International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 05 , May - 2026

A Machine Learning Framework for Personalized Job Recommendation

[1] Rumana Hasinullah Shaikh,[2] Subhalaxmi Nayak
[1],[2] Department of Computer Science Engineering ,GIFT Autonomous ,
Bhubaneswar, Corresponding Author :

Abstract - Resume Based Company Recommendation System streamlines and improves the manual placement processes. In today's competitive job market, the efficient classification of resumes plays a pivotal role in streamlining recruitment processes. The model helps new students to find best fit companies for them. Also, this research investigates the effectiveness of different machine learning models in classifying resumes focusing on two distinct methodologies for splitting the dataset: manual and automatic. The model is evaluated using three machine learning algorithms which are, K-Nearest Neighbors (KNN), Support Vector Machine (SVM), and Random Forest. Labelled datasets comprising resumes from multiple sources are utilized, with one manually splitting technique into training and testing sets and the other automatically spitting technique by the model. This model focuses on automating the process of matching students resumes with company. When a new resume is uploaded then the model recommends a company to that student. The study compares the accuracy of these models under both splitting methodologies and analyzes the impact of dataset partitioning on classification performance. Experimental results demonstrate varying degrees of accuracy across the models, with Random Forest achieving the highest accuracy manual splitting of training and testing dataset and SVM achieving the highest accuracy automatic splitting of training and testing dataset. The findings underscore the importance of dataset splitting techniques and provide insights into the selection of appropriate machine learning models for resume classification tasks. This research contributes to the optimization of recruitment processes and informs practitioners and researchers about effective strategies for resume classification leveraging machine learning techniques.

Key words: Machine Learning algorithms, Support Vector System, Random Forest, Splitting techniques

I. INTRODUCTION

In today's competitive job market, connecting talented individuals with suitable employment opportunities can be challenging. To address this, we've implemented A Resume Based Company Recommendation System. This system uses Machine Learning Technique to streamline the process of recommending companies to student based on their resume. Traditional recruitment processes involve manual sorting of numerous resumes, which is time-consuming and requires a lot of resources. By using machine learning, we can automate this process, making it faster and more efficient. This study involve the implementation of a model using different machine learning algorithms such as K-Nearest Neighbors (KNN), Support Vector Machine (SVM), and Random Forest and different dataset splitting techniques such as manual and automatic to compare the accuracy of the model, and also recommendation of company to a new uploaded student resumeThe K-Nearest Neighbors (KNN) algorithm is a supervised machine learning approach for classification and regression tasks. It works by finding the k closest data points (neighbors) in the training data to a new, unseen data point. Support Vector Machine (SVM) is another supervised machine learning algorithm used for classification and regression tasks. SVM constructs a hyperplane that best separates different classes of data points. This hyperplane is positioned in a way that maximizes the margin between the classes, making SVM effective in handling complex datasets. Random Forest is also a supervised machine learning algorithm used for classification and regression tasks. Random Forest operates by building multiple decision trees during training and outputs the mode (for classification) or average prediction (for regression) of the individual trees. This ensemble approach enhances the robustness and accuracy of the model by reducing overfitting and improving generalization. Moreover, the method employed to partition datasets into training and testing subsets greatly influences the performance of machine learning models. While manual partitioning is conventionally utilized, its effectiveness in comparison to automatic partitioning methods remains largely unexplored. Therefore, there exists a critical knowledge gap regarding the influence of dataset partitioning techniques on model accuracy and generalization. Additionally, we aim to evaluate the accuracy of company recommendations generated by our system. According to the research, Random Forest gives more accuracy results in manually splitting of datasets and SVM gives more accuracy in automatic splitting of datasets. Also, the model automates student resume matching with company. It dynamically adjusts recommendations based on real-time updates and trends, ensuring relevance and accuracy. This empowers students to make

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 05 , May - 2026

informed career decisions. Our model offers a streamlined pathway for students to identify job opportunities aligned with their resumes. By analyzing the recommendations made for new student resumes, we can measure the system's ability to match candidates with suitable employment opportunities effectively. This assessment will help validate the practical utility of our approach in real-world. A comprehensive analysis of different machine learning algorithms and different dataset partitioning strategies is essential to optimize model performance and enhance the efficiency, recommendation of company for a new resume make the model suitable for real time scenarios.

II. PROBLEM STATEMENT

This project is aimed at implementing A Resume Based Company Recommendation System for recommendation of company to a new student based on resume and comparing the efficiency of different models with different dataset splitting techniques. In manual recruitment practices, poses a significant challenge for effective candidate screening and selection. Manual evaluation of resumes not only consumes valuable time and resources but also introduces subjective biases and inconsistencies. Consequently, there is a need to devise automated approaches to reduce the labour work and get more accurate decisions. Furthermore, the process of splitting the dataset into training and testing subsets significantly influences the performance of machine learning models. While manual splitting is commonly used, its efficiency compared to automatic splitting methods is less. Hence, there is a critical knowledge gap regarding the impact of dataset splitting techniques on model accuracy. These addressed challenges, requires the development and evaluation of machine learning models capable of recommending company name to new students based on their resume. Moreover, a comprehensive analysis of different dataset splitting strategies is crucial to optimize model performance. By addressing these issues, this research aims to make decisions easy for students by recommending company name which is best suited and also analyses the accuracy of different models and how spitting techniques can affects the performance of the models.

III. LITERATURE REVIEW

The paper presents a comprehensive approach towards automating the process of CV screening and recommendation, aiming to alleviate the challenges faced by recruiters in selecting suitable candidates efficiently. The approach involves extracting relevant information from CVs and job descriptions using machine learning techniques, followed by comparing the extracted data to find the best-matched CVs using various similarity metrics. The paper examines several different ways to compare resumes to job descriptions (cosine similarity). This is a good thing because it explores different options to find the best fit [1].

The paper presents a detailed overview of various systems proposed for resume screening and sorting using natural language processing (NLP) and machine learning techniques. To improve these systems, the researchers suggest that people who understand AI and people who understand hiring should work together to make sure these systems can be used in the real world for many different jobs. Even though there are some challenges, these new resume sorting systems seem to be very accurate at finding the right candidates for different jobs. This could make hiring much faster in the future [2].

The proposed machine learning-based resume ranking system presents a promising approach to addressing the challenges faced by recruiters in filtering through large volumes of resumes efficiently. This model presents a cost-effective and timesaving solution for recruiters, offering significant advantages in terms of efficiency, accuracy, and user experience. future scope, the system could further be enhanced by incorporating features such as interview scheduling and offer management, which would provide end-to-end support for the recruitment process [3].

This paper represents the automated resume screening and ranking system outlined in the paper show promising results in terms of efficiency and accuracy, it's important to approach its implementation with a critical eye. The system's reliance on algorithms like K-NN and cosine similarity raises concerns about potential biases embedded in the data and the limitations of content-based filtering methods. The reliance on content-based filtering methods may limit the system's ability to identify candidates with unique skills or experiences that are not explicitly stated in the job description. Finally, while the system integrates a candidate screening component using MCQ-based tests, it's essential to recognize the limitations of such assessments in evaluating candidates' true capabilities and fit for the role [4].

The paper provides a comprehensive overview of various machine learning and natural language processing (NLP) techniques employed in the recruitment industry, particularly in the Indian context. While these technologies have undoubtedly improved efficiency and reduced manual effort in resume screening, critical considerations remain regarding potential biases and fairness in the hiring process. Continued research and development are essential to overcome the limitations

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 05 , May - 2026

of current systems and to ensure that automated recruitment processes contribute to a more equitable job market [5].

IV. PROPOSED SOLUTION AND BLOCK DIAGRAM : -

PROPOSED SOLUTION

This project is a resume based company recommendation system designed to assist students in finding suitable companies based on their resume and evaluates different machine learning algorithm and different dataset splitting techniques.

Manual Dataset Splitting Technique:-

• Training the Model:

This involves importing a labeled resume dataset for training the model and the model lists the student resume name with corresponding placed company name, and then the model displays a message that the model is trained.

• Testing the Model:

This involves importing a labeled resume dataset for testing the model and evaluating the model's accuracy.

• Uploading New Resume

This involves selecting and uploading a new resume of a student.

• Company recommendation for the new resume:

Based on student resume, training and testing dataset the model recommends company name to the student.

Automatic Dataset Splitting Technique:-

• Dataset Splitting:

This involves importing a labelled resume dataset which the model itself splits into training and testing dataset. The model is trained using training dataset and tested using testing dataset.

• Training The Model:

This involves training the model using the splitted training dataset and the model lists the student resume name with corresponding placed company name, and then the model displays a message that the model is trained.

• Testing The Model:

This involves testing the model using the splitted tested dataset and evaluating the model's accuracy.

•Uploading New Resume

This involves selecting and uploading a new resume of a student.
• Company recommendation for the new resume:
Based on student resume, training and testing dataset the model recommends company name to the student.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 05 , May - 2026

BLOCK DIAGRAM

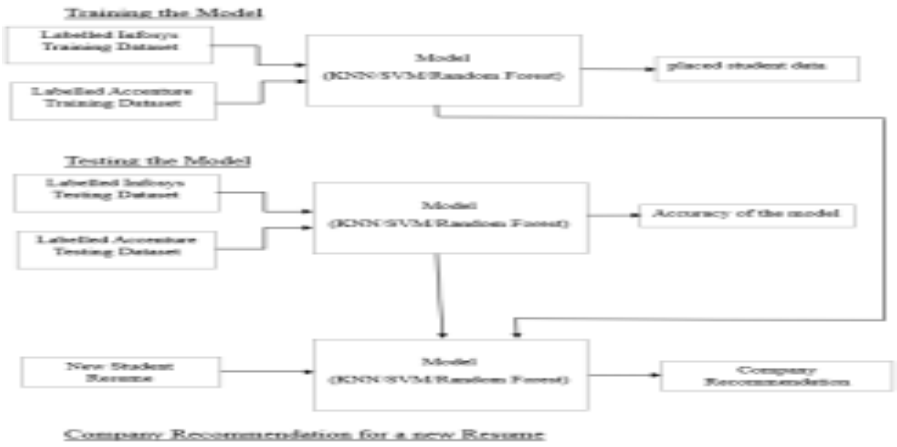

Fig 1: Block Diagram of Manually Splitting of Training and Testing Dataset

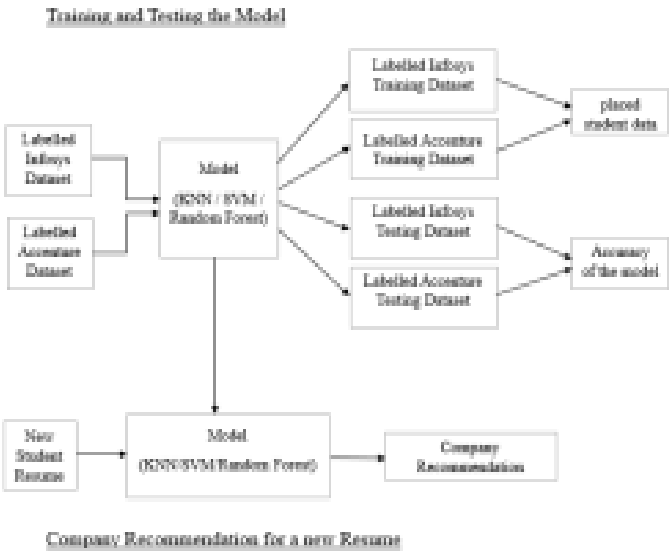

Fig 2: Block Diagram of Automatically Splitting of Training and Testing Dataset

V. WORKING OF THE MODEL

Training the Model:

The labelled resume dataset which comprises resumes from various sources, for companies Infosys and Accenture, is used to train the model. The model counts and displays the total number of resumes for each company. Then these resumes are preprocessed to extract relevant information. The preprocessing step involves converting the textual content of resumes into numerical representations using TF-IDF (Term Frequency Inverse Document Frequency) vectorization, then the model displays the placed students named with respective company as resume name and placed company name. After preprocessing the text data and applying TF-IDF vectorization, the code prints the shape of the resulting TF-IDF matrix for the training dataset. This output confirms the dimensions of the matrix, indicating the number of resumes (rows) and the number of unique features (columns) extracted from the text data. It serves as a checkpoint to ensure that the TF-IDF vectorization process is successful and that the matrix is ready for model training. Following the training of the model, a confirmation message is printed to indicate that the model training process is complete. This message provides assurance that the model is ready for evaluation and recommendation tasks.

Testing the Model:

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 05 , May - 2026

Similarly, after preprocessing and vectorizing the text data from the testing dataset, the code prints the shape of the TF-IDF matrix. This output verifies the dimensions of the testing dataset matrix, ensuring consistency with the training dataset matrix. It confirms that the testing data is properly transformed and ready for model evaluation. Finally, the code calculates the accuracy of the trained model on the testing dataset and prints the result. This accuracy score quantifies the performance of the model in correctly predicting company affiliations based on the resumes in the testing dataset. It provides a measure of the model's effectiveness and helps assess its practical utility in real-world scenarios.

Uploading New Resume:

A new resume can be uploaded and the model based on the student resume, training and testing dataset recommends a company to the new student which will fit best. The function named "upload_new_resume()" is used to facilitate the uploading of a new resume file for recommendation. Upon function invocation, a Tkinter root window is created and immediately hidden using "root.withdraw()". The code prompts the user to select a resume file using the "filedialog.askopenfilename()" method. The file dialog window allows users to navigate their filesystem and choose a PDF file containing the resume. If a file is selected (i.e., the file_path variable is not empty), the selected file's path is printed to the console. This serves as feedback to the user, confirming the successful selection of a resume file. If no file is selected (i.e., the file_path variable is empty), an error message is displayed, indicating the user to select a resume file. The code extracts the text content from the selected PDF resume file using the "extract_text_from_pdf()" function. The extracted text is then preprocessed to ensure consistency. The preprocessed resume text is transformed into a TF-IDF vector using the "tfidf_vectorizer.transform()" method. This step converts the text data into a numerical format suitable for input to the trained machine learning model. The TF-IDF vectorized resume is passed to the model to recommend company name. Finally, the function "upload_new_resume()" is called to initiate the resume upload process when needed.

VI. MODEL DISCUSSION

KNN Model

The K-Nearest Neighbors (KNN) algorithm is a supervised machine learning approach for classification and regression tasks. It works by finding the k closest data points (neighbors) in the training data to a new, unseen data point. KNN is employed to classify resumes based on their similarity to other resumes in the training dataset. It calculates the distance between the input resume and all other resumes and assigns the most frequent class label among the k nearest neighbors. KNN operates by measuring the similarity between resumes based on their TF-IDF (Term Frequency-Inverse Document Frequency) representations. Each resume is treated as a point in a high-dimensional space, where the distance between points reflects their similarity. When a new resume is uploaded, KNN identifies its k nearest neighbors in the training dataset and assigns the class label that appears most frequently among these neighbors. The choice of k determines the granularity of classification; smaller values of k may lead to more flexible decision boundaries but are more susceptible to noise, while larger values may lead to smoother boundaries but risk oversimplification. Thus, selecting an appropriate value of k is crucial for achieving optimal classification performance. KNN is useful in this project as it doesn't require explicit training and can directly classify new resumes based on similarity metrics.

SVM Model

Support Vector Machine (SVM) is supervised machine learning algorithm used for classification and regression tasks. SVM constructs a hyperplane that best separates different classes of data points. This hyperplane is positioned in a way that maximizes the margin between the classes, making SVM effective in handling complex datasets. SVM is applied for binary classification tasks, separating resumes into two classes (Infosys or Accenture) based on the learned hyperplane that maximizes the margin between the classes. By finding the optimal hyperplane, SVM can effectively classify resumes into different company categories. SVM works by finding the optimal hyperplane that separates the resumes belonging to different companies, Infosys or Accenture in the TF-IDF feature space. SVM learns to identify the best decision boundary by maximizing the margin between the two classes while minimizing classification errors. By transforming the input data into a higher-dimensional space using the kernel trick such as linear kernel, SVM constructs a decision boundary that effectively separates the resumes into distinct categories. SVM is useful in this project as it is effective in handling high-dimensional data and its robustness to outliers, where resumes are represented as TF-IDF vectors, also it accurately classifies resumes even in complex feature spaces.

Random Forest Model

Random Forest is a supervised machine learning algorithm used for classification and regression tasks. Random Forest operates by building multiple decision trees during training and outputs the mode (for classification) or average prediction (for regression)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 05 , May - 2026

of the individual trees. Random Forest is chosen for its ability to handle high-dimensional data and provide robust classification performance. It constructs multiple decision trees during training and outputs the mode of the class labels as the prediction. Random Forest operates by aggregating the predictions of multiple decision trees, each trained on a random subset of the training data and features. Random Forest constructs an ensemble of decision trees based on the TF-IDF representations of resumes. During training, each decision tree learns to partition the feature space based on the importance of different terms in distinguishing between Infosys and Accenture resumes. The final prediction is determined by a majority vote among the decision trees, resulting in a robust and stable classification model. Random Forest is suitable for this project as it can handle noise and overfitting, making it robust in classifying resumes into the appropriate company categories based on their TF-IDF representations. Also, it can provide insights into feature importance, aiding in understanding the significant factors influencing the classification decision.

VII. RESULT AND DISCUSSION

Table 1: Results of Manual Splitting of Training and Testing Technique

Model Name	Accuracy
K - Nearest Neighbor Model	40.83 %
Support Vector Machine Model	41.67 %
Random Forest Model	45.00 %

In manual splitting of training and testing dataset technique, Random Forest gives more accuracy comparing to KNN and SVM.

Table 2: Results of Automatic Splitting of Training and Testing Technique

Model Name	Accuracy
K - Nearest Neighbor Model	80.00 %
Support Vector Machine Model	80.83 %
Random Forest Model	78.33 %

In Table 2, automatic splitting of training and testing dataset technique, SVM gives more acuracy comparing to KNN and Random Forest.

This result tells us that the different algorithms used to train the mode can give different accuracy while working with same dataset and also, the splitting technique used to split training and testing dataset can give different accuracy results.

Training and Testing of Model

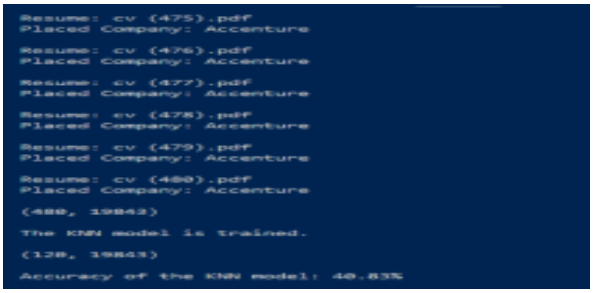

Fig 3: KNN Model using Manually Splitted Dataset

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 05 , May - 2026

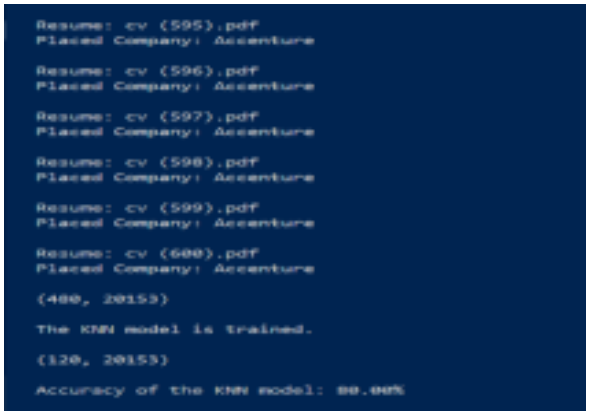

Fig 4: KNN Model using Automatically Splitted Dataset

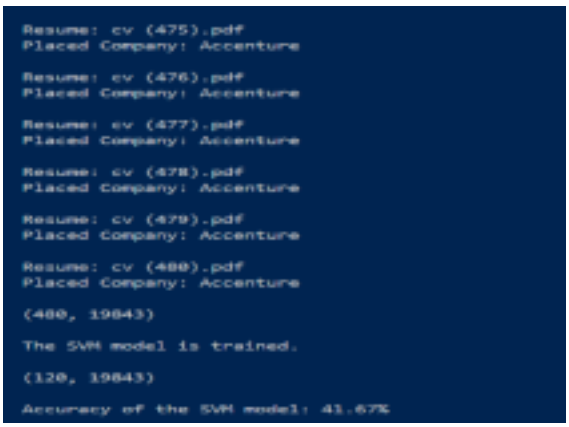

Fig 5: SVM Model using Manually Splitted Dataset

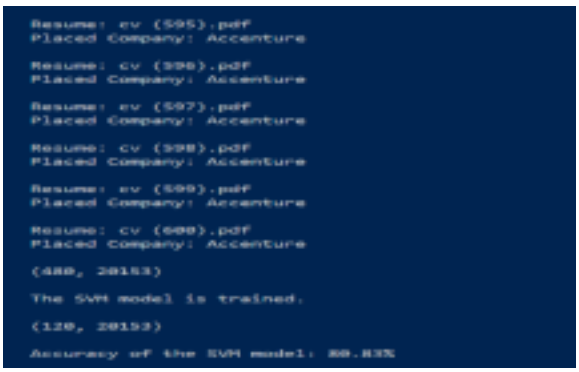

Fig 6: SVM Model using Automatically Splitted Dataset

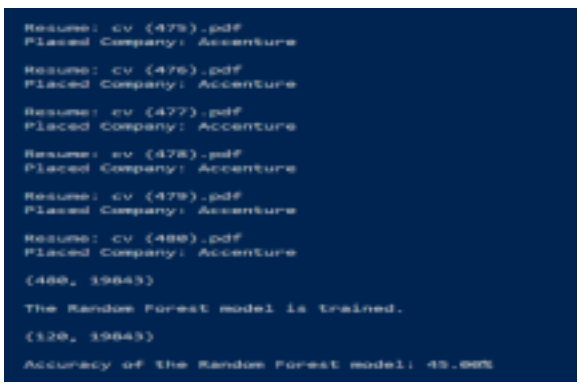

Fig 7: Random Forest Model using Manually Splitted Dataset

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 05 , May - 2026

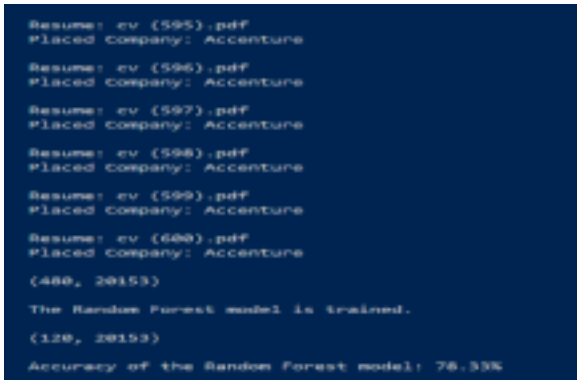

Fig 8: Random Forest Model using Automatically Splitted Dataset

Fig 9: Company Recommendation for a New Resume using KNN Model

Fig 10: Company Recommendation for a New Resume using SVM Model

Fig 11: Company Recommendation for a New Resume using Random Forest Model

VIII. CONCLUSION

In summary, this research underscores the varying effectiveness of machine learning models and different splitting techniques, and recommendation companies based on resume. The experimental results reveal Random Forest's superiority over KNN and SVM in manual datasesplitting and SVM's superiority over KNN and Random Forest in automatic dataset splitting scenarios.

Moreover, the study underscores the significance of dataset partitioning techniques, with automatic splitting demonstrating higher accuracies compared to manual methods. These findings underscore the importance of selecting appropriate machine learning models and dataset splitting methodologies for resume classification tasks. The model can recommend companies to students, students can upload the resume and basis on model results and make clear decisions ahead and this make the model more applicable for real world.

IX. FUTURE SCOPE

Future research could involve including additional factors influencing classification performance and explore advanced machine learning techniques, deep learning techniques to further optimize the model.

X. REFERENCES

[1] S. M. Shahriar Ferdous Shovon, Md. Mahir Absar Bin Mohsin(B), Kanij Tamema Jahan Tama, Jannatul Ferdaous, and Sifat Momen, "CVR: An Automated CV Recommender System Using Machine Learning Techniques", Springer (2023).

[2] Riza Tanaz Fareed, iza Tanaz Fareed, Sharadadevi Kaganurmath, "Resume Classification and Ranking using KNN and Cosine Similarity", IJERT (2021).

[3] Anushka Lad, Siddhi Ghosalkar, Balkrishna Bane, Krutika Pagade, Anupama Chaurasia, "Machine Learning Based Resume Recommendation System", IJMDES (2022).

[4] Tejaswini K, Umadevi V, Shashank,M Kadiwal, Sanj ay Revanna,"Design and development of machine learning based resume ranking system" , KeAi (2022).