ISSN 2063-5346

Resume Screening Classification using Artificial Intelligence and Natural Language Processing

Arvind Kumar Sinha[1], Md. Amir Khusru Akhtar [2], Mohit Kumar[3], Shrikant Upadhyay

[1,2] Faculty of Computing and Information Technology, Usha Martin University, Ranchi, India

[3] Department of IT, MIT Art Design and Technology University, Pune, India

Department of Electronics & Communication Engineering, Cambridge Institute of Technology, Ranchi, Jharkhand, India

doi: 10.48047/ecb/2023.12.si4.1309

ABSTRACT

Resume screening is the process of assessing candidates' resumes to determine their suitability for a particular position. The purpose of resume screening is to identify the most qualified candidates who meet the requirements for the job. Conventionally, resume screening has been a manual process, with hiring managers spending significant time reviewing each resume individually. Besides the fact that it is a time-consuming procedure, there are also unknown biases. Therefore, it is important to research the methods for automating resume screening using Artificial Intelligence and Machine Learning. To address this, this paper proposes a two-phase model named "Prospect" based on feature extraction and matching using machine learning. The first phase pre-processes the dataset and extracts resume content by using feature extraction. The second phase applies "selection" and "rejection" classification by applying a matching score algorithm and custom logic. To validate its approach, this paper also designs a unique Prospect dataset with approximately 5,000(thousand) resumes, which incorporates different data sets to generate an unbiased classification output. Experimental result shows that the Prospect model categorizes the resume in "selected" and "rejected" categories with a 93.5% accuracy which improves the overall accuracy by 19.5% compared to convolutional neural network models.

Keywords:

Levenshtein Distance
Cosine Similarity
FuzzyWuzzy
Artificial Intelligence
Prospect

1. INTRODUCTION

A Resume is unstructured data and extracting information is a complex process. Natural language processing with complex patterns/language analysis techniques have been used for feature extraction[1]. Feature extraction, Cosine similarity, and Levenshtein Distance have been mainly used to get the classification result. Matching with a similarity score is further provided to a model and custom logic is overlayed to get the final result. There is some level of flexibility purposely added to avoid resume rejection in favor of selection, which further with supervised learning can be reduced. The whole of the Artificial Intelligence/Machine Learning implementation is done using Python. Some of the important libraries that are used are sklearn, fuzzyWuzzy, numPy, pandas and spaCy. The level of flexibility is made configurable and customizable for the organization, which provides further flexibility for using the Resume Screening application.

To perform the resume screening, it is important to make it user-friendly and to expose the Artificial Intelligence and Natural Language Processing based model to the web. It is achieved through Django in Python, with this it is just a few clicks that predicts the resume results. A bulk model is also built which performs the matching for "n" number resume at a time. This provides speed to the tool and the whole screening for a hiring drive can be done in just a few minutes which conventionally takes days and weeks. Additionally, a dashboard is also built which does the relative ranking for the resume. This product is being piloted at some esteemed organizations for further learning for machines. There are a few features in a resume that are important for decision-making, below is the list of considered features.

1. Basic Information: This is mainly to maintain uniqueness across the resume list screened by the machine. This includes name, address, phone, and email address. It is not for the decision-making process; however, decisions are tagged to these unique attributes.

2. Decision-making features: These are the attributes that are used for the decision-making process whether the resume is shortlisted or not, against the job description provided. This includes features like skills, education, experience, x /boost factors like if someone has studied from premium institutes or has experience with top x companies. A custom model is built to consider these attributes and a score is derived for each resume. This score with some configurable flexibility is used to predict whether the resume classification is in a "selected" or "rejected" bucket.
3. Externalized Configuration: The SOLID principle has one of the important principles, called Open-Closed. As per this principle, classes should be open for extension and closed for modification. In line with this, all important decision points are externalized and it is up to the organization to configure them as per the requirement. Below are the configuration items that the organization needs to modify, however, the default configuration is also made available.
4. Boost value parameters: Configurable parameters have been defined to mark positive or negative scores for education and work experience. For example, if someone has studied at Indian Insttitute of Technology, National Intstitute of Technology, or Massachusetts Intstitute of Technology or someone has previous experience with companies like Google, Amazon, Apple, etc. the organization may consider it as a plus point while calculating the suitability for resume screening. Organizations may not consider this parameter at all as well, by providing the configuration. Job Description for a given position is fed to the resume screening system by the organization's HR. The margin of error percentage in favor of the candidate section is configurable.

The paper has leveraged the Artificial Intelligence and Machine Learning language [28-32] understanding and applied a custom algorithm for matching and predicting, which makes the application unique. The configurable parameter makes the application very flexible and customizable to different organizations without changing the code.
The structure of the remaining paper has various sections, the second section discussed literature review the third section named as is "methods and algorithms" which discusses the research method, alogorithm and open-source libraries that have been used in resume screening classification. Section four presents result & discussions. It discusses the results, output and it's visualization obtained from the study. The final Section concludes the paper with future scope.

2. Literature Review

As part of the research, we have done an extensive literature review and tried to get insight into text analysis problems. Below are some of the paper and work that has helped to arrive at the optimal solution to a complex problem at hand.
D. W. Otter, J. R. Medina and J. K. Kalita in the paper "A Survey of the Usages of Deep Learning for Natural Language Processing," in IEEE Transactions on Neural Networks and Learning Systems[2] have provided insight on deep learning in the field of natural language processing. It has also highlighted a plethora of recent studies and summarizes a large variety of relevant contributions. It has also analysed research areas including several core linguistic processing issues in addition to many applications of computational linguistics. It discusses further the current state of the art and recommendations for future research in this field.
P. Lavanya and E. Sasikala in paper "Deep Learning Techniques on Text Classification Using Natural Language Processing (NLP) In Social Healthcare Network: A Comprehensive Survey"[3] have provided good insight on deep learning (DL) techniques applied to classify the text. The paper has taken up the processing of unstructured data gathered from social media. It focuses on the models of deep learning (DL) techniques applied to classify the text in social media healthcare networks. The main intention of this paper is to provide an insight for training the data and to classify the text by analyzing and extracting the raw input and producing the output with the help of Natural language processing (NLP). It further focuses on the performance of the text classifier based on effectiveness to improve accuracy and text processing speed by using a suitable methodology in order to produce the promising results in the future.
Professor G. Prasad and K. K. Fousiya in the paper "Named entity recognition approaches: A study applied to English and Hindi language" made a comparative study of different approaches for Named Entity Recognition applied to English and Hindi Language corpus. This study is conducted with a view to extend the work and developing an efficient method for named entity recognition in other native languages.D. F. Mujtaba and N. R. Mahapatra in the paper "Ethical Considerations in Artificial Intelligence-Based Recruitment"[4] published in 2019 IEEE International Symposium on Technology and Society (ISTAS) have provided deep insight on several toolkits to mitigate biases and interpret black box models developed in an effort to promote fair algorithms. This paper presents an overview of fairness definitions, methods, and tools as they relate to recruitment and establishes ethical considerations in the use of machine learning in the hiring space.
T. P. Nagarhalli, V. Vaze and N. K. Rana in the paper "Impact of Machine Learning in Natural Language Processing: A Review"[5] have highlighted and provided insight of important role played by the machine learning

techniques in improving the efficiency of natural language processing.A. Galassi, M. Lippi and P. Torroni in the paper "Attention in Natural Language Processing" published in IEEE Transactions on Neural Networks and Learning have stressed a focus on machine learning with vector representations of the textual data. They have a taxonomy of attention models according to four dimensions: the representation of the input, the compatibility function, the distribution function, and the multiplicity of the input and/or output.

Minhu, Junmei Sun , Xiumei Li and Lei Xiao in the paper[6] "Scoring Mechanism of Defect Report Based on Text Similarity" proposes a software defect repetitive detection method based on the combination of Term Frequency-Inverse Document Frequency and cosine similarity and proposes a defect report scoring mechanism based on the difficulty level and number of defects found by students. Comparing the manually scores of the defect reports by the teachers, the scoring mechanism is reasonable and effective, and greatly improves the work efficiency of the teachers.

Komang Rinartha , Wayan Suryasa and Luh Gede and Surya Kartika in the paper "Comparative Analysis of String Similarity on Dynamic Query Suggestions" have compared few of the similarity scoring pattern for dynamic query suggestions. Comparison of Jaccard similarity, MySQL pattern matching, Levenshtein distance and MySQL Fulltext Index with process time comparison parameters, proximity of data search by suggestion, calculate the proximity rank of data and sorted from the nearest data has been highlighted in the paper. Malgorzata Pikies and Junade Ali in the paper "String similarity algorithms for a ticket classification system"[7] have highlighted the importance of fuzzy searching where in the matching strings to be compared and extracted from bodies of text that are useful in systems which automatically extract and process documents. Resume screening has a similar use case for performing resume matching with the given job description. It has compared various existing algorithms for achieving string similarity measures like Longest Common Subsequence (LCS), Dice coefficient, Cosine Similarity, Levenshtein distance and Damerau distance.

Pinky Sitikhu et al, in the paper "A Comparison of Semantic Similarity Methods for Maximum Human Interpretability"[8] provides the details on the inclusion of semantic information in any similarity measures improves the efficiency of the similarity measure and provides human interpretable results for further analysis. P Sowmya and Madhumita Chatterjee in the paper "Detection of Fake and Clone accounts in Twitter using Classification and Distance Measure Algorithms" [9], a detection method has been proposed which can detect Fake and Clone profiles in Twitter. For Profile Cloning detection two methods are proposed, one using Similarity Measures and the other using C4.5 decision tree algorithm. In Similarity Measures, two types of similarities are considered - Similarity of Attributes and Similarity of Network relationships. C4.5 detects clones by building decision tree by taking information gain into consideration.

Liyan Wang and Yves Lepage in the paper "Vector-to-Sequence Models for Sentence Analogies"[10] have discussed a decoder system to transform sentence embedding vectors back into sequences of words. To generate the vector representations of answer sentences, a linear regression network which learns the mapping between the distribution of known and expected vectors. It leverages this pre-trained decoder to decode sentences from regressed vectors. Dewi Soyusiawaty and Yahya Zakaria in the paper "Book Data Content Similarity Detector With Cosine Similarity"[11] have tried to solve library problem for finding alternate book by implementing Cosine Similarity method in detecting similarity. The research subject is similarity detector system of the book data content in the library by using Cosine Similarity method.

One of the phases in resume screening is breaking down resumes into sections and perform text analysis. Text analysis is an Artificial Intelligence and Natural Language Processing learning technique that lets application analyze unstructured text data like resumes [33-34]. Further, "Text classification"[12] is used for mapping pre-defined tags to random text. It is regarded as an important process for dealing with natural language because it is versatile and can edit and parse any type of text to bring logical facts and issue resolution. Natural Language Processing[13],[14] is a subsystem of Artificial Intelligence and Machine learning, where the system can decipher as well as comprehend text in the same way that humans do.

The topic analysis is also useful in text classification, which organizes text by topic or subject. Intent classification is used to understand the intent rather than looking for literal meaning[15]. The steps included, extraction of features from the given resume and performing text analysis using a hybrid approach. The hybrid approach leveraged rules-based pattern recognition and Artificial Intelligence and Machine Learning -based learning technique. The resume has been read from the standard format like word or pdf based on the extension and converted into plain text. The cleansing process then applied on top of extracted text like stop-word, punctuation, etc. The features from the resume are extracted using a combination of Machine Learning techniques and libraries like tensor flow Natural Language Processing libraries and rule-based systems. This extracted key information are leveraged for further decision-making and finally predicting the result for the given resume.

The key informations like skills, qualifications, years of experience, and organization someone has worked in are extracted from the text data of the resume. Spacy and transformers are used to train the model. Post the model is trained, the unknown data from the new resume are passed into the custom library developed as part of the research program. Labeled trained data available from external sources has been used for training and testing. Table 1 below has done comparative feature analysis of Prospect system with other applications that are published in different papers.

Table 1. Shows PROSPECT features comparison with other papers (DOI included)

Feature	Resume Screener (PROSPECT)	DOI 10.17577/IJERTV10IS080057[16]	DOI 10.1109/ICSCT53883.2021.9642652[17]
Support PDF	Yes	No	Yes
Support CSV	Yes	Yes	Yes
Custom Matching	Yes	No	No
Dashboard for Insight	Yes	No	No
Continuous Improvement	Yes	Yes	Yes
Accuracy	93.5%	NA	74%

3. Methods and Algorithms

Below is the alogorithm followed by high level flow depicted in Figure 1.
Algorithm 1: Resume Screening
Data: Kaggle, External Source (MTIS)
Prospect model for resume classification:

// Phase 1: data set preparation
1 Resume_Screening_data set ← collection (Kaggle, External Source (MTIS)) // Resume Data sourced from Keggle and consulting organization called MTIS
2 Resume_Screening_data set ← pre-process (Resume_Screening_data set) // Data set pre_processing // Phase 2: feature engineering
3 L F ← extract (Resume_Screening_data set) // Linguistic feature extraction
4 FS ← selection (LF) // Feature selection
5 Job_Description_data set ← collection (Kaggle, External Source (MTIS)) // Resume Data
6 Job_Description_data set ← preprocess (Job_Description_data_set) // Data set pre_processing // Phase 2: feature engineering
7 LF2 ← extract (Job_Description_data_set) // Linguistic feature extraction
8 FS2 ← selection (LF2) // Feature selection
9 LF-LF2-MATCH ← MATCH (FS1, FS2) // in loop
10 MATCH_SCORE ← LF-LF2-MATCH (Prospect_data set) // for each feature
10 AVERAGE_MATCH_SCORE ← WEIGHTED_AVERAGE(n) // Weightage average
11 SELECTION_RESULT(SR) ← FINAL_RESULT(AVERAGE_MATCH_SCORE) // Based on configuration, returns Boolean result
// Machine Learning model tuning and voting classifier based on Feedback
12 Model ← bestModel(MNB (SR), RF (SR), LR (SR), XG Boost (SR), LCVC(SR))
// To find the best model

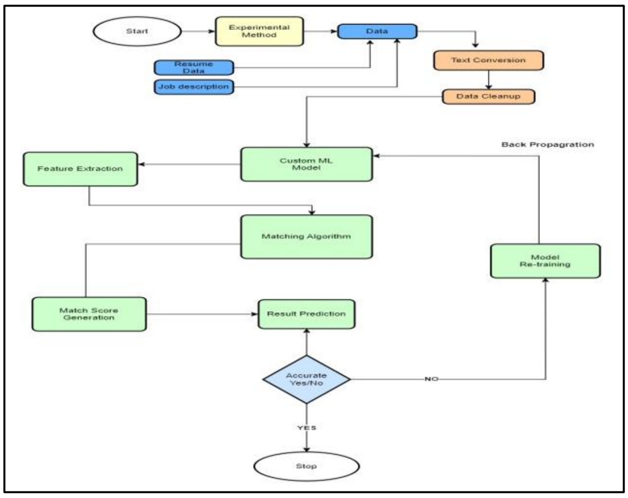

Figure 1 High level prospect flow design and experimental method applied

We review in this section a few Machine Learning methods used for resume classification in the prospect model

1. Random Forest(RF): Random Forest[18] is an ensemble learning method for regression and classification that functions by building a host of decision trees at training time and combining the result of individual trees. The dataset for resume and actual feedback is passed to RF for training and testing , followed by passing unknown resume data.

2. Multinomial Naive Bayes(MNB): Naive Bayes classifiers[19] are a type of simple "probabilistic classifiers" based on Bayes' theorem with strong assumptions between the features. Here also the dataset for resume and actual feedback is passed to MNB for training and testing , followed by passing unknown resume data.

3. Logistic Regression(LR): Logistic Regression[20] usages a logistic function to model a binary dependent variable. The resume data is passed to LR as well and the accuracy is recorded for a new resume.

4. Linear Support Vector Classifier(LCVC): A SVM[21] is a supervised machine learning classifier which is defined by a separating hyperplane. In two-dimensional space, a hyperplane is a line which separates a plane into two separate planes, where each plane belongs to a Class. The resume is also passed to LCVC and a result is recorded.

5. XG Boost[22]: It stands for extreme gradient boosting, it is an implementation of Gradient Boosting Decision Tree. In this, decision tree are created in sequence and weight plays a critical role. Weight is assigned to all independent variables which are then fed into decision tree. The wrong predicted weights are further increased and passed to the next decision tree. The classifier ensembles to provide a precise and powerful model.The ensemble approach has been used and all five algorithms has provided accuracy of more than 82% while XG Boost has performed with an accuracy of 93.5% listed in Table 2.

Table 2. Shows the model comparison and accuracy achieved

Machine Learning Model	Dataset	Accuracy
Random Forest (RF)	Keggle, External Source	87%
Multinomial Naive Bayes (MNB)	Keggle, External Source	89%
Logistic Regression (LR)	Keggle, External Source	82.5%
Linear Support Vector Classifier (LCVC)	Keggle, External Source	88%
XG Boost	Keggle, External Source	93.5%

Open-Source Library/API

The Artificial Iearning and Natural Language Processing based solution is built on Python language with rich set of libraries providing effective and performant ways to solve problems. Below are some of the important libraries used for resume screening solution.

1. NLTK: Natural Language Toolkit library also called as NLTK, is a cutting-edge library of analyzing text. Using the Rapid Automatic Keyword Extraction Rake algorithm and the NLTK tool kit, we have leveraged NLTK to extract key information from resume as it provides powerful keyword extraction tool

2. SpaCy: SpaCy is a statistical Natural Language Processing library[23] for industry. It enhances the standard features by combining in-depth learning with multimedia neural network models. Spacy is a free and open library that makes use of Natural Language Development (NLP) . Along with Spacy, transformers have been used for training the resume screening model.

3. Scikit-learning: Scikit-learning is an advanced Python Data science and Machine Learning library built on top of SciPy, NumPy and matplotlib, that provides very good performance in building text-based analysis models. Resume screening application has leveraged scikit learning in a big way for data extraction and visualization.

4. FuzzyWuzzy: FuzzyWuzzy is a key library that is used in this implementation. It has built on top of some of the matching algorithm like K nearest neighour and Levenshtein Distance[24]. This library has been extensively used for matching resume job description with the extracted features from resume.

5. Django: This framework is used mainly to expose Artificial Intelligence and Machine Learning based learning and the API for the Web. It allowed the application to be user friendly by making the functionality available on web browser. The user friendly interface for resume screening and various dashboards are the key differentiator with other open source product available.

Algorithm for pattern matching:

Below are a few matching algorithms that are used to match the extracted keywords from the resume with the job description.

1. Levenshtein Distance: Levenshtein distance is a string metric for measuring the difference between two sequences. The Levenshtein distance between two words is the minimum number of single-character edits required to change one word into the other. In Levenshtein Distance, we have three possible operations, replace, insert or delete. This is to transform one string into another with a minimal number of operations. This is of the key algorithm used in fuzzyWuzzy library that is implemented in resume screening.

2. Cosine Similarity and Cosine Distance[25]: Given if we have two points P1 and P2, similarity and distance are inversely proportional, meaning if the distance is increased the similarity decreases and vice versa. It ranges between -1 to 1. It is an angle between two points represented as Cos θ.

3. Euclidian Distance: Euclidean distance[26] between two points in Euclidean space is the length of a line segment between the two points.

Euclidian Distance =$\sqrt{(x2 - x1)^2 + (y2 - y1)^2}$

The principle and distance algorithms are used to find the similarity score between the skills of the candidate and the required skills as per the job description. This method is used for matching qualification required and candidate's actual qualification.

4. K Nearest Neighbour[27]: It is used for machine learning classification problems and resume screening qualifies for the classification use case. A few key points while implementing KNN is that K should be an odd value and on top of that we have eliminated the outliers to get optimum results from this algorithm. The nearest neighbour is identified based on the distance. KNN is extensively used in resume screening matching algorithm to classify if the new resume falls into "Selected" or "Rejected" category.

4. RESULTS AND DISCUSSION

Table 3 shows the resume screening result, where blue represents matching percentage, while orange represents non-matching percentage, Resume are passed to the tool, in one case where there is a match, the tool predicts "shortlisting" result while in another case when match against the Job description is low, the tool rejects the applicant. These are the results obtained for positive and negative use cases along with the visualization.

Table 3. Shows the results obtained for two different resumes against same JD

Positive Case	Negative Case
Skill Similarity Score: 88	Skill Similarity Score: 55
Qualification Similarity Score: 100	Qualification Similarity Score: 20
Weighted Average 91.6	Weighted average 44.5
Average 94.0	Average 37.5
Candidate is shortlisted	Sorry, Candidate is NOT shortlisted
	

5. CONCLUSION

In this paper, we have covered step by step two phase implementations for resume screening. The first phase extracts the features from resume using Artificial Iearning and Natural Language Processing. The second phase use matching algorithm to derive a score for the given resume. This paper also demonstrated the final result with machine performing human like decision. Further, it elaborates on how the training is provided on the extracted data with different machine learning algorithm that provides insights for key decision making. Spacy , transformers and other key libraries usage have been explained in details. The paper also highlights the key differentiator that this tool offers in comparison with other available tools in terms of features and accuracy. As part of next step, further regression to be performed to arrive at better accuracy in terms of shortlisting score. There is additional tunning in the dataset by web scrapping of professional network sites to be considered. There will be some outliers that will require to be eliminated for better outcome.

REFERENCES

[1] A. K. Sinha, Md. Amir Khusru Akhtar, and A. Kumar, "Resume Screening Using Natural Language Processing and Machine Learning: A Systematic Review," in Machine Learning and Information Processing, Singapore, 2021, pp. 207–214. doi: 10.1007/978-981-33-4859-2_21.

[2] D. W. Otter, J. R. Medina, and J. K. Kalita, "A Survey of the Usages of Deep Learning for Natural Language Processing," IEEE Transactions on Neural Networks and Learning Systems, vol. 32, no. 2, pp. 604–624, Feb. 2021, doi: 10.1109/TNNLS.2020.2979670.

[3] PM. Lavanya and E. Sasikala, "Deep Learning Techniques on Text Classification Using Natural Language Processing (NLP) In Social Healthcare Network: A Comprehensive Survey," in 2021 3rd International Conference on Signal Processing and Communication (ICPSC), 2021, pp. 603–609. doi: 10.1109/ICSPC51351.2021.9451752.

[4] D. F. Mujtaba and N. R. Mahapatra, "Ethical Considerations in AI-Based Recruitment," in 2019 IEEE International Symposium on Technology and Society (ISTAS), Nov. 2019, pp. 1–7. doi: 10.1109/ISTAS48451.2019.8937920.

[5] T. P. Nagarhalli, V. Vaze, and N. K. Rana, "Impact of Machine Learning in Natural Language Processing: A Review," in 2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV), Feb. 2021, pp. 1529–1534. doi: 10.1109/ICICV50876.2021.9388380.

[6] M. Zhu, J. Sun, X. Li, and L. Xiao, "Scoring Mechanism of Defect Report Based on Text Similarity," in 2019 10th International Conference on Information Technology in Medicine and Education (ITME), Aug. 2019, pp. 182–185. doi: 10.1109/ITME.2019.00050.

[7] H. V, L. J, S. R. A, N. Divya Bharathi, S. Upadhyay and V. R, "Energy Efficient Data Management in Health Care," 2023 Third International Conference on Artificial Intelligence and Smart Energy (ICAIS), Coimbatore, India, 2023, pp. 682-688, doi: 10.1109/ICAIS56108.2023.10073796.

[8] P. Sitikhu, K. Pahi, P. Thapa, and S. Shakya, "A Comparison of Semantic Similarity Methods for Maximum Human Interpretability," in 2019 Artificial Intelligence for Transforming Business and Society (AITB), Nov. 2019, vol. 1, pp. 1–4. doi: 10.1109/AITB48515.2019.8947433.

[9] P. Sowmya and M. Chatterjee, "Detection of Fake and Clone accounts in Twitter using Classification and Distance Measure Algorithms," in 2020 International Conference on Communication and Signal Processing (ICCSP), Jul. 2020, pp. 0067–0070. doi: 10.1109/ICCSP48568.2020.9182353.

[10] L. Wang and Y. Lepage, "Vector-to-Sequence Models for Sentence Analogies," in 2020 International Conference on Advanced Computer Science and Information Systems (ICACSIS), Oct. 2020, pp. 441–446. doi: 10.1109/ICACSIS51025.2020.9263191.

[11] D. Soyusiawaty and Y. Zakaria, "Book Data Content Similarity Detector With Cosine Similarity (Case study on digilib.uad.ac.id)," in 2018 12th International Conference on Telecommunication Systems, Services, and Applications (TSSA), Oct. 2018, pp. 1–6. doi: 10.1109/TSSA.2018.8708758.

[12] Upadhyay, S.; Kumar, M.; Upadhyay, A.; Verma, S.; Kavita; Kaur, M.; Khurma, R.A.; Castillo, P.A. Challenges and Limitation Analysis of an IoT-Dependent System for Deployment in Smart Healthcare Using Communication Standards Features. Sensors ***2023****, 23, 5155. https://doi.org/10.3390/s23115155*

[13] T. M. Harsha, G. S. Moukthika, D. S. Sai, M. N. R. Pravallika, S. Anamalamudi, and M. Enduri, "Automated Resume Screener using Natural Language Processing(NLP)," in 2022 6th International Conference on Trends in Electronics and Informatics (ICOEI), 2022, pp. 1772–1777. doi: 10.1109/ICOEI53556.2022.9777194.

[14] S. Bharadwaj, R. Varun, P. S. Aditya, M. Nikhil, and G. C. Babu, "Resume Screening using NLP and LSTM," in 2022 International Conference on Inventive Computation Technologies (ICICT), 2022, pp. 238–241. doi: 10.1109/ICICT54344.2022.9850889.

[15] M. Alamelu, D. S. Kumar, R. Sanjana, J. S. Sree, A. S. Devi, and D. Kavitha, "Resume Validation and Filtration using Natural Language Processing," in 2021 10th International Conference on Internet of Everything, Microwave Engineering, Communication and Networks (IEMECON), 2021, pp. 1–5. doi: 10.1109/IEMECON53809.2021.9689075.

[16] R. V, R. T. Fareed, and S. Kaganurmath, "Resume Classification and Ranking using KNN and Cosine Similarity," International Journal of Engineering Research & Technology, vol. 10, no. 8, Aug. 2021, doi: 10.17577/IJERTV10IS080057.

[17] M. F. Mridha, R. Basri, M. M. Monowar, and Md. A. Hamid, "A Machine Learning Approach for Screening Individual's Job Profile Using Convolutional Neural Network," in 2021 International Conference on Science & Contemporary Technologies (ICSCT), 2021, pp. 1–6. doi: 10.1109/ICSCT53883.2021.9642652.

[18] Aditi Kumari, Neha Gandotra, Pankaj Josh et. al., "Impact of Node Movement on MANET Using Different Routing Protocol for QoS Improvement under Different Scenario", International Journal of Smart Sensors and Ad Hoc Networks (IJSSAN) ISSN No. 2248‐ 9738 Volume‐1, Issue‐ 3, pp. 95-101, 2012.

[19] Md. A. Rahman and Y. A. Akter, “Topic Classification from Text Using Decision Tree, K-NN and Multinomial Naïve Bayes,” in 2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT), 2019, pp. 1–4. doi: 10.1109/ICASERT.2019.8934502.

[20] C. Wen, J. Wu, and D. Chen, “Analysis of Text Emotion Based on Logistic Regression Model,” in 2022 IEEE 5th International Conference on Automation, Electronics and Electrical Engineering (AUTEEE), 2022, pp. 891–895. doi: 10.1109/AUTEEE56487.2022.9994346.

[21] M. Jamaluddin and A. D. Wibawa, “Patient Diagnosis Classification based on Electronic Medical Record using Text Mining and Support Vector Machine,” in 2021 International Seminar on Application for Technology of Information and Communication (iSemantic), 2021, pp. 243–248. doi: 10.1109/iSemantic52711.2021.9573178.

[22] M. T, S. Upadhyay, R. Babitha Lincy, J. R. J, R. Beaulah Jeyavathana and A. Gopatoti, "Big Data Analytics with Deep Learning based Intracranial Haemorrhage Diagnosis and Classification Model," 2022 International Conference on Augmented Intelligence and Sustainable Systems (ICAISS), Trichy, India, 2022, pp. 877-883, doi: 10.1109/ICAISS55157.2022.10010826.

[23] S. JUGRAN, A. KUMAR, B. S. TYAGI, and V. ANAND, “Extractive Automatic Text Summarization using SpaCy in Python & NLP,” in 2021 International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), 2021, pp. 582–585. doi: 10.1109/ICACITE51222.2021.9404712.

[24] S. Upadhyay, S. K. Sharma, A. Upadhyay, " Comparative Study and Analysis of Different Feature Extraction Technique for Speaker Verifivation and Identification: A Review", INROADS-An InternationaL Journal of Jaipur National University, Vol. 4, Issue 2, ISSN: 2277-4912 pp. 101-106, 2015.

[25] Upadhyay S, Kumar M, Upadhyay A, Verma S, Kavita, Hosen ASMS, Ra I-H, Kaur M, Singh S. Digital Image Identification and Verification Using Maximum and Preliminary Score Approach with Watermarking for Security and Validation Enhancement. Electronics. 2023; 12(7):1609. https://doi.org/10.3390/electronics12071609

[26] A. R. Lubis, S. Prayudani, Al-Khowarizmi, Y. Y. Lase, and Y. Fatmi, “Similarity Normalized Euclidean Distance on KNN Method to Classify Image of Skin Cancer,” in 2021 4th International Seminar on Research of Information Technology and Intelligent Systems (ISRITI), 2021, pp. 68–73. doi: 10.1109/ISRITI54043.2021.9702826.

[27] R. Ransing, A. Mohan, N. B. Emberi, and K. Mahavarkar, “Screening and Ranking Resumes using Stacked Model,” in 2021 5th International Conference on Electrical, Electronics, Communication, Computer Technologies and Optimization Techniques (ICEECCOT), 2021, pp. 643–648. doi: 10.1109/ICEECCOT52851.2021.9707977.

[28] Kumar, A., Kumar, M., Mahapatra, R. P., Bhattacharya, P., Le, T. T. H., Verma, S., & Mohiuddin, K. (2023). Flamingo-Optimization-Based Deep Convolutional Neural Network for IoT-Based Arrhythmia Classification. Sensors, 23(9), 4353

[29] Gupta, Abhishek, Mohit Kumar, Anamika Rangra, Vikas Kumar Tiwari, and Pravesh Saxena. "Network intrusion detection types and analysis of their tools." Department of Computer Science and Information Technology, Jaypee University of Information Technology, India (2012).

[30] Upadhyay, S., Kumar, M., Kumar, A., Ghafoor, K. Z., & Manoharan, S. (2022). SmHeSol (IoT-BC): smart healthcare solution for future development using speech feature extraction integration approach with IoT and blockchain. Journal of Sensors, 2022

[31] Pratap, A., Kumar, A., & Kumar, M. (2021, May). Analyzing the Need of Edge Computing for Internet of Things (IoT). In Proceedings of Second International Conference on Computing, Communications, and Cyber-Security: IC4S 2020 (pp. 203-212). Singapore: Springer Singapore

[32] Kumar, M., Kumar, D., & Akhtar, M. A. K. (2021). A modified GA-based load balanced clustering algorithm for WSN: MGALBC. International Journal of Embedded and Real-Time Communication Systems (IJERTCS), 12(1), 44-63

[33] Akhtar, Md Amir Khusru, Mohit Kumar, and Gadadhar Sahoo. "Automata for santali language processing." In 2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI), pp. 939-943. IEEE, 2017

[34] Akhtar, Amir Khusru, Gadadhar Sahoo, and Mohit Kumar. "Digital corpus of Santali language." In 2017 International Conference on Advances in Computing, Communicationsand Informatics (ICACCI), pp. 934-938. IEEE, 2017.