

The Use of Recurrent Nets for the Prediction of e-Commerce Sales

Eman Aldhahri

Department of Computer Science and Artificial Intelligence, College of Computer Science and Engineering, University of Jeddah, Saudi Arabia
eaal-dhahery@uj.edu.sa (corresponding author)

Received: 16 April 2023 | Revised: 27 April 2023 and 28 April 2023 | Accepted: 28 April 2023

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | DOI: <https://doi.org/10.48084/etasr.5964>

ABSTRACT

The increase in e-commerce sales and profits has been a source of much anxiety over the years. Due to the advances in Internet technology, more and more people choose to shop online. Online retailers can improve customer satisfaction using sentiment analysis in comments and reviews to gain higher profits. This study used Recurrent Neural Networks (RNNs) to predict future sales from previous using the Kaggle dataset. A Bidirectional Long Short Term Memory (BLTSM) RNN was employed by tuning various hyperparameters to improve accuracy. The results showed that this BLTSM model of the RNN was quite accurate at predicting future sales performance.

Keywords-e-commerce; sales; deep learning; prediction; RNNs

I. INTRODUCTION

The term e-commerce refers to any commercial transaction that takes place over the Internet. E-commerce is divided into "non-payment e-commerce", which does not involve the exchange of money or the delivery of products to customers, and "payment e-commerce", which enables customers to make purchases and arrange shipments of items online [1]. Along with the rest, "non-payment e-commerce", includes actions performed by delivery services and activities performed by vendors related to the sale of items [2]. E-commerce can be broken down into four distinct sub-industries: business-to-business (B2B), business-to-consumer (B2C), business-to-government (B2G), and consumer-to-government (C2G). The first two types of electronic business transactions that come to mind are business-to-business (B2B) and business-to-consumer (B2C) [3]. E-commerce is often used to describe a new style of business operations that is distinguished from traditional by the fact that customers buy products or services using a browser and a server-based application hosted on an Internet platform [1]. Since 2013, there is a substantial rise in the use of the Internet to deliver high-quality services and the frequency of electronic commercial transactions continues to grow [4]. Traditional methods of sales forecasting are based on time series analysis and usually accept only input sales data from the most recent quarters. Using these tactics, it is possible to effectively manage commodities whose sales trends are either consistent or seasonal [5]. In contrast to this, the prices of commodities sold through e-commerce are much higher and the sales trends of these commodities are highly unpredictable. In addition, the accuracy of the forecasts obtained using traditional methods is often excessively poor [6].

It is feasible to boost the accuracy of forecasts using the huge amount of data available in the e-commerce industry. Feature engineering is the first step in traditional machine-learning approaches. This process involves manually extracting valuable features from easily accessible data [7]. The accuracy of a forecast model can be strongly impacted, both positively and negatively, by the quality and number of attributes contained within the model. However, the process of generating useful features is both challenging and time-consuming. In addition, these attributes are often acquired on an individual basis for one-of-a-kind business scenarios, which makes it difficult to reuse the models if the data or requirements change in the future. When gathering further data for online sales, feature engineering should be carried out to include the information contained in the new data in the prediction model. It is possible that feature learning could eliminate the requirement to manually engineer features [6]. Numerous studies have been conducted to extract features from unstructured data, such as photos, audio, text, and others, using deep neural networks. [8].

Many studies have investigated reliable and accurate sales prediction, and most of them agreed that it is a challenging issue in the field of e-commerce. In [2], a method was presented to use information based on cross-series using the unified model. A Long Short-Term Memory (LSTM) network was trained using the relationships based on the non-linearity in e-commerce product variety. Moreover, a framework was presented for systematic preprocessing to address e-commerce challenges. Several product grouping strategies have also been introduced to improve LSTM learning. In [3], an LSTM-based model was presented for consumer comments sentiment analysis to predict the short-term demand for products. The LSTM model was carried out based on a sequence of time

series and the rating of sentiment comments to predict future sales. Due to minimal historic data, the decision maker is needed to take proper actions and react toward the market. It was also suggested that the sentiment rating weight adjustment can improve the prediction accuracy. The proposed LSTM approach performed well in terms of accuracy.

In [7], a model called Sen-BERT-CNN was presented to analyze customer reviews, based on CNN and pre-trained Bidirectional Encoder Representations from Transformers (BERT) to extract the sentiment information of a sentence. In [8], a method was used that combined the patterns of two graphs to predict the community trends on an item attribute. In [9], the LSTM and the Seasonal Autoregressive Integrated Moving Average (SARIMA) were used to predict the demand for a product based on previous sales and satisfy future customer requirements. In [10], genetic algorithms were used in conjunction with a profit-maximizing logistic model to forecast client attrition. In a similar strategy in [11], data-driven agent-based analysis of customer behavior was used, while in [12] a study was carried out to estimate and infer heterogeneous treatment effects utilizing random forests to determine the extent to which sales are associated with e-commerce. In the same spirit, machine learning was used in [13] to interpret Internet of Things data. In [14], a feature selection method was presented based on an artificial bee colony and a gradient-boosting decision tree. In [15], the accurate prediction of protein contact maps was presented by integrating residual two-dimensional BLTSM with Convolutional Neural Networks (CNNs). This method was

supported by a large number of additional studies that followed a similar line of reasoning. In [16], a prediction model was presented using CNNs and RNNs. In [17], a learning framework was described that incorporated various levels of knowledge representation for the analysis of aspect-based sentiments. In [18], a semantic study of social networks in educational settings was presented along with recommendations. In [19], a context was established in photo albums by understanding and modeling user behavior in clustering and selection. Authors in [20] focused on convolutional neural networks and their applications in data mining and sales forecasting for e-commerce. Some studies used deep neural networks, LSTM, and Particle Swarm Optimization (PSO) to predict long-term sales in e-commerce companies, as earlier studies that estimated short-term product demands found it possible to predict long-term future sales. Sales remain a crucial variable in e-commerce. Furthermore, there are several studies analyzing the ways a community uses a product based on some characteristics.

II. METHODOLOGY

Figure 2 shows a flow diagram of the proposed method to predict e-commerce sales with the use of CNNs to select features from a dataset and employing a BLTSM model of RNNs for the prediction. This process began with the building of the model, the gathering and the pre-processing of the dataset, and the experimental training and testing for the prediction.

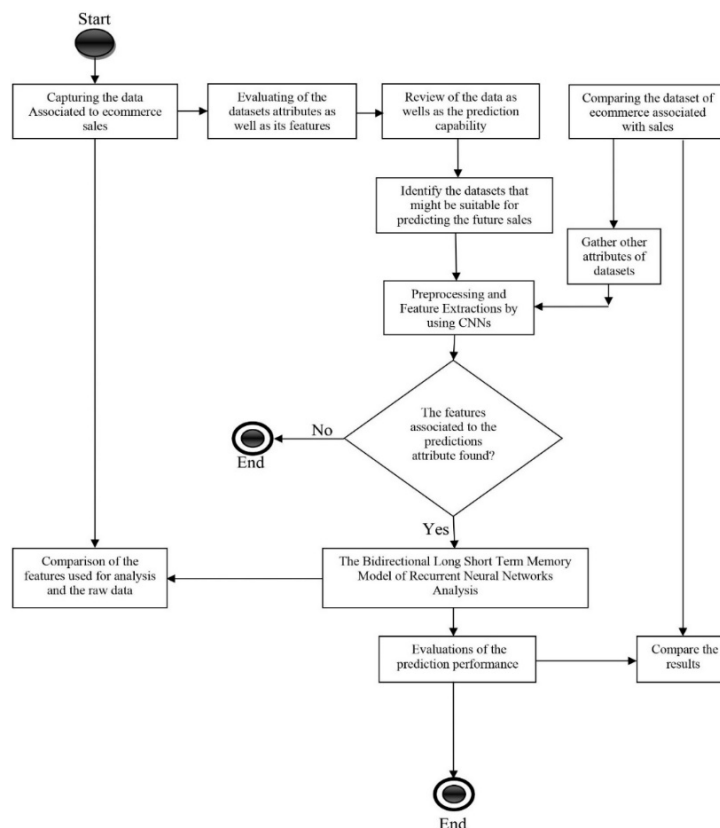


Fig. 1. The flow process of the prediction experimental analysis and result presentation.

A. The Bidirectional Long-Short-Term Memory (BLTSM) Model of Recurrent Neural Networks (RNNs)

This study adopted a bidirectional LSTM RNN model to predict the growth of sales for all products based on multiple factors [21]. A typical neural network does not function well when the data sequence is essential. Language translation, word recognition, and sentiment analysis are only a few examples of this situation. However, RNNs solve this problem, and BLSTM RNNs are sophisticated in such problems [22]. This model operates in a fashion where neurons are set as the basic building blocks of RNNs, and their strength is conveyed by weights. The RNNs typically consist of input, hidden, and output layers, with each layer's neuron coupled to aid in the retracement of past computations. RNNs excel in processing sequential data, especially when ordering is critical. LSTM solves the gradient problem either by vanishing or exploding in RNNs. The vanishing gradient problem results in an error that prevents RNNs from learning when there are more than 5-10 time steps between the events that are input and the signals that are being targeted [23]. To keep the sequence going, the RNN cell takes into account not just the input cells, but also the output of the current cell. Figure 2 shows an RNN cell. Given an input feature sequence $x = (x_1, \dots, x_T)$, the model can be constructed as a standard LSTM RNN for computing the hidden vector $h=(h_1, \dots, h_T)$ and the output vector sequence $y=(y_1, \dots, y_T)$ [23]. This can be determined by solving the following equations repeatedly, starting with time t equal to 1 and going up to T :

$$(h_t, c_t) = H(x_t, h_{t-1}, c_{t-1}) \quad (1)$$

$$y_t = W_{hy}h_t + h_y \quad (2)$$

where $W_{hy}h_t$ is the hidden-output weight matrix of the LSTM-RNN, and c is the cell activation vector that has the same size as the hidden vector h [23]. The W terms describe the weight matrices and the b terms denote the bias vectors, e.g., b_y is the output bias vector.

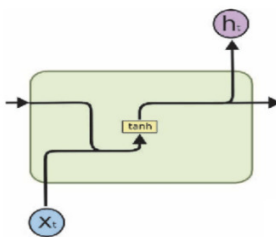


Fig. 2. The Architectural frame of an RNN –cell.

The deep LSTM-RNN has properties that are indicative of DNNs in general. It is feasible to generate it by stacking some LSTM-RNN hidden layers one atop another, with the output sequence of one layer functioning as the input for the succeeding in the stack. This process can be repeated as many times as necessary to generate it. Tanh is a function frequently used as an activation for RNNs. These relatively low values constrain the ultimate gradient while backpropagation takes place, but they have no impact on the weights. Because of the vanishing gradient problem, the process of training the model

goes more slowly. Memory cells are added to the buried layers of the RNN to find a solution to this problem, and the original design of the RNN is transformed into a long- and short-term memory system [25]. The LSTM architecture is made up of memory cells, an input gate, a forget gate, and an output gate. The input gate in the LSTM is responsible for determining whether the existing activation of the cell should be altered by the addition of new information obtained from the input network in the existing cell. The output gate has the responsibility to choose and pass pertinent information from the currently active cell to the next. The forget gate receives information from previous states, resets the action values, and contributes to the process of determining what information from the previous state should be discarded and what should be maintained. These gates alleviate the problem of disappearing gradients that can occur in LSTMs [26]. This is a similar approach to [27] on the detection and recognition of danger signs, to [28] on prediction, and to [29] on the planning of routes using RNNs. The LSTM is guided and uses the historical context maintained by the forget gate. Using this context in both directions allows for accurate transcription to be completed. As a direct consequence, this study developed bidirectional LSTMs by fusing two LSTMs (forward and backward). It is possible to combine many BLSTM layers to achieve a greater sense of depth and abstraction. This method used two LSTM layers, with 256 hidden cells in each. Following this step, this study also used a dropout layer and Adam as an optimizer.

B. Dataset Generation and Preprocessing

The dataset consisted of e-commerce sales data from the Kaggle repository, containing 23486 rows and 10 different feature variables. Customer age, title, rating, recommended flag, and reviews were some of the factors that were included in the dataset along with other variables such as clothing identification, which was a categorical variable that refers to the item in question. Anaconda and Google Collab were used to carry out the analysis. In addition, the recommended binary *IND* variable was used as a flag to indicate whether the client recommends the product (1) or not (0). As a result, ratings, reviews, and the recommended flag were taken into account when creating the training dataset for the products. Following this step, the dataset was cleaned out using several distinct forms of data preprocessing methods, such as tokenization and stop word removal, among others.

Tokenization is the first stage in any Natural Language Processing (NLP) pipeline, as the results of this step have a considerable influence on the subsequent steps. Using a tool called a tokenizer, unstructured data and text written in natural language are both sliced up into chunks of information that can be considered independent elements. A tokenization method was used to generate a vector that accurately represents this data, and then a technique known as the removal of stop words was also applied. These preprocessing steps are the most frequently used in a variety of NLP applications. The basic idea is to filter out any words present in every review in the dataset.

Both these terms are meaningless in the context of the prediction and recommendation task. Figures 3 and 4 show a subset of the dataset and the relevant statistics.

TABLE I. A DATASET SNAPSHOT

Count	Rating	Recommended IND	Positive Feedback Count
0	4	1	0
1	5	1	4
2	3	0	0
3	5	1	0
4	5	1	6
5	2	0	4
6	5	1	1
7	4	1	4
8	5	1	0
9	5	1	0
10	3	0	14
11	5	1	2
12	5	1	2
13	5	1	0
14	3	1	1
15	4	1	3
16	3	1	2
17	5	1	0
18	5	1	0
19	5	1	0
20	4	1	2
21	4	1	14
22	2	0	7
23	3	1	0
24	5	1	0
25	3	0	0
26	2	0	0
27	4	1	0
...
23480	5	1	0
23481	5	1	0
23482	3	1	0
23483	3	0	1
23484	3	1	2
23485	5	1	22

```

Unnamed: 0  Clothing ID  Age  ...  Division Name  Department Name  Class Name
0           0           767  33  ...  Intimates      Intimate  Intimates
1           1           1080  34  ...  General        Dresses  Dresses
2           2           1077  60  ...  General        Dresses  Dresses
3           3           1049  50  ...  General Petite  Bottoms  Pants
4           4           847   47  ...  General        Tops    Blouses
    
```

Fig. 3. Sample of the dataset.

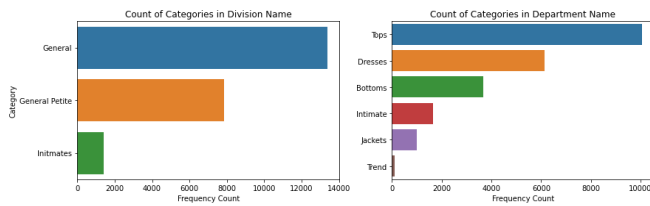


Fig. 4. Sample and statistical analysis of the dataset.

III. EXPERIMENTAL ANALYSIS AND RESULTS

This study used Python v.3.6 and the Scikit learn and Tensor Flow machine and deep learning libraries as backends to Keras [30-31]. Training and prediction were both developed with the help of the NVIDIA CUDA parallel computing platform and API. In addition, the root mean square prediction error was calculated to compare the various outcomes of the proposed BLSTM model, as it provides an all-encompassing perspective of the error distribution. The dataset was

partitioned into two subsets, namely 70% for training and 30% for testing. The model was trained for a total of 32 iterations to prevent overfitting. The model run with various numbers of epochs that ultimately resulted in a 91% accuracy. This shows that the factors used to predict e-commerce sales have a strong prediction performance. Figures 5 and 6 show the model's accuracy and loss, respectively.

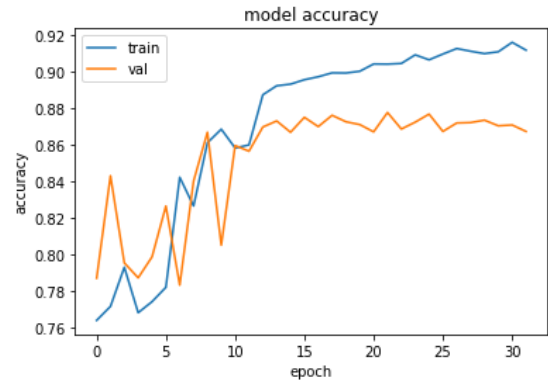


Fig. 5. Model accuracy as a function of epoch number.

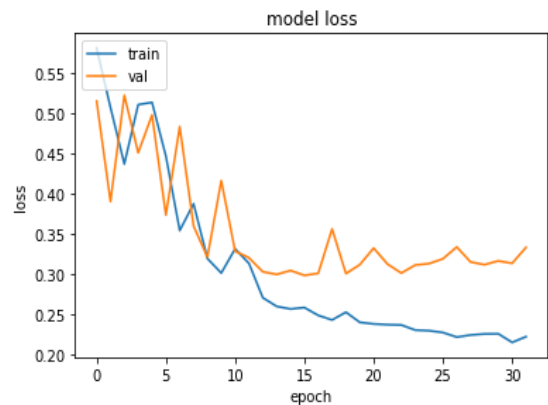


Fig. 6. Model loss as a function of epoch number.

IV. CONCLUSION

Machine learning methods are increasingly used in many industries, including speech recognition, vehicles, clustering, and data classification, as well as the identification of anomalies affecting everyday life. Deep learning has made significant strides in e-commerce, specifically in forecasts and sale marketing, using sentiment analysis to increase customer satisfaction. This study used CNNs to pick features from a publicly available online retail sales dataset and the Bidirectional Long Short Term memory model of RNNs (BLSMT-RNN) to predict future sales. The results revealed that the proposed model had a 91% accuracy, showing a strong prediction performance. Future work will investigate the effectiveness of the various iterations of RNNs.

REFERENCES

[1] Y. Qi, C. Li, H. Deng, M. Cai, Y. Qi, and Y. Deng, "A Deep Neural Framework for Sales Forecasting in E-Commerce," in *Proceedings of the 28th ACM International Conference on Information and Knowledge*

- Management, New York, NY, USA, Aug. 2019, pp. 299–308, <https://doi.org/10.1145/3357384.3357883>.
- [2] K. Bandara, P. Shi, C. Bergmeir, H. Hewamalage, Q. Tran, and B. Seaman, "Sales Demand Forecast in E-commerce Using a Long Short-Term Memory Neural Network Methodology," in *Neural Information Processing*, Sydney, NSW, Australia, 2019, pp. 462–474, https://doi.org/10.1007/978-3-030-36718-3_39.
- [3] Y. S. Shih and M. H. Lin, "A LSTM Approach for Sales Forecasting of Goods with Short-Term Demands in E-Commerce," in *Intelligent Information and Database Systems*, Yogyakarta, Indonesia, 2019, pp. 244–256, https://doi.org/10.1007/978-3-030-14799-0_21.
- [4] Q. Q. He, C. Wu, and Y. W. Si, "LSTM with particle Swarm optimization for sales forecasting," *Electronic Commerce Research and Applications*, vol. 51, Jan. 2022, Art. no. 101118, <https://doi.org/10.1016/j.elerap.2022.101118>.
- [5] W. Dong, Q. Li, and H. V. Zhao, "Statistical and Machine Learning-based E-commerce Sales Forecasting," in *Proceedings of the 4th International Conference on Crowd Science and Engineering*, New York, NY, USA, Jul. 2019, pp. 110–117, <https://doi.org/10.1145/3371238.3371256>.
- [6] Z. Wang, P. Gao, and X. Chu, "Sentiment analysis from Customer-generated online videos on product review using topic modeling and Multi-attention BLSTM," *Advanced Engineering Informatics*, vol. 52, Apr. 2022, Art. no. 101588, <https://doi.org/10.1016/j.aei.2022.101588>.
- [7] F. Wu, Z. Shi, Z. Dong, C. Pang, and B. Zhang, "Sentiment Analysis of Online Product Reviews Based On SenBERT-CNN," in *2020 International Conference on Machine Learning and Cybernetics (ICMLC)*, Adelaide, Australia, Sep. 2020, pp. 229–234, <https://doi.org/10.1109/ICMLC51923.2020.9469551>.
- [8] J. Yuan *et al.*, "Community Trend Prediction on Heterogeneous Graph in E-commerce," in *Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining*, New York, NY, USA, Oct. 2022, pp. 1319–1327, <https://doi.org/10.1145/3488560.3498522>.
- [9] K. F. Islam, M. Rahman, and S. A. Hossain, "Local Inventory Demand Forecasting of E-Commerce with Mapreduce Framework," in *International Conference on STEM and the Fourth Industrial Revolution*, Khulna, Bangladesh, Nov. 2022, pp. 474–483, <https://doi.org/10.53808/KUS.2022.ICSTEM4IR.0082-se>.
- [10] E. Stripling, S. vanden Broucke, K. Antonio, B. Baesens, and M. Snoeck, "Profit maximizing logistic model for customer churn prediction using genetic algorithms," *Swarm and Evolutionary Computation*, vol. 40, pp. 116–130, Jun. 2018, <https://doi.org/10.1016/j.swevo.2017.10.010>.
- [11] D. Bell and C. Mgbemena, "Data-driven agent-based exploration of customer behavior," *Simulation*, vol. 94, no. 3, pp. 195–212, Mar. 2018, <https://doi.org/10.1177/0037549717743106>.
- [12] S. Wager and S. Athey, "Estimation and Inference of Heterogeneous Treatment Effects using Random Forests," *Journal of the American Statistical Association*, vol. 113, no. 523, pp. 1228–1242, Jul. 2018, <https://doi.org/10.1080/01621459.2017.1319839>.
- [13] M. S. Mahdavejad, M. Rezvan, M. Barekatin, P. Adibi, P. Barnaghi, and A. P. Sheth, "Machine learning for internet of things data analysis: a survey," *Digital Communications and Networks*, vol. 4, no. 3, pp. 161–175, Aug. 2018, <https://doi.org/10.1016/j.dcan.2017.10.002>.
- [14] H. Rao *et al.*, "Feature selection based on artificial bee colony and gradient boosting decision tree," *Applied Soft Computing*, vol. 74, pp. 634–642, Jan. 2019, <https://doi.org/10.1016/j.asoc.2018.10.036>.
- [15] J. Hanson, K. Paliwal, T. Litfin, Y. Yang, and Y. Zhou, "Accurate prediction of protein contact maps by coupling residual two-dimensional bidirectional long short-term memory with convolutional neural networks," *Bioinformatics*, vol. 34, no. 23, pp. 4039–4045, Dec. 2018, <https://doi.org/10.1093/bioinformatics/bty481>.
- [16] E. Liberis, P. Veličković, P. Sormanni, M. Vendruscolo, and P. Liò, "Parapred: antibody paratope prediction using convolutional and recurrent neural networks," *Bioinformatics*, vol. 34, no. 17, pp. 2944–2950, Sep. 2018, <https://doi.org/10.1093/bioinformatics/bty305>.
- [17] D. H. Pham and A. C. Le, "Learning multiple layers of knowledge representation for aspect based sentiment analysis," *Data & Knowledge Engineering*, vol. 114, pp. 26–39, Mar. 2018, <https://doi.org/10.1016/j.datak.2017.06.001>.
- [18] A. Khaled, S. Ouchani, and C. Chohra, "Recommendations-based on semantic analysis of social networks in learning environments," *Computers in Human Behavior*, vol. 101, pp. 435–449, Dec. 2019, <https://doi.org/10.1016/j.chb.2018.08.051>.
- [19] D. Kuzovkin, T. Pouli, O. L. Meur, R. Cozot, J. Kervec, and K. Bouatouch, "Context in Photo Albums: Understanding and Modeling User Behavior in Clustering and Selection," *ACM Transactions on Applied Perception*, vol. 16, no. 2, pp. 11:1–11:20, May 2019, <https://doi.org/10.1145/3333612>.
- [20] H. Pan and H. Zhou, "Study on convolutional neural network and its application in data mining and sales forecasting for E-commerce," *Electronic Commerce Research*, vol. 20, no. 2, pp. 297–320, Jun. 2020, <https://doi.org/10.1007/s10660-020-09409-0>.
- [21] Z.-T. Liu, M.-T. Han, B.-H. Wu, and A. Rehman, "Speech emotion recognition based on convolutional neural network with attention-based bidirectional long short-term memory network and multi-task learning," *Applied Acoustics*, vol. 202, Jan. 2023, Art. no. 109178, <https://doi.org/10.1016/j.apacoust.2022.109178>.
- [22] J. Sun, X. Zhang, and J. Wang, "Lightweight bidirectional long short-term memory based on automated model pruning with application to bearing remaining useful life prediction," *Engineering Applications of Artificial Intelligence*, vol. 118, Feb. 2023, Art. no. 105662, <https://doi.org/10.1016/j.engappai.2022.105662>.
- [23] Z. Yang, R. Jia, P. Wang, L. Yao, and B. Shen, "Supervised Attention-Based Bidirectional Long Short-Term Memory Network for Nonlinear Dynamic Soft Sensor Application," *ACS Omega*, vol. 8, no. 4, pp. 4196–4208, Jan. 2023, <https://doi.org/10.1021/acsomega.2c07400>.
- [24] X. Xie, M. Huang, Y. Liu, and Q. An, "Intelligent Tool-Wear Prediction Based on Informer Encoder and Bi-Directional Long Short-Term Memory," *Machines*, vol. 11, no. 1, Jan. 2023, Art. no. 94, <https://doi.org/10.3390/machines11010094>.
- [25] A. Mubarak, M. Asmelash, A. Azhari, F. Y. Haggos, and F. Mulubrhan, "Machine Health Management System Using Moving Average Feature With Bidirectional Long-Short Term Memory," *Journal of Computing and Information Science in Engineering*, vol. 23, no. 3, Aug. 2022, <https://doi.org/10.1115/1.4054690>.
- [26] Y. Luo *et al.*, "Fast Response Prediction Method Based on Bidirectional Long Short-Term Memory for High-Speed Links," *IEEE Transactions on Microwave Theory and Techniques*, pp. 1–13, 2023, <https://doi.org/10.1109/TMTT.2022.3233303>.
- [27] W. Ali, G. Wang, K. Ullah, M. Salman, and S. Ali, "Substation Danger Sign Detection and Recognition using Convolutional Neural Networks," *Engineering, Technology & Applied Science Research*, vol. 13, no. 1, pp. 10051–10059, Feb. 2023, <https://doi.org/10.48084/etasr.5476>.
- [28] S. Tahzeeb and S. Hasan, "A Neural Network-Based Multi-Label Classifier for Protein Function Prediction," *Engineering, Technology & Applied Science Research*, vol. 12, no. 1, pp. 7974–7981, Feb. 2022, <https://doi.org/10.48084/etasr.4597>.
- [29] N. T. T. Vu, N. P. Tran, and N. H. Nguyen, "Recurrent Neural Network-based Path Planning for an Excavator Arm under Varying Environment," *Engineering, Technology & Applied Science Research*, vol. 11, no. 3, pp. 7088–7093, Jun. 2021, <https://doi.org/10.48084/etasr.4125>.
- [30] F. M. Miranda, N. Köhnecke, and B. Y. Renard, "HiClass: a Python Library for Local Hierarchical Classification Compatible with Scikit-learn," *Journal of Machine Learning Research*, vol. 24, no. 29, pp. 1–17, 2023.
- [31] A. Géron, *Hands-on machine learning with Scikit-Learn and TensorFlow: concepts, tools, and techniques to build intelligent systems*, First edition. Springfield, MI, USA: O'Reilly Media, 2017.