2022 International Conference on Knowledge Engineering and Communication Systems (ICKES)

Natural Language Processing using Graph Neural Network for Text Classification

1st V. Suresh Kumar
Department of Information Technology
Vel Tech Multi Tech Dr. Rangarajan
Dr. Sakunthala Engineering College
Chennai, India
sureshkumar@veltechmultitech.org

2nd Ahmed Alemran
Physics department
Medicine college-Misan university
Al Amarah, Iraq
ahmed.mcm@uomisan.edu.iq

3rd Dimitrios A. Karras
General Department
National and Kapodistrian University of Athens (NKUA)
Psachna, Evia
dimitrios.karras@gmail.com

4th Shashi Kant Gupta
Department of CSE
Integral University
Lucknow, India
raj2008enator@gmail.com

5th Chandra Kumar Dixit
Department of Physics
Dr. Shakuntala Misra National Rehabilitation University
Lucknow, India
ckparadise@gmail.com

6th Bhadrappa Haralayya
Department of MBA
Lingaraj Appa Engineering College
Bidar, India
bhadrappabhavimani@gmail.com

2022 International Conference on Knowledge Engineering and Communication Systems (ICKECS) | 978-1-6654-5637-1/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICKECS56523.2022.10060655

***Abstract*— The boom of the technological area has given rise to numerous new applications that actually rule the entire world. Some of them mainly are the social media networks like the Twitter, Gmail and Facebook. Lot of text and content are received to millions of people across the work. It has become very essential to be surer whether the content is fake, or isit malicious or is it a really important content. To classify these contents, text classification techniques are being widely used across the globe. This has invented a new technical world of Natural Language Processing where it becomes very important to process the texts and images before we actually use them. In this paper, we have introduced the use of Graph Neural Networks (GNN) to classify the text according to their content. The use of GNN is done as they work well for 2D vectors and texts are in two dimensions. Computation of Self-Organizing Maps (SOM) are done to compute the nearest neighbors in the graphs and it calculates the actual distances between the neighbors. This process is used in classifying the text and performs better when compared to the existing techniques.**

***Keywords—Natural Language Processing, Graph Neural Network, Self-Organizing Maps, Nearest Neighbors, Text Classification*.**

I. Introduction

With social media at its peak, the need of Natural Language Processing (NLP) has become very important [1]. NLP is way of extracting the information from the texts. As there are lot and lot of information being shared across the globe through various communication channels like mobile phones and computers, it becomes very difficult for the user to know whether the information is genuine or fake. The user needs to know only the text that are from genuine senders. Their always remains a probability that the text could be spam and clicking on it could be malicious [2]. To enhance the model working on NLP numerous researchers are working across the globe to output a model that could classify texts very precise [3]. The connection of NLP with the machine learning and AI is depicted in Fig. 1. The texts are mostly classified into spam or genuine. Most of the NLP models consider the basic units to be words. But it is always a question to how to represent the text while classifying. To answer this, we have introduced the use of Graph Neural Networks (GNN) [4].

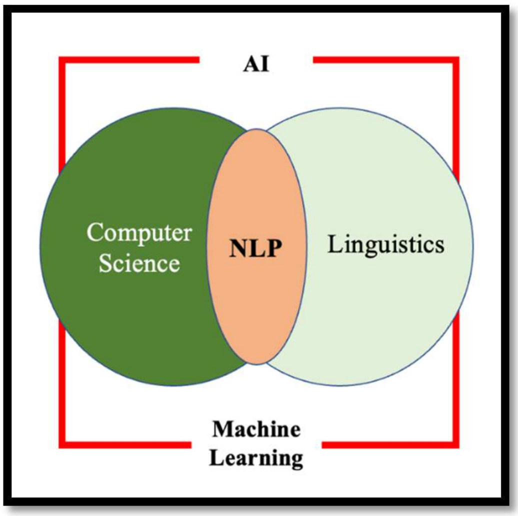

Fig. 1. Natural Language Processing

GNN works the same way as that of the common neural network. The growth of the neural networks has reached such a position that their remains no application without the need of neural networks. Neural Networks are generally used as an alternative to the way back traditional methods as they are quick and they will be able to handle large volume of data at a time. Same as these networks, GNN also can handle large volume of data [5]. Then what makes the GNN different from the traditional neural network? The answer is the deployment of the application. Some applications make use of 1-Dimensional arrays for processing, while others make use of 2-Diemsional and 3-Dimensional. As text is represented using 2D GNN prove to work better than the rest of the networks [6]. Moreover, GNNs are designed to handle the data that are in the form of graphs. When it comes to social connectivity of the users, they can be well represented using the graphs and hence, we have considered the use of GNN in this proposed work. Fig. 2 describes how the GNN is used for analyzing the Social User Networks.

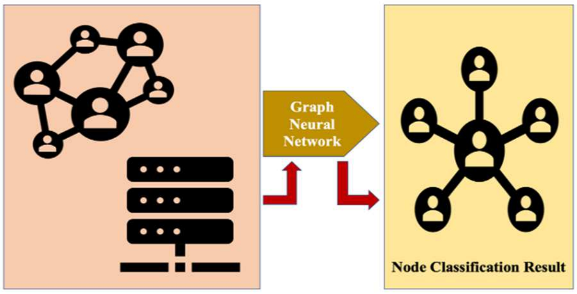

Fig. 2. Use of GNN for Social User Networks

One of the major drawback of NLS architectures are that they need very deep architectures to actually understand the complete sentences and this cannot be done only with a neural network. In addition to that we have introduced the scheme of using the Self-Organizing Maps (SOMs)[7] along with the GNN. The SOMs are capable changing its own dimensionality according to the applications it is deployed in. AAs the neural network need to have deep architecture to understand the underlying meaning of the sentence, with the help of SOMs they will be able to do it in ease. SOMs change their dimensionality which is a unsupervised machine learning technique. It changes itself from high dimension to lower dimension. The proposed paper shows the results obtained while classifying the text using SOM based neural network and the results are discussed.

The subsequent sections discuses on the various state of art techniques used to classify text and that are used in NLP. Section 3 discusses on the methodology used and Section 4 comprehends the experimental results obtained. Chapter 5 concludes the entire proposed work along with future directions.

II. RELATED WORKS

Most natural language processing (NLP) activities are now modelled in one of many major paradigms in the age of deep learning [8]. Recent years have seen an increase in paradigm shifts, when one NLP job is reformulated in a new paradigm so that it may be solved in a different way. The potential for several of these paradigms to unify NLP activities has been shown. Speech changes may be one of the early indicators of cognitive impairment, sometimes surfacing years before other deficiencies [9]. Neuropsychological language testing in this setting are confusing. Natural Linguistic Processing (NLP) approaches may be used to identify language changes in probable patients by analyzing spoken language products.

Traditional supervised learning relies on a single dataset to train a single prediction model for a specific problem [10]. Data from new domains or tasks may be used to train a model with improved generalisation capabilities using the transfer learning technique. Modern NLP transfer learning approaches will be discussed. It is getting more difficult for a person to communicate information in a natural language [11]. The goal of the automated NLP is to do this task as well and accurately as a person would. This chapter discusses the difficulties of NLP and the progress that has been done in this area so far.

The categorization of legal texts is the first step in creating an intelligent legal system. Text categorization receives less consideration in legal texts in the United States [12]. Random forests are used as the classifier in a machine learning technique we devised that uses domain ideas as features. Compared to a system constructed on numerous pre-trained word embeddings and deep neural networks, it beats this system dramatically. Image recognition, audio recognition, text recognition, and pattern recognition all benefit from deep neural networks. Such networks are subject to assault by adversarial instances [13]. When a model is trained on an adversarial example, it will incorrectly classify a sample of text if particular words are modified. There has been an alarming growth in the use of obscene language in social media material published by the general public [14]. An individual's or a community's sentiments might be damaged by this kind of rhetoric. Modular cleaning and tokenizer, three embedding approaches, and eight classifiers form the basis of our model for text categorization.

In [15], order to locate words with comparable semantic meaning, an upgraded text classification approach employs a word. The proposed approach outperforms current dimension reduction strategies in classification tests conducted on three different datasets. For embedding real-world graphs into low-dimensional spaces, graph neural networks are a useful toolset [16]. Systematic and complete examination of the graph neural networks is the goal of this study More and more researchers are likely to make use of them in their work. According to the premise of graph neural networks (GNNs), nodes belonging to the same class are more likely to be linked [17]. The performance of custom homophilic GNNs is severely hampered by heterophily, i.e., the tendency for nodes with different labels to be connected. Traditional neural networks may be applied to graph-structured data using Graph Neural Networks (GNNs) [18]. Graph data with a tree-like structure and a power-law distribution may now be processed using hyperbolic space, which was previously uncommon. The geometry of Euclidean models still constrains and limits their performance in graph-related learning.

Today, Deep Learning is one of the most widely used techniques in the field of Artificial Intelligence [19]. Learning tasks across a wide range of applications may be handled most effectively by Graph Neural Networks (GNNs). While GNNs have been studied extensively, there is still a lot of unanswered questions about how they work and how they may be used. The vast bulk of current renewable energy research focuses on electricity production at the macro level [20]. Solar, geothermal, and biomass heating potentials are examined as part of an integrated strategy in this article. On a county-by-county basis, the heating needs and potentials of renewables are evaluated.

Using machine learning methods known as emergent self-organizing maps (ESOM) [21]. There is no one characteristic that seems to regulate REE distributions, except for Eu, which has a different pattern of distribution. For the purpose of simulating monthly inflows into a reservoir based on satellite-estimated gridded rainfall time series, self-organizing maps (SOMs) were created. Inflow data homogeneity was shown to have an impact on the rainfall-streamflow modelling results. During the calibration phase, the models typically performed better, although the results differed based on the SOM structure used [22]. Processes based on data must make room for AI-driven frameworks [23]. The development of technology is focused on reducing the amount of data that must be transmitted and the amount of network traffic that must be handled [24].

III. Methods

The problem under consideration is to classify texts according to the information present. To obtain this we have made use of GNN that classifies the text accordingly. The GNN can comprise of several layer. The layers include the input layer, the hidden layers and the output layer. In our proposed approach we have taken 10 hidden layers. The activation function used here is ReLU as it classifies an output of either 0 or 1. The GNN working is same as that of the simple neural network but the difference lies in the architecture where the information is connected to each other via nodes. The GNN can extract many information from the text that it needs to classify. The main features extracted here is the node degree. This feature gives the number of nodes connected to one particular node and it will give us the answer of whether the text is really informative or not. The overall flowchart of the proposed method is shown in Fig. 3.

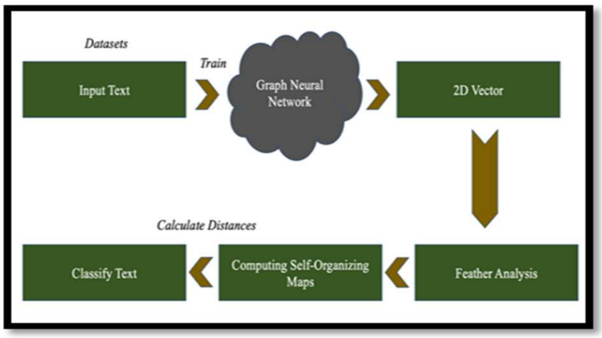

Fig. 3. Flowchart of the Proposed Method

From the flowchart it is known that initially the GNN is trained using the text dataset. With this initial training the model is able to know the keywords that are present when the txt is malicious or spam. The database consists of two categories. The model gets trained with both these categorical samples of being malicious or being informative. The model can handle 2D vector. With the training, the model is trained with 2D vectors. The training of the model for 2D vector is shown in the Fig. 4 and 5 respectively. In Fig. 4 the training epoch is of only till 54 with an elapsed time of about 14seconds whereas when we compare it with Fig. 5, the training epoch is maximum of about 250 with an elapsed time of about 26seconds. As the number of training epoch increases the time elapsed also increases subsequently. The iteration is stopped as the value in the graph does not change and remain constant for a longer period of iterations. The loss learning rate of the model is given as 0.01 and is depicted using Fig. 6.

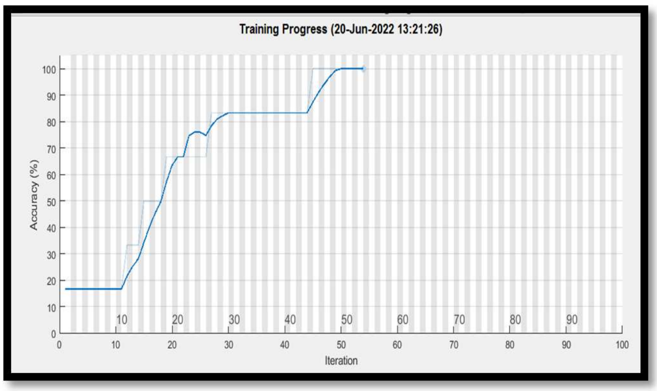

Fig. 4. 2D vector - 54 epoch Training Progress

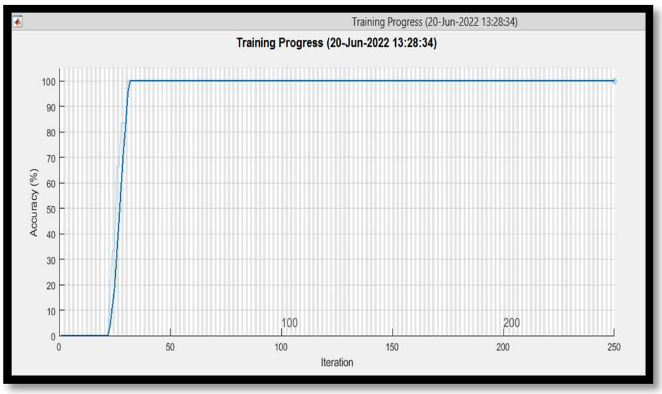

Fig. 5. 2D vector - 250 epoch Training Progress

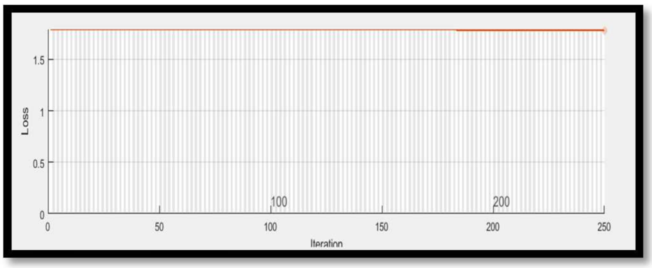

Fig. 6. 2D vector - Loss Leaning Rate

IV. Experimental Results

The experimental results are obtained by implementing the network in MatLab R2017B version. The results obtained while training the network with the SOM are charted in this section. The need for the usage of computing the SOM for the text is to reduce the dimensionality of the feature. By introducing SOM into the network, the model will be able to calculate the distances between the nodes and which will in turn help in the classification of the text. The feather analysis of the text input is shown in Fig. 7. This analysis gives the statistical way of observing the input given to the model. This graph gives the logarithmic value for the input given to the model.

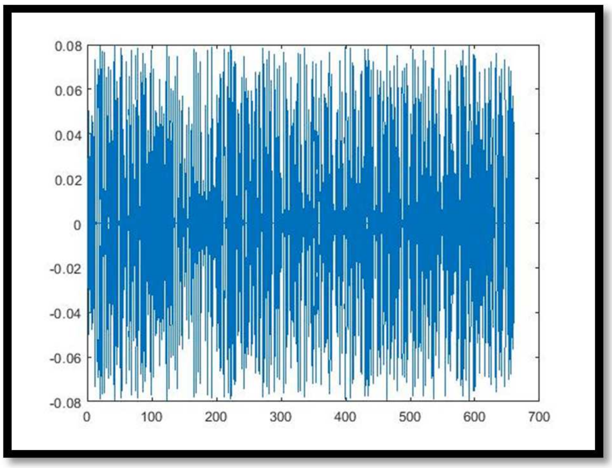

Fig. 7. Feather Analysis

Fig. 8 gives the histogram which is the graphical representation of the input image given. This histogram of various input images given as the train data is used in turn by the network for the training process. Once both the histogram and feather analysis is performed we have calculated the semilogx of the image and is depicted in Fig. 9. This graph gives the values that are in the range of base 10.

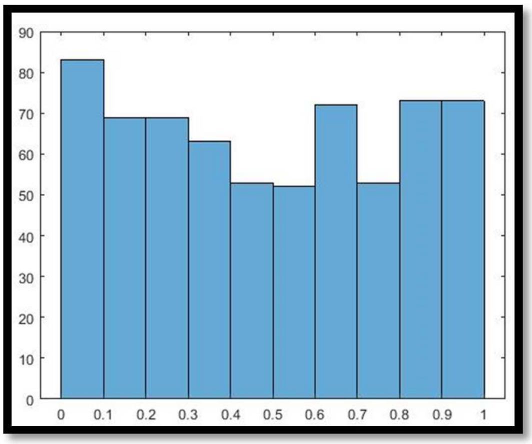

Fig. 8. Histogram

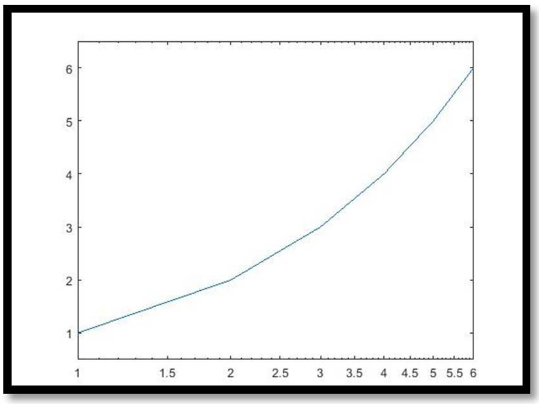

Fig. 9. Semilogx

Once the graphs are plotted, now our network needs to be embedded with the SOMs. So, we have taken two values of neighbors. The first case consists of 10 neighbors and the second case consists of 20 neighbors. Fig. 10 gives the SOM neighbor distance for 10 neighbors and its weights plane is depicted in Fig. 11. Same as this, Fig. 12 gives the SOM for the neighbor distance for 20 neighbors and its weight plane is depicted in Fig. 13. With the help of these maps, we will be able to identify the text relationship which would further help in the classification of the text.

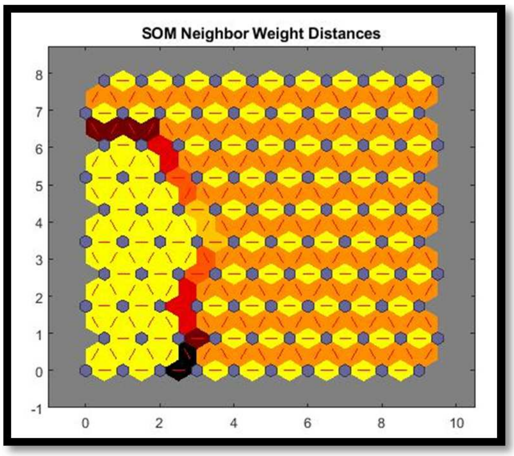

Fig. 10. SOM Neighbor Distances - 10 neighbors

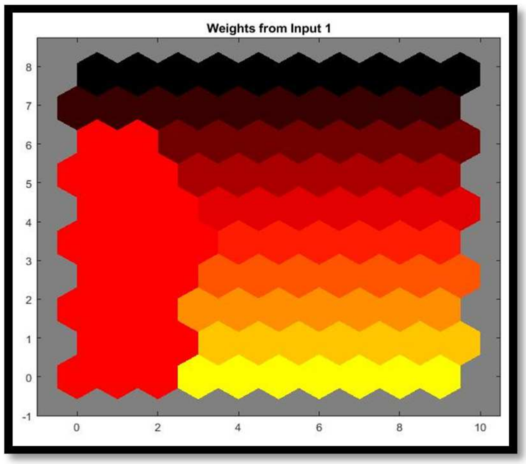

Fig. 11. SOM Weight Planes- 10 neighbors

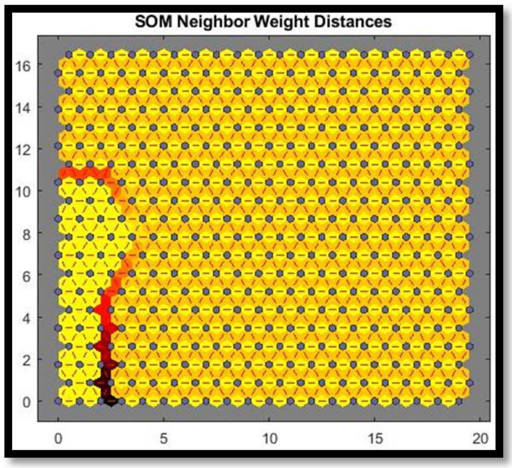

Fig. 12. SOM Neighbor Distances - 20 neighbors

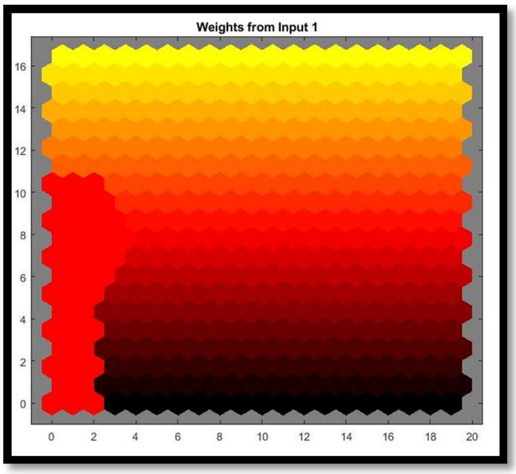

Fig. 13. SOM Weight Planes- 20 neighbors

V. Conclusions

Deep neural networks have recently grown significantly in popularity in many Artificial Intelligence applications due to the emergence of high computing devices. GNN is a powerful tool for graph data analysis, and it is still a relatively active field that requires further research focus to address many computer vision applications. The variety of GNN applications in computer vision is unrestricted and keeps growing. New approaches and prospects for development have emerged with the development of new technologies, and these technologies also address ongoing problems in the field of NLP. There is now substantial research being done on the design and implementation of graph neural networks (GNN) for NLP applications. Several questions remain unresolved. It has been proposed to explicitly mimic interactions between agents, objects, and their surroundings using a spatiotemporal graph neural network architecture. In this paper, a deep learning based model is developed with self-organizing maps. The SOMs are useful for text classification and have the nature of reducing the dimensionality when required. The efficiency of the model was calculated by computing the distance matrices. The weights were observed while the neighbors were equal to 10 and 20. This method generalizes the previous structured models for video comprehension and may implicitly or explicitly characterize things.

REFERENCES

[1] Otter, D. W., Medina, J. R., & Kalita, J. K. (2020). A survey of the usages of deep learning for natural language processing. IEEE transactions on neural networks and learning systems, 32(2), 604-624.

[2] Hasan, M. M., Zaman, S. M., Talukdar, M. A., Siddika, A., & Alam, M. G. R. (2021, December). An Analysis of Machine Learning Algorithms and Deep Neural Networks for Email Spam Classification using Natural Language Processing. In 2021 IEEE International Conference on Service Operations and Logistics, and Informatics (SOLI) (pp. 1-6). IEEE.

[3] Zhang, W. E., Sheng, Q. Z., Alhazmi, A., & Li, C. (2020). Adversarial attacks on deep-learning models in natural language processing: A survey. ACM Transactions on Intelligent Systems and Technology (TIST), 11(3), 1-41.

[4] Pradhyumna, P., & Shreya, G. P. (2021, August). Graph neural network (GNN) in image and video understanding using deep learning for computer vision applications. In 2021 Second International Conference on Electronics and Sustainable Communication Systems (ICESC) (pp. 1183-1189). IEEE.

[5] Hu, J., Qian, S., Fang, Q., Wang, Y., Zhao, Q., Zhang, H., & Xu, C. (2021, October). Efficient graph deep learning in tensorflow with tf_geometric. In Proceedings of the 29th ACM International Conference on Multimedia (pp. 3775-3778).

[6] Shreyas, E., & Sheth, M. H. (2021, August). 3D Object Detection and Tracking Methods using Deep Learning for Computer Vision Applications. In 2021 International Conference on Recent Trends on Electronics, Information, Communication & Technology (RTEICT) (pp. 735-738). IEEE.

[7] Ali, M. N. Y., Sarowar, M. G., Rahman, M. L., Chaki, J., Dey, N., & Tavares, J. M. R. (2019). Adam deep learning with SOM for human sentiment classification. International Journal of Ambient Computing and Intelligence (IJACI), 10(3), 92-116.

[8] Sun, T. X., Liu, X. Y., Qiu, X. P., & Huang, X. J. (2022). Paradigm shift in natural language processing. Machine Intelligence Research, 19(3), 169-183.

[9] Ruder, S., Peters, M. E., Swayamdipta, S., & Wolf, T. (2019, June). Transfer learning in natural language processing. In Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: Tutorials (pp. 15-18).

[10] Beltrami, D., Gagliardi, G., Rossini Favretti, R., Ghidoni, E., Tamburini, F., & Calzà, L. (2018). Speech analysis by natural language processing techniques: a possible tool for very early detection of cognitive decline?. Frontiers in aging neuroscience, 10, 369.

[11] Meera, S., & Geerthik, S. (2022). Natural language processing. Artificial Intelligent Techniques for Wireless Communication and Networking, 139-153.

[12] Chen, H., Wu, L., Chen, J., Lu, W., & Ding, J. (2022). A comparative study of automated legal text classification using random forests and deep learning. Information Processing & Management, 59(2), 102798.

[13] Kwon, H., & Lee, S. (2022). Ensemble transfer attack targeting text classification systems. Computers & Security, 117, 102695.

[14] Hajibabaee, P., Malekzadeh, M., Ahmadi, M., Heidari, M., Esmaeilzadeh, A., Abdolazimi, R., & James Jr, H. (2022, January). Offensive language detection on social media based on text classification. In 2022 IEEE 12th Annual Computing and Communication Workshop and Conference (CCWC) (pp. 0092-0098). IEEE.

[15] Singh, K. N., Devi, S. D., Devi, H. M., & Mahanta, A. K. (2022). A novel approach for dimension reduction using word embedding: An enhanced text classification approach. International Journal of Information Management Data Insights, 2(1), 100061.

[16] Zhou, Y., Zheng, H., Huang, X., Hao, S., Li, D., & Zhao, J. (2022). Graph Neural Networks: Taxonomy, Advances, and Trends. ACM Transactions on Intelligent Systems and Technology (TIST), 13(1), 1-54.

[17] Zheng, X., Liu, Y., Pan, S., Zhang, M., Jin, D., & Yu, P. S. (2022). Graph neural networks for graphs with heterophily: A survey. arXiv preprint arXiv:2202.07082.

[18] Yang, M., Zhou, M., Li, Z., Liu, J., Pan, L., Xiong, H., & King, I. (2022). Hyperbolic Graph Neural Networks: A Review of Methods and Applications. arXiv preprint arXiv:2202.13852.

[19] Wu, L., Cui, P., Pei, J., Zhao, L., & Song, L. (2022). Graph neural networks. In Graph Neural Networks: Foundations, Frontiers, and Applications (pp. 27-37). Springer, Singapore.

[20] Zheng, X., Yang, X., Miao, H., Liu, H., Yu, Y., Wang, Y., ... & You, S. (2022). A factor analysis and self-organizing map based evaluation approach for the renewable energy heating potentials at county level: A case study in China. Renewable and Sustainable Energy Reviews, 165, 112597.

[21] Engle, M. A., Nye, C. W., Neupane, G., Quillinan, S. A., McLaughlin, J. F., McLing, T., & Martín-Fernández, J. A. (2022). Predicting Rare Earth Element Potential in Produced and Geothermal Waters of the United States via Emergent Self-Organizing Maps. Energies, 15(13), 4555.

[22] do Nascimento, T. V. M., Santos, C. A. G., de Farias, C. A. S., & da Silva, R. M. (2022). Monthly streamflow modeling based on self-organizing maps and satellite-estimated rainfall data. Water Resources Management, 1-19.

[23] Navaneetha Krishnan Rajagopal, Naila Iqbal Qureshi, S. Durga, Edwin Hernan Ramirez Asis, Rosario Mercedes Huerta Soto, Shashi Kant Gupta, S. Deepak, "Future of Business Culture: An Artificial Intelligence-Driven Digital Framework for Organization Decision-Making Process", Complexity, vol. 2022, Article ID 7796507, 14 pages, 2022. https://doi.org/10.1155/2022/7796507

[24] Eshrag Refaee, Shabana Parveen, Khan Mohamed Jarina Begum, Fatima Parveen, M. Chithik Raja, Shashi Kant Gupta, Santhosh Krishnan, "Secure and Scalable Healthcare Data Transmission in IoT Based on Optimized Routing Protocols for Mobile Computing Applications", Wireless Communications and Mobile Computing, vol. 2022, Article ID 5665408, 12 pages, 2022. https://doi.org/10.1155/2022/5665408