

Research Article

Prediction of News Popularity Based on Deep Neural Network

Yan Cai ¹ and Zhiqiang Zheng²

¹Shanghai University of Sport School of Media and Arts, Shanghai 200438, China

²Shanghai Dongfang Network Video Co., Ltd, Shanghai 200438, China

Correspondence should be addressed to Yan Cai; caiyan@sus.edu.cn

Received 26 January 2022; Revised 5 March 2022; Accepted 7 March 2022; Published 31 March 2022

Academic Editor: Baiyuan Ding

Copyright © 2022 Yan Cai and Zhiqiang Zheng. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Due to the timeliness and short life cycle of news, the postrelease prediction is limited, and the prerelease prediction also faces huge challenges due to the diversity and difficulty of defining influencing factors. This paper uses the foreclosure. This paper proposes a news popularity prediction method based on GRU deep neural network. Firstly, a web crawler was designed to obtain news data of different types and structures from 10 information security portal websites in China. After data preprocessing, the Word2Vec method was used to extract features, extract key news sentences and construct a subset of content features. Establish a GRU neural network regression prediction model to predict hot news on the Internet. The experimental results show that, compared with the traditional processing method, the model can process the multi-source rough data set in this paper and greatly reduce the prediction error. At the same time, because the threshold recurrent unit structure used in this paper is simpler than the long-short-term memory network structure, it can shorten the prediction time and improve the computing performance.

1. Introduction

The Internet is subtly changing the way of life of human beings, and has transformed from the original entertainment to an indispensable part of people's daily work and life [1]. News popularity prediction is an estimate of the number of page views or shares obtained after news is released. By predicting the number of news page views or shares, it can help content producers to better evaluate the quality of news in advance, and then help to rank news and recommended [2–4].

Popularity is one of the important characteristics of online news, it represents the spread of news and the possible social impact of news. By studying the relationship between online news and its popularity, it is possible to understand the reasons why online users forward and comment on news, and to dig out the key factors that determine news popularity, which is helpful for the reasonable release of online information and the grasp of public opinion trends [5, 6]. News with higher popularity tends to gain more attention from the public and become hot news, which has a greater impact on society and people's lives. News with less

popularity tends to lack attention and have less impact. Major news websites release all kinds of news all the time. Due to the limited time and energy of the public, they cannot read every news. Only a very small number of news will attract widespread public attention and become hot news. Therefore, it is timely. It is necessary to predict which news will become hot [7, 8]. In addition, online hot news prediction has important application value; firstly, it can enable the government to grasp the trend of public opinion in a timely manner, which is convenient for the government to manage public opinion and grasp and handle sudden public events; secondly, it can help news websites manage the release locations of different news, Put hot news in the area that users pay more attention to, thereby increasing the influence of news websites; at the same time, it promotes the public to pay attention to the current hot news in a timely manner, and triggers thinking about daily life from the news, thereby improving the quality of life. For example, when hot news related to telecommunication fraud occupies the homepage of major fine-textured websites, it can increase people's attention to and beware of telecommunication fraud, and help people learn the relevant knowledge of

preventing telecommunication fraud [9–12]. Network news has become the main source of network waves and public opinion, and it is of great theoretical and applied value to accurately predict hot news and attract public attention and discussion [13–15].

With the development of Web 2.0 and various self-media and online social networks at home and abroad, it has greatly changed the way users generate and consume content. Whether it is for content consumers, or for companies, content providers and self-media content producers, online content is a valuable asset and a major attraction and competitiveness on the Internet. At the same time, user-generated content has exploded due to the ease and lower cost of content creation. For example, users around the world send more than 300,000 tweets on Twitter, share more than 680,000 pieces of content on Facebook, and upload 100 hours of video every minute [16]. In this case, determining the quality of the content will become very important. Online users overwhelmed with information can reduce clutter and focus on the information most relevant to them, helping content consumers focus on the most valuable resources in the online world. For content distributors, they can rely on the popularity prediction after content distribution to actively allocate resources according to the needs of future users. The news popularity prediction studied in this paper is to predict the number of pageviews or retweets that may be obtained after the news is released. It can help journalists to better evaluate the quality of news and rank news, so as to conduct news delivery more reasonably [17–20]. Online news popularity prediction is an extremely challenging task. Due to the richness and difficulty of mining text content, it is important for researchers to measure various factors that affect popularity (such as the quality of content or relevance to users) and how to choose an appropriate predictive model that can adapt to different datasets and It is difficult to fit the popularity value after news release more accurately [21–26].

Aiming at the problems in this field, this paper proposes a news popularity prediction model based on deep neural network, taking financial hot and nonhot news crawled from the Sohu News website as the research object, extracting features from the content of online news, and establishing a GRU neural network. The news popularity prediction regression model combined with the network and the fully connected layer, thereby shortening the prediction time and improving the computing performance.

2. Word Vector Representation of Text

2.1. Word Vector Representation. A word vector refers to a low-dimensional real number vector (usually 50-dimensional or 100-dimensional) to represent words. For example, the word “Xi’an” can be represented as a vector of the form [0.692, -0.877, -0.157, 0.119, -0.532, . . .], the text “I go to school in Xi’an” is represented by word segmentation as “I,” “zai,” “Xi’an,” and “going to school,” where each word can be trained as a $1 * n$ -dimensional tensor, and the subtext can be Represented as a $4 * n$ -dimensional tensor.

Traditional text representations are based on bag-of-words models. The shortest answer is the one-hot model, which arranges all the words in the corpus in a column. For a certain document, the position of the words it contains in the corpus is k , then the k -th position is set to 1, for this document The words that are not included are set to 0; in order to show the importance of different words, the TF-IDF weight is used instead of 0 or 1, which is the traditional text space vector representation. The space vector model of text classification often results in the problem of too large text vector dimension and sparse space vector. The use of principal component analysis and singular value decomposition can achieve the purpose of reducing the dimensionality. However, in the process of dimensionality reduction, important information is often lost, such as Word order information. In addition, the spatial vector representation in this paper considers that any two words are completely independent, but adjacent words in text data are often associated. Using word vectors to represent text can not only solve the problem of high dimension and sparse vectors, thereby reducing the training difficulty of the model, but also word vectors can better reflect the semantic characteristics of words, reduce the distance between similar words or synonyms, and at the same time It can better represent the text content considering the order of words in the text.

2.2. Word2vec Theory. The vectorized representation of words includes One-Hot representation and word vector representation. The One-Hot representation represents each word as a long vector containing only 0s and 1s. The dimension of the vector is the size of the dictionary, and the position of 1 corresponds to the position of the word in the dictionary. There are some problems with this approach:

- (1) Since the dimension of the vector is determined by the size of the dictionary. Then when the dictionary is large, it will cause the dimension of the vector to be too high and cause dimension disaster.
- (2) When the vectorized representation of each word is performed, most of the vector values are 0, which will cause the word vector to be a high-dimensional sparse vector.
- (3) In addition, the One-Hot method does not consider the semantic and contextual relationship between words.

To solve the above problems, Google in 2013 proposed a word vector representation that uses contextual relationships to map words to low-dimensional, dense vectors—Word2vec. Word2vec is a deep learning based tool released by Google in 2013. The tool mainly adopts two model architectures—CBOW model (Continuous Bag-of-Words) and Skip-Gram (Continuous Skip-Gram) model to learn the vector representation of words. The CBOW model predicts the word output at the current position based on the words in the window around the current word; Skip-Gram, on the contrary, predicts the word output around it based on the words at the current position. Converting the text dataset

into the input form accepted by the two models of Word2vec, CBOW and Skip-Gram, can get its corresponding word vector as the output. The basis of the vectorization of the two models of Word2vec is to construct a word list based on the training text data set, and learn the vector representation of the words in the word list. Word vectors can perform mathematical operations and are widely used in natural language processing and other related fields, such as calculating the similarity and semantic correlation between two words. The features learned through the generated word vectors can be used in fields such as text classification, named entity extraction, clustering, and sentiment analysis.

$$P(w_t | w_{t-k}, w_{t-(k-1)}, \dots, w_{t-1}, w_{t+1}, w_{t+2}, \dots, w_{t+k}). \quad (1)$$

The training goal of the CBOW model is to find the word that maximizes the probability shown in formula (1), where k is the window size, and the current word is predicted according to the context word.

2.3. Word Vector Training. Word2Vec is Google's open source natural language processing training tool. It represents all words in the corpus with vectors, so that the relationship between words can be quantitatively measured by distance measurement. Word2Vec uses the Distributed representation to replace the One-hot model. Through training, each word is mapped to a specified lower-dimensional vector, and then the relationship between words can be studied by ordinary statistical methods. The words in the document The use of vector representation can fully consider word semantics and word order. Word2Vec is often used for some problems in the field of natural language processing, such as machine translation, relation mining, part-of-speech tagging, etc. It is computationally efficient and can train a large number of corpora. Word2Vec generally uses a neural network without hidden layers for training, and the definitions of input and output include Skip-Gram and CBOW training models. The structure of the CBOW model is shown in Figure 1.

The input information of the CBOW model is the word vector corresponding to the context-related words of the target word in the vocabulary, and the output is the word vector of the target word. For example, the words "related," "network," "hot spots," "news," "of," "research," "has," "important," and "meaning," the target word is "news," and the word vector of the word is output; when the window size is 5, the context value is 2, then the word vectors of the first two words "network," "hot spot" and the last two words "de" and "research" adjacent to the target word are the input of the model; the training goal is to make The softmax probability of the target word in the training sample is the largest. In the CBOW neural network model corresponding to the above example, there are 4 neurons in the input layer, and the number of neurons in the output layer is equal to the number of words in the vocabulary. The parameters of the model can be obtained by the BP algorithm, and the word vectors corresponding to all the words in the vocabulary are obtained at the same time. In the figure, $w(t)$ represents the

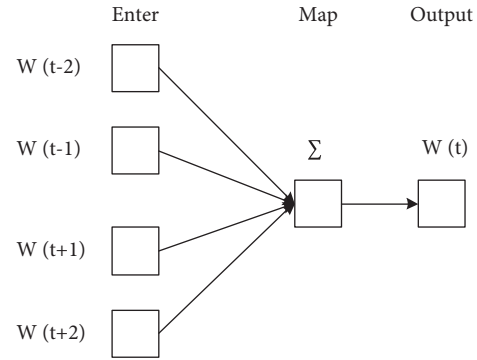


FIGURE 1: CBOW model structure.

position t of the target word in the sentence. In the window (the window size is 5), the upper and lower words except the target word together constitute the context.

3. GRU Neural Network

The LSTM network structure can solve the long-distance dependence problem of the traditional RNN network, but the internal structure of the LSTM is very complex, the training time is long, and there are many parameters. Therefore, various variants have appeared on the basis of LSTM. For example, in order to improve the training speed, the SRU (Simple Recurrent Unit, SRU) proposed by Lei et al. A widely popular variant of LSTM is that in 2014, CHOK et al. proposed a simplified structure of LSTM—GRU (Gated Recurrent Unit, GRU).

GRU and LSTM also use the gating mechanism, but compared with LSTM, GRU only retains two gating units reset gate (reset) and update gate (update). At the same time, the cell state and output gate are abandoned, and the cell state is combined with the hidden state. Compared with the LSTM structure, the update gate in the GRU plays the role of the input gate and the forgetting gate, controlling how much of the memory of the previous moment is retained at the current moment, that is, z_t in Figure 2 where \hat{h}_t is the memory state at the current moment, and h_t is the hidden state at the current moment.

Similar to the LSTM structure, the external input x_t at the current moment and the hidden state h_t at the previous moment jointly generate the input of the reset gate and the update gate. The update gate plays the role of "connecting the past and linking the future," it controls the discarding of the state at the previous moment and the information retained at the current moment; the reset gate controls the proportion of the information at the previous moment being forgotten. The cooperation of the reset gate and the update gate enables the learning of long-term memory. Since the GRU structure was proposed, it has been favored by researchers at home and abroad, and has been widely used in different fields. Various variants have also emerged based on their basic structure. The GRU structure has an advantage over the LSTM structure in terms of computational performance due to its simpler structure and fewer parameters, on the premise of ensuring the same effect as the LSTM.

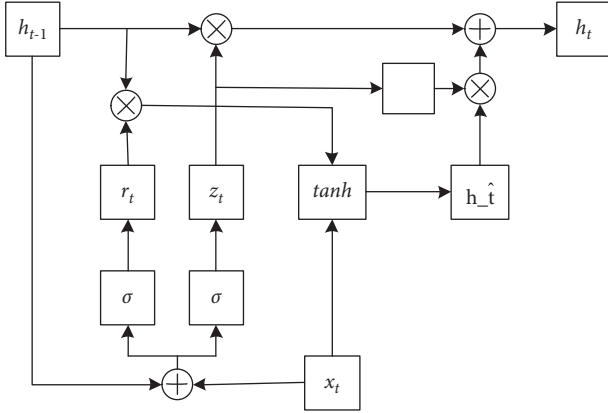


FIGURE 2: GRU network structure diagram.

The GRU structure not only overcomes the gradient dispersion problem existing in the traditional RNN structure and can learn long-distance dependencies, but also has a simpler structure and fewer parameters than LSTM. These advantages of GRU make it one of the widely used variants of the traditional RNN structure.

$$\hat{h}_t = \tanh(W^x x_t + U(r \otimes h_{t-1})). \quad (2)$$

4. News Popularity Prediction Model based on GRU Neural Network

The main process of the framework to achieve news popularity prediction is as follows:

- (1) Data preprocessing: there is a lot of “dirty data” in the news data set obtained by designing web crawlers. Data preprocessing needs to be carried out first, which mainly includes deduplication of data and judgment of empty data. In the preprocessing of data It also includes word segmentation and stop word removal operations for news headlines to prepare for text feature extraction.
- (2) Feature extraction process of surface information: after the data preprocessing in step (1), the news is decomposed into title, category, author, release time and text, and the custom features are derived from the title, category, author and release time. Sentiment polarity analysis and named entity extraction are performed on news headlines as “text features;” news headline lengths, publishing time conversions, author scores and category scores are calculated to construct a set of “metadata features.”
- (3) Key sentence extraction and content features: before using the key sentence extraction algorithm to extract key sentences from the news text, it is necessary to perform preprocessing of sentence segmentation, word segmentation and stop word removal. The preprocessed news uses the key sentence extraction algorithm to complete the extraction of key sentences. The named entity feature extraction of the key sentences of the news mainly includes the number,

place name and organization name to construct a subset of content features.

- (4) Feature set construction: after steps (2) and (3), the surface feature extraction and key sentence feature extraction process of news are completed, and text features, metadata features and content features are obtained respectively. The three feature subsets are connected and fused into a feature set, that is, the final feature set participating in model training.
- (5) Model training: in this paper, the regression model is trained with the simplified structure GRU of LSTM, and the fully connected layer is added after the GRU network layer to output the prediction result. In order to prevent overfitting, a dropout layer is added to control the random deactivation of neural units and the L2 regular term during model training, and the ReLU function is used as the activation function. Take MSE as loss.
- (6) Result output: based on this framework, the multi-feature extraction and fusion of news are realized, and the GRU structure is used to connect the fully connected layer to train a regression prediction model to predict news popularity.

4.1. Data Set. This paper describes the popularity prediction of news as a regression prediction problem, which predicts the number of pageviews received in the future after the news is published. Design a web crawler based on the crawler framework Scrapy, and use xpath to match by viewing the source code of the web page to obtain data and save it as a csv file. This article obtains some news from 2014 to 2018 through the crawler, and uses the number of news views as the prediction target of this article. In order to achieve faster convergence during training, logarithmically transform the news pageviews, and the transformed pageviews obey a Gaussian distribution.

4.2. Algorithm Evaluation Index. The research on the news popularity prediction framework in this paper belongs to regression prediction, so the MAE (Mean Absolute Error, MAE), MSE (Mean Squared Error, MSE), R^2 (R-squared, R^2) and RMSE (Root Mean Squared Error, RMSE) as metrics to measure the performance of the framework.

$$MSE = \frac{1}{n} \sum_{i=1}^n (f(x_i) - y_i)^2,$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |f(x_i) - y_i|,$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (f(x_i) - y_i)^2}{n}}, \quad (3)$$

$$R^2 = 1 - \frac{(\sum \text{Squared-Residuals}/N)}{\text{Variance}_{Y_{\text{true}}}}.$$

Among them, MAE, MSE, and RMSE indicators reflect the prediction error of the framework and are the main evaluation indicators; R^2 is the discriminant coefficient, which is used to indicate that the interpretability of the framework is a secondary indicator. Lower values of MAE, MSE and RMSE indicate smaller prediction errors of the framework.

5. Analysis of Results

The data set used in this article is the news data set of the security news portal website obtained through web crawlers. The experimental environment is the Python3 integrated environment Anaconda 3. The tools used mainly include the Python scientific computing package Numpy, Pandas, the word segmentation tool stuttering word segmentation and visualization Toolkit Matplotlib, Seaborn.

The construction of the model is based on Keras with Theano as the backend, performing 5 repeated experiments and taking the average result of the 5 experiments as the final result. At the same time, in order to more accurately evaluate the prediction effect of the framework during training, the data set is divided into training set, test set and validation set according to the ratio of 7:2:1. The experimental results shown in this article are the test set results. When building the GRU network, the Keras sequential model is used. In order to prevent overfitting, a Dropout layer is added, and the random deactivation ratio of neurons in the hidden layer is set to 0.5. At the same time, the L2 regularization is added to the GRU layer as a penalty term for model complexity; the ReLU function is used to calculate the model complexity. As an activation function, because the ReLU function does not involve a large number of exponential operations compared to the sigmoid or tanh function, the model converges faster. At the same time, ReLU can randomly inactivate some neurons, that is, "dead neurons," which can alleviate overfitting and There will be no gradient dispersion problems. The loss function used in this paper is MSE, stochastic gradient descent (SGD) is used as the optimization method, and the epoch is determined to be 6 by using grid search. Figure 3 shows the training process of the network. In the early stage of network training, with the increase of epoch, the loss function of the training set and the validation set decreases rapidly, indicating that the network learning speed is faster at this time; when the epoch continues to increase, the loss function decreases slowly, the network learning speed slows down; until epoch = 6, the loss function of the training set and the validation set basically does not change, and the network almost stops learning.

The results obtained by repeating the experiment three times and taking the average of the results are shown in Table 1.

The GRU and multi-feature fusion method proposed in this paper outperforms LSTM plus custom feature combination, ordinary linear regression-based, and random forest-based prediction methods in regression prediction, three indicators MSE, RMSE and MAE. The deep learning-based security news popularity prediction framework proposed in this paper shows low prediction error on rough and irregular

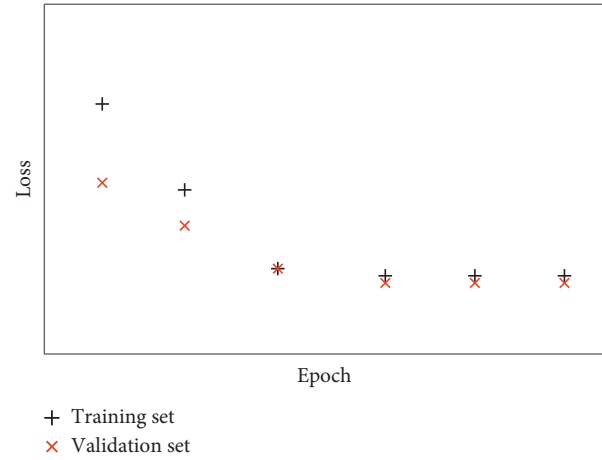


FIGURE 3: Model training loss comparison chart.

TABLE 1: Comparison of experimental results.

Model	MSE	MAE	RMSE	R^2
GRU	0.0152	0.0865	0.1098	0.6001
LSTM	0.2114	0.3352	0.4576	0.6298
Linear regression	0.2976	0.3621	0.5423	0.2651
Random forest	0.3521	0.4108	0.5901	0.1365

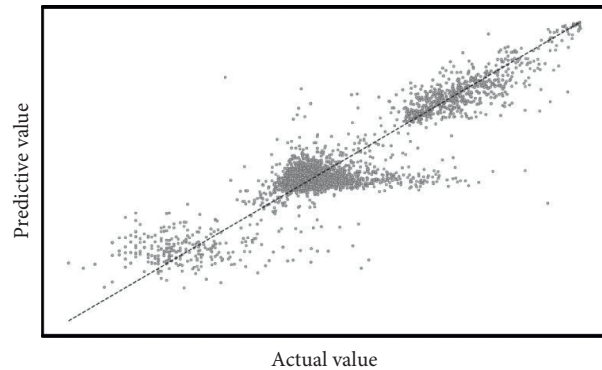


FIGURE 4: Model fitting effect diagram.

TABLE 2: Comparison of computing performance.

Index	GRU	LSTM
Time	10689	12351

datasets, while the prediction method of LSTM plus custom features does not achieve good results when dealing with rough multi-source datasets effect. At the same time, the discriminant coefficient R^2 of the prediction method proposed in this paper is 0.60349, which is slightly lower than the prediction method of LSTM plus custom features. R^2 is a secondary indicator, indicating that the model has better data fitting ability. Figure 4 is the fitting result of the predicted value and the actual value.

Compared with LSTM, GRU requires less training time due to its simpler structure and fewer parameters. Table 2 shows the comparison of computing performance using GRU structure and LSTM structure respectively. When

faced with larger-scale data sets, the advantage of GRU structure in computing performance will be more obvious.

6. Conclusion

News popularity prediction before news release can predict the number of clicks, comments or reposts of news in the future. Based on the prediction of popularity, news quality evaluation, news ranking, news recommendation and news retrieval can be carried out. At the same time, news popularity prediction plays an important role in alleviating the information explosion and information overload caused by the rapid development of the Internet and social media. Prerelease forecasting also faces huge challenges due to the diversity and difficulty of defining influencing factors. Compared with the existing work, this paper proposes a news popularity prediction framework that can process multi-source rough data sets and greatly reduce the prediction error by extracting multi-feature fusion and combining with neural network structure threshold recurrent unit training regression prediction model.

This paper proposes a news popularity prediction method based on GRU deep neural network. Firstly, a web crawler was designed to obtain news data of different types and structures from 10 information security portal websites in China. After data preprocessing, the Word2Vec method was used to extract features, extract key news sentences and construct a subset of content features. Establish a GRU neural network regression prediction model to predict hot news on the Internet. The experimental results show that, compared with the traditional processing method, the model can process the multi-source rough data set in this paper and greatly reduce the prediction error. At the same time, because the threshold recurrent unit structure used in this paper is simpler than the long-short-term memory network structure, it can shorten the prediction time and improve the computing performance.

Data Availability

The dataset can be accessed upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

- [1] Y. Yang, Y. Liu, X. Lu, J. Xu, and F. Wang, "A named entity topic model for news popularity prediction," *Knowledge-Based Systems*, vol. 208, 2020.
- [2] U. Douglas Zanatta, B. J. Luiz, M. C. Bonato, and S. Eusebio, "Exploring the effects of different clustering methods on a news recommender system," *Expert Systems with Applications*, p. 183, 2021.
- [3] B. Ş. KemalAkyol, "Modeling and predicting of news popularity in social media sources," *Computers, Materials & Continua*, vol. 61, no. 1, 2019.
- [4] A. S. Wicaksono and A. A. Supianto, "Hyper parameter optimization using genetic algorithm on machine learning methods for online news popularity prediction," *International Journal of Advanced Computer Science and Applications*, vol. 9, no. 12, 2018.
- [5] T. T. Esther, J. A. Kumar, P. Narayanasamy, and J. Vidya, "Attention-based C-BiLSTM for fake news detection," *Applied Soft Computing Journal*, vol. 110, 2021.
- [6] N. Moniz, L. Torgo, M. Eirinaki, and P. Branco, "A framework for recommendation of highly popular news lacking social feedback," *New Generation Computing*, vol. 35, no. 4, 2017.
- [7] D. Chen, F. Qin, K. Cai, and Y. Shen, "Detecting and classifying typhoon information from Chinese news based on a neural network model," *Sustainability*, vol. 13, no. 13, 2021.
- [8] P. Saigal and V. Khanna, "Multi-category news classification using Support Vector Machine based classifiers," *SN Applied Sciences*, vol. 2, no. 2, 2020.
- [9] S. A. Kanekar, A. Sharma, G. S. Patkar, K. Amey, and S. Tilve, "Building semantically annotated corpus for text classification of indian defence news articles," *International Journal of Information Technology*, 2021.
- [10] M. A. Mouriño-García, R. Pérez-Rodríguez, L. Anido-Rifón, and M. Vilares-Ferro, "Wikipedia-based hybrid document representation for textual news classification," in *Proceedings of the 2016 3rd International Conference on Soft Computing & Machine Intelligence (ISCMCI)*, Dubai, UAE, November 2018.
- [11] A. Abid, A. Ansar, K. Adel, F. Muhammad Shoaib, R. Iqbal, and F. Uzma, "An architectural framework for information integration using machine learning approaches for smart city security profiling," *International Journal of Distributed Sensor Networks*, vol. 16, no. 10, 2020.
- [12] Y. Zhang and K. Lin, "Predicting and evaluating the online news popularity based on random forest," *Journal of Physics: Conference Series*, vol. 1994, no. 1, 2021.
- [13] M. Naseri and H. Zamani, "Analyzing and Predicting News Popularity in an Instant Messaging Service," in *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, New York, NY, USA, July 2019.
- [14] H. Young Baek, D.-K. Kim, and J. W. Kim, "Management earnings forecasts and adverse selection cost: good vs. bad news forecast," *International Journal of Accounting and Information Management*, vol. 16, no. 1, 2008.
- [15] P. Kittisak and V. Peerapon, "Stock trend prediction using deep learning approach on technical indicator and industrial specific information," *Information*, vol. 12, no. 6, 2021.
- [16] J. L. O. Hui, G. K. Hoon, and W. M. N. W. Zainon, "Effects of word class and text position in sentiment-based news classification," *Procedia Computer Science*, vol. 124, 2017.
- [17] L. N. H. Nam and H. B. Quoc, "The hybrid filter feature selection methods for improving high-dimensional text categorization," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 25, no. 2, 2017.
- [18] A. K. Tiwari, *Deep Learning and its Applications*, Nova Science Publishers, Inc., Suite N Hauppauge, NY, USA, 2021.
- [19] S. Desai Prathamesh, "News sentiment informed time-series analyzing AI (SITALA) to curb the spread of COVID-19 in houston," *Expert Systems with Applications*, vol. 15, p. 180, 2021.
- [20] F. Majeed, A. Muhammad Waqas, A. Hassan Muhammad, S. Ali, and M. I. Lali, "Social media news classification in healthcare communication," *Journal of Medical Imaging and Health Informatics*, vol. 9, no. 6, 2019.
- [21] Z. Zhang, G. Zhang, and C. Zhao, "PeikeWang. Online Chinese news classification system based on Python reptile and convolutional neural network," *International Journal of*

- Intelligent Information and Management Science*, vol. 7, no. 6, 2018.
- [22] D. Mehta, D. Aniket, P. Arunabha, and M Anand Kumar, "A transformer-based architecture for fake news classification," *Social Network Analysis and Mining*, vol. 11, no. 1, 2021.
- [23] D. Dangol, R. Dahi Shrestha, and A. Timalina, "Automated news classification using N-gram model and key features of Nepali language," *Scitech Nepal*, vol. 13, no. 1, 2018.
- [24] K. S. Kyaw and S. Limsirratana, "Optimization of text feature selection process based on advanced searching for news classification," *International Journal of Swarm Intelligence Research*, vol. 11, no. 4, 2020.
- [25] Y. Zhang, Y. Dang, H. Chen, M. Thurmond, and C. Larson, "Automatic online news monitoring and classification for syndromic surveillance," *Decision Support Systems*, vol. 47, no. 4, 2009.
- [26] Z. Wang and B. Song, "Research on Hot News Classification Algorithm Based on Deep learning," in *Proceedings of the 2019 IEEE 3rd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC)*, pp. 2376–2380, Chengdu, China, March 2019.