International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10
Article Received: *17 August 2023* ***Revised:*** *22 September 2023* ***Accepted:*** *19 October 2023*

Domain Classification for Marathi blog articles using Deep learning

[1]Kiran N. Girase, [2]Dr. Geetanjali V. Kale, [3]Kimaya R. Urane
[1]Department of Computer Engineering
SCTR'S Pune Institute of Computer Technology
Pune, India
kirangirasemai@gmail.com
[2]Department of Computer Engineering
SCTR'S Pune Institute of Computer Technology Pune, India gvkale@pict.edu
[3]Department of Computer Engineering
SCTR'S Pune Institute of Computer Technology Pune, India krurane@pict.edu

Abstract— Nowadays the exponential growth of online content, particularly in the form of blog articles is tremendous, the need for effective techniques to automatically categorize them into relevant domains has become increasingly important. To overcome the challenges the domains like natural language processing (NLP), machine learning (ML) and deep learning (DL)are being working as booster effect to emerge out with solutions. In this proposed system methodology-based NLP and DL domain the long short-term memory (LSTM) classifier for domain classification and compared the existing multiclass classification techniques with having accuracy around 94% and 91% by long short-term memory (LSTM) model using two different data sets one is Marathi new article and another one Financial article data set. The proposed model is being compared with multiple other models like naïve bayes (NB), XGBoost, support vector machine (SVM) and random forest (RF). The final estimated result achieved is best combination of dataset and deep learning algorithm LSTM.

Keywords- Marathi Natural Language Processing, Text Classification, Marathi BERT, Domain Classification, LSTM, Marathi Datasets.

I. INTRODUCTION

All The increasing volume of online content, particularly in the form of Marathi blog articles, calls for effective techniques to categorize them into relevant domains. Domain classification of Marathi blog articles plays a crucial role in content organization, personalized recommendations, and targeted advertising. However, classifying Marathi blog articles presents unique challenges due to the language's linguistic and structural characteristics. In this survey, the state-of-the-art explored with multiclass classification techniques for domain classification of Marathi blog articles. By examining various approaches, feature extraction methods, and classification algorithms, this survey aims to provide researchers and practitioners with valuable insights into the existing techniques, their strengths, limitations, and potential areas for improvement.

In addition, the survey identifies challenges and open research questions in the domain classification of Marathi blog articles. These challenges may include handling noisy or incomplete data, addressing class imbalance, and incorporating domainspecific knowledge. The survey also sheds light on potential directions for future research, suggesting avenues for improving the accuracy, efficiency, and robustness of multiclass classification techniques for Marathi blog article domain classification.

Overall, this survey plans to be a comprehensive resource. for researchers and practitioners interested in multiclass classification techniques for domain classification of Marathi blog articles. By providing an overview of the existing techniques, highlighting their strengths and limitations, and identifying areas for future research, this survey contributes to the advancement of domain classification in the Marathi language and facilitates the development of more accurate and efficient models in this domain.

Marathi has become the third most spoken language in India, with roughly 83 million people speaking it. The language is indigenous to Maharashtra and is the third most widely used language after Hindi and Bengali. [1]. Despite Maharashtra being the educational and industrial hub of India, Marathi natural language processing (NLP) has received an inadequate amount of attention from academic circles or industry. [2]. In this dissertation, implementing a model for multi-label classification on Marathi blog articles. The feature contribution of this model is to help classify the Marathi blog articles domain-wise. Text-based natural language processing (NLP) has become mainstream with libraries like natural language toolkit (NLTK) and Spacy. These frameworks support both

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10
Article Received: *17 August 2023* ***Revised:*** *22 September 2023* ***Accepted:*** *19 October 2023*

basic rule-based features like tokenization, stemming, lemmatization, and so on, as well as more advanced machine learning-based features like classification, named entity recognition, POS tagging, and so on. Transformers that are used supplies APIs to support cutting-edge PyTorch or TensorFlow models for complex operations such as summarization, machine translation, question answering, sentence similarity, and so on. While production-ready libraries and models for English are readily available, the same cannot be said for low-resource languages. [3]. In this dissertation work, more specifically focus is on Indian low resource language i.e. Marathi.

In this paper the detail study explains about each sections. Section I is introduction part were the whole picture of domain is been explained in detail. Section II is literature part were all the related work done before had been discussed and introfied with different new techniques to fulfil the research work. In section III Proposed methodology is been explained with the detail structure and system architecture so that each concept will be understood to implement the solution easily. Also, we have come up with block diagram that it self-explanatory. Followed by Data collection part, were information of dataset is been discussed. Next Data pre-processing were how the data is been separated from raw data and cleaned out to have as per our expectation. Next follows Feature Engineering part were each attribute and features from the dataset is been processed to get out the best results from the model. Last part is training the model to get accurate output. In section IV the detail understanding to algorithm used with the mathematical explanation of it. In section V there is result and statistic part of the output with caparison of different inputs to get expected output. In section VI final conclusion part were the whole and sole part of paper and achieved insights from the work is depicted.

II. LITERATURE STUDY

Wherever Recently, there has been a tremendous amount of work in single and multilanguage Natural language processing (NLP) applications. A number of attempts have been made to construct monolingual model for solving predictive, classification, and analytical problems in the Marathi language. The Bidirectional Encoder Representations from Transformers (BERT) is currently one of When it comes to NLP tasks like text classification, the most effective language models are those that perform the best. Previous research papers have shown how BERT captures the language context in an efficient way [4][5][6]. In [7] author presented a comparison between monolingual and multilingual transformer-based models, particularly the variants of BERT. They tested these models on datasets for hate speech detection and text classification. They used standard multilingual models namely mBERT, indicBERT, and xlm-RoBERTa for evaluation. They highlight the need for better sentence embedding models . In [8] author proposed L3Cube-MahaNLP supports tasks like tokenization, word vectors, monolingual BERT models, sentiment analysis, named entity recognition, next token prediction, and hate speech detection. L3CubeMahaCorpus, MahaBERT, and MahaFT. The MahaCorpus is a Marathi monolingual corpus and is a significant addition to the existing monolingual corpus. The Marathi BERT is trained in three different flavors namely MahaBERT, MahaRoBERTa, and MahaAlBERT. The MahaFT is the Marathi fast text word embeddings. These resources are exclusively trained on Marathi monolingual corpus. The models are evaluated on downstream Marathi classification and NER tasks work better than their multi-lingual counterparts [9][10].

The author of [11] conducted experiments with and evaluated various deep learning models for the task of Marathi text classification. They present a comparison of various input representations and model combinations. They demonstrate that when used with Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) models, IndicNLP FastText embeddings outperform random initialization and original Facebook FastText embeddings.There is comparative study in the Hindi language where authors approach multiple ways for word embedding. In [12] author did the classification of Hindi poetries where they tests 5 different machine learning algorithms. The accuracy of Random Forest is 56%, K-Nearest Neighbors is 52%, Decision Tree algorithm is 44%, Naïve Bayes is 64% and Support Vector Machine is 52%. Naïve Bayes and Random Forest algorithms are giving better results as compared to other algorithms for the given dataset. In [13] author presents the analysis of Hindi speeches delivered by leaders. These speeches consist of topics such as social, and economic issues, and many more. They followed text preprocessing steps where they add manually stop words that are not present in standard Hindi stop word modules. In this paper, the author used LM Classifier to achieve better accuracy. As study and survey work done so far in Marathi and Hindi languages. As compared there is work done in other languages also.One of them, as demonstrated by the author of [14], revealed how the supervised deep learning neural network classification model can be used to solve the problem of multiclass, single-label Arabic text classification. In addition, a feedforward deep learning (DL) neural network was implemented. The TF-IDF vectors of the most frequent terms in the dataset [14][15][16] were fed into the first layer. The first layer's output was used as the input for the next layer. In addition, the Adam optimizer was used to reduce the error rate. And used two datasets of Arabic documents to run a series of experiments to validate the approach. As a base classifier, the supervised logistic regression classification model is used. When compared to the logistic regression model, the

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10
Article Received: *17 August 2023* ***Revised:*** *22 September 2023* ***Accepted:*** *19 October 2023*

experimental results showed a significant improvement in classification accuracy and time-building model in favour of the deep learning model. The findings indicate that deep learning classification models are very promising for the problem of Arabic text classification. The author of proposed a classification system with multiple classes for the Arabic corpus in [17], which was used to train the XGBoost algorithm. So far, work in the Arabic language has introduced new rich and unbiased datasets for both single-label and multi-label Arabic text categorization tasks [17][18]. They compare several deep learning (DL) models for Arabic text categorization in depth. The LINGO [Label Induction Grouping] algorithm was proposed by the author in [19] for text categorization of Marathi documents. The VSM [Vector Space Model] underpins LINGO. The LINGO algorithm has several advantages over other clustering algorithms, including dynamic clustering based on user queries, identifying cluster labels first and then assigning the document to that cluster, and many others [18][19]. Lingo begins by extracting the user-readable and frequently occurring words/phrases from the input documents. Furthermore, we reduce the term-document matrix using the Singular Value Decomposition (SVD) method, and then we find the labels of clusters and assign documents to those labels based on the similarity value[20].

The common text pre-processing approach I found in many papers that mention the steps for Marathi text pre-processing. Input validation: [21] author they evaluated the given input document, which was written in Devanagari script. Words that are not valid in the Devanagari script are removed. Tokenization is the process of converting a continuous stream of text into words, phrases, symbols, or other meaningful elements known as tokens. In this project, we took input in the form of a Marathi text document and tokenized it using blank spaces
[18][19][20][21].

Stop Word Removal: The Author [22] concentrated on frequently occurring stop words in the document in this step. They are manually maintaining a stop word list in this process, and this list is used to search for a specific keyword using a query. It detects and removes stop-words in input documents. Stopwords add noise to the results and produce poor results, so stopwords should be removed from documents [20][21][22]. Stemming: The process of removing a suffix from a word and obtaining the root word from a complex word is known as stemming. The word is composed of the following elements: word = root word /stem +infection +suffix Suffix is composed of three parts.

TABLE I LITERATURE SURVEY

Sr. No	Paper name		Language	Techiques	Corpus	Limitation
1	Text Classification for Marathi Documents using Supervised Learning Methods -2016 [11]		Marath0i	NB,KNN,MKNN	Selfcreated Documents	KNN and MKNN are time consuming and less accurate.
2	Automatic Text Categorization		Marathi	LINGO	Selfcreated Documents	SVD do not help to reduce the zeros that increase the complexity of the document.
	Methods for Automated TELUGU text categorization with effective classifiers (2013) [7]					
	of Marathi Language Documents -2016 [18]					sparsity of matrix
3	Automatic Text Categorization of Marathi		Marathi	LINGO	Online documents	Limitation of stemming is that it will not return the root word which is available in diction ary form.

	Documents Using Clustering Technique -2014 [19]				
4	Comparison of Marathi Text Classifiers -2014 [2]	Marathi	NB, KNN, Cen-troid and MKNN	Selfcreated Marathi document	As the input set of Marathi document s are increasing it will reduce the performance of KNN and MKNN.
5	Supervised Learning Methods for Bangla Web Document Categorization -2014 [5]	Bengali	DT, KNN, NB and SVM	Selfcreated document	In pre-processing technique root words are not accessible.
6	Indian Language Text Representation and Categorization Using Supervised Learning Algorithm -2013 [8]	Telugu	K-NN, DT, Nave Bayes	Selfcreated document	Feature extraction is not implemented
7	A Comparative study on term weighting	Telugu	SVM, Nave Bayes, K-NN	Selfcreated document	Document represents numeric values of vector which consist of

III. Proposed Methodology

There are multiple types of text classification which include binary and multiclass text classification. So far good research and development have been done in the English language as well as Arabic and Hindi language. In the above section II mention so far work done in the Marathi domain like sentiment analysis, text categorization, and text classification based on the ML approach like Support Vector Classification (SVC), Naive Bayes, and Random Forest (RF). In Table (II) There is modelwise accuracy comparison.

In this paper, the Marathi text pre-processing steps and classification model along with good accuracy is performed. The following steps represent data scraping to model building.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10
Article Received: *17 August 2023* ***Revised:*** *22 September 2023* ***Accepted:*** *19 October 2023*

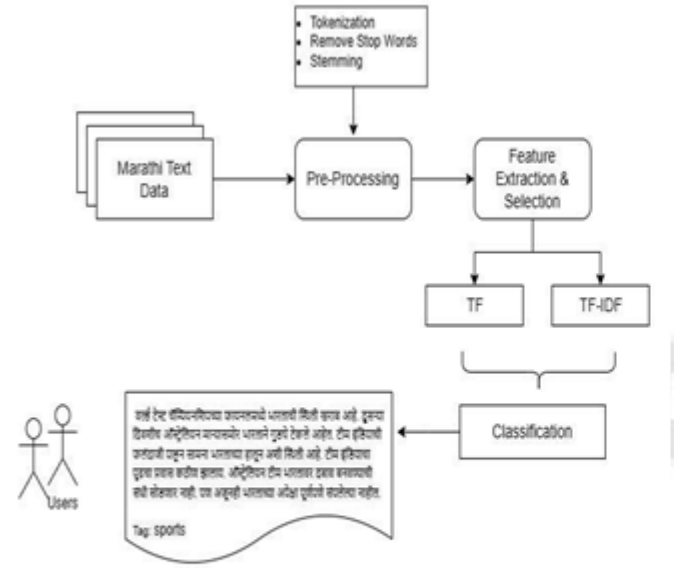

Fig. 1. Block diagram of proposed method

A. Data Collection

The Scraping data from the Marathi new article website. In this step, proposed system had used a web scraping Python module named beautiful soup to help collect data from web sources. And save in the form of Excel or CSV format. There are two columns name Headlines and Label, Where the Headline represents the title of the new and the label represents the domain name of the new article.

B. Data Pre-processing

After scrapping and loading the data it is described in the label column to understand the count of each domain using the bar chart.

1) The pre-processing step is built from scratch where tokenizing the sentence to represent in the form of words. It helps to understand the word frequency of more weighted words.

2) After tokenization there was removing the special character or symbol from tokenized data.

3) Collect the manual list of Marathi stop words that closely belong to Marathi news articles as well as the finance domain. After preparing the list of Marathi stop words match the stop words with tokenized words and if the match is found drop those words from the list of tokenized words.

C. Feature Extraction

In this step, the TFIDF method is used to extract featu
main reasons for using feature extraction and sele
algorithms are that it reduces the dimensionality of the
speeds up training, and improves accuracy. Several sta
feature extraction and selection methods used by vario
researchers in the field of text classification have yielde
positive results [14].
The Term Frequency (TF) uses the following equation
calculate the number of occurrences of a term t in a sp
document d.

$$TF(t, d) = (\text{Number of times term } t \text{ appear in document } d) / (\text{Total number of term in document } d) \quad (1)$$

The Inverse Document Frequency (IDF) metric is used
determine the importance of a term t in the overall dat
Equation 2 is used to calculate it, where N is the total
of text documents.

$$IDF(t) = \log \frac{N}{DF(t)} \quad (2)$$

4) For stemming purposes, collecting the Marathi suffix list of words. It helps to find the Marathi root word.

5) After text processing concatenate the all words.

TFIDF (Term Frequency-Inverse Document Frequency) - It computes the weight of each term t in document d using the equation's product of term frequency and inverse document frequency.

$$TF\text{-}IDF(t) = TF(t) * IDF(t) \quad (3)$$

For selecting features, that is, for the TF-IDF scheme, a normalized value between 0 and 1 has been considered to get more accurate feature values [14].

D. Long Short-Team Memory(LSTM)

Long Short-Term Memory (LSTM) is a type of recurrent neural network (RNN) architecture designed to overcome the vanishing gradient problem in traditional RNNs. LSTMs are widely used for tasks involving sequential data, such as natural language processing and time series analysis.
An LSTM cell consists of several components: an input gate (i), a forget gate (f), a cell state (c), an output gate (o), and a hidden

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10
Article Received: *17 August 2023* ***Revised:*** *22 September 2023* ***Accepted:*** *19 October 2023*

state (h). The equations governing the behavior of an LSTM cell are as follows:

1. Input Gate (i): i = sigmoid (Wix + Whi + bi) (1)
2. Forget Gate (f): f = sigmoid (Wfx + Whf + bf) (2)
3. Cell State Update (g): g = tanh (Wgx + Whg + bg) (3)
4. Cell State (c): c = f * cprev + i * g (4)
5. Output Gate (o): o = sigmoid (Wox + Who + bo) (5)
6. Hidden State (h): h = o * tanh(c) (6)

W and b in the preceding equations represent weight matrices and bias vectors, respectively. Input, forget, cell state update, and output gates are denoted by the subscripts i, f, g, and o, respectively. The input and hidden states are represented by the subscripts x and h, respectively.

The implementation steps for an LSTM network

1) Data Preprocessing: Prepare your data by normalizing or scaling it appropriately. Split the data into training and testing sets.

2) Initialize LSTM Parameters: Define the input size, hidden size, output size, and any other necessary hyperparameters. Initialize the weights and biases for the LSTM cells.

3) Define LSTM Cell Operations: Implement the LSTM cell operations, including the sigmoid and tanh activation functions. These operations involve calculating the forget gate, input gate, cell state, output gate, and the updated hidden and cell states.

4) LSTM Forward Pass: Implement the forward pass function that takes a sequence of inputs and propagates them through the LSTM cells. Iterate over the inputs, calling the LSTM cell operation for each time step and storing the outputs.

5) Training: Define a loss function appropriate for your task (e.g., mean squared error, cross-entropy). Calculate the loss by comparing the LSTM outputs to the ground truth targets.

6) Backpropagation Through Time (BPTT): Calculate the gradients of the loss with respect to the LSTM parameters using the chain rule and backpropagation through time. Start from the last time step and work backward, accumulating the gradients.

7) Parameter Update: Update the LSTM parameters (weights and biases) using an optimization algorithm such as gradient descent. Adjust the parameters in the opposite direction of the gradients, scaled by the learning rate.

8) Repeat: Repeat steps 4-7 for multiple epochs or until convergence, iterating over the training data to improve the model's performance.

9) Prediction: Once the LSTM is trained, you can use it to make predictions on new, unseen data. Pass the new inputs through the LSTM using the forward pass function and obtain the predicted outputs.

10) Evaluation: Evaluate the performance of your LSTM model on the testing set by calculating metrics such as accuracy, precision, recall, or any other suitable evaluation metric for your specific task.

IV. RESULT AND STATISTIC

In result and statistic, we have made comparison with different algorithms so that we have different analysis and understanding of each algorithm. Different algorithms like Naïve Bayes, XGBoost, Random Forest, Support Vector Classifier (SVC) and LSTM. Above each algorithm plays its vital role with finite dataset in achieving highest accuracy for the model.

TABLE II COMPARATIVE ANALYSIS OF ALGORITHM USING DIFFERENT MODELS

Classifier Algorithm	**Name of Data set**	**Model Accuracy**
Naive Bayes	Marathi News Articles	76%
	financial Articles	71%
XGBoost	Marathi News Articles	78%
	financial Articles	80%
Random Forest	Marathi News Articles	79%
	financial Articles	87%
Support Vector Classifier (SVC)	Marathi News Articles	80%
	financial Articles	82%

LSTM	Marathi News Articles	94%
	financial Articles	91%

V. CONCLUSION

In the research study, proposed system refers to several authors' research papers where there are different approaches for Marathi text classification. They use the short Marathi text documents for model training purposes, and based on that they conclude the results and accuracy of the model. In this paper, the system proposed the method for Marathi text classification where Marathi text pre-processing using added manual Marathi stop words had incorporated, and to find the root word manual suffix Marathi keywords were implemented. Based on the implementation model it gives the accuracy of 94% and 91% using two different datasets as Marathi New Article and other one Financial Article. The classifier models have made comparison study with different algorithms like naïve bayes, XGBoost, support vector machine and random forest. So, the proposed system came up with the best results and statistics using LSTM statistical comparison with other classifier models. The future scope can been done by increasing accuracy in terms of using transformers.

. REFERENCES

[1] A. Velankar, H. Patil and R. Joshi, "Mono vs Multilingual BERT for Hate Speech Detection and Text Classification: A Case Study in Marathi," [Online]. DOI: https://doi.org/10.1007/978-3-031-20650-4 10, (2022).

[2] R. Joshi, "L3Cube-MahaNLP: Marathi Natural Language Processing Datasets, Models, and Library," [Online]. DOI: https://doi.org/10.48550/arXiv.2205.14728, (2022).

[3] R. Joshi, "L3cube-maha corpus and mahabert: Marathi monolingual corpus, Marathi bert language models, and resources." [Online]. DOI: https://doi.org/10.48550/arXiv.2202.01159, (2022).

[4] K. Pal and B. V. Patel, "Automatic Multiclass Document Classification of Hindi Poems using Machine Learning Techniques", 2020 International Conference for Emerging Technology (INCET), DOI: https://doi.org/10.1109/INCET49848.2020.9154001, (2020). [5] S. Pundlik, P. Dasare, P. Kasbekar, and A. Gawade."Multiclass classification and class-based sentiment analysis for Hindi language", 2016 International Conference on Advances in Computing, Communications and Informatics (ICACCI), DOI: 10.1109/ICACCI.2016.7732097, (2016).

[6] Katrina Sundus, Fatima Al-Haj and Bassam Hammo, "A Deep Learning Approach for Arabic Text" 2019 2nd International Conference on new Trends in Computing Sciences (ICTCS), DOI:10.1109/ICTCS.2019.8923083, (2019).

[7] Nouh Sabri Elmitwally, "Building A Multi-class XGBoost Model for Arabic Figurative Language," 2020 2nd International Conference on Computer and Information Sciences (ICCIS), DOI: 10.1109/ICCIS49240.2020.9257669, (2020).

[8] Ashraf Elnagar, Ridhwan Al-Debsi and Omar Einea, "Arabic text classification using deep learning models", [Online]. DOI: https://doi.org/10.1016/j.ipm.2019.102121, (2019) [9] Jyoti U. Kundale and Nikita J. Kulkarni, "Language Independent Multi-Class Sentiment Analysis", 2019 5th International Conference On Computing, Communication, Control And Automation (ICCUBEA), DOI:10.1109/ICCUBEA47591.2019.912838, (2020).

[10] M. Sadanandam, A. Nagesh, V. Kamakshi Prasad and V. Janaki, "Text Independent Language Recognition System for Indic Languages With new Features", 2012 IEEE International Conference on Computational Intelligence and Computing Research, DOI: 10.1109/ICCIC.2012.6510219, (2013)

[11] Shalini Puri and Satya Prakash Singh, "An Efficient Hindi Text Classification Model Using SVM", [Online], DOI: https://doi.org/10.1007/978981-13-7150-9 24, (2019)

[12] Omar Einea, Ashraf Elnagar and Ridhwan Al Debsi, "SANAD: Singlelabel Arabic News Articles Dataset for automatic text categorization", [Online], DOI: https://doi.org/10.1016/j.dib.2019.104076, (2019).

[13] Apurva D. Dhawale, Sonali B. Kulkarni, Vaishali M. Kumbhakarna, "Apurva D. Dhawale, Sonali B. Kulkarni, Vaishali M. Kumbhakarna", International Journal of Engineering and Advanced Technology (IJEAT), (2020).

[14] Ankita Dhar, Niladri Sekhar Dash, and Kaushik Roy, "Classification of Bangla Text Documents based on Inverse Class Frequency", 978-15090-6785-5/18 IEEE, (2018). [15] Shraddha A. Narhari and Rajashree Shedge, "Text Categorization of Marathi Documents using Modified LINGO", 978-1-5386-3852-1/17/ IEEE, (2017).

[16] Rajnish M. Rakholia1 and Jatinderkumar R. Saini, "Classification of Gujarati Documents using Na¨ıve Bayes Classifier", Indian Journal of Science and Technology, Vol 10(5), DOI: 10.17485/ijst/2017/v10i5/103233, (2017). [17] Pooja Bolaj and Sharvari Govilkar, "Text Classification for Marathi Documents using Supervised Learning Methods", International Journal of Computer Applications (0975 – 8887) Volume 155 – No 8, (2016).

[18] Jaydeep Jalindar Patil and Prof. Nagaraju Bogiri, "Automatic Text Categorization Marathi documents", International Conference on Energy Systems and Applications (ICESA 2015) Dr. D. Y. Patil Institute of Engineering and Technology, Pune, India, (2015).

[19] Mrs. Sushma R. Vispute and Prof. M. A. Potey, "Automatic Text Categorization of Marathi Documents Using Clustering Technique", 978-1-4673-2818-0/13/ ©2013 IEEE, (2013).

[20] G. Jawahar, B. Sagot, and D. Seddah, "What does BERT learn about the structure of language?" in Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Florence, Italy: Association for Computational Linguistics, Jul. 2019, pp. 3651–3657. [Online]. Available: https://aclanthology.org/P19-1356

[21] I. Tenney, D. Das, and E. Pavlick, "BERT rediscovers the classical NLP pipeline," in Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Florence, Italy: Association for Computational Linguistics, Jul. 2019, pp. 4593–4601. [Online]. Available: https://aclanthology.org/P19-1452

[22] W. de Vries, A. van Cranenburgh, and M. Nissim, "What's so special about BERT's layers? a closer look at the NLP pipeline in monolingual and multilingual models," in Findings of the Association for Computational Linguistics: EMNLP 2020. Online: Association for Computational Linguistics, Nov. 2020, pp. 4339–4350. [Online]. Available:https://aclanthology.org/2020.findings-emnlp.389

[23] Divate, Manisha Satish. "Sentiment analysis of Marathi news using LSTM." International journal of Information technology 13, no. 5 (2021): 2069-2074.

[24] Shilaskar, Swati, Shivam Ghadge, Mugdha Dhopade, Janhvi Godle, and Mihika Gulhane. "English to Marathi Text Translation using Deep learning." In 2022 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT), pp. 1-5. IEEE, 2022.