A Discretized Recurrent Deep Learning Classifier based on Stochastic Gradient ChatGPT to Improve Lead Conversion Rate

Raghavendra M. Ichangi
Visvesvaraya Technological University, Belagavi-590018, Karnataka, India | Data Science, Sphoorthy Engineering College Hyderabad, Telangana, India
rmichangi@gmail.com (corresponding author)

Shrinivasrao B. Kulkarni
Department of Computer Science and Engineering, SDM College of Engineering and Technology, Dharwad Affiliated to Visvesvaraya Technological University, Belagavi-590018 Karnataka, India
sbkulkarni_in@yahoo.com

Received: 9 November 2024 | Revised: 7 December 2024 and 19 December 2024 | Accepted: 21 December 2024

 DOI: https://doi.org/10.48084/etasr.9549

ABSTRACT

In the vast domain of digital marketing, lead generation forms the foundation for business development. Business strategies depend on converting the leads into customers. It has become very crucial and challenging to choose an appropriate digital platform for marketing. The proposed method, called Stochastic Gradient ChatGPT-based Discretized Recurrent Deep Learning Classification (SG-CDRDLC), employs an efficient way for lead conversion based on influencing feature keywords. The DL classifier with two hidden layers allows companies to determine the popularity of the keywords in the first layer. The second layer measures the keyword density based on a variety of user queries to evaluate and enhance the conversion rate. The proposed model was trained and tested on three datasets and compared against existing methods using accuracy, precision, recall, and training time.

Keywords-digital marketing; deep learning; stochastic gradient; accuracy; precision; recall

I. INTRODUCTION

Search engines are now more equipped to handle complicated search queries and improve search results. This is due to advances in artificial intelligence and semantic strategies. This emphasizes how crucial it is to correctly discover keywords and connect them to the content of web pages. On a wide scale, search engines have emerged as an essential resource for matching user requests with online content. Furthermore, content producers utilize several techniques, including Search Engine Optimization (SEO), to ensure that their information appears prominently in search results.

To increase user awareness of specific content, SEO is crucial for ranking websites and content in search engine results. In [1] two distinct methods, namely Random Forest (RF) and Long Short-Term Memory (LSTM), were applied to handle the prediction of entities. Strongly represented entities were handled by LSTM, whereas weakly represented entities were handled by RF. Social networks are essential for disseminating information and evaluating public and government perspectives. Many previous studies have examined information distribution and opinion analysis. LSTM networks are suitable for speech recognition, natural language processing, and series analysis because they can identify long-term dependencies and patterns. Numerous decision trees are used by Random Forest (RF), combining their predictions, which can also be used to rank the importance of features. In [2], a unique hybrid fuzzy method was developed for opinion analysis and information dispersion. In this case, "hybrid fuzzy" meant combining the concepts of forest fire, fuzzy c-means, and cuckoo search. First, a forest fire was used to obtain diffusers and non-diffusers in social networks. Twitter data were then classified into multiple opinion classes using fuzzy c-means and cuckoo searches, and the change in opinion or lead conversion was determined. The views could be neutral, favorable, or negative.

The process of converting a latent customer, or lead, into a paying customer is known as lead conversion. The number of leads divided by the total number of visits is known as the lead conversion rate. A traditional search engine is unable to develop a brief and ambiguous question based only on the semantic relationship and meaning of keywords. In [3], a Machine Learning (ML) technique, based on Support Vector Machines (SVM), was used to increase web page traffic and continuously improve accuracy over time. In [4], a comprehensive search engine technique combined the benefits of a semantic ontology-based search engine with a keyword-based search engine. Preprocessing was applied to extract keywords. The semantic score was calculated using the ranking approach based on a fuzzy membership function. Notable results were grouped, increasing the total accuracy rate, but training time was not considered. In [5], many ML techniques were presented, using search engine settings to extract highly significant information. The binary classification algorithm found the most relevant attributes for the search engine results. However, precision was not taken into account. In [6], a comprehensive evaluation of fine-tuning methods using encoders and decoders was presented to increase the quality of text summarization. In [7], a novel approach was proposed to classify ML algorithms and identify factors to predict the SEO quality of a web page.

This study presents SG-CDRDLC to address these aspects. This approach uses an ML model to define the challenge of lead generation. This model reduces the total training time by generating leads in a formative manner using an ML algorithm based on stochastic gradient keyword extraction. Furthermore, a conversion model that uses the ChatGPT-based discretized recurrent Deep Learning (DL) classification method intends to greatly increase recall and precision. The proposed SG-CDRDLC method aims to address these problems, offering the following contributions:

- Improves lead generation and converts it to users using two main processes: lead generation and conversion.
- Uses an information gain ratio stochastic function in conjunction with a stochastic gradient keyword extraction ML model to extract relevant feature keywords from three distinct datasets. This shortens the training period and improves the accuracy of the lead conversion rate. The proposed SG-CDRDLC technique uses a ChatGPT-based discretized recurrent DL classifier model to obtain an accurate lead-to-sale conversion rate.
- The DL classifier uses multiple layers for a thorough analysis of the most important keywords. Then, by extensively optimizing the lead conversion rate, the max-out activation function gives the predicted results.
- A detailed experimental evaluation details how the proposed method outperforms traditional approaches in terms of precision, recall, accuracy, and training time.

The importance of web page elements that affect search engine ranking has been well-studied in the field of SEO. Research processes that are specifically used in the context of SEO are based on assessments of high-ranking web pages and their characteristics, including how content modifications affect rank. In [8], latent semantic analysis was used for keyword extraction in SEO, but lead generation was not considered. SEO is useful not only for consumable items but also for the medical industry. In [9], a comprehensive correlation-based SEO mechanism was presented for disease diagnosis. In [10], a deep CNN was used to perform SEO for biomedical images. The similarity results were retrieved by combining the vector space model with the CNN biomedical image query. In [11], ML and DL were used to predict Influenza-Like Illnesses (ILI), using Google Trends (GT) data and keywords based on symptoms. Time-series modeling was employed to produce forecast results more accurately. In [12], a graphical CNN-based autoencoder was proposed to extract complex connectivity patterns, and a Min-Max game was employed to improve the recommendations. In [13], analytical models were used to identify optimal sales compensation designs to solve issues related to multichannel attribution. In [14], the focus was on recommendation systems that use DL techniques for various domains, including social networks. In [15], another ontology-based recommendation method focused on the domain of online social networks to reduce the budget and time consumed in strategy definition. In [16], a systematic review of social recommendation systems was presented, employing DL techniques for social network users.

This study proposes a method based on the SG-CDRDLC approach, focusing on both social media and website SEO.

II. METHODOLOGY

The selection of keywords for search engine positioning or keyword extraction is crucial for drawing in customers when it comes to digital marketing planning. The associated expenses are becoming a more significant portion of the marketing budget. The entire project is divided into two parts: lead creation through keyword extraction and user acquisition through classification. This study examined and verified datasets from two social media platforms (Twitter and Instagram) and a website SEO dataset. Figure 1 presents the flow diagram for the SG-CDRDLC approach. Keywords are the first step in the SEO process. When searching for specific information in search engines, internet users enter one or more keywords, as the use of keywords is still more common than the use of questions. When using a search engine, the user may occasionally enter phrases in place of words or keywords. Therefore, the core principle of SEO still revolves around determining the keywords that would elevate the website's position in Search Engine Result Pages (SERP).

This study used three datasets, namely website SEO [17], the Instagram Influencer Dataset [18], and a Twitter dataset [19], to train and test the model. The first dataset includes SEO sample data with numerous features that influence the ranking of selected words in the search engine. The Instagram influencer dataset consists of 33,935 Instagram influencers classified into nine classes, namely beauty, family, fashion, fitness, food, interior, pet, travel, etc. To perform the simulation, 300 posts per influencer were collected, resulting in a total of 10,180,500 Instagram posts. The Instagram dataset has two file types, post metadata and image files. Post metadata files include information such as caption, user tags, hashtags, timestamp, sponsorship, likes, and comments. In addition, it includes 12,933,406 image files, since a post can have more than one image. Finally, the Twitter dataset [19] provides a resource to investigate the dynamics of social media communication, providing a comprehensive analysis of the sentiments, trends, and behavioral patterns of users within the Twitterverse. The dataset presents the evolution of hashtags and influential users, providing awareness of the oscillating landscape of online conversations. In addition, it includes significant tweet information, with each row denoting a single tweet, such as tweet ID, user name, tweet, text, count, timestamp, and retweet count.

The proliferation of information available to customers and the accelerating growth of search makes it harder to find data over the Internet. The topics that are popular on Twitter as well as Instagram are ever-changing. Moreover, numerous elements influence search engine ranking, making it a difficult and time-consuming process to improve search engine ranking. This study builds a decision-making model based on stochastic gradient keyword extraction, which can assist website, Instagram, and Twitter administrators improve their SEO. It does this by employing Twitter, Instagram influencers, and SEO sample data as an empirical example. To improve website performance and satisfy customer demands, the stochastic gradient ML model initiates a search engine ranking determination procedure, as shown in Figure 2. To determine the significant weights and relationships between SEO indicators, as well as the performance gaps that need to be closed to reach the desired level, this study employs an ML model based on stochastic gradient extraction as its foundation.

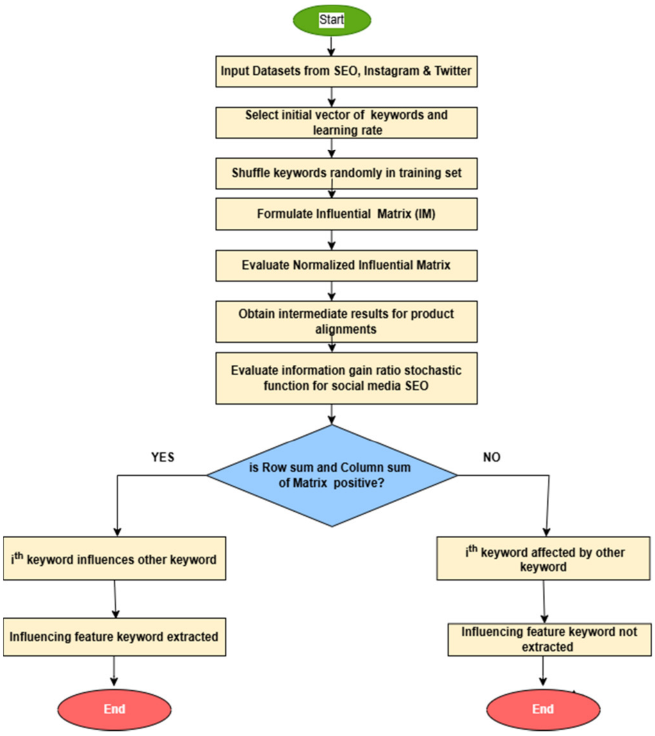

Fig. 1. Flowchart for the SG-CDRDLC method.

Using the raw data gathered from the website, Instagram, and Twitter, the Influential Matrices (IM) are initially generated as:

$$IM[WS] = \frac{1}{l^2}\sum_{i=1}^{u}\sum_{j=1}^{n}\frac{\left[AL_{lj}^{n}[WS]-AI_{lj}^{n-1}[WS]\right]}{AI_{lj}[WS]} * 100 \quad (1)$$

$$IM[T] = \frac{1}{l^2}\sum_{i=1}^{u}\sum_{j=1}^{n}\frac{\left[AL_{lj}^{n}[T]-AI_{lj}^{n-1}[T]\right]}{AI_{lj}[T]} * 100 \quad (2)$$

$$IM[Ins] = \frac{1}{l^2}\sum_{i=1}^{u}\sum_{j=1}^{n}\frac{\left[AL_{lj}^{n}[Ins]-AI_{lj}^{n-1}[Ins]\right]}{AI_{lj}[Ins]} * 100 \quad (3)$$

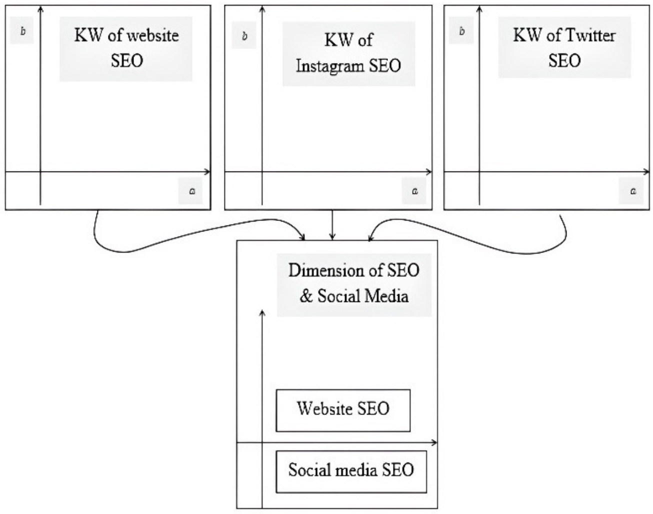

Fig. 2. Analytical framework of the SG-CDRDLC model.

The corresponding IM for the website $IM[WS]$ is formulated based on the number of sample n keywords, and AL_{lj}^{n} represents the average influence of l keyword on j. Although similar mathematical formulations are made for the website and two social media platforms (i.e., Instagram and Twitter), the keywords are different (i.e., words in the case of a website, re-tweets in the case of Twitter, and categories in the case of Instagram). Using the previously acquired resultant IM, the Normalized IM ($NormIM$) is evaluated separately for social media (Twitter and Instagram) and website SEO as:

$$NormIM\ (WS) = IR(WS) * IM(WS) \quad (4)$$

$$NormIM\ (T) = IR(T) * IM(T) \quad (5)$$

$$NormIM\ (Ins) = IR(Ins) * IM(Ins) \quad (6)$$

The Interim Results (IR) in $NormIM$ can be generated as follows:

$$IR(WS) = min\left[\frac{1}{max\sum_{i=1}^{m}[IM_{ij}]}, \frac{1}{max\sum_{j=1}^{n}[IM_{ij}]}\right] \quad (7)$$

$$IR(T) = (IG(RK[T], KW)/SI(RK[T])) \quad (8)$$

$$IR(Ins) = (IG(RK[Ins], KW)/SI(RK[Ins])) \quad (9)$$

The Information Gain (IG) ratio and Split Information (SI) are determined as:

$$IG(RK[T], KW) = H(RK[T]) - H(RK[T]|KW) \quad (10)$$

$$SI(RK[T]) = -\sum_{j=1}^{n}\frac{N(RK_j[T])}{N(RK[T])} * log_2\frac{N(RK_j[T])}{N(RK[T])} \quad (11)$$

$$IG(RK[Ins], KW) =$$
$$H(RK[Ins]) - H(RK[Ins]|KW) \quad (12)$$

$$SI(RK[Ins]) =$$

$$-\sum_{k=1}^{u}\frac{N(RK_j[Ins])}{N(RK[Ins])} * log_2 \frac{N(RK_j[Ins])}{N(RK[Ins])} \quad (13)$$

Finally, the overall influencing keywords are given by the Total IM (TIM), mathematically represented as:

$$a = [a_i]_{n*1} = \left[\sum_{i=1}^{m} TIM_{ij}(WS)\right]_{n*1} \quad (14)$$

$$b = [b_j]_{n*1} = \left[\sum_{j=1}^{n} TIM_{ij}(WS)\right]_{1*n} \quad (15)$$

where a_i and b_j denotes the row sum and column sum of the i^{th} row and j^{th} column in matrix TIM.

The proposed classifier is based on Recurrent Deep Learning (RDL) and is designed to analyze and validate the conversion rate of leads to customers or users, including website SEO and social media SEO along with image and multimedia content. With RDL composed of interconnected layers of nodes, called neurons, that exercise and transmit information, the ChatGPT-based Discretized RDL Classifier takes a string of text as input (i.e., influencing feature keywords) and produces a response as output. Also, since RDL classifiers are complex mathematical functions that require numerical data as input, the influencing feature keywords in ChatGPT are initially encoded into numerical data before being fed into the network.

To start with the influencing feature keywords obtained from the stochastic gradient keyword extraction-based machine learning model as given in the above section, $IK[WS], IK[T], IK[Ins]$ is stored in ChatGPT as given below for further processing:

$$IK[WS], IK[T], IK[Ins] \rightarrow ChatGPT \rightarrow Input\ Vector \quad (16)$$

The influencing feature keywords of both website SEO and social media SEO are stored in the form of an Input Vector in ChatGPT through Bidirectional Encoder Representations from Transformers (BERTs).

III. EXPERIMENTAL SETUP AND RESULTS

The proposed method considers different numbers of samples in the range of 1000 to 10000 for the experimental process. Table I shows the hyperparameters for the proposed model. The SEO sample data, Instagram influencer dataset, and Twitter dataset were divided into two sets for training and testing. For each dataset, 70% is used for training, and the remaining 30% is used for testing purposes. The performance of the SG-CDRDLC method was determined in terms of precision, recall, accuracy, and training time. Precision refers to the ratio of relevant instances (i.e., leads converted into users) among the retrieved instances (i.e., retrieved leads), and Recall refers to the ratio of relevant instances that were retrieved. Confusion matrices were used for all three datasets for a sample size of 1000.

$$Precision = \frac{TP}{TP+FP} \quad (17)$$

$$Recall = \frac{TP}{TP+FN} \quad (18)$$

TABLE I. HYPER-PARAMETERS

	Hyperparameter	Description
1	Number of layers	Three layers
2	Number of hidden layers used	Two hidden layers. First layer: Popularity Second layer: Keyword density
3	Function used in the output layer	Maxout activation function
4	Learning rate	0.01
5	Number of epochs	10
6	Weight	0.1
7	Batch size	5000 as samples for simulation.

TABLE II. CONFUSION MATRIX FOR SEO SAMPLE DATASET

Predicted values	Actual values		Total samples
	Positive	Negative	
Positive	TP=920	FP=30	950
Negative	FN=10	TN=40	50
Total samples	930	70	1000

As shown in Table II, 920 samples out of 1000 were classified as true positive, which results in 96.82% precision and 98.92 % recall for the SEO sample dataset. Similarly, Tables III and IV show the confusion matrices for the other two datasets. Precision and recall were 94.73% and 98.57% for the Instagram influencer dataset and 95.78% and 98.69% for the Twitter dataset, respectively.

TABLE III. CONFUSION MATRIX FOR THE INSTAGRAM INFLUENCER DATASET

Predicted values	Actual values		Total samples
	Positive	Negative	
Positive	TP=900	FP=50	950
Negative	FN=13	TN=37	50
Total samples	913	87	1000

TABLE IV. CONFUSION MATRIX FOR THE TWITTER DATASET

Predicted values	Actual values		Total samples
	Positive	Negative	
Positive	TP=910	FP=40	950
Negative	FN=12	TN=38	50
Total samples	922	78	10000

The proposed SG-CDRDLC model was compared against Long Short-Term Memory and Random Forest (LSTM-RF) and Hybrid Fuzzy methods for various sample sizes ranging from 1000 to 10000 on the three datasets. Table V shows the precision values. The proposed model exhibits better performance compared to the other two methods. Similarly, Table VI shows the recall for the proposed and the other two methods for sample sizes ranging from 1000 to 10000. In the case of recall, the SG-CDRDLC model also performs well. This study focused on precision and recall since the datasets are imbalanced. Other metrics, such as accuracy and training time were also used to evaluate the performance of the proposed over the other two methods.

TABLE V. PRECISION ESTIMATION

Samples	Precision								
	SEO data			Instagram			Twitter		
	SG-CDRDLC	LSTM-RF	Hybrid fuzzy	SG-CDRDLC	LSTM-RF	Hybrid fuzzy	SG-CDRDLC	LSTM-RF	Hybrid fuzzy
1000	0.96	0.9	0.9	0.94	0.9	0.9	0.95	0.9	0.9
2000	0.93	0.9	0.8	0.9	0.8	0.8	0.8	0.7	0.7
3000	0.9	0.8	0.8	0.87	0.8	0.8	0.77	0.7	0.7
4000	0.88	0.8	0.8	0.85	0.8	0.7	0.75	0.7	0.6
5000	0.85	0.8	0.8	0.82	0.7	0.7	0.72	0.6	0.6
6000	0.83	0.8	0.7	0.8	0.7	0.7	0.7	0.6	0.6
7000	0.85	0.8	0.7	0.82	0.7	0.7	0.72	0.6	0.6
8000	0.87	0.8	0.7	0.84	0.7	0.7	0.74	0.6	0.6
9000	0.88	0.8	0.7	0.85	0.8	0.7	0.75	0.7	0.6
10000	0.92	0.8	0.8	0.89	0.8	0.7	0.79	0.7	0.6

TABLE VI. RECALL ESTIMATION

Samples	Recall								
	SEO sample data			Instagram influencer			Twitter		
	SG-CDRDLC	LSTM-RF	Hybrid fuzzy	SG-CDRDLC	LSTM-RF	Hybrid fuzzy	SG-CDRDLC	LSTM-RF	Hybrid fuzzy
1000	0.98	0.9	0.9	0.98	0.9	0.9	0.98	0.9	0.9
2000	0.96	0.9	0.8	0.93	0.8	0.8	0.83	0.7	0.7
3000	0.93	0.9	0.8	0.9	0.8	0.7	0.8	0.7	0.7
4000	0.91	0.8	0.8	0.88	0.8	0.7	0.78	0.7	0.7
5000	0.88	0.8	0.8	0.85	0.8	0.7	0.75	0.7	0.6
6000	0.85	0.8	0.7	0.82	0.7	0.6	0.72	0.6	0.6
7000	0.88	0.8	0.8	0.79	0.7	0.6	0.69	0.6	0.6
8000	0.89	0.8	0.8	0.82	0.7	0.6	0.72	0.6	0.6
9000	0.91	0.8	0.8	0.84	0.7	0.7	0.74	0.6	0.6
10000	0.93	0.9	0.8	0.85	0.8	0.7	0.75	0.7	0.6

Accuracy determines the correct predictions by the model, which can be calculated by:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{19}$$

Table VII shows the accuracy for all three models, where SG-CDRDLC outperformed the others.

TABLE VII. ACCURACY COMPARISON

Samples	Accuracy (%)								
	SEO sample data			Instagram influencer			Twitter		
	SG-CDRDLC	LSTM-RF	Hybrid fuzzy	SG-CDRDLC	LSTM-RF	Hybrid fuzzy	SG-CDRDLC	LSTM-RF	Hybrid fuzzy
1000	96	93	90	93	91	89	94	92	89
2000	94	90	85	90	82	72	92	84	74
3000	92	88	83	87	79	69	89	81	81
4000	90	86	81	84	74	64	86	76	76
5000	88	84	79	82	73	63	84	75	75
6000	87	83	78	80	72	62	82	74	74
7000	85	81	76	78	70	60	80	72	72
8000	89	85	80	81	73	63	83	75	75
9000	91	87	82	83	75	65	85	77	77
10000	93	89	84	85	77	67	87	79	79

The training time (TT) was measured based on the samples considered for experimentation, denoted as S_i, and the influencing feature keywords for the corresponding datasets. It was measured in terms of milliseconds (ms).

$$TT = \sum_{i=1}^{m} S_i * Time[IK[WS]] \tag{20}$$

Table VIII shows the training time for the SG-CDRDLC model, which requires substantially less time compared to the other two models for various sample sizes.

TABLE VIII. TRAINING TIME FOR THE MODEL

Samples	Training time (ms)								
	SEO sample data			Instagram influencer			Twitter		
	SG-CDRDLC	LSTM-RF	Hybrid fuzzy	SG-CDRDLC	LSTM-RF	Hybrid fuzzy	SG-CDRDLC	LSTM-RF	Hybrid fuzzy
1000	35	42	51	39	47	55	44	53	62
2000	45	60	70	53	70	85	58	65	70
3000	55	75	95	65	95	115	65	90	100
4000	75	90	100	80	115	135	80	115	123
5000	85	115	125	95	130	145	95	140	155
6000	105	135	148	115	145	168	105	155	168
7000	125	155	175	123	155	175	125	170	185
8000	140	170	205	145	170	195	145	195	215
9000	155	195	225	165	190	213	155	215	240
10000	170	205	240	180	215	225	180	225	255

IV. DISCUSSION

Figure 3 shows that there was no direct or inverse relationship between the precision rate and the input samples. Additionally, the three methods had greater precision on the SEO sample dataset than the social media datasets. From Figure 4, it can be deduced that recall was neither rising nor falling with growing input samples. Furthermore, it was discovered that the recall rates for all three methods for the SEO sample dataset were superior to those of the other two. The highly influential keywords or leads produced by the machine learning model were used to create the generated leads, which were then divided into tokens and kept in the input layer of the ChatGPT-based discretized recurrent network.

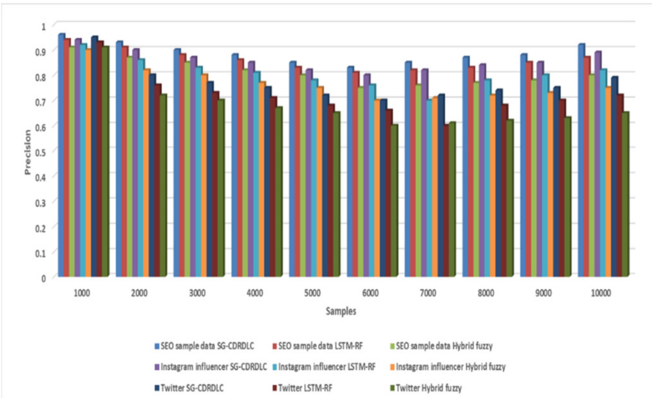

Fig. 3. Precision vs samples.

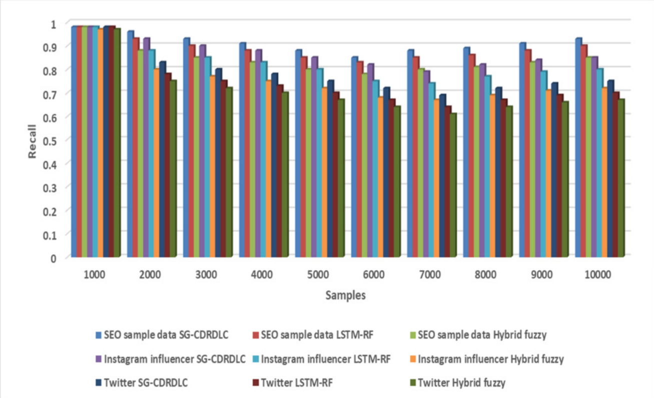

Fig. 4. Recall vs samples.

The lead conversion to user precision for the SEO dataset utilizing the SG-CDRDLC approach was improved by 3% and 10% over LSTM-RF and Hybrid Fuzzy, respectively. Additionally, using the Instagram influencer dataset, improvements of 7% and 12% were noticed compared to LSTM-RF and Hybrid Fuzzy, and finally, 8% and 14% improvements were noticed on the Twitter dataset. Compared to LSTM-RF and Hybrid Fuzzy using the SEO sample data, the Instagram influencer dataset, and the Twitter dataset, SG-CDRDLC had an improved lead conversion rate in terms of recall by 3%, 9%, and 6%, respectively.

V. CONCLUSION

This study proposed a stochastic gradient model to extract influencing feature keywords based on user preferences to improve lead conversion. The proposed SG-CDRDLC model uses a stochastic function to choose the maximum occurrences of the keywords in a computationally efficient manner. This significantly minimizes the time to extract products aligning with user preferences compared to traditional methods such as LSTM-RF and Hybrid Fuzzy. The model also ensures increased accuracy in lead conversion rates by identifying and reducing the ratio of incorrectly identified keywords. The deep learning-based classifier examines the influential keywords to boost the lead conversion rate by incorporating two layers that emphasize the popularity and density of the keywords. The Max-out activation function is employed to provide prediction results by calculating the keyword density. The model was compared for various performance metrics, such as precision, recall, accuracy, and training time, and more efficient than LSTM-RF and Hybrid Fuzzy models.

REFERENCES

[1] B. U. D. Abbasi, I. Fatima, H. Mukhtar, S. Khan, A. Alhumam, and H. F. Ahmad, "Autonomous schema markups based on intelligent computing for search engine optimization," *PeerJ Computer Science*, vol. 8, Dec. 2022, Art. no. e1163, https://doi.org/10.7717/peerj-cs.1163.

[2] S. E. Uthirapathy and D. Sandanam, "Predicting opinion evolution based on information diffusion in social networks using a hybrid fuzzy based approach," *International Journal of Information Technology*, vol. 15, no. 1, pp. 87–100, Jan. 2023, https://doi.org/10.1007/s41870-022-01109-2.

[3] J. More, V. Gaikwad, J. Lahange, S. Shinde, and J. Y. Kapadnis, "Design and Implement Search Engine Optimization Prediction System Using Machine Learning," *International Journal of Innovative Research in Engineering & Multidisciplinary Physical Sciences*, vol. 11, no. 3, May 2023.

[4] M. M. El-Gayar, N. E. Mekky, A. Atwan, and H. Soliman, "Enhanced Search Engine Using Proposed Framework and Ranking Algorithm Based on Semantic Relations," *IEEE Access*, vol. 7, pp. 139337–139349, 2019, https://doi.org/10.1109/ACCESS.2019.2941937.

[5] W. K. Portier, Y. Li, and B. A. Kouassi, "Feature Selection using Machine Learning Techniques Based on Search Engine Parameters," in *Proceedings of the 2020 3rd International Conference on Signal Processing and Machine Learning*, Beijing, China, Oct. 2020, pp. 28–34, https://doi.org/10.1145/3432291.3432308.

[6] C. Anastasiu, H. Behnke, S. Lück, V. Malesevic, A. Najmi, and J. Poveda-Panter, "DeepTitle -- Leveraging BERT to generate Search Engine Optimized Headlines." arXiv, 2021, https://doi.org/10.48550/ARXIV.2107.10935.

[7] G. Matošević, J. Dobša, and D. Mladenić, "Using Machine Learning for Web Page Classification in Search Engine Optimization," *Future Internet*, vol. 13, no. 1, Jan. 2021, Art. no. 9, https://doi.org/10.3390/fi13010009.

[8] M. R. Lam, G. N. Manion, and B. K. Young, "Search engine optimization and its association with readability and accessibility of diabetic retinopathy websites," *Graefe's Archive for Clinical and Experimental Ophthalmology*, vol. 262, no. 9, pp. 3047–3052, Sep. 2024, https://doi.org/10.1007/s00417-024-06472-3.

[9] M. Amara, R. Zaghdoud, O. B. Rhaiem, and E. Althubiti, "Enhancing Image SEO using Deep Learning Algorithms: A Research Approach," *Engineering, Technology & Applied Science Research*, vol. 14, no. 4, pp. 15194–15200, Aug. 2024, https://doi.org/10.48084/etasr.7473.

[10] R. Mishra and S. P. Tripathi, "Deep learning based search engine for biomedical images using convolutional neural networks," *Multimedia Tools and Applications*, vol. 80, no. 10, pp. 15057–15065, Apr. 2021, https://doi.org/10.1007/s11042-020-10391-w.

[11] S. O. Olukanmi, F. V. Nelwamondo, and N. I. Nwulu, "Utilizing Google Search Data With Deep Learning, Machine Learning and Time Series Modeling to Forecast Influenza-Like Illnesses in South Africa," *IEEE Access*, vol. 9, pp. 126822–126836, 2021, https://doi.org/10.1109/ACCESS.2021.3110972.

[12] J. Yu, H. Yin, J. Li, M. Gao, Z. Huang, and L. Cui, "Enhancing Social Recommendation With Adversarial Graph Convolutional Networks," *IEEE Transactions on Knowledge and Data Engineering*, vol. 34, no. 8, pp. 3727–3739, Aug. 2022, https://doi.org/10.1109/TKDE.2020.3033673.

[13] S. Banerjee and P. Bhardwaj, "Aligning marketing and sales in multi-channel marketing: Compensation design for online lead generation and offline sales conversion," *Journal of Business Research*, vol. 105, pp. 293–305, Dec. 2019, https://doi.org/10.1016/j.jbusres.2019.06.016.

[14] C. Li, I. Ishak, H. Ibrahim, M. Zolkepli, F. Sidi, and C. Li, "Deep Learning-Based Recommendation System: Systematic Review and Classification," *IEEE Access*, vol. 11, pp. 113790–113835, 2023, https://doi.org/10.1109/ACCESS.2023.3323353.

[15] M. Arafeh, P. Ceravolo, A. Mourad, E. Damiani, and E. Bellini, "Ontology based recommender system using social network data," *Future Generation Computer Systems*, vol. 115, pp. 769–779, Feb. 2021, https://doi.org/10.1016/j.future.2020.09.030.

[16] M. Alrashidi, A. Selamat, R. Ibrahim, and O. Krejcar, "Social Recommendation for Social Networks Using Deep Learning Approach: A Systematic Review, Taxonomy, Issues, and Future Directions," *IEEE Access*, vol. 11, pp. 63874–63894, 2023, https://doi.org/10.1109/ACCESS.2023.3276988.

[17] M. Varli, "SEO Sample Data." Kaggle, [Online]. Available: https://www.kaggle.com/datasets/muhammetvarl/seo-sample-data.

[18] S. Kim, J. Y. Jiang, M. Nakada, J. Han, and W. Wang, "Multimodal Post Attentive Profiling for Influencer Marketing," in *Proceedings of The Web Conference 2020*, Taipei Taiwan, Apr. 2020, pp. 2878–2884, https://doi.org/10.1145/3366423.3380052.

[19] A. Goyal, "Twitter-Dataset." Kaggle, [Online]. Available: https://www.kaggle.com/datasets/goyaladi/twitter-dataset.