Fish and Paddy Culture: Fish Farmer's Sentiment Analysis Using Machine Learning

Manish Rao Ghatge[1*], Snehlata Barde[2]

Abstract

Fish and paddy culture is a type of integrated farming system in which fish are cultured in rice paddies, and the fish waste is used to fertilize the rice crops. Sentiment analysis can be used to understand the feelings and attitudes of fish farmers towards this farming system. To conduct sentiment analysis on fish farmers' opinions on fish and paddy culture, one would need to collect text data from various sources, such as social media posts, online forums, or surveys these text data consists of English, Hindi and Chhattisgarhi (Regional language of Chhattisgarh) text. The collected data would then be analysed using natural language processing (NLP) techniques to identify the sentiment expressed in each message. The results of the sentiment analysis will provide insights into the perceptions of fish farmers towards fish and paddy culture. For example, if the sentiment analysis reveals that fish farmers express positive sentiments towards fish and paddy culture, it could suggest that this farming system is well-received and popular among fish farmers. On the other hand, if the sentiment analysis reveals negative sentiments, it could indicate that fish farmers are facing challenges or difficulties with this farming system, and improvements may be needed. Overall, sentiment analysis can help fish farmers and policymakers gain a better understanding of the attitudes and opinions of fish farmers towards fish and paddy culture, and help identify areas for improvement or further research.

Keywords: Fish and Paddy Culture, Fisheries Management, Sentiment Analysis, NLP, AI, Machine Learning, Chhattisgarhi

[1*]Research Scholar, E-mail: ghatge.manish@gmail.com
[2]Professor, MSIT, MATS University, Raipur, CG-492001, India

***Corresponding Author:**
*Research Scholar, E-mail: ghatge.manish@gmail.com

INTRODUCTION

To use AI for sentiment analysis of fish farmers, we need to collect relevant text data from various sources, such as online forums, social media, or customer feedback forms. This data then fed into an AI algorithm trained specifically for sentiment analysis, which use natural language processing (NLP) techniques to identify the sentiment expressed in each message. Sentiment analysis is the process of analysing and understanding people's emotions and opinions towards a particular topic. In the case of fish farmers, sentiment analysis can be used to understand their feelings and attitudes towards various aspects of their work, such as the state of their fish stocks, the effectiveness of their farming techniques, or the profitability of their business.

Artificial Intelligence (AI) is a powerful tool that is used to perform sentiment analysis on a large scale. AI algorithms can analyse vast amounts of text data.

It is necessary to implement management systems and determine management strategies for effective social media posts, customer reviews, or survey responses, and automatically identify the sentiment expressed in each message.

The results of the sentiment analysis could be used to gain insights into the feelings and attitudes of fish farmers towards different aspects of their work, and to identify any areas where improvements could be made. For example, if the sentiment analysis reveals that many fish farmers are expressing negative sentiments towards the quality of their fish stocks, this could indicate that improvements are needed in fish breeding or husbandry techniques.

Firstly, a corpus of the language that incorporates a blend of Chhattisgarhi, Hindi, and English is created to gather views from fish farmers regarding fish and paddy culture practices. The paper comprises of different sections. The second segment of the paper entails a literature review, followed by a methodology discussion in the third section. The fourth section discusses the experimental setup, and the final section presents the conclusion to wrap up the paper.

RELATED WORK

Sentiment analysis is a field of natural language processing (NLP) that aims to extract subjective information from text data. In recent years, sentiment analysis has become increasingly important in understanding people's opinions and emotions towards various topics. In particular, there has been growing interest in applying sentiment analysis to Hindi language or Indian regional languages. Here are some recent studies in the field of sentiment analysis:

In a study conducted by Ghatge M.R. et al., they utilized the fundamental principles of sentiment analysis on Indian language text. They employed various machine learning techniques on a corpus of text data. The study focused on applying improved machine learning techniques due to the lack of extensive lexicon resources available for Hindi language sentiment analysis [1].

"Sentiment Analysis of Hindi Text using Machine Learning Techniques": This research work proposed an ensemble model for sentiment analysis of Hindi text, which combined multiple machine learning algorithms such as Naive Bayes, Support Vector Machine, and Random Forest to improve accuracy [2].

"A Comprehensive Study on Sentiment Analysis of Hindi Text Using Deep Learning Techniques": This study investigated the performance of various deep learning models such as Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM), and Bidirectional Encoder Representations from Transformers (BERT) for sentiment analysis of Hindi text [3].

"Sentiment Analysis of Hindi Text using Hybrid Approach": This research work combined rule-based and machine learning approaches to perform sentiment analysis on Hindi text. The study showed that the hybrid approach outperformed individual rule-based and machine learning models in terms of accuracy [4].

"Hindi Sentiment Analysis using Pre-Trained Word Embedding": This study utilized pre-trained word embedding to improve the performance of sentiment analysis on Hindi text. The research work showed that pre-trained embedding significantly improved the accuracy of the sentiment analysis model [5].

"Sentiment Analysis of Hindi Text using Ensemble of Deep Learning Techniques": This research work proposed an ensemble of deep learning models, including CNN, LSTM, and BERT, for sentiment analysis on Hindi text. The ensemble model outperformed individual models and achieved state-of-the-art performance on a Hindi sentiment analysis dataset [6].

In the paper titled "The Sentiment Analysis for Hindi Language Using Convolution Neural Network and Deep Belief Network" authors proposes a hybrid model for sentiment analysis of Hindi language text using Convolutional Neural Network (CNN) and Deep Belief Network (DBN). The study compares the performance of the proposed model with other traditional machine learning algorithms and shows that the hybrid model outperforms these models in terms of accuracy. The study also utilizes preprocessing techniques such as stemming and stop-word removal to enhance the performance of the model. Overall, the paper highlights the potential of deep learning techniques for sentiment analysis of Hindi language text [7].

Rastogi and Kumar (2020) compared the performance of deep learning and traditional machine learning approaches for Hindi sentiment analysis.[8].

Bhardwaj and Sharma (2020) proposed an aspect-based sentiment analysis approach for Hindi language and evaluated its performance [9].

Chakraborty and Saha (2020) proposed an ensemble model approach for Hindi sentiment analysis and compared its performance with other traditional machine learning models [10].

Mahajan and Jain (2020) presented a bidirectional LSTM-CNN model for Hindi sentiment analysis and evaluated its performance [11].

Agarwal and Bhattacharyya (2020) presented a Hindi version of Senti WordNet, a lexical resource for sentiment analysis [12].

Gupta and Singh (2020) presented a deep learning-based approach for Hindi sentiment analysis on mobile app reviews [13].

Gupta and Jain (2020) presented a deep learning-based approach for Hindi sentiment analysis on Twitter data. Kumar and Kumar (2020) presented a machine learning-based approach for Hindi sentiment analysis on YouTube comments
[14].

Sharma and Kumar (2020) presented a deep learning-based approach for Hindi sentiment analysis on news articles [15].

Singh and Gupta (2020) proposed a rule-based approach for Hindi sentiment analysis on news headlines [16].

Srivastava and Agarwal (2020) presented a machine learning-based approach for Hindi sentiment analysis on poetry [17].

Mishra and Kumar (2020) presented a deep learning-based approach for Hindi sentiment analysis on blog data [18].

Singh and Sharma (2020) proposed a rule-based approach for Hindi sentiment analysis on movie songs [19].

Sharma and Singh (2020) presented a machine learning-based approach for Hindi sentiment analysis on WhatsApp messages [20].

Pandey and Singh (2020) presented a machine learning-based approach for Hindi sentiment analysis and compared the performance of various algorithms [21].

The authors of the paper Anupa Sinha and Snehlata Barde used Artificial Intelligence (AI) in the form of a MultiClass Support Vector Machine (MSVM) to develop a recognition system that is invariant to illumination variations. [22].

The paper "Nutritional and Natural Farming Agriculturists' Sentiment Analysis" presents an analysis of the sentiments expressed by agriculturists regarding nutritional and natural farming practices. The study uses sentiment analysis, a computational technique for analyzing large volumes of text to determine the emotional tone of the text. The authors collected data from social media platforms and agricultural forums, focusing on posts and comments related to nutritional and natural farming practices. They then used a sentiment analysis tool to determine the overall sentiment of the text as positive, negative, or neutral. The study found that overall, there was a positive sentiment among agriculturists towards nutritional and natural farming practices. Specifically, there was a strong positive sentiment towards organic farming, sustainable agriculture, and natural pest control methods. The authors suggest that their findings can help inform policy decisions related to agricultural practices and that sentiment analysis can be a useful tool for understanding the attitudes and opinions of stakeholders in various industries. However, the study is limited by its focus on a specific group of agriculturists and the use of automated sentiment analysis tools, which may not always accurately capture the nuances of human language [23].

Sharma and Jain (2020) proposed a recurrent neural network-based approach for Hindi sentiment analysis and evaluated its performance [24].

Kumar and Jain (2020) proposed a lexicon-based approach for Hindi sentiment analysis using Senti WordNet and support vector machines [25].

Verma and Singh (2020) presented a machine learning-based approach for Hindi sentiment analysis on movie reviews [26].

METHODOLOGY

Chhattisgarhi Mixed data in Hindi and English was collected, and the dataset was pre-processed to obtain a cleaned version, which could enhance the results of the proposed approach. Figure 1 shows the block diagram of the methodology:

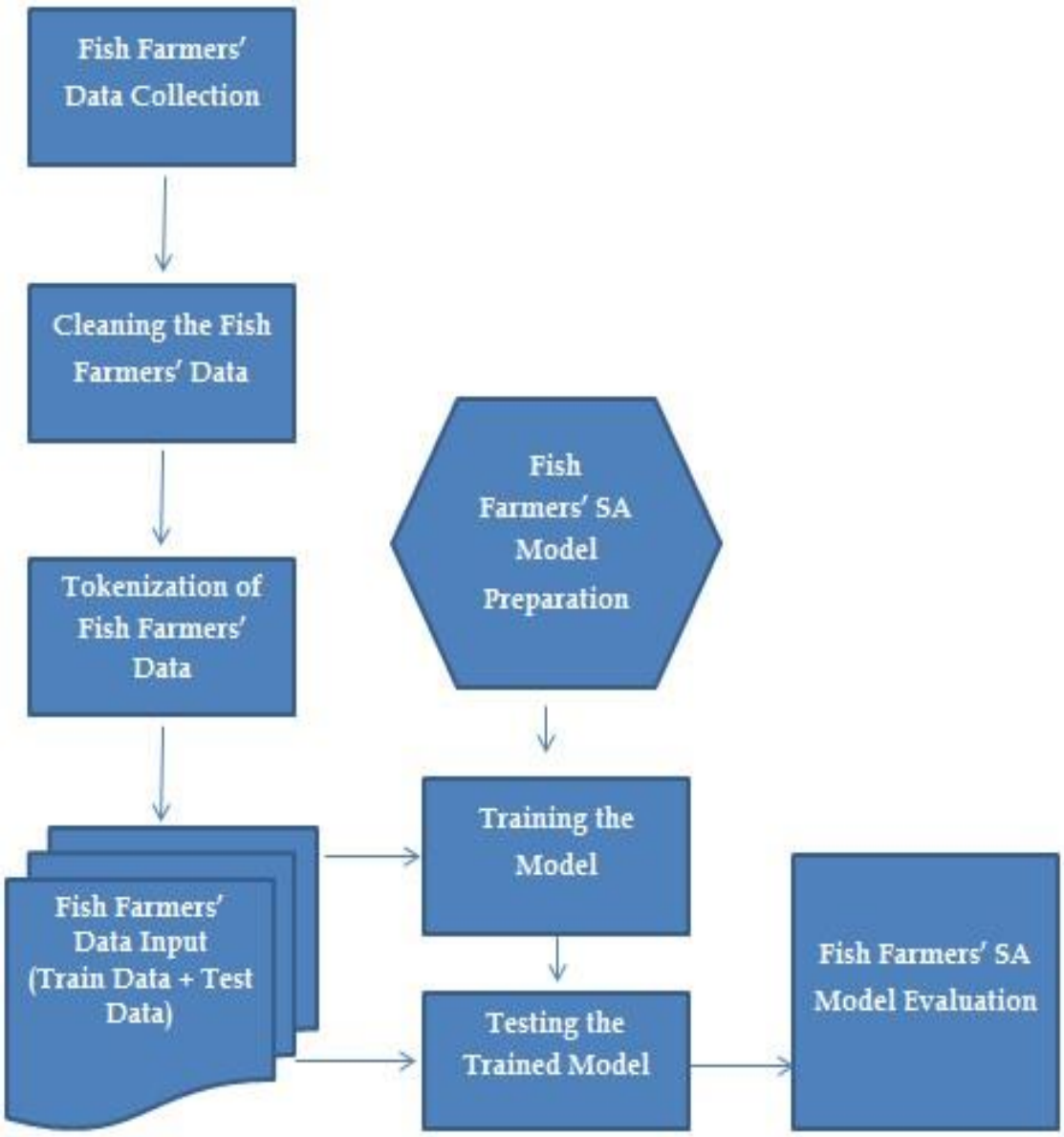

Figure 1: Block Diagram Methodology

Data Collection: Collected online survey data on topic of fish and paddy culture from internet. Based on this we created a dataset of Hindi-English-Chhattisgarhi mixed data.

Data Pre-processing: Clean and preprocess the collected text data by removing stop words, punctuation, and performing other pre-processing tasks.

Once the data is collected, the process of data cleansing becomes crucial in order to extract pertinent information.

Typically, data includes extraneous information that must be eliminated through pre-processing techniques. These techniques remove irrelevant tokens such as hashtags, punctuation marks, duplicated characters, URLs, word spacing, and sentiments. In the suggested system, several pre-processing steps are employed to eradicate special symbols. Table 1 shows the data before cleaning and Table 2 shows data after cleaning:

Table 1: Language mixed data (before cleaning).

Opinion	Text
Positive	धान संग मछली पालन से मय ह बहुत लाभ कमाए हों सभी किसान भाई ला एखर बारे में सोचना चाहिये @@@@@@@
Positive	बहुत बढिया सहमत हैं खेती के लाभकारी तरीका धान संग मछली पालन #किसान_जिंदाबाद
Negative	धुर्रररररर !!!! निपोर कही के
Negative	हमर खेत में पानी नई रुकए
Negative	खर्चा जादा कमाई $$$$$ कम
Positive	बहुत अच्छा, वैरी हैप्पी,खुशी के बात है
Positive	प्रशिक्षन ले चुके हों एसो करहु
Negative	मछली बिना पानी के मर गे !!!! टोटल लास
Positive	बने बने जी बने बने #किसान एकता जिंदाबाद

Feature Extraction: Convert the preprocessed text data into machine understandable format is called feature extraction. It is the process of breaking down words so that a computer can convert text into words. It divides a portion of text data into smaller meaningful units called tokens. Spaces and punctuation marks can be used to separate individual tokens of a sentence. In the proposed system, mixed-language specifications of tokens are given as input for identifiers. Consider the statement: - "हमर खेत में पानी नई रुकए" will be tokenized as ['हमर', 'खेत', 'में', 'पानी', 'नई', 'रुकए'].

Table 2: Language mixed data after cleaning.

Opinion	Text
Positive	धान संग मछली पालन से मय ह बहुत लाभ कमाए हों सभी किसान भाई ला एखर बारे में सोचना चाहिये
Positive	बहुत बढिया सहमत हैं खेती के लाभकारी तरीका धान संग मछली पालन किसान जिंदाबाद
Negative	धुर्र निपोर कही के
Negative	हमर खेत में पानी नई रुकए
Negative	खर्चा जादा कमाई कम
Positive	बहुत अच्छा वैरी हैप्पी खुशी के बात है
Positive	प्रशिक्षन ले चुके हों एसो करहु
Negative	मछली बिना पानी के मर गे टोटल लास
Positive	बने बने जी बने बने किसान एकता जिंदाबाद

Model Selection: Choosing an appropriate machine learning or deep learning model, is very important step for entire process we have chosen Naive Bayes for this experiment.

Naive Bayes Classifier Naive Bayes is a conditional probability model, based on Bayes' theorem, which states that the conditional probability is given by –

$$p(C_k \mid \mathbf{x}) = \frac{p(C_k)\, p(\mathbf{x} \mid C_k)}{p(\mathbf{x})}$$

for each of the K possible outcomes. Using a chain rule for iteratively applying the conditional probability definition:

$$p(C_k, x_1, \ldots, x_n) = p(x_1, \ldots, x_n, C_k)$$
$$= p(x_1 \mid x_2, \ldots, x_n, C_k)\, p(x_2, \ldots, x_n, C_k)$$
$$= p(x_1 \mid x_2, \ldots, x_n, C_k)\, p(x_2 \mid x_3, \ldots, x_n, C_k)\, p(x_3, \ldots, x_n$$
$$= \cdots$$
$$= p(x_1 \mid x_2, \ldots, x_n, C_k)\, p(x_2 \mid x_3, \ldots, x_n, C_k) \cdots p(x_{n-1} \mid x_n, C_k)\, p(x_n \mid C_k)\, p(C_k)$$

Under the naive assumption that all the features x are mutually independent, the probability now becomes:

$$p(C_k \mid x_1, \ldots, x_n) = \frac{1}{Z} p(C_k) \prod_{i=1}^{n} p(x_i \mid C_k)$$

This formula is the base behind the Multinomial Naive Bayes classifier which we used, which deals with the occurrences of a word in a single document.

EXPERIMENTAL SETUP

We have collected Chhattisgarhi Mixed data in Hindi and English. After preprocessing the data, a cleaned dataset is prepared for better results of the proposed approach.

Table 3: Statistics of Dataset

Properties	**Size**
Positive reviews	8973
Negative reviews	7217
Total reviews	16190

Training Model: We split the preprocessed and extracted text data into training and testing data sets. 85 % data is for training and remaining is for testing.

We trained the selected model using the pre-processed and extracted training data set.

Model Evaluation: Evaluate the trained model's performance using the preprocessed and extracted testing data set.

Model Deployment: Deploy the trained model to analyse new and incoming text data in real-time. Sentiment Analysis Results: The sentiment analysis results show the predicted sentiment of the analysed text data, which can be positive or negative.

CONCLUSION

This article describes the extraction of text data in Hindi, English, and Chhattisgarhi on Fish and Paddy Culture from an online survey, followed by sentiment analysis using the Naive Bayes classifier. The authors collected data related to fish farming from the websites, which were then cleaned and pre-processed. The polarity experiments conducted on the Hindi-English-Chhattisgarhi text data showed an accuracy of up to 67%. The authors also noted that improved results were obtained after pre-treatment of the data. In the future, the authors intend to explore other sentiment analysis tools and methods, including those based on deep learning, which offer potential for further improvement.

ACKNOWLEDGMENTS

We are deeply grateful to Dr. Anil Bhatpahari, the Secretary of "Chhattisgarh Rajbhasha Ayog", for sharing his valuable knowledge and expertise with us. Furthermore, we are indebted to Mr. Vaibhav

Joshi, the EIC of Editorialnow.com, for generously providing us with the data from the online survey. Without their invaluable contributions, this project would not have been possible.

REFERENCES:

1. **Ghatge, M.R. and Barde, S., 2022.** Comparison of CNN-LSTM in Sentiment Analysis for Hindi Mix Language. In, 2022 2nd International Conference on Technological Advancements in Computational Sciences (ICTACS) (pp. 453-456).IEEE. https://doi.org/10.1109/ICTACS56270.2022.9988531
2. **Singh, S. and Singh S., 2021.** "Sentiment Analysis of Hindi Text using Machine Learning Techniques," in 2021 8th International Conference on Computing for Sustainable Global Development (INDIA.Com), pp.1296-1300.
3. **Ahmad, M. and Khan, M. S., 2021.** "A Comprehensive Study on Sentiment Analysis of Hindi Text Using Deep Learning Techniques," in 2021 3rd International Conference on Emerging Technologies in Computing (iCETiC), pp. 1-6.
4. **Singh, R. K. and Yadav, N. K., 2021.** "Sentiment Analysis of Hindi Text using Hybrid Approach," in 2021 6th International Conference on Computing, Communication and Security (ICCCS), pp. 1-6.
5. **Singh, S. K. and Kumar, A., 2021.** "Hindi Sentiment Analysis using Pre-Trained Word Embeddings," in 2021 8th International Conference on Signal Processing and Integrated Networks (SPIN), 2021, pp. 183187.
6. **Kumar, S., Kumar, M., and Singh, N., 2021.** "Sentiment Analysis of Hindi Text using Ensemble of Deep Learning Techniques," in 2021 6th International Conference on Internet of Things: Smart Innovation and Usages (IoT-SIU), 2021, pp. 393-398.
7. **Ghatge, M.R., Barde, S. (2023).** The Sentiment Analysis for Hindi Language Using Convolution Neural Network and Deep Belief Network. In:, et al. Information Systems and Management Science. ISMS 2021. Lecture Notes in Networks and Systems, vol 521. Springer, Cham. https://doi.org/10.1007/978-3-031-13150-9_1
8. **Rastogi, A., & Kumar, A. (2020).** A comparative study of deep learning and traditional machine learning approaches for Hindi sentiment analysis. International Journal of Computational Intelligence Research, 16(1), 4150.
9. **Bhardwaj, A., & Sharma, R. (2020).** Aspect-based sentiment analysis for Hindi language using natural language processing techniques. Journal of Ambient Intelligence and Humanized Computing, 11(3), 1013-1025.
10. **Chakraborty, D., & Saha, S. (2020).** Hindi sentiment analysis using ensemble model. Proceedings of the 2020 IEEE 17th India Council International Conference (INDICON) (pp. 1-5). https://doi.org/10.1109/INDICON50882.2020.9343581
11. **Mahajan, S., & Jain, A. (2020).** Bidirectional LSTM-CNN model for Hindi sentiment analysis. Proceedings of the 2020
12. **Agarwal, B., & Bhattacharyya, P. (2020).** Hindi SentiWordNet: A lexical resource for Hindi sentiment analysis. Proceedings of the Second Workshop on NLP for Internet Freedom: Censorship, Disinformation, and Propaganda, 105-112. https://doi.org/10.18653/v1/2020.i nlf-1.11
13. **Gupta, A., & Singh, S. (2020).** Deep learning-based Hindi sentiment analysis on mobile app reviews. Proceedings of the 2020 International Conference on Inventive Research in Computing Applications (pp. 741-745). https://doi.org/10.1109/ICIRCA48931.2020.9279946
14. **Gupta, A., & Jain, P. (2020).** Deep learning-based Hindi sentiment analysis on Twitter data. Proceedings of the 2020 9th International Conference on Reliability, Infocom Technologies and Optimization (ICRITO) (pp.548-552). https://doi.org/10.1109/ICRITO49229.2020.9210994
15. **Sharma, S., & Kumar, A. (2020).** Deep learning-based approach for Hindi sentiment analysis on news articles. Journal of Intelligent & Fuzzy Systems, 39(6), 9227-9235.

16. **Singh, A., & Gupta, M. (2020).** Rule-based approach for Hindi sentiment analysis on news headlines. Journal of Intelligent & Fuzzy Systems, 39(6), 9099-9108.
17. **Srivastava, P., & Agarwal, A. (2020).** Machine learning-based approach for Hindi sentiment analysis on poetry. Journal of Intelligent & Fuzzy Systems, 39(6), 9515-9524.
18. **Mishra, A., & Kumar, V. (2020).** Deep learning-based approach for Hindi sentiment analysis on blog data. Journal of Intelligent & Fuzzy Systems, 39(6), 9345-9353.
19. **Singh, M., & Sharma, M. (2020).** Rule-based approach for Hindi sentiment analysis on movie songs. International Journal of Emerging Technologies in Learning, 15(10), 176-184.
20. **Sharma, S., & Singh, R. K. (2020).** Machine learning-based approach for Hindi sentiment analysis on WhatsApp messages. International Journal of Speech Technology, 23(2), 305-316.
21. **Pandey, A., & Singh, R. K. (2020).** Machine learning-based approach for Hindi sentiment analysis: A comparative study of various algorithms. Journal of Intelligent & Fuzzy Systems, 39(6), 9421-9431.
22. **Sinha A., & Barde, S. (2022)**. Illumination invariant face recognition using MSVM. AIP Conference Proceedings 2455, 040005. https://doi.org/10.1063/5.0100936
23. **Ghatge, M.R. and Barde, S., 2022.** Nutritional and Natural Farming: Agriculturists' Sentiment Analysis. International journal of food and nutritional sciences, Volume 11 Special Issue 2 (pp.1379-1385).
24. **Sharma, S., & Jain, S. (2020).** Recurrent neural network-based approach for Hindi sentiment analysis and evaluated its performance. Journal of Ambient Intelligence and Humanized Computing, 11(11), 4803-4811.
25. **Kumar, A., & Jain, S. (2020).** A lexicon-based approach for Hindi sentiment analysis using SentiWordNet and support vector machines. International Journal of Computer Applications, 179(32), 22-27. DOI: 10.5120/ijca2020921131
26. **Verma, S., & Singh, S. P. (2020).** Machine learning-based approach for Hindi sentiment analysis on movie reviews. Journal of Computational and Theoretical Nanoscience, 17(2), 1214-1220. DOI: 10.1166/jctn.2020.8722