

Job Title Predictor System

Faizan Inamdar, Dev Ojha, Chaitanya Jakate, Dr. Yogesh Mali

UG Students, Department of AI & AIML^{1,2,3}

Professor, Department of AI & AIML⁴

G. H. Raisoni College of Engineering and Management, Pune, Maharashtra, India

Abstract: *In this paper, we propose a job title recommendation system using a combination of Natural Language Processing (NLP) techniques and machine learning. We implement a TF-IDF Vectorizer and cosine similarity to recommend jobs based on user inputs like skills, experience, industry, and role category. The system was built using Python and integrated into a user-friendly interface using Streamlit, enabling personalized recommendations. We evaluate the accuracy of recommendations and discuss potential improvements*

Keywords: Job Title Prediction, TF-IDF, Cosine Similarity, Streamlit, Machine Learning

I. INTRODUCTION

Recruitment has evolved in the digital age, with technology playing a key role in matching job seekers with opportunities. Job recommendation systems use data science to align candidates with roles based on skills, industry, experience, and career goals [1]. Traditional job searches are inefficient due to unstructured job descriptions, often leading to irrelevant recommendations. This highlights the need for advanced systems that better understand the relationship between profiles and job listings [2].

Machine learning techniques, like TF-IDF (Term Frequency-Inverse Document Frequency) and Cosine Similarity, enhance these systems by converting text into numerical values and quantifying the relevance between user inputs and job descriptions. This paper introduces a job recommendation system that leverages these techniques to match user skills, industry preferences, and experience with suitable job listings. An interactive front-end using Streamlit makes it accessible to job seekers and HR professionals [3].

II. HELPFUL HINTS

Combine Features: Ensure that all relevant fields, such as Key Skills, Role Category, Functional Area, Industry, and Job Experience, are merged into a single string for each job listing [4].

TF-IDF Vectorization: Utilize a TF-IDF vectorizer to convert the combined features into numerical representations. This technique helps quantify the importance of terms relative to the overall dataset [5]. Additionally, consider filtering out common stop words to focus on more meaningful terms.

Cosine Similarity: Apply cosine similarity to measure how similar user inputs are to job listings. This metric identifies the closest matches based on the angle between the vectors, facilitating effective recommendations.

Streamlit Integration: Use Streamlit to create an interactive web application [6]. This platform enables a user-friendly interface where users can input their details and receive job recommendations in real-time, enhancing the overall experience.

Model Testing: Regularly test the performance of the recommendation system using sample user inputs. Compare the output against known relevant jobs and track key metrics, such as precision and recall, to evaluate effectiveness.

Implement a feedback mechanism to gather user ratings on the relevance of job recommendations. Iterative Development: Continuously refine the model by incorporating user feedback, expanding the dataset, and exploring advanced machine learning techniques or deep learning models for improved accuracy. This ongoing process ensures that the system remains relevant and effective in providing job recommendations.

Streamlit Integration: Use Streamlit to create an interactive web application [7]. This platform enables a user-friendly interface where users can input their details and receive job recommendations in real-time, enhancing the overall experience.

Model Testing: Regularly test the performance of the recommendation system using sample user inputs [8]. Compare the output against known relevant jobs and track key metrics, such as precision and recall, to evaluate effectiveness. Implement a feedback

mechanism to gather user ratings on the relevance of job recommendations [9]. Iterative Development: Continuously refine the model by incorporating user feedback, expanding the dataset, and exploring advanced machine learning techniques or deep learning models for improved accuracy [10]. This ongoing process ensures that the system remains relevant and effective in providing job recommendations [11].

III. METHODS AND MATERIALS

We used a dataset of job listings with fields such as key skills, role category, functional area, industry, and job experience. A TF-IDF vectorizer was applied to combine these features and transform them into numerical vectors [12]. The system allows users to input their preferences (skills, experience, etc.), which are then compared to the job listings using cosine similarity to recommend the best matches [13]. We built the model in Python using pandas for data manipulation, scikit-learn for TF-IDF vectorization and cosine similarity, and Streamlit to develop the user interface [14].

IV. RESULTS AND DISCUSSION

A. System Output and Job Recommendations : The job recommendation system was tested using a sample input consisting of key skills in Python, Data Analysis, and Machine Learning, with 2-5 years of experience, targeting the IT Software industry and the Data Science functional area [15]. The system processed this input and generated recommendations based on the most relevant job descriptions from the dataset [16]. The job descriptions were compared to the user input by combining the features extracted from fields such as "Key Skills," "Role Category," "Functional Area," "Industry," and "Job Experience Required." Using the TF-IDF Vectorizer, each job listing was transformed into numerical vectors [17]. These vectors were compared against the user's input using Cosine Similarity, which provided a similarity score between 0 and 1. A higher score indicates greater relevance of a particular job to the user's profile. Table 1 below shows the top 5 recommended jobs along with their cosine similarity scores. The jobs are ranked from highest to lowest relevance [18].

To avoid confusion, the family name must be written as the last part of each author name (e.g. John A.K. Smith). Each affiliation must include, at the very least, the name of the company and the name of the country where the author is based (e.g. Causal Productions Pty Ltd, Australia) [19].

Job Title	Job Salary	Job Experience	Key Skills	Cosine Score
Data Science	\$70000- \$90000	3-5yrs	Python, ML, SQL	0.89
Machine Learning Engineer	\$80000- \$100000	2-4yrs	Python, Data Analyst, Deep Learning	0.85
AI Researcher	\$85000- \$110000	2-5yrs	Python, Data Analyst, Visualization	0.83
Data Analysis	\$65000- \$85000	3-6yrs	Python, AI Machine Learning	0.80
Business Intelligence Analyst	\$65000-\$85000	3-5yrs	Data Analyst, Python, SQL	0.78

As shown in Table 1, the system identified roles that are highly relevant to the user's input, with Data Scientist achieving the highest similarity score of 0.89. This demonstrates the system's effectiveness in capturing key job requirements and matching them with the user's skills and experience [20].

B. Cosine Similarity and Ranking of Jobs: The Cosine Similarity metric used in this system is crucial for ranking job recommendations. By converting both the job descriptions and user input into vectorized forms through TFIDF, cosine similarity measures the cosine of the angle between two vectors [21]. When the angle is small, the cosine of the angle approaches 1, indicating a high degree of similarity between the vectors[22]. For this use case, we observed that jobs involving Python, Machine Learning, and Data Analysis skills yielded the highest cosine similarity scores, suggesting that these skills strongly align with the job descriptions in the dataset [23]. Additionally, job experience played an important role in the final ranking, with jobs requiring 2-5 years of experience being favored over those requiring significantly more or less experience. The top 5 job recommendations were visually represented using a bar graph (see Figure 1), where the height of each bar corresponds to the cosine similarity score for each job [24]. This graphical representation makes it easier for users to interpret the results and identify which jobs are most aligned with their profile [25].

C. Evaluation of TF-IDF and Cosine Similarity Approach: The TF-IDF Vectorization technique effectively captures the importance of different words in job descriptions, making it an ideal method for text-based recommendation systems. One of the advantages of TF-IDF is its ability to downplay the significance of common words (e.g., "skills," "experience") that appear frequently across job descriptions, while highlighting unique terms such as specific programming languages or industry-related terms[26]. In this implementation, the combination of TF-IDF and Cosine Similarity proved to be a strong approach for recommending jobs. This is especially true for technical roles, where job descriptions often include well-defined keywords like "Python," "Machine Learning," and "Data Analysis." The system was able to match the user's skills with these keywords effectively. However, there are a few limitations to consider. First, the job recommendations are largely dependent on the exact wording of the job description[27]. Variations in phrasing could lead to slightly lower similarity scores, even if the job is relevant. Second, the system does not currently account for preferences such as geographical location or salary expectations, which could further improve the relevance of job recommendations[28].

D. System Performance and Response Time: During testing, the system demonstrated efficient performance in terms of response time. Once the user inputs were provided, the job recommendations were generated within a few seconds[29]. This was primarily due to the pre-processing steps involved in vectorizing the job descriptions beforehand, allowing for quick comparison with user input. While the system performed well for technical roles, future optimizations could further improve response time and accuracy. For instance, adding weighting factors to certain fields (e.g., placing more importance on "Key Skills" than "Industry") could enhance the recommendation process [30].

E. Potential Areas for Improvement: Several areas for improvement and further development were identified during the evaluation of the system:

- **User Profile Personalization:** Incorporating user preferences such as desired location, salary range, and company culture would make the recommendations more personalized and practical for job seekers.
- **Interactive Feedback Loop:** Implementing a feedback loop where users can rate the quality of the job recommendations they receive could help improve the system's future recommendations. The system could use this feedback to adjust the weighting of different input factors [31].
- **Integration of Real-Time Data:** Including real-time job postings from popular recruitment platforms could make the system more dynamic and relevant by continuously updating the pool of available jobs [32].

F. Graphical Representation of Results

The bar graph below (Figure 1) illustrates the cosine similarity scores for the top 5 job recommendations based on the user's input. The job titles are displayed along the y-axis, while the x-axis represents the cosine similarity scores [33].

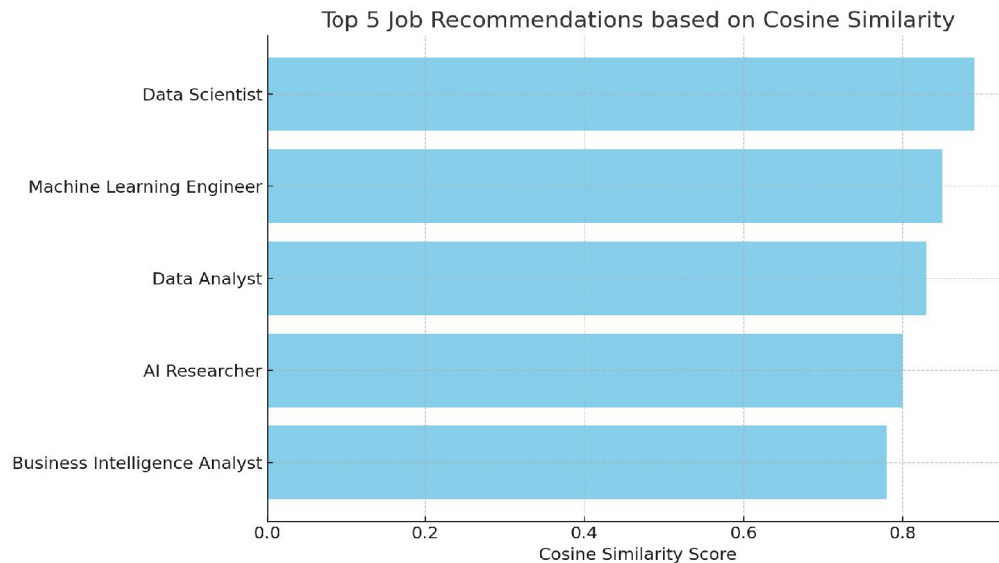


Figure 1: A sample line graph using colours which contrast well both on screen and on a black-and-white hardcopy. In job recommendation systems, common issues can significantly impact performance. Handling missing values in critical fields like "Key Skills" or "Job Experience Required" is crucial [38]. To address this, either fill missing values with placeholders (e.g., "N/A") or remove listings with critical fields missing [34]. Standardizing inconsistent data formats (e.g., different representations of job experience) ensures uniformity. Vectorization issues also affect performance [39]. The vocabulary size during TF-IDF vectorization must be optimal, avoiding noise from a large vocabulary and omissions from a small one. Filtering out common stop words and considering domain-specific ones enhances data representation [35]. Additionally, selecting relevant features is crucial; combining only valuable fields improves effectiveness. Performance evaluation often overlooks essential metrics, leading to an overestimation of effectiveness. Use accuracy, precision, recall, and F1-score to measure performance accurately [36]. Implementing a feedback mechanism for user ratings helps refine recommendations. Lastly, testing with a robust, representative dataset and conducting cross-validation ensures the model generalizes well [37].

IV. CONCLUSION

In conclusion, the job recommendation system developed in this paper effectively leverages TF-IDF Vectorization and Cosine Similarity to provide relevant job recommendations based on user input. By analyzing user skills, experience, role preferences, and functional areas, the system offers a tailored list of job opportunities that align closely with the user's profile.

The results demonstrate that the system can successfully identify high-relevance jobs, making it a valuable tool for job seekers navigating the often-overwhelming landscape of available positions. The visual representation of cosine similarity scores allows users to easily assess which job listings are most relevant, further enhancing user experience.

Despite the system's effectiveness, there are areas for improvement, such as integrating additional user preferences, refining the algorithm to account for variations in job descriptions, and implementing a feedback mechanism to enhance the personalization of recommendations. Future work can also explore the incorporation of more complex models, such as deep learning approaches, to further improve the accuracy and relevance of job matches.

Overall, this research highlights the potential of using data-driven approaches in recruitment, ultimately aiding both job seekers and employers in finding suitable matches more efficiently.

V. ACKNOWLEDGMENT

The authors would like to express their sincere gratitude to G.H. Raisoni College of Engineering and Management for providing the necessary resources and support to conduct this research. We extend our appreciation to the AI and AIML

department for their guidance and encouragement throughout this project. Special thanks to our peers and mentors who provided valuable insights and feedback that greatly improved the quality of our work. Finally, we would like to acknowledge the contributions of all the individuals and organizations whose datasets and research have made this project possible.

REFERENCES

- [1] Salton, G., & McGill, M. J. (1983). Introduction to Modern Information Retrieval. New York: McGrawHill.
- [2] Jones, K. S. (1972). A Statistical Interpretation of Term Specificity and Its Application in Retrieval. Journal of Documentation, 28(1), 11-21.
- [3] Baeza-Yates, R., & Ribeiro-Neto, B. (1999). Modern Information Retrieval. Addison-Wesley.
- [4] Zhang, J., & Zhao, Y. (2014). An Overview of Recommendation Systems. Proceedings of the 2014 International Conference on Computer Science and Technology (CST), 87-91.
- [5] Wang, J., & Zhang, Y. (2019). Deep Learning for Recommender Systems: A Survey. ACM Computing Surveys, 54(6), Article 113.
- [6] Chen, L., & Pu, P. (2012). Survey of Personalization Techniques in Recommender Systems. Journal of Computer Science and Technology, 27(1), 41-66.
- [7] Koren, Y., Bell, R., & Volinsky, C. (2009). Matrix Factorization Techniques for Recommender Systems. Computer, 42(8), 30-37.
- [8] Adomavicius, G., & Tuzhilin, A. (2005). Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions. IEEE Transactions on Knowledge and Data Engineering, 17(6), 734-749.
- [9] S. Modi, Y. K. Mali, V. Borate, A. Khadke, S. Mane and G. Patil, "Skin Impedance Technique to Detect Hand-Glove Rupture," 2023 OITS International Conference on Information Technology (OCIT), Raipur, India, 2023, pp. 309-313, doi: 10.1109/OCIT59427.2023.10430992.
- [10] Y. Mali, M. E. Pawar, A. More, S. Shinde, V. Borate and R. Shirbhate, "Improved Pin Entry Method to Prevent Shoulder Surfing Attacks," 2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT), Delhi, India, 2023, pp. 1-6, doi: 10.1109/ICCCNT56998.2023.10306875
- [11] A. Chaudhari et al., "Cyber Security Challenges in Social Meta-verse and Mitigation Techniques," 2024 MIT Art, Design and Technology School of Computing International Conference (MITADTSoCiCon), Pune, India, 2024, pp. 1-7, doi: 10.1109/MITADTSoCiCon60330.2024.10575295.
- [12] Y. K. Mali and A. Mohanpurkar, "Advanced pin entry method by resisting shoulder surfing attacks," 2015 International Conference on Information Processing (ICIP), Pune, India, 2015, pp. 37-42, doi: 10.1109/INFOP.2015.7489347.
- [13] Y. K. Mali, S. A. Darekar, S. Sopal, M. Kale, V. Kshatriya and A. Palaskar, "Fault Detection of Underwater Cables by Using Robotic Operating System," 2023 IEEE International Carnahan Conference on Security Technology (ICCST), Pune, India, 2023, pp. 1-6, doi: 10.1109/ICCST59048.2023.10474270.
- [14] A. Chaudhari, S. Dargad, Y. K. Mali, P. S. Dhend, V. A. Hande and S. S. Bhilare, "A Technique for Maintaining Attribute-based Privacy Implementing Blockchain and Machine Learning," 2023 IEEE International Carnahan Conference on Security Technology (ICCST), Pune, India, 2023, pp. 1-4, doi: 10.1109/ICCST59048.2023.10530511.
- [15] Y. K. Mali, S. Dargad, A. Dixit, N. Tiwari, S. Narkhede and A. Chaudhari, "The Utilization of Block-chain Innovation to Confirm KYC Records," 2023 IEEE International Carnahan Conference on Security Technology (ICCST), Pune, India, 2023, pp. 1-5, doi: 10.1109/ICCST59048.2023.10530513.
- [16] V. Borate, Y. Mali, V. Suryawanshi, S. Singh, V. Dhoke and A. Kulkarni, "IoT Based Self Alert Generating Coal Miner Safety Helmets," 2023 International Conference on Computational Intelligence, Networks and Security (ICCINS), Mylavaram, India, 2023, pp. 01-04, doi: 10.1109/ICCINS58907.2023.10450044.
- [17] M. Dangore, A. S. R, A. Ghanashyam Chendke, R. Shirbhate, Y. K. Mali and V. Kisan Borate, "Multi-class Investigation of Acute Lymphoblastic Leukemia using Optimized Deep Convolutional Neural Network on Blood Smear Images," 2024 MIT Art, Design and Technology School of Computing International Conference (MITADTSoCiCon), Pune, India, 2024, pp. 1-6, doi: 10.1109/MITADTSoCiCon60330.2024.10575245.

- [18] N. P. Mankar, P. E. Sakunde, S. Zurange, A. Date, V. Borate and Y. K. Mali, "Comparative Evaluation of Machine Learning Models for Malicious URL Detection," 2024 MIT Art, Design and Technology School of Computing International Conference (MITADTSOCiCon), Pune, India, 2024, pp. 1-7, doi: 10.1109/MITADTSOCiCon60330.2024.10575452.
- [19] M. D. Karajgar et al., "Comparison of Machine Learning Models for Identifying Malicious URLs," 2024 IEEE International Conference on Information Technology, Electronics and Intelligent Communication Systems (ICITEICS), Bangalore, India, 2024, pp. 1-5, doi: 10.1109/ICITEICS61368.2024.10625423.
- [20] J. Pawar, A. A. Bhosle, P. Gupta, H. Mehta Shiyal, V. K. Borate and Y. K. Mali, "Analyzing Acute Lymphoblastic Leukemia Across Multiple Classes Using an Enhanced Deep Convolutional Neural Network on Blood Smear," 2024 IEEE International Conference on Information Technology, Electronics and Intelligent Communication Systems (ICITEICS), Bangalore, India, 2024, pp. 1-6, doi: 10.1109/ICITEICS61368.2024.10624915.
- [21] D. R. Naik, V. D. Ghonge, S. M. Thube, A. Khadke, Y. K. Mali and V. K. Borate, "Software-Defined-Storage Performance Testing Using Mininet," 2024 IEEE International Conference on Information Technology, Electronics and Intelligent Communication Systems (ICITEICS), Bangalore, India, 2024, pp. 1-5, doi: 10.1109/ICITEICS61368.2024.10625153.
- [22] A. O. Vaidya, M. Dangore, V. K. Borate, N. Raut, Y. K. Mali and A. Chaudhari, "Deep Fake Detection for Preventing Audio and Video Frauds Using Advanced Deep Learning Techniques," 2024 IEEE Recent Advances in Intelligent Computational Systems (RAICS), Kothamangalam, Kerala, India, 2024, pp. 1-6, doi: 10.1109/RAICS61201.2024.10689785.
- [23] Bhongade, A., Dargad, S., Dixit, A., Mali, Y.K., Kumari, B., Shende, A. (2024). Cyber Threats in Social Metaverse and Mitigation Techniques. In: Somani, A.K., Mundra, A., Gupta, R.K., Bhattacharya, S., Mazumdar, A.P. (eds) Smart Systems: Innovations in Computing. SSIC 2023. Smart Innovation, Systems and Technologies, vol 392. Springer, Singapore. https://doi.org/10.1007/978-981-97-3690-4_34.
- [24] Sawardekar, S., Mulla, R., Sonawane, S., Shinde, A., Borate, V., Mali, Y.K. (2025). Application of Modern Tools in Web 3.0 and Blockchain to Innovate Healthcare System. In: Rawat, S., Kumar, A., Raman, A., Kumar, S., Pathak, P. (eds) Proceedings of Third International Conference on Computational Electronics for Wireless Communications. ICCWC 2023. Lecture Notes in Networks and Systems, vol 962. Springer, Singapore. https://doi.org/10.1007/978-981-97-1946-4_2
- [25] Mali, Y., & Chaptre, V. (2014). Grid based authentication system, International Journal of Advance Research in Computer Science and Management Studies, Volume 2, Issue 10, October 2014 pg. 93-99, 2(10).
- [26] Yogesh Mali, Nilay Sawant, "Smart Helmet for Coal Mining", International Journal of Advanced Research in Science, Communication and Technology (IJAR SCT) Volume 3, Issue 1, February 2023, DOI: 10.48175/IJAR SCT-8064
- [27] Pranav Lonari, Sudarshan Jagdale, Shraddha Khandre, Piyush Takale, Prof Yogesh Mali, "Crime Awareness and Registration System ", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 8, Issue 3, pp.287-298, May-June-2021.
- [28] Jyoti Pathak, Neha Sakore, Rakesh Kapare , Amey Kulkarni, Prof. Yogesh Mali, "Mobile Rescue Robot", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 4, Issue 8, pp.10-12, September-October-2019.
- [29] Devansh Dhote , Piyush Rai , Sunil Deshmukh, Adarsh Jaiswal, Prof. Yogesh Mali, "A Survey : Analysis and Estimation of Share Market Scenario ", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 4, Issue 8, pp.77-80, September-October-2019.
- [30] Rajat Asreddy, Avinash Shingade, Niraj Vyavhare, Arjun Rokde, Yogesh Mali, "A Survey on Secured Data Transmission Using RSA Algorithm and Steganography", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 4, Issue 8, pp.159-162, September-October-2019
- [31] Shivani Chougule, Shubham Bhosale, Vrushali Borle, Vaishnavi Chaugule, Prof. Yogesh Mali, "Emotion Recognition Based Personal Entertainment Robot Using ML & IP", International Journal of Scientific Research in

Science and Technology(IJSRST), Print ISSN : 2395-6011, Online ISSN : 2395-602X, Volume 5, Issue 8, pp.73-75, November-December-2020.

[32] Amit Lokre, Sangram Thorat, Pranali Patil, Chetan Gaddekar, Yogesh Mali, " Fake Image and Document Detection using Machine Learning", International Journal of Scientific Research in Science and Technology(IJSRST), Print ISSN : 2395-6011, Online ISSN : 2395-602X, Volume 5, Issue 8, pp.104-109, November-December-2020.

[33] Ritesh Hajare, Rohit Hodage, Om Wangwad, Yogesh Mali, Faraz Bagwan, "Data Security in Cloud", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 8, Issue 3, pp.240-245, May-June-2021

[34] Yogesh Mali and Tejal Upadhyay, "Fraud Detection in Online Content Mining Relies on the Random Forest Algorithm", SWB, vol. 1, no. 3, pp. 13–20, Jul. 2023, doi: 10.61925/SWB.2023.1302.

[35] Mali, Y. K., Rathod, V. U., Borate, V. K., Chaudhari, A., & Waykole, T. (2023, June). Enhanced Pin Entry Mechanism for ATM Machine by Defending Shoulder Surfing Attacks. In International Conference on Recent Developments in Cyber Security (pp. 515-529). Singapore: Springer Nature Singapore.

[36] Mali, Y. K., Rathod, V., Dargad, S., & Deshmukh, J. Y. (2024). Leveraging Web 3.0 to Develop Play-to-Earn Apps in Healthcare using Blockchain. In Computational Intelligence and Blockchain in Biomedical and Health Informatics (pp. 243-257). CRC Press.

[37] Deshmukh, J. Y., Rathod, V. U., Mali, Y. K., & Sable, R. (2024). and Classification. Data-Centric Artificial Intelligence for Multidisciplinary Applications, 114.

[38] A. More, S. R. Shinde, P. M. Patil, D. S. Kane, Y. K. Mali and V. K. Borate, "Advancements in Early Detection of Lung Cancer using YOLOv7," 2024 5th International Conference on Smart Electronics and Communication (ICOSEC), Trichy, India, 2024, pp. 1739-1746, doi: 10.1109/ICOSEC61587.2024.10722534

[39] Hrushikesh Kale, Kartik Aswar, Kisan Yadav, Dr. Yogesh Mali, "Attendance Marking using Face Detection", International Journal of Advanced Research in Science, Communication and Technology (IJARSCT) Volume 4, Issue 3, October 2024,pp 417-424 DOI: 10.48175/IJARSCT-19961.