Applied Mathematics & Information Sciences
An International Journal

http://dx.doi.org/10.18576/amis/180217

Deep Learning-Based Mathematical Modelling For Predictive Analysis in Media Consumer Behaviour

Abd Elmotaleb A. M. A. Elamin

Department of Mathematics, College of Science and Humanities, Prince Sattam bin Abdulaziz University, Al-Kharj 16273, Saudi Arabia

Received: 3 Nov. 2023, Revised: 22 Jan. 2023, Accepted: 26 Jan. 2023
Published online: 1 Mar. 2024

Abstract: Advanced predictive models are required to understand and predict consumer behavior due to the rapid evolution of media consumption patterns. This work aims to improve the accuracy of predictive analyses in media consumer behavior by introducing a new method as Bayesian optimized Long-Short Term Memory (LSTM)-Based Mathematical Modelling. The proposed model uses Bayesian optimization techniques to improve performance in capturing temporal dependencies within media consumption data by optimizing LSTM networks for hyperparameter tuning. Models based on long short-term dependencies in sequential data are based on recurrent neural networks, a class of networks well-known for this capacity. To ensure that the LSTM model is precisely tuned to the particular features of media consumption datasets, the Bayesian optimization framework makes it easier to tune hyperparameters automatically. A more accurate and efficient representation of the complex patterns present in media consumer behavior is made possible by combining LSTM networks and Bayesian optimization. The mathematical model based on Bayesian optimized LSTM is increased the accuracy with 99%, which is 9.62% higher than the accuracy of Random Forests, RNN Based Click Stream Model and Gradient Tree Boosting Method. In an era of constant technological and content evolution, the results of this work adds to the growing field of predictive analytics by providing a potent tool for comprehending and forecasting the dynamic nature of media consumer behavior.

Keywords: Bayesian Optimization, Long Short-Term Memory, Predictive Analysis, Media Consumer Behavior, Hyperparameter Tuning, Recurrent Neural Networks.

1 Introduction

Online shopping and other knowledge-based economies have long existed since the development of the Internet. Implicit knowledge may be extracted from the online retailers' logs using machine learning techniques [1]. Businesses and industries use the information to gain a better understanding of consumer behavior, as well as the associated opportunities and threats. Microsoft uses machine-learning techniques to evaluate customer-shopping behavior automatically and offer relevant services. Machine learning techniques leverage data analytic tools to identify customer behavioural patterns [2]. Internet-based purchasing sites offer a plethora of data regarding occasions, connections, and mindsets. To achieve analysis of sentiment, data extraction, and user influence analysis, a number of technologies are utilized, including statistical theory, text mining, association analysis, and visualization [3].

Understanding consumer behavior can be useful in analysing product relationships, consumer characteristics, and other related topics. Building consumption structures based on the purchase behavior records of various consumers is, therefore, a very useful study. Social media refers to web pages and services that are designed to allow users to consume content in small chunks too quickly, effectively, and in real-time. It has changed the mode we enter and the mode in which we perform an activity [4]. It can share images, reviews, activities, and more in real-time. Retailers who rely heavily on social media for their advertising strategy typically pay a measurable price. However, the key to successful social media marketing and advertising is to stop treating it like an extra accessory and instead treat it with the same respect, admiration, and curiosity as the rest of your marketing and advertising campaigns. Social media marketing refers to using various channels to engage with your target audience in order to build your brand, increase

* Corresponding author e-mail: aa.alameen@psau.edu.sa

revenue, and improve website traffic density. In the digital age, social media has completely changed how people connect, communicate, and share information. It includes a wide range of internet-based platforms that make it easier for people to create, share, and interact with user-generated content [5]. Through the sharing of opinions, personal experiences, and multimedia content, users of these platforms create virtual communities and networks that are not limited by geography. Social media has developed into a potent tool that companies, influencers, and organizations can use to connect and interact with their target audiences in real time and in a dynamic way [6]. However, the widespread use of social media gives rise to worries about misinformation, privacy, and the effect of digital interactions on mental health. As a result, discussions regarding platform regulations and responsible usage are continuing. Social media is essentially a powerful and dynamic force that constantly shifts the way individuals and societies communicate and share information in the contemporary digital landscape [7]. This entails posting noteworthy content on your social media profiles to draw in more followers and enhance the impact of your studies, as well as jogging social media classified ads. The widely used social media platforms, including ”YouTube,” ”Instagram,” and ”Facebook” [8].The increasing number of online platforms and technological advancements are driving a rapid evolution of media consumption landscapes [9].

The current media landscape is dynamic and constantly changing for those in the media industry due to the proliferation of digital media platforms and the exponential growth of information available. The difficulty lies not only in producing visually appealing content but also in figuring out the complex web of customer preferences, responses, and decision-making processes. While marketers strive to create strategies that captivate and retain viewers in an oversaturated digital space, content creators aim to create narratives that resonate with a diverse range of audiences. Conversely, media strategists struggle to predict trends and quickly adjust to changes in consumer behavior, so understanding and anticipating these changing patterns is critical to success in this cutthroat market. As such, the necessity of deciphering the intricacies of consumer behavior in media consumption has emerged as a central undertaking for all significant industry stakeholders. Customers use digital channels to navigate through a wide variety of content as active participants rather than as passive recipients [10]. Because of this, conventional approaches to studying consumer behavior are finding it difficult to fully capture the complex and multidimensional nature of contemporary media engagement. This emphasizes the necessity of creative strategies that can adjust to the complexity of the modern media environments [11].

One of the most important tools for tackling the problems caused by the changing nature of media consumption is predictive analysis. One of the most important tools for tackling the problems caused by the changing media consumption landscape is predictive analysis. The utilization of sophisticated algorithms and data analytics enables media practitioners to predict audience preferences, trends, and content performance. By facilitating strategic decision-making, this foresight aids media companies in customizing their content to satisfy changing consumer needs. Predictive analysis also helps with marketing strategy optimization, which results in more successful audience engagement and focused outreach. In the end, proactive insights obtained from predictive analysis enable media organizations to maintain their relevance to their audiences and stay ahead of the curve in a constantly changing environment. Using deep learning, a subfield of machine learning that is skilled at identifying complex patterns in large-scale datasets, this work aims to develop sophisticated mathematical models that go beyond what is possible with traditional analytics. Neural networks used in deep learning have layered and hierarchical structures that allow the system to automatically learn hierarchical representations of data features, improving the system’s ability to recognize subtle patterns.

The goal of the project is to model complex relationships within large datasets with unprecedented levels of accuracy and sophistication by utilizing deep learning, which will improve analytical and predictive capacities. When faced with complex and non-linear relationships, where traditional analytical methods might not be as effective, this approach is especially beneficial. This study’s use of deep learning holds the potential to advance predictive modelling, providing new opportunities for more complex insights and better decision-making in data-intensive fields. These models will help stakeholders stay ahead of the curve in the ever-changing media landscape by predicting future trends in addition to describing past consumer behavior. Understanding and projecting consumer behavior in the context of media consumption has become a critical challenge for marketers, media strategists, and content creators in the age of digital media and information abundance. Because customer preferences are constantly changing and there is a huge amount of content available, sophisticated analytical techniques that can identify complex patterns and offer useful insights are required. This research uses deep learning-based mathematical modelling for predictive analysis to attempt to close the knowledge gap between state-of-the-art technology and media consumer insights. With the help of Long Short-Term Memory (LSTM) neural networks, a particular kind of recurrent neural network (RNN) renowned for its capacity to recognize long-term dependencies in sequential data, this research focuses on building a predictive model for media consumer behavior. The use of Bayesian optimization techniques to optimize and fine-tune the hyperparameter of the LSTM model, thereby improving its predictive accuracy, is what makes this research novel.

The Bayesian optimization approach efficiently navigates the high-dimensional and complex space of hyperparameter, making it especially useful in situations where traditional methods might not be effective. A Computational Model to Predict Consumer Behaviour was proposed by Safara [12]. Chaudhary et al. [13] proposed big data analytics combined with machine learning-based mathematical modelling to forecast social media user behavior. Zhoa et al. [14] proposed analysing consumer behavior to advance business. Guo et al. [15] proposed a multivariate real-time sequence analysis-based model for predicting consumer behavior. Wibowo et al. [16] proposed Consumer Behavior as a Product of Social Media Marketing: The Contribution of Customer Experience and Social Media Marketing Activities. In order to gauge the quality of the customer-business relationship, which can influence the behavioural outcomes of the customer, such as intent to buy, loyalty intent, and participation intention, this study looked at social media marketing activity (SMMA) and the client experience (CX). David et al. [17] proposed a Neuromarketing analysis using an adaptive machine learning approach. The purpose of this study is to determine which features of package design matter more to consumers when they buy educational toys.

This study employed the idea of big data technology to handle and evaluate data in order to forecast user behavior on social media. We have examined the actions of customers on social media sites using a number of standards and guidelines. We examined how customers felt and viewed the social media network. We use a variety of data pre-processing techniques to identify errors, noise, outliers, and duplicate records in order to obtain high-quality results. Using machine learning, we created mathematical modelling to forecast user behavior on social media. The purpose of this model is to forecast user behavior on social media platforms. The maximum accuracy across the board is 0.98, and the mean square error is 156556.45. This model will produce very subpar results if it is applied to daily data. This model's drawback is that it cannot process consumer data on a daily basis. From the reviewed literature the problem suggests that the computational models that are currently in use to predict consumer behavior using big data analytics and machine learning show encouraging accuracy [18].

Nevertheless, there are still issues that need to be resolved, such as the requirement for different classifiers, effective management of daily customer data, and better understanding of both short- and long-term consumption patterns. This calls for additional study and development in order to resolve these problems and improve predictive models' overall efficacy in predicting and comprehending customer behavior. The need for an efficient and precise predictive model to comprehend and forecast consumer behavior in the media industry is addressed. Utilizing Long-Short-Term Memory (LSTM) as the mathematical model that underpins media consumption data analysis presents a challenge in identifying sequential patterns and dependencies. This entails creating a strong predictive framework that can represent the ever-changing nature of media consumption habits. By giving media professionals and marketers insights into future consumer preferences, engagement patterns, and content consumption trends, the aim is to improve their decision-making processes. Optimizing the model architecture, managing temporal dependencies, and making sure the LSTM-based approach is scalable and flexible enough to accommodate a range of media consumption scenarios are important factors to take into account.

The goal of this work is to improve the LSTM-based model's robustness and efficiency by implementing Bayesian optimization. This will enable the model to more effectively adapt and generalize to the varied and dynamic nature of media consumption patterns. In this work, we develop and implemented a strong media consumer behavior prediction model by utilizing Long Short-Term Memory's (LSTM) gated structure to efficiently handle sequential patterns and long-term dependencies. The main objective is to develop a flexible architecture that can manage various scenarios of media consumption and guarantee reliable and accurate predictions in a variety of situations. Model parameter selection should be stable and well-informed by using Bayesian Optimization (BO) for careful hyperparameter tuning. When making decisions about media consumer behavior analysis, use the BO-LSTM method to produce precise forecasts and insightful analysis. Predictive analysis can be used to predict media consumption trends, which can help media strategists and content producers anticipate and adjust to changing consumer preferences.

2 Proposed BO-LSTM Methodology

The methodology involves collecting data from Kaggle's dataset on consumer buying behavior. In order to enable customized offerings based on particular needs and preferences, the dataset is analyzed to better understand ideal clients. To handle different ranges, numerical features are subjected to Min-Max normalization, which guarantees consistent scaling for machine learning models. RNN use the Long Short-Term Memory architecture to process sequential data, capturing sequential features and predicting subsequent data points. Long-term dependency problems are solved by gated RNNs, particularly LSTM, where gates regulate the flow of information. Input, forget, and output gates are the building blocks of the LSTM model, which controls data input, storage, and output. Bayesian Optimization (BO) is applied to hyperparameter optimization in order to improve the performance of LSTM models over grid and random search techniques. By choosing hyperparameter sets that show promise, BO optimizes the objective function while requiring fewer evaluations. For efficient LSTM-based modelling in media consumer behavior prediction, this method guarantees stable selections, increases prediction accuracy, and sheds light on hyperparameter relationships. The proposed BO-LSTM methodology's block diagram is given in the Figure 1.

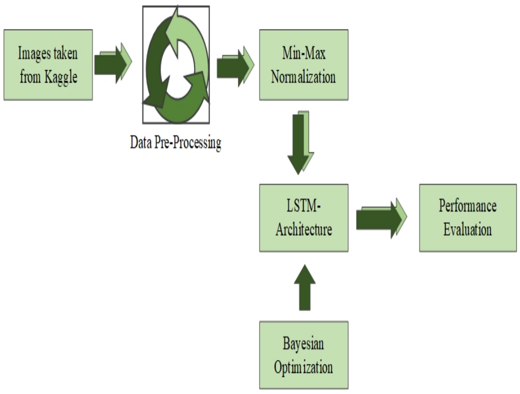

Fig. 1: The proposed BO-LSTM methodology's block diagram.

2.1 Data Collection

The dataset was collected from the Kaggle dataset website [19] .A detailed investigation of a company's ideal customers is conducted through analysing consumer purchasing behavior. It helps a business better understand its customers and makes it easier for them to customize their products to meet the particular needs, preferences, and concerns of different customer segments. Utilizing this analysis, which considers a variety of client categories, a business can modify its product based on its target market. Instead of spending money marketing a new product to every customer in the database, a company could identify which client segment is most likely to buy it and then concentrate only on that particular segment of the market.

2.2 Min-Max Normalization for Data Pre-Processing

Data normalization is a pre-processing approach carried out to mathematical features prior to utilizing techniques for classifying or clustering, which are primarily developed to handle numerical characteristics. In particular, when those values have significantly divergent ranges, the normalization technique aims to avoid some features from concealing the impact of other variables. However, selecting the normalization method and normalizing range (interval) is considered a critical step in the pre-processing stage due to the modification that affects the inspected data and, consequently, the outputs of the machine learning method that will be employed after pre-processing [20].

A method for converting mathematical data into a range, usually between 0 and 1, termed features scaling, also known as min-max normalization. Every feature, or column, in the dataset follows this process. [21]. Scaling numerical data within a given range, usually between 0 and 1, is known as min-max normalization. This approach guarantees, in the absence of an explicit equation, that the lowest and maximum values in the dataset are converted to 0 and 1, respectively, and that all other information points are rescaled linearly with respect to this range [22]. The normalization method scales each data point separately by determining the lowest and greatest values of the feature or column, removing the minimum value, and divided by the range of values. The Eqn. (1) represents min-max normalization:

$$Z_{norm} = \frac{Z - Z_{min}}{Z_{max} - Z_{min}} \tag{1}$$

This guarantees that the values in between are linearly scaled in accordance with the transformation of the minimum value to 0 and the greatest value to 1.In order to provide consistency across the features and support machine learning model performance during training, this normalization strategy is especially helpful when the features have varying scales.

In the data pre-processing stage, min-max normalization is important because it ensures consistent scaling of features like media consumption patterns before feeding them into the RNN. The use of min-max normalization, which addresses the sensitivity of RNNs to the scale of input features and improves training and convergence, is a critical step in the process. Reversing the min-max normalization on the predicted values is also a necessary step in output scaling for a meaningful interpretation of the predictions. The overall effect of min-max normalization includes boosting model convergence during training, which makes RNNs more efficient at identifying and forecasting patterns in the behavior of media consumers. In the predictive analysis of media consumption behavior, min-max normalization is a useful and essential pre-processing step that enhances the performance and interpretability of RNN-based models.

2.3 Architecture of Long Short-Term Memory for Predictive Analysis

Neural network designs with hidden states, or recurrent neural networks (RNNs), process a series of data points that ultimately inform the final output through feedback loops. RNN models can therefore identify sequential features in the data and assist in forecasting the subsequent likely data point in the data sequence. RNN use cases typically involve language models or time-series data analysis, utilizing the capabilities of sequential data processing. Nonetheless, a number of well-known RNN structures have been applied in various experimental contexts, ranging from deep RNN to LSTM and Simple RNN.

Let Y = [y_1,y_2, · · ·, y_t] represent the sequence of data, with t representing the state label. The data at the first state is represented by Y_1, and the data at the t state is represented by Y_t. One way to formulate the Markov assumption in Eqn. (2).

$$P(y_t \mid y_1, \cdots, y_{t-1}) = P(y_t \mid y_{t-1}) \quad (2)$$

Where the conditional probability is expressed by P (·). RNN and HMM are comparable in that they both rely on an earlier state for computation of the current state. Unlike traditional ANNs, the RNN processes sequential data in a circular manner, meaning that each data instance in the sequence will receive the same processing, with each state's outcome depending on the state before it. Nevertheless, assuming sequential data Y = [y_1,y_2, · · ·, y_t] the hidden state r_t can be expressed in Eqn. (3).

$$D_t = f_d(X_{dy}y_t + X_{dd}D_{t-1} + b_t) \quad (3)$$

The bias variable is b_t. The nonlinear function of activation is represented by f_d (·), and the hidden state at time step t is indicated by D_t.Eqn. (4) and the calculation of output at state t are very similar.

$$Z_t = f_z(X_{dz}D_t + b_z) \quad (4)$$

Where X_{dy}represents the weight matrix connecting the output and hidden state. $f_Z(\cdot)$ is the nonlinear activation function, and by is the bias. Since the hidden state (D_t) is computed via forward propagation using the previous state as a basis, it can be thought of as the RNN model's memory. Sequential data from earlier states are also taken into account in the interim. Unlike a traditional neural network, some parameters in such propagation in forward such as the three distinct weight matricesX_{dd}, X_{dy},, and X_{dz}, are shared throughout all steps. By reducing the number of trainable parameters, the parameter-sharing scheme improves the efficiency of the entire computation. Gated RNNs were proposed as a solution to this long-term dependency issue. One kind of gated RNN that was proposed in 1997 is called long short-term memory. LSTM has proven to be an effective model in numerous applications, including image captioning, machine translation, speech recognition, and more, because of its ability to remember long-term dependencies. The RNN's outer recurrence is complemented by an internal loop in the LSTM. The inner loop gradients that occur are context-dependent rather than fixed, and they can flow for extended periods of time. Each cell has the same input and output as a regular RNN, but it also has a system of gated units to regulate the information flow.

LSTM was created using advanced recurrent neuron. Every recurrent neuron in an LSTM can be thought of as a single cell state [23]. The LSTM uses three gates: the forget gate, update gate, and output gate. A particular kind of recurrent neural network identified as the LSTM was created to address the disappearing and explode gradient issue that conventional RNNs faced. The idea was first put forth by Schmidhuber and Hochreiter. Because this model can handle individual data points, it is useful when applied to sequential datasets. The input layer of our LSTM model establishes the dimensionality of the provided data features. Three hidden layers were used, each with 128 memory cells to record the input sequence's long-term dependencies.

An LSTM network has the ability to recall and make connections between data gathered in the past and current. Three gates are coupled with LSTM: an input gate, a forget gate, and an output gate [24].The input is denoted by D_t andD_{t-1}, denotes new and last state respectively, and the current and prior outputs by Z_t and Z_{t-1}.

$$i_t = \sigma(X_i \cdot [Z_{t-1}, y_t] + b_i) \quad (5)$$

$$\tilde{D}_t = \mathrm{tanZ}(X_i \cdot [Z_{t-1}, \mathrm{y}_t] + \mathrm{b}_i) \quad (6)$$

$$\mathrm{D}_t = f_t \mathrm{D}_{t-1} + i_t \tilde{D}_t \quad (7)$$

Where Eqn. (5) determines which piece of data should be added by passing Z_{t-1} and y_t through a sigmoid layer. When Z_{t-1} and $y_t$have travelled through the tanZ layer, Eqn. (6) is then used to get new information. In Eqn. (7), the long-term storage data D_{t-1} into $D_t$and the present moment information, $\tilde{D}_t$ are merged. X_i Denotes a sigmoid output, while $\tilde{D}_t$ stands for tanZ output. Here, b_t stands for the LSTM input gate bias while $X_i$stands for weight matrices. The LSTM's forget gate then enables the dot product and sigmoid layer to selectively pass information. With a certain probability, the choice of whether to delete relevant data from an earlier cell is carried out. Eqn. (8) is used to determine whether or not to retain relevant information from a preceding cell with a particular chance. X_f Stands for weight matrix, b_f for offset, and σ for sigmoid function.

$$f_t = \sigma\ (X_f \cdot [Z_{t-1}, y_t] + \mathrm{b}_f) \quad (8)$$

The output gate of the LSTM ascertains the necessary states for the subsequent Eqns. (9) and (10) states provided by the Z_{t-1} and $y_t$inputs. After obtaining the final output, the state decision vectors that send fresh dataD_t, via the tanz layer are multiply by it.

$$P_t = \sigma(X_o \cdot [Z_{t-1}, y_t] + b_o) \quad (9)$$

$$Z_t = P_t\ \mathrm{tanZ}(\mathrm{D}_t) \quad (10)$$

Where the weighted matrices X_o and the biasb_o, respectively.

The LSTM model, complete with memory cells and gating mechanisms, is depicted in Figure 2 to highlight its complex architecture. Input, output, and forget gates highlight the model's applicability for tasks requiring temporal dependencies by indicating its capacity to extract and store pertinent information over sequential data. The network's ability to address vanishing or exploding gradient problems that are frequently seen in deep learning architectures is made evident by the visualization, which offers a clear understanding of the network's internal workings.

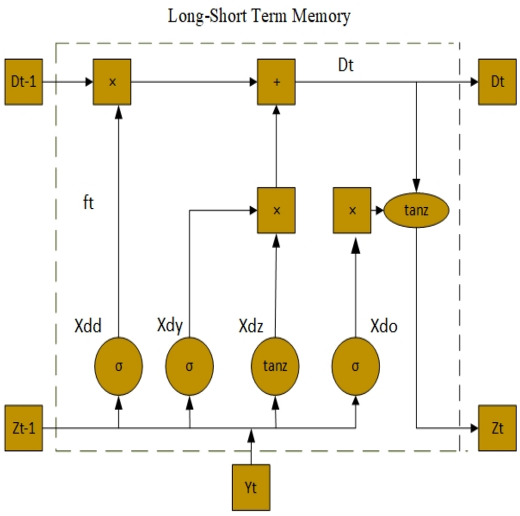

Fig. 2: Architecture of long-short term memory.

2.3.1 Bayesian optimization

Hyper parameter optimization can be used to enhance the effectiveness of algorithms that employ machine learning, particularly deep learning-based predictive models. The most popular techniques for determining the ideal set of hyper parameters to generate more accurate models are grid search and random search. Grid search, however, becomes less effective as the amount of hyper parameters increases, and its computational complexity rises exponentially with each new parameter because it requires an increasing number of evaluations. However, random search can be inefficient in locating the ideal hyper parameter points for some sophisticated models since it combines randomly selected variables based on their statistical distribution.
Without a closed-form expression, Bayesian optimization (BO) is an effective method for solving computationally demanding functions. By selecting only the most promising set of hyper parameters, it reduces the amount of times the goal function needs to be run by building a probability model of the objective function to help determine the optimum hyper parameters in an informed manner. There are two main components to the BO approach. Based on the ideas of Bayes' theorem, BO uses a probabilistic surrogate designs to effectively approximate a desired function at each iteration. The Gaussian process is shown to be a very useful surrogate model that helps identify sets of hyper parameters that show promise for assessment in the true objective function. The objective function is estimated by the surrogate model, which directs further sampling. In order to determine the points in the area of search should be evaluated, BO uses an acquisition function. This allows it to provide information about the ideal setting of an objective function. This function strikes a balance between exploration, which deliberately explores less-explored regions of the search space, and exploitation, which concentrates on areas likely to improve the current approach based on the substitute model. The surrogate model and acquisition function work in tandem to coordinate an iterative process that helps optimize complex functions by skilfully navigating the hyperparameter space and identifying the most promising sets for assessment.

$$x^* = arg\min_{x\varepsilon P} f(x) \tag{11}$$

Eqn. (11) gives an explanation of the way BO solves problems. It does this by identifying those parameters that reduce the objective function in a limited area, with lower and upper bounds on each variable. The score that needs to be minimized is denoted by ?? (??), the domain of the hyper parameter values is P, and the combination of hyper parameters that yields the lowest score ??∗ is$\hat{f}(x)$. Lastly, to attain better performance, BO can be combined with LSTMs. Bayesian Optimization streamlines hyper parameter tuning and outperforms grid and random search techniques as a more effective and efficient way to fine-tune an LSTM model's parameters in media consumer behavior prediction. BO's smart selection of evaluation hyper parameters, based on previous evaluations, improves overall performance and speeds up convergence. This optimization strategy not only improves prediction accuracy but also offers insightful information about the complex relationships between various hyper parameters and the performance of the LSTM model, leading to a deeper comprehension of its behavior. Moreover, the LSTM-based model's hyperparameter selection stability is greatly enhanced by the application of Bayesian Optimization. By methodically examining the hyperparameter space and using probabilistic models to direct the search, BO reduces the possibility of making less-than-ideal decisions, which might otherwise impair the performance of the model. This methodical approach reduces the risks associated with making arbitrary or poorly informed hyperparameter selections, while also improving the efficiency of the mathematical modelling process. Thus, the addition of BO greatly enhances the general dependability and forecast precision of the LSTM-based model for media consumer behavior analysis. The suggested BO-LSTM (Bayesian Optimized Long-Short

Term Memory) method's flowchart is shown in Figure 3. This diagram shows the sequential steps in the methodology and offers a guide for applying the model into practice. Key stages like data pre-processing, model architecture design, hyperparameter tuning, and training are probably depicted in the flowchart.

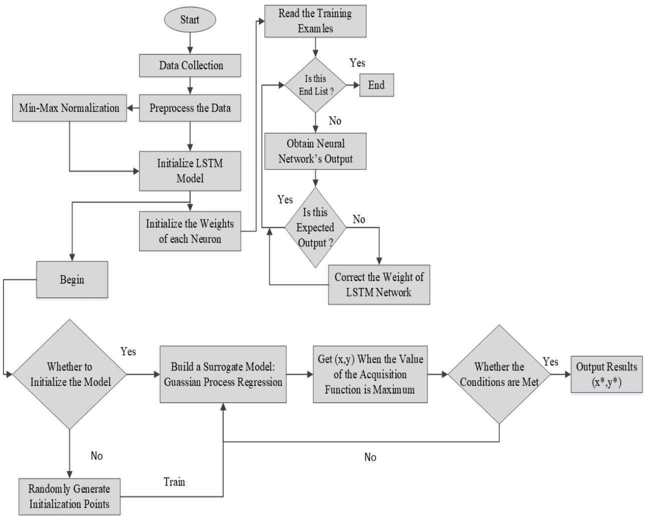

Fig. 3: Flowchart of the BO-LSTM method.

3 Results and Discussion

In the results and discussion section, we delve into the performance evaluation of the proposed Bayesian Optimized-Long Short-Term Memory (BOLSTM) model for predictive analysis in media consumer behavior. The robustness of the model was further demonstrated by the precision, recall, and F1-score metrics. The BOLSTM method has produced notable improvements when compared to traditional approaches. The results' interpretation goes beyond simple numerical measurements to explore the implications for comprehending the behavior of media consumers. The results provide insightful information that can be applied to current research and theoretical frameworks, illuminating the effectiveness of the BOLSTM model in identifying complex patterns and improving forecasts in the ever-changing media consumption landscape. Discussions regarding the model's real-world applications and possible influence on media industry refining strategies are grounded in this thorough analysis.

3.1 Performance Metrics

The predictive analysis of media customer behaviour criteria were segmented using precision, recall, F1-score, and accuracy for comparison. True Positives are the instances that the model correctly classified as belonging to the positive category. Situations that could have been classified as positive but were mistakenly labelled as negative outcomes are known as False Negatives. Since losing a positive occurrence (i.e., a false negative) comes at a significant cost, significant recall is preferred when accurately classifying every case of a particular class. These parameters were used to evaluate the model. They're displayed below.

3.1.1 Accuracy

Comparing the ground truth (actual) labels for your test dataset with the predicted class labels produced by your RNN in order to determine the accuracy. If the projected label matches the actual label for an image in the test dataset, increase the "Number of Correct Predictions." then divide this count by the "Total Number of Predictions" after processing all the test photos to determine the accuracy. Accuracy is commonly determined by applying the following Eqn. (12).

$$Accuracy = \frac{RN + RP}{RP + AP + RN + AN} \tag{12}$$

Where, 'RN' means true negative ;'RP' means true positive ; 'AP' means false positive ; 'RN' means true negative; 'AN' means false negative.Although accuracy offers a clear indicator of the overall performance of the model, it might not be the best metric when the dataset is unbalanced (i.e., one class is significantly more frequent than others).

3.1.2 Precision

By calculating the proportion of correctly predicted positive cases among the predicted positive cases, it assesses the model's accuracy in making positive predictions. The formula for precision in Eqn. (13) is as follows:

$$Precision = \frac{True\ Positives}{(True\ Positives + False\ Positives)} \tag{13}$$

The ratio of accurately predicted positive observations to the total number of predicted positives is the definition of precision, put more simply. When the cost of false positives is high, it offers valuable insights into the model's ability to prevent false positives.

3.1.3 Recall

"Recall" often refers to one of the outcome indicators used to assess the model's efficacy. Sensitivity and true positive rate are other names for recall.. The ratio of true positives to the total of false negatives and true positives is the way it is defined. Recall can be defined as mathematically in Eqn. (14)

$$Recall\,(sensitivity) = \frac{True\ Positives}{True\ Positives + False\ Negatives} \tag{14}$$

The model's sensitivity to detecting real vulnerabilities is revealed by recall. The model's ability to identify and produce exploits for a greater percentage of known vulnerabilities is demonstrated by a higher recall value. In cybersecurity applications, a thorough assessment of a model's performance necessitates striking a balance between precision and recall.

3.1.4 F1-Score

A popular metric for assessing a model's performance in a binary system and information retrieval is the F1 score. It provides a balance between recall and precision by combining both into a single value. The F1 score is particularly useful in datasets that are unbalanced meaning that one class significantly outnumbers the other. The following Eqn. (15) is used to determine the F1 score

$$F1\ Score = 2 * \frac{(Precision * Recall)}{(Precision + Recall)} \tag{15}$$

It has a score between 0 and 1, with a higher F1 score denoting better model performance and a perfect score of 1 denoting optimal precision and recall. Because it takes into account both false positives and false negatives in its computation, the F1 score is especially helpful in situations where there is an uneven class distribution.

Table 1: The suggested method's performance metrics are compared to those of existing methods.

Method	Accuracy (%)	Precision (%)	Recall (%)	F1Score (%)
Random Forests [25]	89	89	89	87
RNN Based Click Stream Model [26]	92.95	92.55	94	88.13
Gradient Tree Boosting Method [27]	89	90.18	85.52	84.63
Proposed BO-LSTM	99	98	98	97

The suggested method's accuracy is displayed in Table 1. It shows the Accuracy (99%), Recall (98%), Precision (98%) and F1-score (97%) of the proposed approach with other methods. The accuracy of the suggested method BO-LSTM (99%) is greater than the traditional approaches Random Forests, RNN Based Click Stream Model and Gradient Tree Boosting Methods.

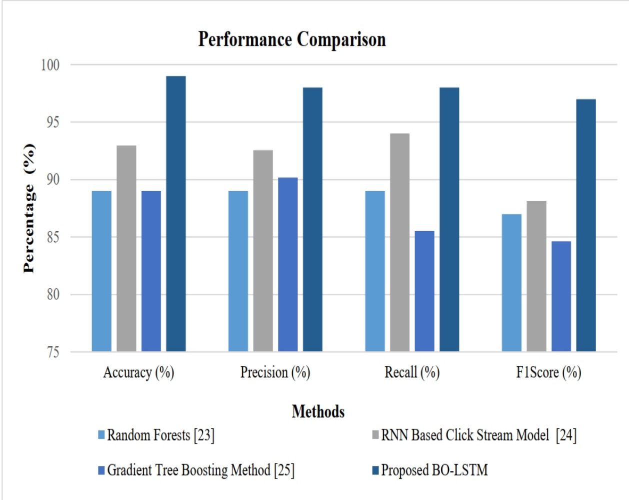

Fig. 4: Visual representation of the performance measures of the suggested BO-LSTM using traditional methods.

Figure 4, shows a graphical representation of the suggested performance metrics in comparison to the current methods. The proposed method BOLSTM demonstrates the greatest accuracy across all the categories Random Forests, RNN Based Click Stream Model and Gradient Tree Boosting Method with 99% high accuracy.

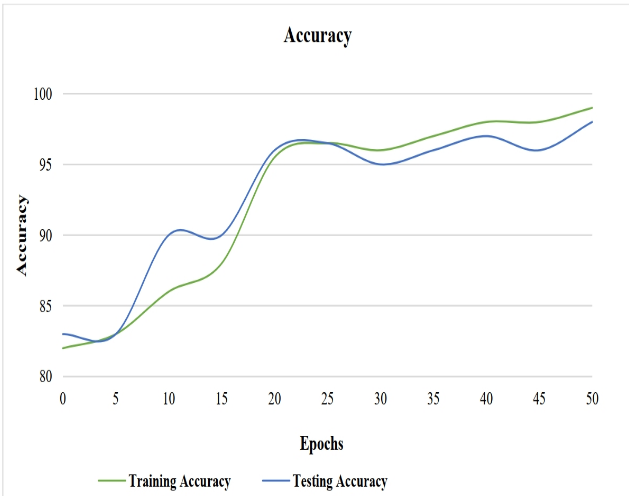

Fig. 5: suggested BO-LSTM method's graphical representation for both training and testing accuracy.

The graphical depiction of the proposed BO-LSTM approach is presented in Figure 5, with particular attention to testing and training accuracy. By displaying the way accuracy changes during training and testing, the visualization offers a clear window into the model's learning process across epochs.

Table 2: The suggested B0-LSTM method's performance metrics.

Proposed BO-LSTM	Percentage (%)
Accuracy	99
Precision	98
Recall	98
F1 score	97

Table 2 presents the excellent results of the BO-LSTM approach, with higher metrics indicating its effectiveness. The outcomes highlight its potential as a reliable solution, indicating that it should be given more thought in a variety of applications.

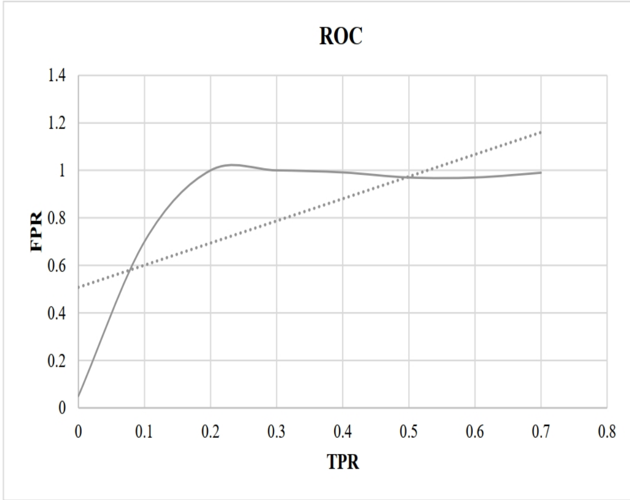

Fig. 7: The suggested BOLSTM method's graphical representation of ROC.

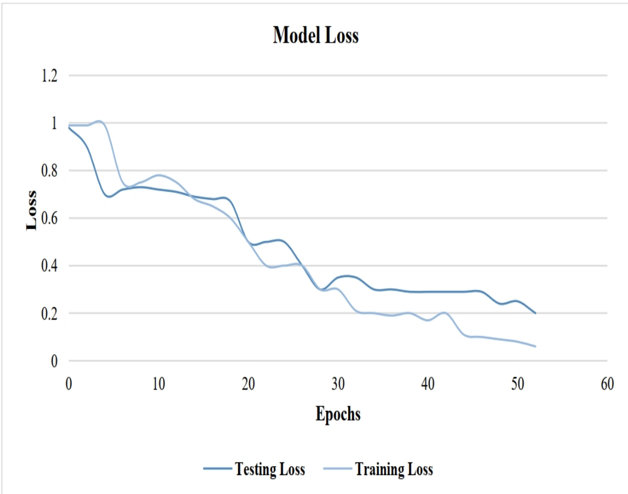

Fig. 6: The proposed BO-LSTM method's training and testing loss is illustrated graphically.

The loss values against number of epochs are shown in Figure 6. It shows the overall loss from the proposed BO-LSTM Model. The training and testing loss graphical representation of the proposed BO-LSTM method gives an understandable overview of the manner in which the model optimizes its performance over epochs and offers a visual insight into the convergence during model training. When evaluating the model's learning dynamics and generalization abilities in media consumer behavior predictive analysis, this graphical representation is valuable.

ROC curve, shown in Figure 7, provides a visual depiction of the effectiveness of the proposed BOLSTM method. The trade-off between true positive rate and false positive rate is graphically represented by the ROC curve, which offers a thorough summary of the discriminatory power of the model. The model's ability to discriminate between positive and negative instances is indicated by its shape and proximity to the upper-left corner of the plot. This information can help practitioners assess and refine the BOLSTM model for the best predictive accuracy in media consumer behavior analysis.
The fitness representation of the proposed BOLSTM technique is shown in Figure 8. The fitness representation functions as a graphic representation of the extent to which the model recognizes trends and forecasts the behavior of media consumers.

3.2 Discussion

A thorough evaluation of the suggested BO-LSTM (Bayesian Optimized-Long Short-Term Memory) approach for media consumer behavior prediction is given by the performance evaluation and graphical representations shown in Figures. With an astounding accuracy of 99%, the BO-LSTM model outperforms conventional techniques like Random Forests [25], RNN Based Click Stream Model [26], and Gradient Tree Boosting Method [27] . Table 1 presents a comparative analysis of accuracy, precision, recall, and F1-score. The BO-LSTM's dominance across all metrics is visually reinforced, which further demonstrates its superiority. The training and testing accuracy, as well as the training and testing loss over epochs, are shown, which offer a

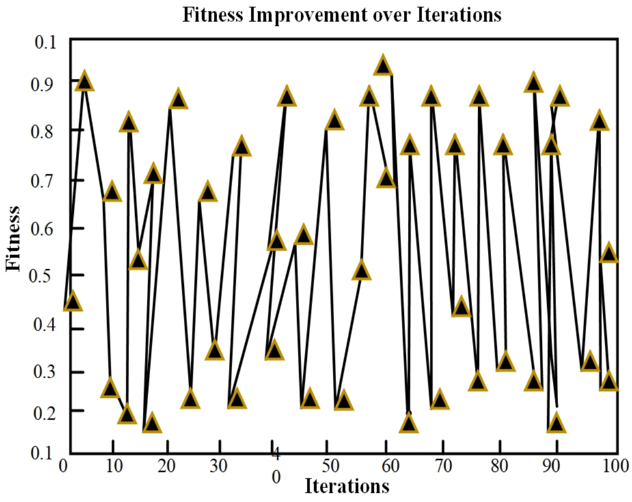

Fig. 8: The suggested BOLSTM method's fitness representation.

detailed understanding of the learning dynamics of the model. The ROC curve, highlighting the BO-LSTM method's ability to discriminate. The model's capacity to identify patterns in media consumer behavior is captured in the fitness representation shown in Figure 8. Overall, these visual aids support the efficacy of the suggested BO-LSTM technique in predictive analysis, providing a solid and trustworthy means of comprehending and forecasting media consumer behavior.

4 Conclusion

In conclusion, the Bayesian Optimized Long Short-Term Memory (BOLSTM) methodology demonstrates itself to be a potent and successful method for modelling and analysing media consumer behavior through predictive analysis. Through the integration of LSTM architecture with Bayesian Optimization for hyperparameter tuning, the proposed method demonstrates significant advancements over traditional approaches. The results underscores the robustness of the BOLSTM model, showcasing improvements in metrics compared to conventional methods. Beyond simple numerical measurements, the results' interpretation offers insightful information about the complex patterns of media consumption behavior. These findings contribute not only to the advancement of research but also offer practical implications for refining strategies within the dynamic landscape of the media industry. Real-time prediction model adaptation would be beneficial to stay up with the ever evolving media consumption patterns, particularly in industry applications that demand fast insights. Ultimately, adding sentiment analysis and user engagement metrics to the analysis would give rise to a more comprehensive predictive model by offering a more nuanced understanding of the emotional factors influencing media consumption decisions. In the ever-changing field of media consumer behavior research, these directions together provide a roadmap for improving and expanding the BOLSTM methodology. By addressing the outlined future scope, researchers can continue to refine and advance our understanding of the intricate dynamics at play in the ever-evolving landscape of media consumption.

Acknowledgments

The authors extend their appreciation to Prince Sattam bin Abdulaziz University for funding this research work through the project number (PSAU/2023/01/25269).

Conflicts of Interest:

The authors declare that they have no conflicts of interest to report regarding the present study.

Data availability

: Data sharing is not applicable to this article as no data sets were generated during the current study.

Ethical approval:

This article does not contain any studies with human participants performed by the authors.

References

[1] C.-M. Leong, A. M.-W. Loi and S. Woon, The influence of social media ewom information on purchase intention, *Journal of Marketing Analytics* **10** (2022) p. 145.
[2] S. Goel, J. M. Hofman, S. Lahaie, D. M. Pennock and D. J. Watts, Predicting consumer behavior with web search, *Proceedings of the National academy of sciences* **107**(41) (2010) 17486–17490.
[3] S. Giglio, E. Pantano, E. Bilotta and T. Melewar, Branding luxury hotels: Evidence from the analysis of consumers'"big" visual data on tripadvisor, *Journal of business research* **119** (2020) 495–501.
[4] T. L. Keiningham, B. Cooil, L. Aksoy, T. W. Andreassen and J. Weiner, The value of different customer satisfaction and loyalty metrics in predicting customer retention, recommendation, and share-of-wallet, *Managing service quality: An international Journal* **17**(4) (2007) 361–384.

[5] P. Rita, T. Oliveira and A. Farisa, The impact of e-service quality and customer satisfaction on customer behavior in online shopping, *Heliyon* **5**(10) (2019).

[6] M. Marghny, R. M. A. ElAziz and A. I. Taloba, Differential search algorithm-based parametric optimization of fuzzy generalized eigenvalue proximal support vector machine, *arXiv preprint arXiv:1501.00728* (2015).

[7] A. Abozeid, R. Alanazi, A. Elhadad, A. I. Taloba, A. El-Aziz, M. Rasha *et al.*, A large-scale dataset and deep learning model for detecting and counting olive trees in satellite imagery, *Computational Intelligence and Neuroscience* **2022** (2022).

[8] D. Gurler and I. Buyukceran, Assessment of the medical reliability of videos on social media: Detailed analysis of the quality and usability of four social media platforms (facebook, instagram, twitter, and youtube), *Healthcare* **10**(10) (2022) p. 1836.

[9] J. E. Siegel, D. C. Erb and S. E. Sarma, A survey of the connected vehicle landscape—architectures, enabling technologies, applications, and development areas, *IEEE Transactions on Intelligent Transportation Systems* **19**(8) (2017) 2391–2406.

[10] M. M. Abdelgwad, T. H. A. Soliman and A. I. Taloba, Arabic aspect sentiment polarity classification using bert, *Journal of Big Data* **9**(1) (2022) 1–15.

[11] P. C. Murschetz, A. Omidi, J. J. Oliver, M. Kamali Saraji and S. Javed, Dynamic capabilities in media management research: a literature review, *Journal of Strategy and Management* **13**(2) (2020) 278–296.

[12] F. Safara, A computational model to predict consumer behaviour during covid-19 pandemic, *Computational Economics* **59**(4) (2022) 1525–1538.

[13] K. Chaudhary, M. Alam, M. S. Al-Rakhami and A. Gumaei, Machine learning-based mathematical modelling for prediction of social media consumer behavior using big data analytics, *Journal of Big Data* **8**(1) (2021) 1–20.

[14] M. Alruily, O. R. Shahin, H. Al-Mahdi and A. I. Taloba, Asymmetric dna encryption and decryption technique for arabic plaintext, *Journal of Ambient Intelligence and Humanized Computing* (2021) 1–17.

[15] L. Guo, B. Zhang and X. Zhao, A consumer behavior prediction model based on multivariate real-time sequence analysis, *Mathematical Problems in Engineering* **2021** (2021) 1–5.

[16] A. Wibowo, S.-C. Chen, U. Wiangin, Y. Ma and A. Ruangkanjanases, Customer behavior as an outcome of social media marketing: The role of social media marketing activity and customer experience, *Sustainability* **13**(1) (2020) p. 189.

[17] A. I. Taloba, A. Elhadad, R. M. A. El-Aziz and O. R. Shahin, Prediction of data threats over web medium using advanced blockchain based information security with crypto strategies, *Journal of Ambient Intelligence and Humanized Computing* (2021) 1–17.

[18] M. Anshari, M. N. Almunawar, S. A. Lim and A. Al-Mudimigh, Customer relationship management and big data enabled: Personalization & customization of services, *Applied Computing and Informatics* **15**(2) (2019) 94–101.

[19] Consumer buying behavior analysis https://kaggle.com/code/aradhanapratap/consumer-buying-behavior-analysis, (2023), Last accessed Dec. 02, 2023.

[20] E. Alshdaifat, D. Alshdaifat, A. Alsarhan, F. Hussein and S. M. F. S. El-Salhi, The effect of preprocessing techniques, applied to numeric features, on classification algorithms' performance, *Data* **6**(2) (2021) p. 11.

[21] S. Yu, J. Wang, J. Liu, R. Sun, S. Kuang and L. Sun, Rapid prediction of respiratory motion based on bidirectional gated recurrent unit network, *IEEE Access* **8** (2020) 49424–49435.

[22] A. Soleimani and S. Khadem, Early fault detection of rotating machinery through chaotic vibration feature extraction of experimental data sets, *Chaos, Solitons & Fractals* **78** (2015) 61–75.

[23] R. M Abd El-Aziz, R. Alanazi, O. R Shahin, A. Elhadad, A. Abozeid, A. I Taloba, R. Alshalabi *et al.*, An effective data science technique for iot-assisted healthcare monitoring system with a rapid adoption of cloud computing, *Computational Intelligence and Neuroscience* **2022** (2022).

[24] M. Z. Islam, M. M. Islam and A. Asraf, A combined deep cnn-lstm network for the detection of novel coronavirus (covid-19) using x-ray images, *Informatics in medicine unlocked* **20** (2020) p. 100412.

[25] I. Ullah, B. Raza, A. K. Malik, M. Imran, S. U. Islam and S. W. Kim, A churn prediction model using random forest: analysis of machine learning techniques for churn prediction and factor identification in telecom sector, *IEEE access* **7** (2019) 60134–60149.

[26] D. Koehn, S. Lessmann and M. Schaal, Predicting online shopping behaviour from clickstream data using deep learning, *Expert Systems with Applications* **150** (2020) p. 113342.

[27] A. Martínez, C. Schmuck, S. Pereverzyev Jr, C. Pirker and M. Haltmeier, A machine learning framework for customer purchase prediction in the non-contractual setting, *European Journal of Operational Research* **281**(3) (2020) 588–596.