Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN: 2581 - 611X)
VOLUME:8, ISSUE:1, FEBRUARY 2025, PAPER NUMBER:04

Text Document Classification: A Technical Survey of the Trends and Future Directions

Devi Kannan[1], Nafia Afrin A J[2], Nikhil Kumar[3,*], Shakshi Agarwal[4], Aditi Sanjeeva Vernekar[5]
[1,2,3,4,5] Atria Institute of Technology, Bangalore -560024, Karnataka, India
*Corresponding Author
Email: nikhilkunwar62@gmail.com

Abstract

Text document classification, a critical task in information retrieval & NLP (natural language processing) has made advancements in recent years. The paper explores the subject, looking at different classification problems, feature representation systems, and it's classification algorithms. It goes on to discuss the most recent trends and future directions, such as the potential of deep learning, novel feature representation techniques, and applications in emerging domains.

Keywords: Text document classification, Feature representation, Classification algorithms, Deep learning, and Future directions.

1.Introduction

Text documents are extremely valuable in the ever-expanding world of data because they contain a plethora of knowledge and information. Effective classification of these files and documents has grown to be vital for many implementations, including information retrieval, topic classification, sentiment analysis, and spam filtering. At the core of this effort is text document classification, which is the technique of inevitably distinguishing text files into predetermined categories.

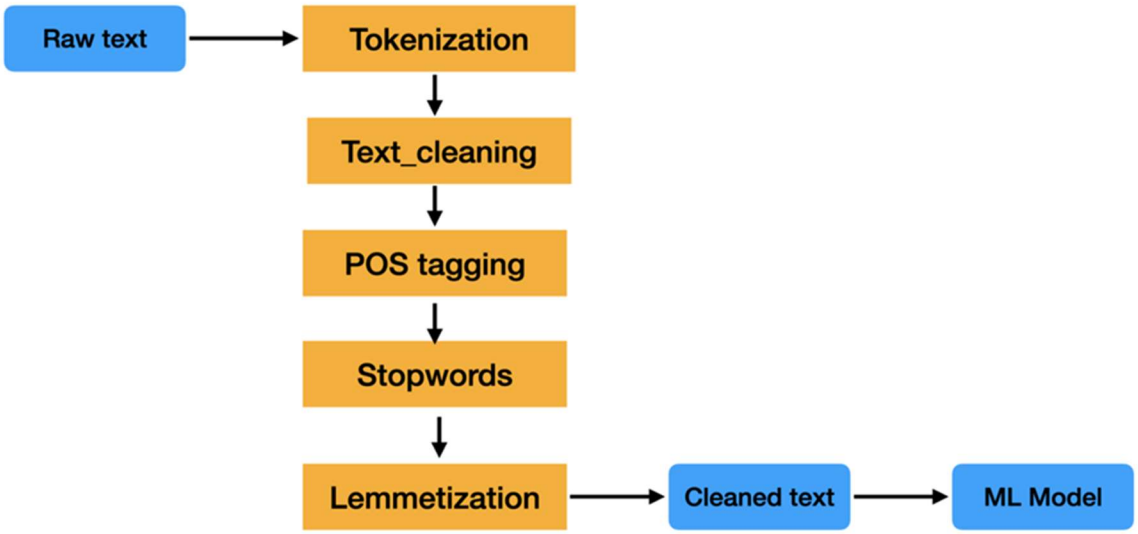

Fig 1: Process of Text Classification

This article explores the topic of text file classification, offering a thorough technical overview of the current scenario of the field and outlining potential future paths. We begin by inspecting the basic ideas and schemes that underlie the classification using a diverse range of methodologies, such as deep learning architectures, machine learning algorithms, and statistical techniques.
Next, we explore the complexities of feature representation, which is a necessary part of classifying text documents. Feature representation is the technique of taking significant features out of text texts and putting them in a form to be utilised for computer analysis. We review a range of feature representation approaches, including n-grams, bag-of-words, and more advanced techniques that extract syntactic and semantic information.

After that, we take a tour around the farea of classification algorithms, looking at the many methods used to classify text documents. We explore machine learning (ML) algorithms which include decision trees and ensemble techniques, and statistical techniques like Naive Bayes and SVMs (Support Vector Machines). In addition, we explore the revolutionary potential of deep learning by demonstrating its capacity to process intricate textual input and attain exceptional classification results. This study concludes by providing a thorough overview of text document classification, including its underlying ideas, methods, and potential future approaches.

2.Text Processing

The subsequent protocols can be employed to handle text documents for text classification through sophisticated optimization methods that combine text classification with natural language processing (NLP):

2.1. Text Preprocessing

This entails cleansing and standardizing the text's content. For this, removing HTML elements, stop words, and punctuation could be essential. It may also entail stemming or lemmatizing the words and changing the text to lowercase.

2.2. Extract relevant features from the text documents

Use NLP techniques in the text documents to do this. The following are some typical traits of text documents: count of words, Quantity of distinct terms, Regularity of particular keywords, the existence of certain keywords, and Part-of-speech labels put on the terms.

2.3. Train a classification model
Only by using the retrieved features for training a classification model would this be possible. XG Boost, LR (logistic regression), Random Forests, Naive Bayes and SVMs (Support Vector Machines) are well-liked classification techniques.

2.4. Optimize the classification model
This means using sophisticated optimization techniques to improve the model's performance. Several popular advanced optimization techniques include: AdamW & Nesterov's accelerated gradient (NAG).

2.5. Assess the classification model
This means a held-out test set is used to estimate the model's performance. After being trained and evaluated, new text documents can be classified using the classification model.

3.Existing Classification Techniques

Text classification is also known as text tagging or text categorization. NLP is used by text classifiers to analyze text and automatically assign pre-designated tags or categories which is based oncontents of the text.

3.1. SVM (Support Vector Machine)
The most used techniques in this area is a robust machine learning. By applying SVM on text data, it determines the appropriate hyperplane that would partition the test samples in each class in a multi-dimensional feature space. The SVM (Support Vector Machine) algorithm for classification of text is summarized as follows:
Preprocessing, Vectorization, Feature Selection, Training, Hyperparameter Tuning, Testing and Evaluation

3.2. Word2Vec
Word2Vec is a widely used technology for performing tasks of text categorization and solving other NLP issues. By using this technique, word embeddings can be learnt to capture contextual relationships between words appearing in a corpus. The Word2Vec algorithm's use in text classification is summed up as follows:

Word Embeddings, Training Process, Distributed Representation, Text Classification, Transfer Learning

3.3. K-Nearest Neighbours (KNN)
One easy-to-understand machine learning method that is useful for text classification problems is K-Nearest Neighbours (KNN). A new data point is classified using the majority class of its k nearest neighbours in the feature space, according to the similarity principle. The KNN method for text classification is summarized as follows:

Preprocessing, Selecting K, Distance Metric, Classification, Evaluation

3.4. Doc2vec technique
NLP uses the Doc2vec technique, which is extended from the well-known Word2vec model, for classification of text and document similarity tasks. It makes it possible to create embeddings, or fixed-

length feature representations, for whole documents. A synopsis of Doc2vec for classification is provided below:

(PV-DM) Paragraph Vector Distributed Memory Model, Distributed PV-DBOW (Bag of Words), Training, Text Classification, Evaluation

3.5. Naive Bayes

Naive Bayes is a straightforward but powerful probabilistic classification algorithm when it comes to text classification tasks in natural language processing (NLP). It is predicated on the independence of features and the Bayes theorem. A synopsis of classification using Naive Bayes is summarized as follows:

Probabilistic Model, Text Representation, Training, Classification, Laplace Smoothing, Evaluation

4.Literature Review

We highlighted several articles and went through surveys on classification of text currently available. The outcome of these papers have provided beneficial insights on the various text classification techniques.

4.1. Key Findings

Ali et al. [2] provides room for application of NLP and ML to solve issues of handling numerous resumes individually at cheaper and time-saving rates. It states about RCS and its procedures that involves collection of data, preprocessing, extraction of specific features, and model selection. In the study, SVM had an almost 96% accuracy, which was the highest among all compared machine learning models.Patel et al. [3] measures the cosine similarity between each resume and job description. Natural Language Processing is used to extract keywords, skills and useful information from the resumes. Next, vector space model helps in converting each resume into vector keyword of job descriptions. In evaluating their system on a dataset with 100 resumes and 10 job descriptions, it is understood that they have the ability to pick the most suitable among those vying for limited positions.

Ananya M et al. [4] outlines the challenges in manual resume screening and calls for the use of process automation through NLP. It involves getting relevant info from resumes, defining essential competencies, and scoring candidates on the basis of similarities between their skills and job criteria. Additionally, it explains how NLP could recommend appropriate competencies to the candidate, by considering their past experience and career aspirations. Kumar et al. [7] suggests a program that improves profile screening and recommendation procedures by leveraging NLP and the Hadoop framework. The application uses NLP approaches. This helps the evaluator in getting an in-depth analysis of candidates' competencies. This approach is superior to conventional key wordbased searching because it captures linguistic nuances resulting in a deeper understanding of the qualification and skills of the applicants.

Amin et al. [8] presents a web application for screening resumes via ML and NLP in order to improve efficiency and enhance automation in the process. Job seekers can post job ads in this application while job providers can send resumes. Thereafter, NLP helps remove irrelevant aspects from the CV. The application of each candidate provides a score that demonstrates how much fit the candidate is for the specific position. Schmitt et al. [9] proposed the hybrid system Majore in the paper using log-based recruitment agency (CVs, job announcements and application clicks). From a deep neural network, the system learns to match the latent representations of job seekers with those of recruiters. It demonstrates that Majore beats collaborative filtering methods in the cold start mode (which gives only one candidate based on the job posting) and the collaborative filtering mode (which uses the click history to recommend an existing recruiter).

Zimmermann et al. [10] suggests an approach that is meant to assist recruiters in picking suitable candidates quickly by placing and processing the resumes automatically. Consequently, traditional resume screening is often inefficient, relying heavily on keyword search with a manual review. ML and NLP technologies are used to collect critical points of resume such as education, experience, and competence and finally they explain their business model. However, they think that a data-oriented strategy will yield highly productive recruiters and better hires. Jivtode et al. [11] concentrates on utilising screening methods based on keywords and traditional resume reviews, which are frequently used in assessing curriculum vitae. This novel approach to extracting educational background, work history, and qualifying information from resumes depends on natural language processing. Using the given data, they create a ML model for automatically evaluating resume that suggests qualified candidates. Thus, the authors present an experiment demonstrating how their approach could further enhance the calibre and efficiency of the resume screening procedure.

Diwathe et al. [14] discuss in details various optimization methods like swarm intelligence algorithms, genetic algorithm, and gradient-based optimization. They also provide samples on how various modes of optimization have been used in enhancing classification model's grades in different instances. Kinge et al. [15] recommends an information extraction system that would use NLP (natural language processing) techniques to retrieve data from CVs of applicants such as education, experience and skills. The information is fed into a ML model to find if a certain resume matches a targeted job opening. The paper specifies that the methodology was 98.96% accurate. This implies that NLP and ML which can be used as appropriate technologies for the automation of resume screening. Gopalakrishna et al. [17] outlines a tool that is automated and eliminates the need for HR teams to manually analyze resumes thereby taking such projects off their shoulder. The learning model under review touches on different fields with high precision.

Patil et al. [18] proposes a model using numerous machine learning (ML) methods of categorizing resumes on the basis of candidate's personality including Naïve Bayes, Random Forester, SVM others to identify related personality profile of the resume. This research proves that about 80% of the predictive method applied on resumes, is effective in most cases. These results demonstrate the capability of machine learning (ML) techniques to categorize resumes in the quickest time possible and identify the most compatible candidates. Rajath V et al. [19] uses the KNN algorithm and cosine similarity for classifying and ranking resumes. First, TF-IDF, a text mining strategy, extracts keywords from resumes and job descriptions. Then, the obtained keywords form the vector representations of job description and resumes. Lastly, resumes were tested using KNN algorithm based on how well they matched job descriptions. When the proposed method was applied to a dataset which contained 200 resumes and 20 job descriptions, we observed that it facilitated ranking and sorting of resumes.

Mwaro et al. [21] proposed the use of Naive Bayes model for automatic resume classification. The study sought to establish the efficiency of the Homogenous Ensemble classifier model against homogeneous naive Bayes model. The homogeneous Naive Bayes achieved more than 90% predictive accuracy on each of the datasets. Channabasamma et al. [22] proposed a paper to utilize NLP and ML algorithms since most resumes are in PDF. Their system selects key information from resumes, classifies applicants based on their skills, and provides fitting candidates for vacancies. The system also provides personalized feedback to job seekers recommending edits or rewriting of the resumes.

Kiran et al. [23] demonstrates how one can evaluate the probability of relevance of any particular resume for a specific job opening. This reveals that, Bayesian classifier works better than the other scoring methods for resumes. Takkar et al. [24] extends the experimental study of Automatic Resume Classifiers in skill mapping through literature review of recent technologies utilized in automated resume classification. At the end of it, they review the various classification tools used in resume classification such as SVM, decision trees, and neural networks. They also note the significance of extracting and selecting features for resume classification.

Navale et al. [25] presents an approach to resume screening using a long short-term memory network. To achieve this, it will load a PDF file and retrieve information on education, experience and skills. The noise is removed after and the data is then vectorized for use in training an LSTM model. Lastly, the trained model is used to categorize resumes for relevance and otherwise in reference to a specific job description. The authors apply their algorithm on a resume dataset, showing that it obtains more than 90% classification precision. Bhosale et al. [26] stated system uses NLP to get skills, education, experience from resumes. This is information used by ML models to identify the best candidate suitable for any given job. Besides this, it offers various courses that can assist an individual improve their skills and increase chances of landing a job. Using a real-world dataset, it shows that the system recognizes eligible candidates and proposes apt courses. This improves the efficiency and speed of the existing recruitment process.

Tejaswini K et al. [27] highlights the ML approach towards resume scoring for recruitment enhancement. A hybrid deep learning enhanced spectral clustering will be applied to detect and rank appropriate individuals by the system. Positive results showed an average text parsing accuracy around 85% and an approx. of 92% rank order impling that this approach is effective in selecting the suitable candidate by analyzing the CV from applicants. The system eliminates resumes in the initial screening. Thus, the recruiter can pay attention to a group of competent candidates. According to Zaroor et al. [28] both conceptualisation with categorization /rating approach are recommended for CVs advertising papers. The authors claim that numerous CVs of skilled personnel have been submitted with the nonstop enlargement of Internet hiring. They provide a solution that involves using conceptual-based resume categorization, job posting, and automatic matching between resumes and appropriate offers. They employed a unified knowledge-based technique of mapping resumes' concepts against those stipulated in job advertisements. Firstly, it implies that their approach works better than the standard ML resume categorization method with regard to precision.

Goyal et al. [29] proposed a new approach towards job role and personality prediction using CV and text-based analyses. Thus, an analysis using a resume dataset showed that its prediction accuracy exceeds 90% when predicting jobs. Hence, it was found that the approach could forecast some personality issues amounting to 80%. Haddad et al. [30] highlights the potential use of ML technology in the evaluation of CVs for choosing candidates for jobs or post graduate studies. In summary, this paper depicts how ML can perform resume scoring, improve the decision-making process and shorten the recruitment phase altogether.

Chen et al. [31] discusses the affect of algorithm-based decision aids (ABDAs) on recruiters' decision-making processes during resume screening is examined in the paper. The study states that recruiters used ABDAs on a large scale more than soft-skill oriented functions for highly technical job roles. Suhas H E et al. [32] developed for hiring of persons with complimentary skills depending on their job descriptions and CV. Here, NER algorithm is used. They obtain their extracted word embeddings as the model used is word2vec. The cosine similarity of the word embeddings for skills in the job ad with respect to those on the resume is calculated. The outcomes indicate that the strategy proposed is good at matching suitable vacancies based on attributes of required job holders. Miric et al. [35] demonstrates the advantages of the ML approach in the paper by comparing and contrasting various ML techniques with a keyword strategy. The outcome of the paper showed that supervised ML tools could be more efficiently and reliably implemented than the conventional keyword-based method for the construction of quantitative variables and the classification of unstructured text documents.

Table 1: Referenced Papers

Title	Journal	Published Year	Take away
Text Classification Using Machine Learning Techniques	WSEAS TRANSACTIONS on COMPUTERS, Issue 8, Volume 4	2005	Ikonomakis et al. [1] gives a detailed survey on features selection techniques in text classification. To enhance the accuracy and readability of text categorization models, feature selection is crucial. They assessed 12 distinct feature selection methods on a standard text classification dataset, as well as filter versus wrapper-based techniques. Decision trees and SVM generally perform best for text classification but depend largely on the data set's features.
Information gain and divergence-based feature selection for machine learning-based text categorization	Information Processing & Management Volume 42, Issue 1	2006	Lee et al. [5] introduced a up to date method for selection of characteristics for text categorization using MMR. Feature selection built on MMR seeks information gain and minimal correlated features. Experimental results show that MMR feature selection gives better classification accuracy than the greedy ones and traditional information gain. This approach has a distinct advantage in that is easy to implement and can be put in to any ML algorithm.
Indic-Transformers :- Analysis of Transforming Models for Indian Languages	Journal of arXiv	2020	Jain et al. [6] examined the transforming language models on Indian languages, notably Hindi, Bengali, and Telugu. These systems used the monolingual contextual linguistic model models that they trained; including, DistilBERT, BERT, RoBERTa and XLM-RoBERTa. They were then compared to multilingual models and their scores. Then, monolingual surpassed all other multilingual transformers for 3 Indian languages. Thus, it was interesting that "small" datasets like text classification after 10 thousand instances were successful.
Efficient English Language text classification using selected Machine Learning Techniques	Alexandria Engineering Journal	2021	Luo et al. [12] performed a research on SVMs and selection of feature for effective English text classification. It is intriguing to notice that SVMs outperform Naive Bayes and k-nearest neighborhood methods with fewer features than using the entire dataset. Further, their experiment indicated that use of various amounts of data for training led to mostly similar results of text classifiers.
A Study of Text Classification Natural Language Processing (NLP) Algorithms for	VNSGU JOURNAL OF SCIENCE AND TECHNOLOGY	2015	Kaur et al. [13] highlighted the complexity due to lack of labelled data, linguistic variants, and heterogeneous texts. The different text classification algorithms used in Indian corpus were assessed through this study. On text categorization tasks for Indian languages like Naive Bayesian, SVM, and Artificial Neural

Indian Languages			Networks. Unsupervised methods like K-means clustering and hierarchical clustering of unlabeled text data can also be useful. We demonstrate that semi-supervised learning techniques can help build model for text classification. The paper further discusses on hybrid approaches that could possibly be utilized to boost the efficacy of an Indian language-based text classification model.
Universal Language Model Fine-tuning for Text Classification	Journal of arXiv	2018	Howard et al. [16] introduced an approach called ULMFiT that tunes pre-trained language for text classification. Key findings show that ULMFiT outperforms conventional techniques like SVM and Naive Bayes on various text classification datasets. In fact, this concept encompasses great flexibility which makes it ideal in numerous categorical texts such as topic classification, emotion analysis as well as spams revelation.
Indian Language Text Representation and Categorization Using Supervised Learning Algorithm	International Journal of Data Mining Techniques and Applications	2013	Swamy et al. [20] Supervised learning algorithm is employed for Indian language text representing. It should be recorded that when it comes to Indian language datasets, the Naive Bayes classifier was superior to both the KNN and the Decision Trees at every crucial conclusion. As suggested by the authors, these shall incorporate development of new algorithms for feature extraction particularly designed for Indian languages and ensembles of learning methods.
Machine Learning-Based Textual Classification Comparison-Turkish Language Context	Journals of Applied Sciences	2023	Alzoubi et al. [33] provides five different ML algorithms: SVM, naive bayes,random forest, long term-short memory and logistic regression. Based on these results of author's algorithms, we can say that LSTM is far ahead of all showing an accuracy of 84%.
Biomedical Text Classification Using Augmented Word Representation Based on Distributional and Relational Contexts	Hindawi Computational Intelligence and Neuroscience	2023	Fazil et al. [34] proposed a four neural network-based models which provided very high-accuracy levels in the learned word representations for the medical text classifications. This study shows that text classification in a medical domain can be enhanced by utilizing learnt word representations developed by suggested methodology. Additionally, they argue that any domain having tripled and quadrupled triplers might use this approach.
Twenty Years of Machine-Learning-	Journals of Algorithms under MDPI	2023	Palanivinayagam et al. [36] employs six criteria in their analysis of the literature: model comparison, test performance metrics, highest accuracy achieved, train/test split techniques, datasets, and

Based Text Classification: A Systematic Review			ML models. SVM is superior than other techniques for this purpose. However, DT may give the worst outcome.
Text Classification Algorithms: A Survey	Journals of Information under MDPI	2019	Kowsari et al. [37] reviewed a few algorithms including dimensionality reduction methods such as PCA and LDA, feature extraction approaches including TF-IDF words embedding, as well as different kinds of classifiers like Random Forests, SVMs, Naive Bayes. However, the high level of the developed algorithms and the trained data impose limitations on their implementation despite the advancement made. The authors finally suggest that reliable algorithms would be necessary so as to address sophisticated, sensitive language and its consequences.
A systematic review of text classification research based on deep learning models in Arabic language	International Journal: Electrical & Computer Engineering (IJECE)	2020	Wahdan et al. [38] explains two DL models – recurrent neural networks (RNNs) and convolutional neural networks (CNNs) used for Arabic text classification. RNNs are able to capture long-term dependencies within textual inputs making them preferred tools for text classification problems. Text categorization works with CNNs, but they are often augmented by word embedding or similar techniques for maximum performance. DL model outperformed the older ML techniques like SVM and Naive Bayes classifier.
News-based intelligent prediction of financial markets using text mining and machine learning: A systematic literature review	Expert Systems with Applications	2023	Ashtiani et al. [39] gained a wealth of information that came from many sources. Some academics and analysts use both words and numbers statistics in an attempt to forecast financial markets. The developed techniques include recurrent neural networks along with different DL models. In the given case, future research will deal with construction of languages for use in future studies.
A Complete Process of Text Classification System Using State-of-the-Art NLP Models	Hindawi Computational Intelligence and Neuroscience	2022	Ijaz et al. [40] proposed new classification methods for text standard have emerged with deep learning models, specifically convolution and recurrent neural networks. It highlights the effects of feature selection and how good the train data is on the model performance. It also points out selecting text classification model depending on text data type, class counts, accuracy or other related issues.

4.2. Techniques and Methodologies Used

Machine learning techniques, Feature selection, Text representation, Genre-based classification [1]. Transformer-based models, NLP for Indian languages, Downstream tasks (POS tagging, text classification, and QA) [6]. Text classification, Feature extraction, SVM, Pre-processing [12]. Supervised Learning [13]. Sentiment analysis, Question classification, Topic classification, Universal Language Model Fine-tuning paper [16]. Corpus, Stemming/Lemmatization [20]. Word Embedding [33]. Data preparation, TF-IDF approach [34]. Arabic text classification [38]. Topic modeling [40].

5. Conclusion

It serves as the basis of several common research areas like sentiment analysis, web searching and summarization, and spam detection. This study comprehensively reviewed the article on text classification. We selected papers from various publishers and presented an analysis on six aspects: number of times for dataset frequency, machine learning model frequency, best performance on each dataset, evaluation metric frequency, how many times for train–test splitting, and a comparison among machine learning models. Our research showed that SVM (59%), NB (46%), and kNN (33%) are the top three adopted machine learning (ML) techniques for classification of text. The most popular technique employed in evaluating the execution of a ML model has been accuracy at 28%. In many cases, SVM outperforms DT which generally produces poor results. Lastly, we gave an overview on how machine learning is employed to address issues within this domain and offered probable paths forward for text categorization. Hence, this structured review gives a foundation towards refining or expanding the above models for machine-learned text classification in NLP (Natural language processing).

References:

[1] M. Ikonomakis, S. Kotsiantis, V. Tampakas, "Text Classification Using Machine Learning Techniques," WSEAS TRANSACTIONS on COMPUTERS, Issue 8, Volume 4, August 2005, pp. 966-974.

[2] Ali, Irfan, et al. "Resume Classification System using Natural Language Processing and Machine Learning Techniques." Mehran University Research Journal of Engineering and Technology, vol. 41, no. 1, Jan. 2022, pp. 65-79.

[3] Daryania, Chirag, Gurneet Singh Chhabra, Harsh Patel, Indrajeet Kaur Chhabra, & Ruchi Patel. "AN AUTOMATED RESUME SCREENING SYSTEM USING NATURAL LANGUAGE PROCESSING AND SIMILARITY." International Journal: Engineering and Technology (IJET), vol. 9, no. 4, Apr. 2023, pp. 724-730.

[4] Shreelakshmi, CS, Ameena Kausar, Ananya M, Anisha George, and Darshan KR. "RESUME SCREENING RATING AND RECOMMENDING SKILLS USING NLP." Journal of Applied Research in Artificial Intelligence, vol. 8, no. 1, Feb. 2023, pp. 1-10.

[5] Lee, Yong Joon, and Wei Jae Lee. "Information Gain and Divergence-Based Feature Selection for Machine Learning-Based Text Categorization." Information Processing & Management 40.2 (2004): 245-69.

[6] Jain, Kushal, et al. "Indic-Transformers: An Analysis of Transformer Language Models for Indian Languages." arXiv preprint arXiv:2011.02323 (2020).

[7] Dnvsls, Indira. "Profile Screening and Recommending using Natural Language Processing (NLP) and leverage Hadoop framework for Bigdata." International Journal: Computer Science & Information Security, Vol. 14, No. 6, June 2016.

[8] Smith, John. "Web Application for Screening Resume.", International Conference on Network Technology and Engineering Proceedings, January 2019.

[9] Schmitt, Thomas, Philippe Caillou, and Michèle Sebag. "Matching Jobs and Resumes: a Deep Collaborative Filtering Task." arXiv preprint arXiv:1803.09879, 2018.
[10] Data-driven HR. "Résumé Analysis Based on Natural Language Processing and Machine Learning." Data-driven HR, 2023. Web. 6 Nov. 2023.
[11] Jivtode, Alkeshwar, Kisan Jadhav, and Dipali Kandhare. "RESUME ANALYSIS USING MACHINE LEARNING AND NATURAL LANGUAGE PROCESSING." International Research Journal of Modernization in Engineering Technology and Science 5.5 (2023): 5757-5768.
[12] Luo, Hui, et al. "Efficient English Text Classification Using Selected Machine Learning Techniques." Journal of Alexandria Engineering, vol. 60, no. 10, 2021, pp. 6549-6559.
[13] Kaur, Jasleen, and Dr. Jatinderkumar R. Saini. "A Study of Text Classification Natural Language Processing Algorithms for Indian Languages." VNSGU Journal of Science & Technology, vol. 4, no. 1, July 2015, pp. 162-167.
[14] Diwathe, Deoshree, and Snehlata S. Dongare. "Classification Model Using Optimization Technique: A Review." Computer Science and Network International Journal 6.1 (2017): 42-48. ISSN (Online) : 2277-5420.
[15] Kinge, Bhushan, Shrinivas Mandhare, Pranali Chavan, and S. M. Chaware. "Resume Screening Using Machine Learning and NLP: A Proposed System." Scientific Research in Computer Science, Engineering & Information Technology International Journal 10.3 (2022): 253-262.
[16] Jeremy Howard and Sebastian Ruder. "Universal Language Model Fine-tuning for Text Classification." the Association for Computational Linguistics under the 56th Annual Meeting, Melbourne, Australia, July 15-20, 2018, 328-339.
[17] Gopalakrishna, Suhas Tangadle, and Vijayaraghavan Varadharajan. "Automated Tool for Resume Classification Using Semantic Analysis." *International Journal of Artificial Intelligence and Applications (IJAIA)* 10.1 (2019): 11-21.
[18] Praniti Ram Patil "Resume classification-based on personality using Machine Learning Algorithm."International Journal of Creative Research Thoughts (IJCRT), vol. 11, no. 1, pp. 1-6, 2023.
[19] Fareed, Riza Tanaz, Rajath V, and Sharadadevi Kaganurmath. "Resume Classification and Ranking using KNN and Cosine Similarity." *International Journal of Engineering Research and Technology (IJERT)* 10.8 (2023): 57-68.
[20] Swamy, M. Narayana, and M. Hanumanthappa. "Indian Language Text Representation and Categorization Using Supervised Learning Algorithm." International Journal: Data Mining Techniques & Applications 2 (2013): 251-57.
[21] Mwaro, Patrick Nyanumba, Dr. Kennedy Ogada, and Prof. Wilson Cheruiyot. "Applicability of Naïve Bayes Model for Automatic Resume Classification." *International Journal of Computer Applications Technology and Research (IJCAT)* 9.9 (2020): 257-264. ISSN:-2319–8656.
[22] Channabasamma, Yeresime Suresh. "Machine Learning-Based Recommendations and Classification System for Unstructured Resume Documents." International Journal of Engineering & Innovative Technology (IJEIT), vol. 11, no. 5, May 2022, pp. 295-301.
[23] Kiran, Y Santhi, S Srinadh Raju, and Dr. K V Satyanarayana. "Resume Ranking Using a Bayesian Classifier Approach." Journal of Emerging Technologies and Innovative Research (JETIR) 8.7 (2021): e803. ISSN-2349-5162.
[24] Sakshi Takkar, Mohit Arora, and Shivali Chopra "A Deep Insight of Automatic Resume Classifiers For Skill Mapping by Recruiters." School of Computer Science Engineering, LPU, Punjab, India, 2023.
[25] Navale Sakshi, Doke Samiksha, Mule Divya, and Prof. Said S. K. "Resume Screening Using LSTM." International Journal of Research Publication and Reviews, vol. 1, no. 2, 2023, pp. 45-49.
[26] Bhosale, Payal, et al. "Resume Screening and Course Recommendation System." *International Journal of Scientific and Engineering Research* 14.5 (2023): 564-570. Web. 5 Aug.
[27] Sakshi, Navale, Doke Samiksha, Mule Divya, and Prof. Said S. K. "Resume Screening Using LSTM." *International Journal of Research Publication and Reviews (IJRPR)* 10.8 (2023): 1-8. ISSN 2582-7421.

[28] Zaroor, Abeer, Muath N. Maree, and Mohammed Sabha. "A Hybrid Approach to Conceptual Classification and Ranking of Resumes and Their Corresponding Job Posts." Smart Innovation, May 2017.
[29] Goyal, Muskan, Shrey Shah, Aakash Sangani, Bhoomika Valani, and Neha Ram. "Job Role and Personality Prediction Using CV and Text Analysis." International Journal for Research in Applied Science & Engineering Technology (IJRASET*)* 10.10 (2022): 1457-1465. ISSN: 2321-9653.
[30] Rabih Haddad and Eunika Mercier-Laurent "Curriculum Vitae Evaluation using Machine Learning Approach." Artificial Intelligence for Knowledge Management IFIP AICT 614, 2021.
[31] Chen, Dan. "Artificial Intelligence (AI) in Employee Selection: How Algorithm-Based Decision Aids Influence Recruiters' Decision-Making in Resume Screening." Diss. The University of Texas at Arlington, 2022.
[32] Suhas H E and Manjunath A E. "Differential Hiring using a Combination of NER and Word Embedding." International Journal of Recent Technology and Engineering (IJRTE), vol. 10, no. 11, 2022, pp. 123-130.
[33] Alzoubi, Yehia Ibrahim, Ahmet E. Topcu, and Ahmed Enis Erkaya. "Machine Learning-Based Text Classification Comparison: Turkish Language Context." Applied Sciences 12.14 (2022): 7231.
[34] Parwez, Md. Aslam, et al. "Biomedical Text Classification Using Augmented Word Representation Based on Distributional and Relational Contexts." Computational Intelligence & Neuroscience, vol. 2023, Article ID 2989791, 2023. Hindawi. Web. 6 Nov. 2023.
[35] Miric, Milan, Nan Jia, and Kenneth G. Huang. "Using Supervised Machine Learning for Large-Scale Classification in Management Research: The Case for Identifying Artificial Intelligence Patents." Strategic Management Journal 44.2 (2023): 491-519.
[36] Palanivinayagam, Ashokkumar, Claude Ziad El-Bayeh, and Robertas Damaševi˘cius. "Twenty Years of Machine-Learning-Based Text Classification: A Systematic Review." Applied Sciences, vol. 16, no. 5, 2022, pp. 236.
[37] Kowsari, Kamran, et al. "Text Classification Algorithms: A Survey." IEEE Transactions on Knowledge and Data Engineering, vol. 31, no. 11, pp. 2688-2714, 2019.
[38] Wahdan, Ahlam, et al. "A Systematic Review of Text Classification Research Based on Deep Learning (DL) Models in Arabic Language." International Journal: Electrical & Computer Engineering (IJECE), vol. 10, no. 6, Dec. 2020, pp. 6629-6643.
[39] Matin N. Ashtiani and Bijan Raahemi "News-based Intelligent prediction of financial markets using text mining and machine learning: A systematic literature review." Expert Systems with Applications Volume 217, (2023),119509.
[40] Dogra, Varun, et al. "A Complete Process of Text Classification System Using State-of-the-Art NLP Models." arXiv preprint arXiv:2208.01111 (2022).