of (CITC). Available online: https://www.cst.gov.sa/en/mediacenter/reports/Documents/PR_REP_013Eng.pdf (accessed on 2 June 2023).
17. Hassounah, M.; Raheel, H.; Alhefzi, M. Digital response during the COVID-19 pandemic in Saudi Arabia. *J. Med. Internet Res.* **2020**, *22*, e19338. [\[CrossRef\]](#)
18. Digital 2019 Saudi Arabia. Available online: <https://www.slideshare.net/DataReportal/digital-2019-saudi-arabia-january-2019-v01> (accessed on 2 June 2023).
19. Brandusoiu, I.; Todorean, G.; Ha, B. Methods for churn prediction in the prepaid mobile telecommunications industry. In Proceedings of the International Conference on Communications, Kuala Lumpur, Malaysia, 22–27 May 2016; pp. 97–100.
20. He, Y.; He, Z.; Zhang, D. A study on prediction of customer churn in fixed communication network based on data mining. In Proceedings of the Sixth International Conference on Fuzzy Systems and Knowledge Discovery, Tianjin, China, 14–16 August 2009; Volume 1, pp. 92–94.

21. Idris, A.; Khan, A.; Lee, Y.S. Genetic programming and AdaBoosting based churn prediction for telecom. In Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics (SMC), Seoul, Republic of Korea, 14–17 October 2012; pp. 1328–1332.
22. Huang, F.; Zhu, M.; Yuan, K.; Deng, E.O. Telco churn prediction with big data. In Proceedings of the ACM SIGMOD International Conference on Management of Data, Melbourne, Australia, 31 May–4 June 2015; pp. 607–618.
23. Makhtar, M.; Nafis, S.; Mohamed, M.; Awang, M.; Rahman, M.; Deris, M. Churn classification model for local telecommunication company based on rough set theory. *J. Fundam. Appl. Sci.* **2017**, *9*, 854–868. [\[CrossRef\]](#)
24. Amin, A.; Anwar, S.; Adnan, A.; Nawaz, M.; Howard, N.; Qadir, J.; Hawalah, A.; Hussain, A. Comparing oversampling techniques to handle the class imbalance problem: A customer churn prediction case study. *IEEE Access* **2016**, *4*, 7940–7957. [\[CrossRef\]](#)
25. Tul, Q.; Ali, M.; Riaz, A.; Noureen, A.; Kamranz, M.; Hayat, B.; Rehman, A. Sentiment analysis using deep learning techniques: A review. *Int. J. Adv. Comput. Sci. Appl.* **2017**, *8*, 424–433. [\[CrossRef\]](#)
26. Pang, B.; Lee, L. Opinion mining and sentiment analysis. In *Foundations and Trends in Information Retrieval*; Alet Heezemans: Rotterdam, The Nederland, 2008; Volume 2, pp. 1–135.
27. Vieira, S.T.; Rosa, R.L.; Rodríguez, D.Z.; Ramírez, M.A.; Saadi, M.; Wuttisittikulkij, L. Q-meter: Quality monitoring system for telecommunication services based on sentiment analysis using deep learning. *Sensors* **2021**, *21*, 1880. [\[CrossRef\]](#)
28. Chiu, S.T.; Susanto, H.; Leu, F.Y. Detection and defense of DDoS attack and flash events by using Shannon entropy. In *Innovative Mobile and Internet Services in Ubiquitous Computing, Proceedings of the IMIS 2022, Kitakyushu, Japan, 29 June–1 July 2022*; Lecture Notes in Networks and Systems; Barolli, L., Ed.; Springer: Cham, Switzerland, 2022; Volume 496.
29. Heru, S.; Leu, F.Y.; Alifya, K.S.S. Genetics algorithm approaches of cheminformatics reengineering process. *J. Biomed. Sci.* **2022**, *4*, 1523–1530.
30. Setiana, D.; Norsarah, S.; Besar, N.; Anna, T.; Nasution, M.; Susanto, H. Technology disruption in the time of the digital ecosystem society’s adoption: Cyber bullying phenomenon—The truth or hoax? In *Handbook of Research on Big Data, Green Growth, and Technology Disruption in Asian Companies and Societies*; IGI Global: Hershey, PA, USA, 2022; pp. 238–255.
31. Alamsyah, A.; Paryasto, M.; Putra, F.J.; Himmawan, R. Network text analysis to summarize online conversations for marketing intelligence efforts in telecommunication industry. In Proceedings of the 2016 4th International Conference on Information and Communication Technology (ICICT 2016), Bandung, Indonesia, 25–27 May 2016.
32. Najadat, H.; Al-Abdi, A.; Sayaaheen, Y. Model-based sentiment analysis of customer satisfaction for the Jordanian telecommunication companies. In Proceedings of the 2018 9th International Conference on Information and Communication Systems (ICICS 2018), Irbid, Jordan, 3–5 April 2018; pp. 233–237.
33. Qamar, A.M.; Ahmed, S.S. Sentiment classification of Twitter data belonging to Saudi Arabian telecommunication companies. *Int. J. Adv. Comput. Sci. Appl.* **2017**, *8*, 395–401.
34. Hang, A.; Li, B.; Wang, W.; Wan, S.; Chen, W. MII: A Novel Text Classification Model Combining Deep Active Learning with BERT. *Comput. Mater. Contin.* **2020**, *63*, 1499–1514.
35. Gabhane, M.D.; Suriya, D.S.B.A. Churn Prediction in Telecommunication Business using CNN and ANN. *J. Posit. Sch. Psychol.* **2022**, *6*, 4672–4680.
36. DiPietro, R.; Hager, G.D. Deep learning: RNNs and LSTM. In *Handbook of Medical Image Computing and Computer Assisted Intervention*; Zhou, S.K., Rueckert, D., Fichtinger, C.A.I., Eds.; The Elsevier and MICCAI Society Book Series; Academic Press: Cambridge, MA, USA, 2020; pp. 503–519.
37. Almuqren, L.; Cristea, A. AraCust: A Saudi Telecom Tweets corpus for sentiment analysis. *PeerJ Comput. Sci.* **2020**, *7*, e510. [\[CrossRef\]](#) [\[PubMed\]](#)
38. Hathlian, N.F.B.; Hafez, A.M. Subjective text mining for Arabic social media. In *Cognitive Analytics: Concepts, Methodologies, Tools, and Applications*; IGI Global: Hershey, PA, USA, 2020; pp. 1483–1495.
39. Sun, W.; Cai, Z.; Li, Y.; Liu, F.; Fang, S.; Wang, G. Data processing and text mining technologies on electronic medical records: A review. *J. Healthc. Eng.* **2018**, *2018*, 4302425. [\[CrossRef\]](#)
40. Webster, J.J.; Kit, C. Tokenization as the initial phase in NLP. In Proceedings of the 14th International Conference on Computational Linguistics (COLING 1992), Nantes, France, 23–28 August 1992; Volume 4.
41. Barabas, P.; Kovacs, L. Efficient encoding of inflection rules in NLP systems. *Acta Marisiensis Ser. Technol.* **2012**, *9*, 11.
42. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [\[CrossRef\]](#)
43. Cambria, E.; White, B. Jumping NLP curves: A review of natural language processing research. *IEEE Comput. Intell. Mag.* **2014**, *9*, 48–57. [\[CrossRef\]](#)
44. Medsker, L.R.; Jain, L.C. Recurrent neural networks. *Des. Appl.* **2001**, *5*, 64–67.
45. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780. [\[CrossRef\]](#) [\[PubMed\]](#)
46. Dey, R.; Salem, F.M. Gate-variants of gated recurrent unit (GRU) neural networks. In Proceedings of the 2017 IEEE 60th International Midwest Symposium on Circuits and Systems (MWSCAS), Boston, MA, USA, 6–9 August 2017; pp. 1597–1600.
47. Huang, Z.; Xu, W.; Yu, K. Bidirectional LSTM-CRF models for sequence tagging. *arXiv* **2015**, arXiv:1508.01991.
48. Gu, J.; Wang, Z.; Kuen, J.; Ma, L.; Shahroudy, A.; Shuai, B.; Liu, T.; Wang, X.; Wang, G.; Cai, J.; et al. Recent advances in convolutional neural networks. *Pattern Recognit.* **2018**, *77*, 354–377. [\[CrossRef\]](#)

49. Almuqren, L.; Alrayes, F.S.; Cristea, A.I. An empirical study on customer churn behaviours prediction using Arabic twitter mining approach. *Future Internet* **2021**, *13*, 175. [[CrossRef](#)]
50. Aftan, S.; Shah, H. Using the AraBERT model for customer satisfaction classification of telecom sectors in Saudi Arabia. *Brain Sci.* **2023**, *13*, 147. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.