

ML Based Resume Classifier System (ResuPro)

SUBMITTED IN PARTIAL FULFILLMENT FOR THE REQUIREMENT OF THE AWARD
OF DEGREE OF

BACHELOR OF TECHNOLOGY

IN

COMPUTER SCIENCE



Submitted by

ABHISHEK VERMA (2100290120010)

ADARSH MISHRA (2100290120012)

ALOK RANJAN DUBEY (2100290120026)

ANURAG KUMAR (2100290120038)

Supervised by

Dr. ABHISHEK GOYAL

(Associate Professor)

Session 2024-25

DEPARTMENT OF COMPUTER SCIENCE

KIET GROUP OF INSTITUTIONS, GHAZIABAD

(Affiliated to Dr. A. P. J. Abdul Kalam Technical University, Lucknow, U.P., India)

May 2025

DECLARATION

I/We hereby declare that this submission is our own work and that, to the best of our knowledge and belief, it contains no material previously published or written by another person nor material which to a substantial extent has been accepted for the award of any other degree or diploma of the university or other institute of higher learning, except where due acknowledgment has been made in the text.

Signature:-

Name:- Abhishek Verma

Roll No.:- 2100290120010

Signature:-

Name:- Alok Ranjan Dubey

Roll No.:- 2100290120026

Signature:-

Name:- Adarsh Mishra

Roll No.:- 2100290120012

Signature:-

Name:- Anurag Kumar

Roll No.:- 2100290120038

Date :- 08/05/25

CERTIFICATE

This is to certify that Project Report entitled “**ML BASED RESUME CLASSIFIER**” which is submitted by **Group Id (PCS-03)** in partial fulfillment of the requirement for the award of degree B. Tech. in Department of Computer Science of Dr. A.P.J. Abdul Kalam Technical University, Lucknow is a record of the candidates own work carried out by them under my supervision. The matter embodied in this report is original and has not been submitted for the award of any other degree.

Date: 08/05/25

Supervisor

Dr. Abhishek Goyal

(Associate Professor)

ACKNOWLEDGEMENT

It gives us a great sense of pleasure to present the report of the B. Tech Project undertaken during B. Tech. Final Year. We owe special debt of gratitude to Professor Abhishek Goyal, Department of Computer Science, KIET, Ghaziabad, for his/her constant support and guidance throughout the course of our work. His/Her sincerity, thoroughness and perseverance have been a constant source of inspiration for us. It is only his cognizant efforts that our endeavors have seen light of the day.

We also take the opportunity to acknowledge the contribution of Dr. Ajay Kumar Srivastava, Head of the Department of Computer Science, KIET, Ghaziabad, for his full support and assistance during the development of the project. We also do not like to miss the opportunity to acknowledge the contribution of all the faculty members of the department for their kind assistance and cooperation during the development of our project.

Last but not the least, we acknowledge our friends for their contribution in the completion of the project.

ABSTRACT

In the current rapid-paced job market, the hiring process is susceptible to various issues, particularly during the early stages of resume screening. Traditional methods by way of manual verifying CVs not only consumes time but also is wasteful and it is based on human prejudices and personal views. Increasing numbers of cases are being filed, and thus an increasing requirement of computer-based systems in an effort to hasten the process.

ResuPro is a machine machine learning (ML)-based resume classifier system for conducting computerized resume screening processes. Sophisticated machine learning processes and natural language processing (NLP), ResuPro classifies resumes into predefined job categories based on content. The system pulls important information like skills, experience, and company type (Startup, Midlevel, Enterprise), to give the complete and proper determination of candidates.

ResuPro organizes resumes by career (i.e., IT, Finance, Marketing) and employs a grading system to categorize resumes based on how well they fit job experience-based, skill-based, and education-based requirements. This experience-based scoring system improves decision-making by ranking resumes, shortlisting the best candidates, and furnishing personalized information on their experience and suitability for the job.

The resumes of the applicants are sorted with 88% accuracy. ResuPro does it automatically and thus saves manual effort, simplifying the process of hiring, and better decision-making by winnowing resumes for relevance, reducing human bias, and providing more balanced evaluation of candidates. This system offers a strong, efficient, and impartial resume screening mechanism that can easily be employed by firms to optimize the hiring process.

The architecture is discussed in the article, methodology, and ResuPro performance, highlighting the strength of ML and NLP technologies that have the potential to significantly improve hiring efficiency, equity, and effectiveness in today's hectic working life.

TABLE OF CONTENTS	Page No.
DECLARATION.....	ii
CERTIFICATE.....	iii
ACKNOWLEDGEMENTS.....	iv
ABSTRACT.....	v
LIST OF FIGURES.....	ix
LIST OF TABLES.....	xi
LIST OF ABBREVIATIONS.....	xii
SDG MAPPING WITH JUSTIFICATION	xii
 CHAPTER 1 INTRODUCTION	 1-5
1.1 Introduction to Project	1
1.2 Project Category	3
1.3 Objectives	3
1.4 Structure of Report	4
 CHAPTER 2 LITERATURE REVIEW	 6-11
2.1 Literature Review	6
2.2 Research Gaps	9
2.3 Problem Formulation	10
 CHAPTER 3 PROPOSED SYSTEM	 12-21
3.1 Proposed System	12
3.2 Unique Features of The System	20
 CHAPTER 4 REQUIREMENT ANALYSIS AND SYSTEM SPECIFICATION	 22-26
4.1 Feasibility Study (Technical, Economical, Operational)	22
4.2 Software Requirement Specification	22
4.2.1 Data Requirement	22
4.2.2 Functional Requirement	23

4.2.3	Performance Requirement	24
4.2.4	Maintainability Requirement	24
4.2.5	Security Requirement	24
4.3	SDLC Model Used	24
4.4	System Design	25
4.4.1	Data Flow Diagrams	25
4.4.2	Use Case Diagrams	26
4.5	Database Design	26
CHAPTER 5	IMPLEMENTATION	27-29
5.1	Introduction Tools and Technologies Used.	27
5.2	Dataset Description	28
CHAPTER 6	TESTING, AND MAINTENANCE	30-33
6.1	Testing Techniques and Test Cases Used	30
CHAPTER 7	RESULTS AND DISCUSSIONS	34-40
7.1	Presentation of Results (Charts/Graphs/Tables)	34
7.1.1	Resume Creation Result	34
7.1.2	Classification Result	35
7.1.3	Scoring Result	36
7.2	Performance Evaluation	38
7.3	Key Findings	40
CHAPTER 8	CONCLUSION AND FUTURE SCOPE	41-43
	REFERENCES	44-46
	Proof of patent publication (Screenshot of Publication)	47
	Plag Report	48-54

LIST OF FIGURES

Figure no.	Description	Page no.
1.	Deployment diagram	18
2.	Data flow diagram	25
3.	Use case diagram	26
4.	ER diagram	26
5.	Classification Dataset Graph	28
6.	Scoring Dataset	29
7.	Resume Creation result	34
8.	Output of classification model	35, 36
9.	Output of scoring model	37, 38
10.	Accuracy of classification model	39
11.	Accuracy of scoring model	39

LIST OF TABLES

Table no.	Description	Page no.
1.	Test cases for classification model	31
2.	Test cases for scoring model	32
3.	Test cases for frontend functionalities	33

LIST OF ABBREVIATIONS

Abbreviation	Full form
AI	Artificial Intelligence
API	Application Programming Interface
CDN	Content Delivery Network
HTML	HyperText Markup Language
CSS	Cascading Style Sheets
JWT	JSON Web Token
NLP	Natural Language Processing

SDG MAPPING WITH JUSTIFICATION

SDG 4: Quality Education

Target: Ensure inclusive and equitable quality education and promote lifelong learning opportunities to all.

Reasoning:

1. **Skill-Based Resume Preparation** – ResuPro enables students and candidates to empower themselves by helping them build formal, skill-based, and field-specific resumes. ResuPro focuses on expressing tangible skills and abilities as qualifications rather than general descriptions. Not only does this make the resume clearer, but it also meets recruiters' expectations better. By causing candidates to reflect on their learning achievements and skills, ResuPro creates self-responsibility and self-awareness in professional development.

This focus directly benefits the improvement of employability and applicable education to real-world job markets. Personalized Resume Scoring Based on Machine Learning ResuPro incorporates intelligent ML-based resume scoring that rates individual resumes on various parameters such as relevance, clarity, organization, and skill alignment. Such customized feedback enables users to gain awareness of their strengths and weaknesses through data-driven insights. Through the provision of constructive and objective feedback, users learn better how to make their resumes relevant to the industry. Such self-awareness results in lifelong learning whereby users will be motivated to upskill and continuously enhance their qualifications to stay relevant.

2. **Closing the Education-to-Employment Gap** – The incompatibility between workforce needs and school education is one of the core issues of developing countries. ResuPro fills this gap by determining resumes with job titles, desired experiences, and sets of skills across different sectors. With intelligent recommendations and clustering

techniques, it helps students and recent graduates correlate their higher education with job opportunities. This practical connection between teaching and employment is meant to make sure that educational education is not divorced from useful requirements, thus injecting efficiency and fairness into education systems.

3. **Evidence-Based Career Counseling and Guidance** – ResuPro provides career guidance using data by mapping the resumes of users against successful industry-specific profiles. It provides users with specific feedback regarding what skills or qualifications they should acquire in order to enter the chosen careers. It helps students make well-informed choices about courses, certification, and internships, thus giving them an individualized learning experience. This particular guidance avoids making people lose their way among the vast job market, thus providing them with that all-important confidence as well as long-term career guidance.
4. **Promoting Digital Competence and Professional Communication** – Through its interactive website and resume templates, ResuPro indirectly encourages digital literacy through acquiring knowledge on how to utilize technology for occupational purposes. Users are presented with the need for formatting, conciseness, cohesion, and professionalism—essential communication skills in the modern times. This aids academic and non-academic learners in being digitally literate to enable and professionally competent

SDG 17: Partners for the Goals

Target: Strengthen means of implementation and reinvigorate commitment to Sustainable Development.

Reasoning:

1. **Partnership with Learning Institutions** – ResuPro collaborates aggressively with universities, vocational schools, e-learning websites, and government-sponsored skill development programs. Integrating the platform into school curricula and career guidance programs, students are prepared early on with the skills to create effective, industry-

related resumes. Such collaborations help ensure that providers of education remain in tune with the changing needs of the employment market, improving the employability value of educational programs. These partnerships foster cooperation between the public, private, and education sectors towards promoting employability—a common task in sustainable development.

2. **Recruitment and Interaction with HR Experts** – ResuPro adds value to recruitment agencies and organizations through its AI-based resume screening, candidate shortlisting, and other capabilities. This collaboration lightens the burden of HR departments while allowing merit-based candidate evaluations without bias. The platform also provides a channel for candidate-HR interaction through mentoring sessions, webinars, and virtual career fairs. These interactions not only make the recruitment method more diversified but also make the hiring process more inclusive, open, and skill-based, which helps in international employment equality and economic integration.
3. **Job Market Intelligence through AI** – Through aggregation and analysis of big resume data and labor market trends, ResuPro is an intelligent partner for policymakers, educators, and employers. It provides actionable intelligence on emerging professions, required skills, industry shortfalls, and available talents. This intelligence supports fact-based decision-making in sectors—whether it is curating educational courses, designing workforce policies, or directing skill-building investments. This market intelligence powered by artificial intelligence facilitates well-planned decision-making and long-term economic development.
4. **Open-Source Contributions & API Integration** – ResuPro is built to work in an open digital network. With open-source, nicely-documented APIs, it will be easy to integrate it with other platforms like career websites, HRMS (Human Resource Management Systems), LMS (Learning Management Systems), and public sector job portals. This makes it possible to enable cross-platform collaboration, scalability, and data exchange between systems—increasing impact through technology-facilitated collaborations. It

also encourages developers and organizations to extend ResuPro, fostering innovations through collaborative efforts.

5. **Contributing to Multistakeholder Collaboration for Sustainable Employment -**

ResuPro serves as a bridge between various stakeholders—students, educators, government, industry professionals, and employers. By providing a common digital platform for interaction, it promotes discussion and collaboration centered on long-term human capital building. This multistakeholder model is critical to the accomplishment of SDG 17 goals, as it combines varied strength and means toward a common objective of inclusive and sustainable development.

6. **Global Accessibility** – Deployed via cloud infrastructure, ResuPro ensures that users from diverse geographic locations—urban or rural—have access to its features without requiring heavy system resources. This global accessibility supports international partnerships, allowing educational institutions, NGOs, and organizations around the world to adopt the platform according to local needs. The consistent user experience and scalability make it an effective tool for promoting global equality in access to career development resources, especially in underrepresented and underserved regions.

CHAPTER 1

INTRODUCTION

1.1 Introduction to Project

In the era of high-speed recruitment, companies typically have piles of resumes and become tired in the process. The traditional process of manually sifting through resumes to recruit the most suitable person for a given job opening is time and effort-consuming and prone to human error. To counter the above disadvantages, we present a cutting-edge solution – the Machine Learning (ML) based Resume Classifier System.

This initiative leverages the power of advanced machine learning technology in the automation of resume screening to make hiring easier and less expensive and to optimize talent acquisition teams. Our algorithm, developed with the use of natural language processing (NLP) and pattern matching, is constructed to rank and sort resumes according to how effectively the resumes match the qualifications of the jobs being offered.

The Resume Classifier System with ML not only streamlines the recruitment process but also minimizes the risk of losing quality candidates. This project is a milestone in the HR tech industry with a scalable and flexible answer to the ever-present looming issues of talent sourcing in the evolving nature of the prevailing work climate. In the following discussion of the details of this system, we will discuss its architecture, functionality, and the level of impact this system can have in reframing organizations' strategy in reaching and acquiring the best talent.

Technology Stack Overview and Explanation

1. Frontend: React + Tailwind CSS

- JavaScript library for building rich interactive user interfaces. In our project, it enables one to create an interactive and responsive frontend where the user can upload his or her

resume with ease, construct them from templates, and be able to edit their profiles. React's component-based nature enables reusable UI components, allowing for clean code and easy maintenance.

- Tailwind CSS is a utility-first CSS framework that complements React by enabling quick UI development with limited amount of custom CSS. Tailwind has existing classes for styling, which aids in design consistency but not at the cost of freedom during customization. The combination of both ensures a modern, mobile-first, and visually appealing user experience.

2. Backend: Node.js + Express + MongoDB

- Node.js is a runtime environment that executes JavaScript outside of the web browser environment, and it is most suitable for scalable backend service development. Node.js allows asynchronous operations, which maximize web application performance by executing several user requests in parallel effectively.
- Express.js is a light and versatile Node.js web application framework that offers a wide array of features to develop web and mobile applications. Express in this context serves as the foundation for routing and server-side activities such as resume file upload, authentication, and database operations.
- MongoDB is a NoSQL database that stores data in a flexible, JSON-based form. It is appropriate for use in applications requiring processing of unstructured information—such as resumes—present in a variety of formats. MongoDB is capable of efficient querying and high scalability, ensuring secure storage and retrieval of user information, resume information, and classification results.

3. Machine Learning Models (Deployed on Streamlit)

The backbone of the system's information is its ML models, which carry out the function of classifying the resumes under applicable job categories and rating them on skills, work experience, and company experience. These models are run with Streamlit, an open-source Python library to facilitate rapid web application development and deployment with ML. Streamlit offers an interactive view for inspecting model output and real-time inference on resumes uploaded.

These models employ state-of-the-art Natural Language Processing (NLP) techniques, such as:

- **Tokenization:** Segmentation of text into informative units such as words or phrases.
- **Lemmatization:** Word normalization to its base or dictionary word to deal with variations.
- **TF-IDF Vectorization:** Word frequency and inverse document frequency-based numerical transformations of text data to help quantify word importance within context.

Each of these helps the system scan semantic content in resumes so that scoring and classification transcend keyword comparisons alone. This facilitates more accurate and meaningful candidate judgments.

1.2 Project Category

- Machine Learning and Web Development.
- Application of NLP techniques to classify resumes based on domain-specific requirements.

1.3 Objectives

With the help of this project, we are hoping to achieve the following main objectives:

1. To create UI/UX in the creation of a candidate's resume.

2. To create a system that is able to screen resumes automatically at the first level, reducing the time and resources put into the manual checking processes.
3. To build algorithms capable of realizing the contextual understanding of job requirements and applicant qualifications in order to enhance the accuracy of the classification task.
4. The new system will be based on Natural Language Processing (NLP) methods, the system strives to understand the semantic intent of resumes, allowing for a more accurate judgment of candidates' skills and fitness for particular jobs.
5. The project endeavors to improve decision-making in the initial stages of recruitment by offering data insights, thus making the selection process better and detecting high-level talent.

1.4 Structure of the Report

The report is organized to provide a concise systematized summary of the ResuPro project addressing the critical areas of research, system development, testing, and implementation. The report is divided into the following chapters:

- **Chapter 2: Literature Review**

This chapter addresses the current state of research and technology included in resume classification systems, recruitment machine learning solutions, and complexities in the HR domain. It also addresses the shortcomings of current solutions and lays the groundwork for the ResuPro system.

- **Chapter 3: Proposed System**

This chapter describes the ResuPro system and its primary features, including the ML-based resume categorization, domain-oriented resume scoring, and automated creation of personalized comments. It also indicates the special features of the system that differentiate it from the current solutions.

- **Chapter 4: System Specification and Requirement Analysis**

This chapter gives a thorough explanation of the system requirements, functional and non-functional specifications. It also consists of the feasibility study (technical, economical, and operational) and an overview of the implemented SDLC model. Topics

such as data flow diagrams, use case diagrams, and database design are addressed as well.

- **Chapter 5: Implementation**

This chapter covers the installation of the ResuPro system, describing the tools and the related technologies, e.g., application of machine learning algorithms to resume classification and scoring. It is also a question of front-end and back-end development processes and deployment of elementary system functionality.

- **Chapter 6: Testing and Maintenance**

The testing processes and techniques employed for the ResuPro project are discussed in this chapter. It is the test plan, testing methods (functional, performance, security, etc.), and test cases utilized to verify the quality of the system. It also displays the outcomes from the different levels of tests conducted, for example, unit testing, integration testing, and user acceptance testing (UAT).

- **Chapter 7: Results and Discussions**

This chapter presents the findings that have been gathered from the execution of the system testing. It features performance measurement, resume scoring outcomes by domain, and feedback on the success of the system. Some of the main findings gathered from the process of testing, including some of the problems that were experienced and how they were addressed, are also covered.

- **Chapter 8: Conclusion and Future Scope**

The last chapter offers an overview of main outcomes of the ResuPro project and the possible influence it can exert on the recruitment business. It also mentions possible fields of further development, such as broadening the capabilities of the system, improving machine learning algorithms, and applying new technologies to automate the recruitment process even further.

CHAPTER 2

LITERATURE REVIEW

2.1 Literature Review

Automated resume classification has undergone tremendous growth, especially with the incorporation of Natural Language Processing (NLP) and Machine Learning (ML) methods. There have been several research studies that have provided meaningful insights into developing intelligent resume screening systems. A detailed review of some of the major contributions in this area is presented below:

1. Surendiran B, Harsha Vardhan Chirumamilla, Maruprolu Naga Raju Reddy – IConSCEPT 2023

Title: An Automated Resume Classification Solution Using Machine Learning

This research brings forth an extensive method for resume classification through the application of ML models like Decision Tree, Random Forest, K-Nearest Neighbors (KNN), and Support Vector Machine (SVM). The authors solve the inefficiencies involved in conventional hiring by automating resume screening, hence minimizing both time and recruitment costs. 3,446 resumes were pre-processed with NLP methods (such as stop word elimination, tokenization, and lemmatization). The models performed very well overall, with visual representations employed to show model performance. This work highlights the efficacy of ML in resume classification, making the assignment quicker and more efficient.

2. Riza Tanaz Fareed et al. – Resume Classification Using Cosine Similarity

This research entails resume listing and job description matching using Cosine Similarity, a metric that approximates text similarity based on vector space models. The authors used an NLP pipeline consisting of removing stop words, lemmatization, and word extraction. TF-IDF vectorization is used to convert text into feature vectors, which are then fed through a KNN classifier. The system demonstrates high accuracy at 98.96%, showing the efficiency of vector

space-based similarity measurements in the task of resume classification.

3. Suhas H E and Manjunath A E – Resume Suggestion Model

The authors propose a resume suggestion system that tries to suggest suitable resumes for a specific job position. The model integrates Named Entity Recognition (NER) for technical skill identification from resumes and uses them to train a word2vec word embedding model. Cosine Similarity is employed to calculate the similarity between resumes and job descriptions. This method was 79.8% accurate, proving the effectiveness of combining semantic word embeddings with similarity measures towards resume suggestions personalized to the user.

4. Sujit Amin et al. – Resume Screening Web Application

The research describes the web application design for auto-screening resumes. The system is divided into three modules: recruiter interface, job applicant interface, and server-side processing. SpaCy is used in the NLP pipeline for operations such as entity recognition and language processing. The server scores resumes based on a score calculator developed specifically for this purpose, and 220 resumes were used for testing and training. The website is centered on modularity and end-to-end automation throughout the process of resume screening.

4. Sujit Amin et al. – Web Application for Resume Screening

This article discusses the development of a web application for resume screening through automated techniques. The system is divided into three modules: server-side processing, recruiter interface, and job applicant interface. The NLP pipeline utilizes SpaCy for entity recognition and language processing. The server ranks resumes based on a score constructed using a custom-built score calculator, and 220 resumes were used for testing and training. The software is designed to value modularity and end-to-end automation of the resume screening.

5. Pradeep Kumar Roy – Recruitment Cost Reduction

The study centers on cost-saving strategies in recruitment through optimization of applicant screening inefficiencies. The model employs n-gram-based text classification, NLP, and NER to

measure the aptitude of applicants. The different ML algorithms were tried and tested, and Linear Support Vector Machine (LSVM) performed best with 78.53% accuracy. The study here indicates how precise resume screening with NLP and ML can result in extreme reductions in hiring costs and enhance candidate quality.

6. Bhushan Kinge et al. – Resume Screening Using NLP and ML

The paper proposes an NLP and ML-based automated resume screening system. The main algorithms used that were experimented with include KNN, Random Forest, and SVM. The process begins with text extraction and preprocessed using NLP-based methods (e.g., tokenization and lemmatization). Of the models, Random Forest provided the highest classification accuracy of 94.5%. Authors conclude that the employment of ensemble learning models in combination with NLP preprocessing improves resume filtering accuracy considerably and accelerates the recruitment process.

7. Tejaswini K et al. – Machine Learning-Based Resume Ranking

The article presents a resume ranking system, which ranks applicants based on job requirement matching. It is compared with Naïve Bayes, KNN, and SVM performance algorithms. The SVM model functioned well on both accuracy and effectiveness, offering an efficient means for candidate ranking. The work brings to fore the ability of ML-powered systems to automate staffing processes in ranking resumes far superior to traditional heuristic algorithms.

8. D. Jagan Mohan Reddy et al. – Job Candidate Acceptance Prediction

This work shifts its interest from the screening of resumes to quantifying a candidate's likelihood of accepting a job offer. It utilizes the synergy between statistical techniques and ML classifiers, including Random Forest, to estimate candidate acceptance probability from attributes such as skill alignment, compensation, and location preferences. The Random Forest model was 94.86% accurate, which means that predictive analytics can not just be used by HR teams for screening but also for offer management and engagement programs.

2.2 Research Gaps

1. Resume Categorization Algorithm Precision and Efficiency

Most resume screening programs use rule-based or keyword-algorithm matching methods. However, the effectiveness of such approaches may still be insufficient regarding the capture of text semantics or context. There remains potential for further improvement in utilizing more advanced machine learning methods (e.g., deep learning, NLP algorithms) to classify and rank resumes more precisely.

Research Gap: Constructing and evaluating more precise, context-oriented resume classification models based on considering differences in candidate qualifications.

2. Domain-Specific Resume Scoring

Some resume rating systems provide scores, but they are mostly generic and not tailored to specific industries or job functions. Scoring algorithms need to be customized to individual domains, with an understanding of the skills, qualifications, and experience that are important for those domains.

Research Gap: Developing more accurate, domain-specific scoring models that effectively screen resumes in line with the specific requirements of individual jobs.

3. Bias in Resume Screening

A common challenge in machine learning models is de-biasing, especially when biased data sets are used for training the algorithms. For instance, previous employee hiring records may contain biases in candidate selection, which can easily be translated to machine systems.

Research Gap: Conducting experiments and studies on reducing biases in resume screening algorithms and ensuring fairness in employee hiring procedures.

4. User Experience and Usability in Resume Building Tools

Resume-building tools are widely available, but many fail to provide a good user experience, particularly for non-technical users. The interfaces of these tools can be confusing or unintuitive, leading to adoption and usability issues.

Research Gap: Investigating improved user interface (UI) and user experience (UX) designs for resume-building tools to make them simpler and more accessible to all users.

5. **Resume Tools' Capacity for Adaptation with Regard to New Job Market Trends**

The labor market is continually evolving, and certain skills and sectors quickly become obsolete. However, many resume builders are not adaptive to these changes and cannot incorporate new skills or qualifications.

Research Gap: Developing adaptive resume-building tools that dynamically adjust templates, categories, and scoring models to accommodate new trends and labor market needs.

6. **Seamless Integration with Other Hiring Tools and HR Platforms**

Most resume-building tools do not function independently and lack proper integration with other hiring systems like job boards, LinkedIn, or Applicant Tracking Systems (ATS).

Research Gap: Conducting research on advanced integration methods for resume-building tools, enabling seamless interactions with HR and recruitment software to create a more efficient recruitment process.

2.3 Problem Statement

With the present scenario of online resume websites and job application platforms, there is a considerable lack of integration between resume assessment tools and job specialty-specific categorization systems.

Online resume websites popular today usually tend to concentrate on either one of two aspects:

- **Resume Libraries:** Providing users with access to large collections of resume templates without giving a relevant assessment of quality or suitability to a specific job profile.
- **Scoring Engines:** Offering resume scores on a general formatting and keyword use basis, and not on a basis that incorporates resume suitability in terms of industry, role, or recruiter requirements.

As a rule, these two aspects seldom go smoothly hand in hand. A job applicant may, for instance, be awarded a "high" resume score from one platform and yet realize that his resume remains incompatible with the precise needs of his target occupation sphere (e.g., Software Engineering, Finance, Data Science).

Likewise, those platforms that pay close attention to resume building tend to ignore smart scoring so that users have no idea how their resume would fare during actual recruitment scans.

This gap creates a number of issues:

- **Confusion for New Job Seekers:** Particularly for new graduates or career changers, traveling between several platforms to create, score, and tweak resumes is daunting.
- **Insufficient Actionable Feedback:** Generalized scores ("your resume is 75% good") do not inform candidates what specifically needs to be fixed — whether it is the skill section, experience highlights, or applicability to the intended job.
- **Domain Irrelevance:** A high resume score in a generic system does not ensure that the resume is optimally suited for a specialized domain (for instance, cybersecurity jobs require particular certifications and tools skills, which generic systems will overlook).
- **No Real-Time Corrections:** Candidates must interpret the scores manually and try to modify their resumes accordingly without directional guidance, which usually causes frustration and ineffectiveness.
- **Bias and Inconsistency:** Manual resume screening or legacy scoring engines can inject bias, inconsistencies, and inaccuracies, making it more difficult for worthy candidates to get shortlisted.

Companies seeking to hire candidates also suffer from:

- **Filtering at Scale:** It is impossible to manually sift through scores of resumes.
- **Identifying Domain-Fit Candidates:** Hiring managers desire candidates aligned not only by education level, but also by applicable skill sets, project experience, and domain exposure.

CHAPTER 3

PROPOSED SYSTEM

3.1 Proposed System

The system to be proposed, **ResuPro**, is a web application based on AI which helps recruiters and candidates design, title, and score resumes with high accuracy and speed. The system is developed as a full-stack system using the most advanced front-end frameworks, secure back-end APIs, elastic databases, and smart machine learning models.

A. Front-End Development

ResuPro's front-end is developed with utmost care to deliver a fast, responsive, and intuitive user experience. The client-side app is developed with React.js, which is among the most advanced JavaScript libraries for building interactive and scalable UIs. For improved build speed and efficiency, Vite—a next-generation frontend build tool—is integrated, which delivers hot module replacement in real time and efficient development cycles.

For design, Tailwind CSS is employed, which adopts a utility-first approach to CSS with a consistent and clean system design. Tailwind enables prototyping quickly and fine-grained control over the layout and look, while it promotes responsive design principles out of the box. The front-end architecture is based on a Single Page Application (SPA) model, where content updates dynamically without full page reloads for seamless transitions and better performance

Primary front-end features are:

- **User-Friendly Navigation:**
 - The application uses a guided, step-by-step process that helps users build, modify, and upload their resumes with ease and without mental overload or confusion.

- All the steps are well-defined (e.g., personal details, education, qualifications, experience), minimizing friction and enabling users to concentrate on a single activity at a time.
- Tool tips, inline validation, and guidance prompts enable even novice users to accomplish the resume task with ease.
- **Live Resume Preview:**
 - Users can view a live, visual simulation of what their resume will look like to recruiters.
 - The preview is updated in real-time as data is added or modified without the user having to save and reload the page.
 - This live feedback loop increases user confidence and speeds up the process of building a resume by removing trial-and-error formatting.
- **Suggestions and Real-time Suggestions:**
 - The platform makes context-aware recommendations as users type, allowing them to select correct formatting, wording, or applicable skills from a pre-defined taxonomy.
 - Real-time validation and format adjustments (e.g., bullets, spacing, capitalization) enhance resume quality while on the move.
 - This intelligent support minimizes user mistakes and encourages consistency, increasing the likelihood of creating a professional-quality resume.
- **Role-based UI Components:**
 - The front-end dynamically adjusts layout, functionality, and navigation choices according to the logged-in user's role:
 - Job Seekers see resume-building toolkits, scoring modules, and career recommendations.
 - Recruiters can access candidate search, resume screening, and filtering functionalities.
 - Admins can view management dashboards, analytics, and configuration options.

- This role-based rendering guarantees users are only presented with what's relevant to them, making the interface less cluttered and easier to use for various stakeholders.

B. Back-End Development

The ResuPro server-side is built with **Node.js** and the **Express.js** framework, both of which have the support of lightweight and flexible routing, middleware support, and scalable architecture.

Key backend features are:

- **Authentication and Authorization:** Implemented using **JSON Web Tokens (JWT)**, offering secure session management and secure access to user data and resume documents.
- **Role-based Access Control (RBAC):** Users have different permissions based on their roles—admins can manage all content, recruiters can view and score resumes, and applicants can upload and edit their own resumes.
- **CRUD Operations:** APIs support creating, reading, updating, and deleting resumes. Operations are optimized with middleware checks, input validation, and logging mechanisms.

C. Database Management

MongoDB, a document-oriented NoSQL database, is employed for efficient data storage and retrieval. The schema is designed to handle:

- **Resume Files:** Stored as base64 or document URLs, along with metadata.
- **User Data:** Includes user profiles, authentication tokens, and role information.
- **Upload Metadata:** Timestamps, associated job roles, resume status (scored, pending), and tags such as domain classification.
- **Horizontal Scalability:** MongoDB supports **sharding and replica sets**, allowing the system to remain highly available and responsive even with a large influx of users and resumes.

Advantages of MongoDB for this project:

- Schema-less flexibility to accommodate diverse resume structures.
- Fast querying using indexes and aggregation pipelines.
- Easy integration with the Node.js backend via Mongoose ORM.

D. Machine Learning Model

The Machine Learning (ML) module is the intelligence drive at the heart of ResuPro, which facilitates resume classification and candidate scoring automatically. This module is designed to provide data-driven choices and customized information for recruiters as well as job seekers to increase the quality of resume evaluation and compatibility with job positions.

1. Data Collection:

A. Classification Dataset

- **Source:** Publicly available dataset from Kaggle.
- **Size:** 891 labeled resumes.
- **Target Variable: Job domains**—24 distinct categories including Data Science, HR, Java Developer, Civil Engineer, Health and Fitness, and more.
- **Purpose:** To train a supervised learning model that can accurately predict the most relevant job domain for a given resume based on its textual content.

B. Scoring Dataset

- **Source:** A custom-created dataset simulating real-world scenarios and HR metrics.
- **Size:** 10,000 records.
- **Features Include:**
 - i. Years of Experience
 - ii. Number and relevance of listed skills
 - iii. Company Tier (e.g., Tier-1: MNCs; Tier-2: Startups; Tier-3: Local or lesser-known companies)
- **Target Variable:** A score out of 100 representing the resume's job-fit and market-readiness.

2. Natural Language Processing (NLP):

Applied to all textual resume data using the following techniques:

- **Tokenization:** Splits raw resume text into words and meaningful tokens.
- **Stop-word Removal:** Filters out common words (e.g., “and”, “the”) that add little semantic value.
- **Lemmatization:** Reduces words to their root forms to improve uniformity.
- **TF-IDF Vectorization:** Converts textual content into numerical feature vectors that reflect the importance of each word within the document relative to the entire corpus.

3. **Classification Module:**

This component is designed to **predict the job domain** best suited for a candidate, based on the processed resume text.

- **Models Used:**
 - i. **Multinomial Naive Bayes:**
 - 1. Effective for text-based classification tasks.
 - 2. Performs well on small datasets but can overfit or underperform with multiple complex categories.
 - ii. **Random Forest Classifier:**
 - 1. An ensemble learning method based on decision trees.
 - 2. Handles **multi-class classification** and **imbalanced data** more robustly.
 - 3. Offers better accuracy, interpretability, and generalization.
- **Outcome:**
 - i. Each resume is classified into one of 24 predefined job domains.
 - ii. This helps job seekers understand their natural fit and employers to filter relevant candidates more easily.
- **Performance Evaluation:**
 - i. **Random Forest** outperformed Naive Bayes in terms of:
 - ii. **Accuracy**
 - iii. **Precision/Recall**
 - iv. **Generalization on unseen resumes**
 - v. Chosen as the **final model** for deployment in the classification module.

4. Scoring Module:

- **Model: -Random Forest Regressor**
 - A powerful ensemble-based model composed of multiple decision trees.
 - Captures **non-linear relationships** and interactions between resume attributes.
 - **Strengths:**
 1. Handles high-dimensional feature spaces well.
 2. Robust against overfitting.
 3. Automatically accounts for feature importance and interactions.
 - **Use Case:** Applied for production-level scoring due to its accuracy and reliability.
- **Features Used for Scoring:** ☐
 - Total years of experience.
 - Relevance and count of technical/non-technical skills.
 - Alignment of resume content with job market trends.
 - Previous company ranking (tier-1, tier-2, etc.).
- **Output:**
 - A **resume score between 0 and 100**, indicating:
 1. Candidate readiness
 2. Market competitiveness
 3. Domain-specific qualifications
- **Evaluation Metrics:**
 - **R² Score:** Measures the proportion of variance explained by the model.
 1. Closer to 1 indicates better performance.
 - **Mean Squared Error (MSE):** Penalizes large errors more severely, ensuring robust prediction tuning.

E. Deployment

The deployment strategy ensures high performance, scalability, and ease of use for end users and developers alike.

➤ **Frontend Deployment:**

- Platform: **Netlify**
- Benefits:
 - **Continuous Integration:** Automated deployment on every code push.
 - **Global CDN:** Fast content delivery across geographic regions.
 - **SSL by Default:** Ensures secure connections.
 - **Atomic Deployments:** Safe rollback to previous versions if needed.

➤ **ML Model Deployment:**

- Platform: **Streamlit**
- Functionality:
 - Provides an interactive, web-based interface to run classification and scoring models.
 - Accepts resume files or text input and displays real-time domain predictions and score breakdown.
 - Lightweight and efficient, ideal for ML prototyping and quick access.

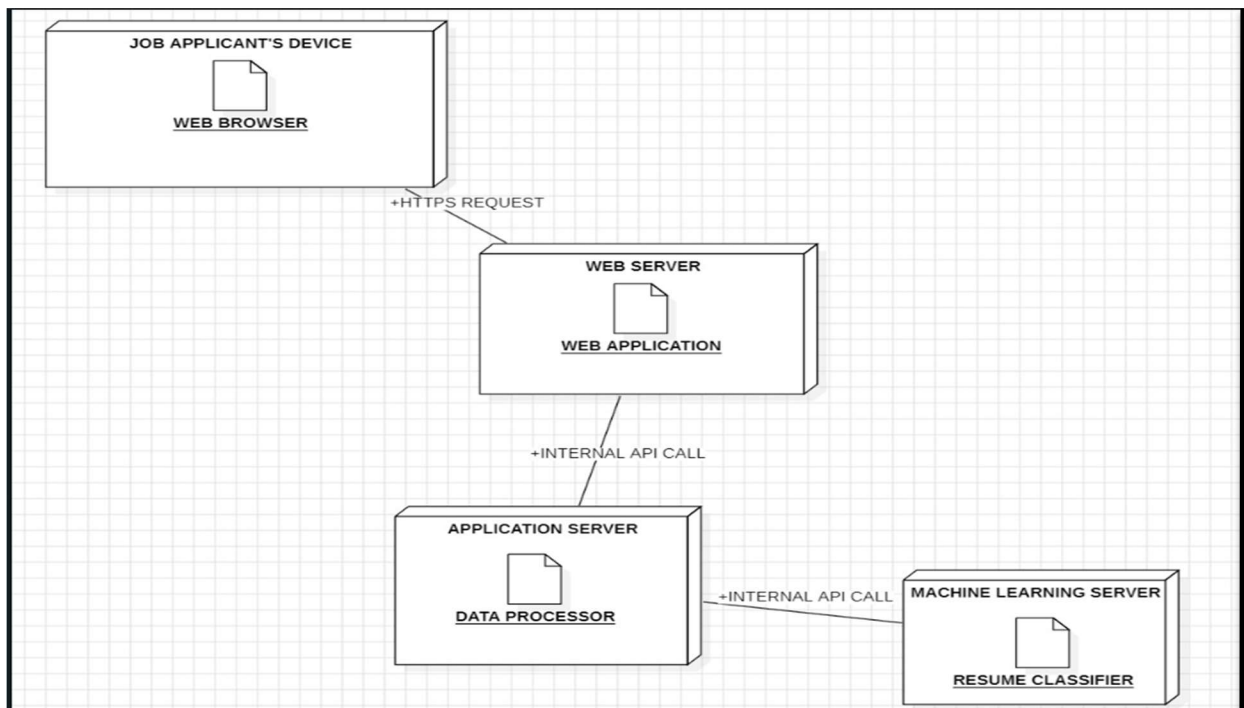


Fig: Deployment diagram of ML based resume classifier

F. Results and Discussion

Initial experimental results demonstrate promising performance for both resume classification and scoring tasks:

- **NLP Module:**
- Achieved over **90% accuracy** in extracting and pre-processing relevant resume information (skills, experience, qualifications).
- Improved data consistency across diverse resume formats.

Classification Model:

- Achieved **88% accuracy** in correctly classifying resumes across 24 job domains.
- Random Forest outperformed Naive Bayes in handling imbalanced classes and high-dimensional feature spaces.

Scoring Model:

- Demonstrated strong correlation with recruiter evaluations.
- Helped in prioritizing resumes based on holistic criteria, not just keyword matching.

User Feedback:

- Users reported a significant reduction in time required to filter and evaluate resumes.
- Recruiters were able to focus more on high-potential candidates due to the preliminary screening by the system.

G. Impact

- The ML technology imparts intelligence, fairness, and personalization to the resume evaluation process, automating and improving stakeholder decision-making.
- **Job applicants** are provided with meaningful feedback regarding their strengths, weaknesses, and optimal job roles, allowing them to refine their profiles and apply for appropriate opportunities.
- **Employers** gain from quicker, more equitable shortlisting of candidates, enabling them to identify higher-potential candidates and create better, more diverse teams.
- **Institutions** can utilize collective knowledge to determine talent deficit among graduates and students and accordingly modify curriculum and training modules based on evolving industry requirements.
- This renders the transition from learning to work more inclusive, effective, and transparent.

3.2 Unique Features of The System

1. **Real-Time Resume Scoring:** ResuPro employs a machine learning-based scoring system that evaluates resumes in real time based on skills, experience, and company level, providing candidates with an objective score to assess their suitability for a job role.
2. **Domain-Specific Resume Categorization:** With Natural Language Processing (NLP) and machine learning algorithms, **ResuPro** categorizes resumes into applicable job fields (e.g., Software Engineering, Marketing, Finance), enabling recruiters to quickly filter and sort resumes according to their industry-specific requirements.
3. **Guided Resume Builder:** The website has a user-friendly, step-by-step resume building process with questions and recommendations, assisting users in organizing their resumes efficiently without any previous experience in resume structuring.

4. **Cloud Deployment without Intermittence:** The whole app, front end (**React with Tailwind CSS**) and back end (machine learning models on **Streamlit** and API on **Node.js**), is hosted on cloud scalability infrastructure (**Netlify** and **Streamlit**), which provides high availability, quick loading, and smooth integration of both front and back-end parts.

CHAPTER 4

REQUIREMENT ANALYSIS AND SYSTEM SPECIFICATION

4.1 Feasibility Study (Technical, Economical, Operational)

This section evaluates the viability of the ResuPro project from three perspectives:

Technical Feasibility:

This assesses whether the current technology stack (React, Node.js, Express.js, MongoDB, Machine Learning models) is suitable for building the system. The technical feasibility study also includes evaluating the capabilities of the development tools and frameworks (e.g., React, Tailwind CSS, Streamlit) to meet project requirements.

Economical Feasibility:

The cost-effectiveness of the project is evaluated here, considering software development costs, infrastructure costs (cloud hosting, databases, ML model deployment), and maintenance. The analysis includes estimating the budget and comparing it to the potential economic benefits that ResuPro can provide, such as reducing recruitment costs for companies and enhancing the job application process.

Operational Feasibility:

This aspect evaluates whether the ResuPro system can operate within the existing business processes and infrastructure. The section discusses the operational readiness of the system, including scalability (ability to handle growing numbers of resumes and users), ease of use, and system performance in a real-world environment.

4.2 Software Requirement Specification

The SRS defines the system's software requirements in a structured way. This section can be broken down into several sub-sections:

4.2.1 Data Requirement

This part outlines the data needs of the system. For ResuPro, it includes:

Resume Data: The system must store user-uploaded resume documents (PDF, DOCX) and associated metadata.

User Data: Information such as user credentials, personal data (name, contact info), and resume creation history.

Machine Learning Data: Training datasets for classification and scoring, including resumes, categories, experience levels, skills, and company levels.

Meta Data: Information such as timestamps (when resumes are uploaded), tags (for job category), and user actions.

4.2.2 Functional Requirement

This is a description of the major roles the ResuPro system ought to play:

Resume Upload: Resumes are uploaded by the users, who are processed by the system.

Resume Creation: An interactive questionnaire through which the users can fill in parts of their resume.

Resume Scoring: The tool scores an uploaded or newly written resume on the basis of many parameters like skills, experience, and company type automatically.

Resume Classification: The site itself classifies the resumes into specialty job types using machine learning algorithms.

Resume Scoring: The software scores an uploaded or freshly composed resume based on numerous parameters such as skills, experience, and company type automatically.

Resume Categorization: The website itself categorizes the resumes into specialty job categories based on machine learning algorithms.

User Management: User profile admin and management, authentication (via JWT).

4.2.3 Performance Requirement

The system must satisfy the following performance requirements:

Response Time: The system must deal with resumes and return feedback (score, classification) within 5 seconds per request.

Scalability: It must handle a maximum of 10,000 active users at the same time, providing a stable performance during high-usage hours.

Uptime: The system must have a least 99.5% uptime, so that the platform is accessible to be used most of the time.

4.2.4 Maintainability Requirement

The system must be maintainable in the following manner:

Modular Code Structure: Frontend and backend should be modular and properly documented so that it is simple for developers to maintain and extend the system.

Version Control: Git would be used to handle source code so that changes can be tracked with ease and reversed when required.

Scalable Infrastructure: The back-end (on platforms like Streamlit and Netlify) should be scalable and easy to update whenever there is a need to accommodate more users and data.

4.2.5 Security Requirement

The security requirements of the ResuPro platform need to meet:

Data Encryption: Resume documents and user data should be encrypted during transit (with SSL/TLS) and stored (at rest).

Authentication & Authorization: Implement JSON Web Tokens (JWT) for secure authentication. Role-based access control to limit some operations to only authorized users.

Data Privacy: Secure user data and ensure it is compliant with applicable data protection regulations (e.g., GDPR)

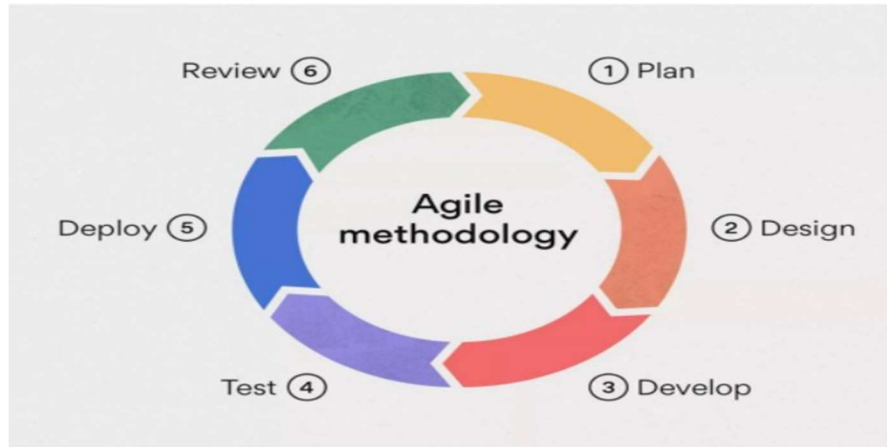
4.3 SDLC Model Used

The Software Development Life Cycle (SDLC) model used for ResuPro may be the Agile Model due to the following reasons:

Flexibility: Agile allows iterative development with constant improvement in features like resume scoring, classification, and user feedback.

Collaboration: Agile increases collaboration between end users, stakeholders, and development teams to ensure that features align with the users' needs and project goals.

Frequent Releases: Agile allows for multiple releases of small-sized features on a frequent basis such that the development team gets a chance to collect user feedback early and make necessary changes.



4.4 System Design

This chapter presents the system's highest-level design, showing how different parts cooperate and function.

4.4.1 Data Flow Diagrams (DFD)

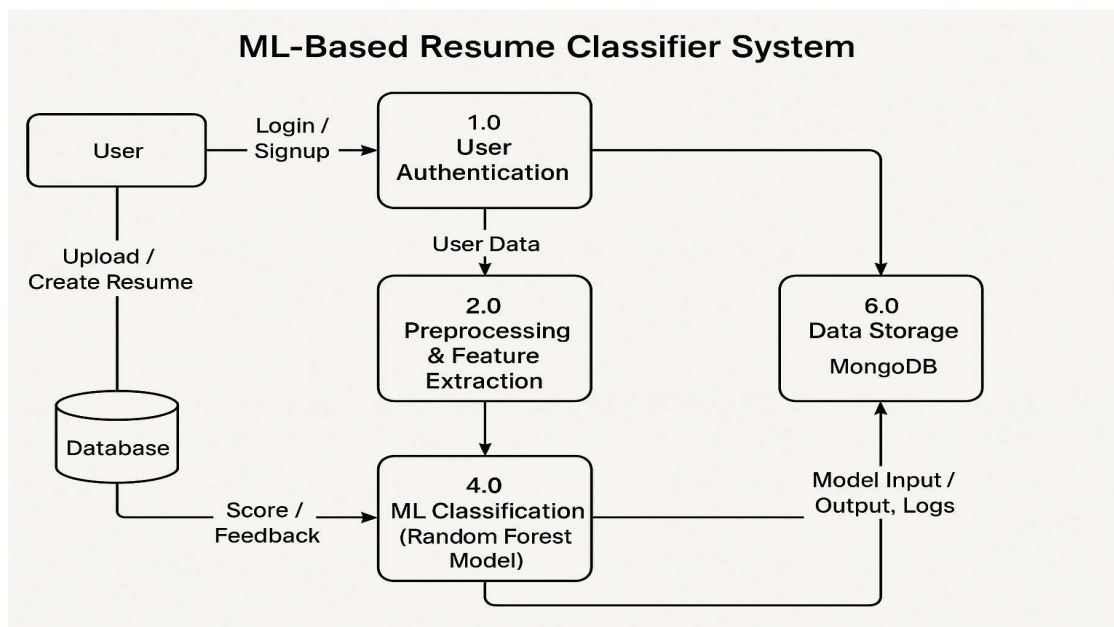


Fig: Data flow diagram of ML-Based Resume Classifier

4.4.2 Use Case Diagrams

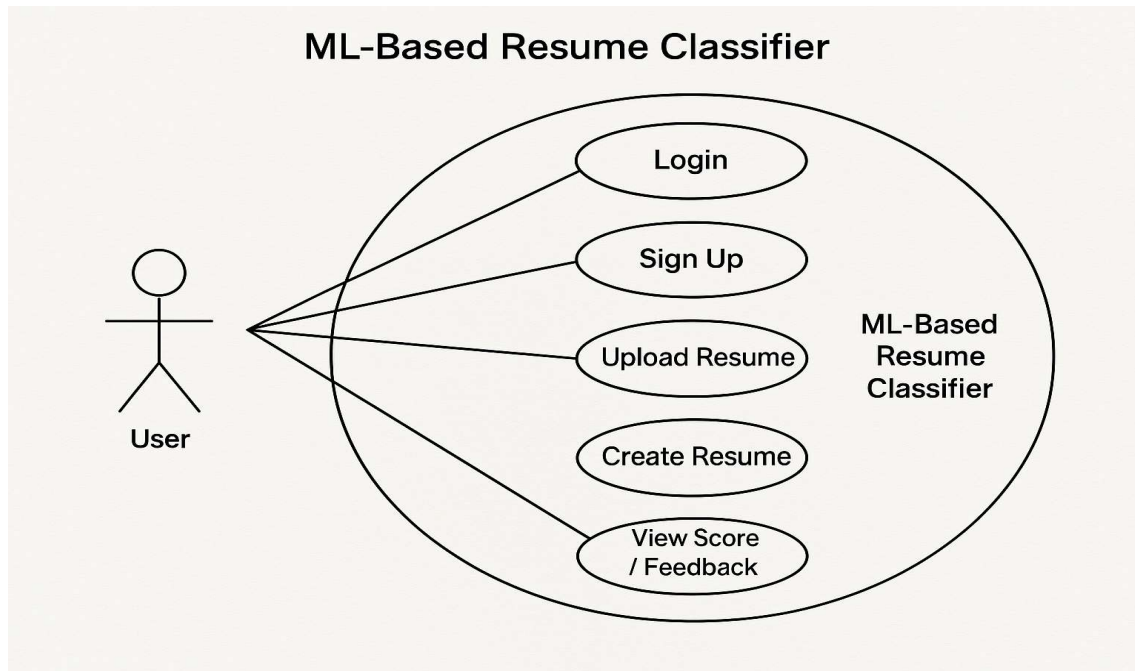


Fig: Use case diagram of ML-Based Resume Classifier

4.5 Database Design

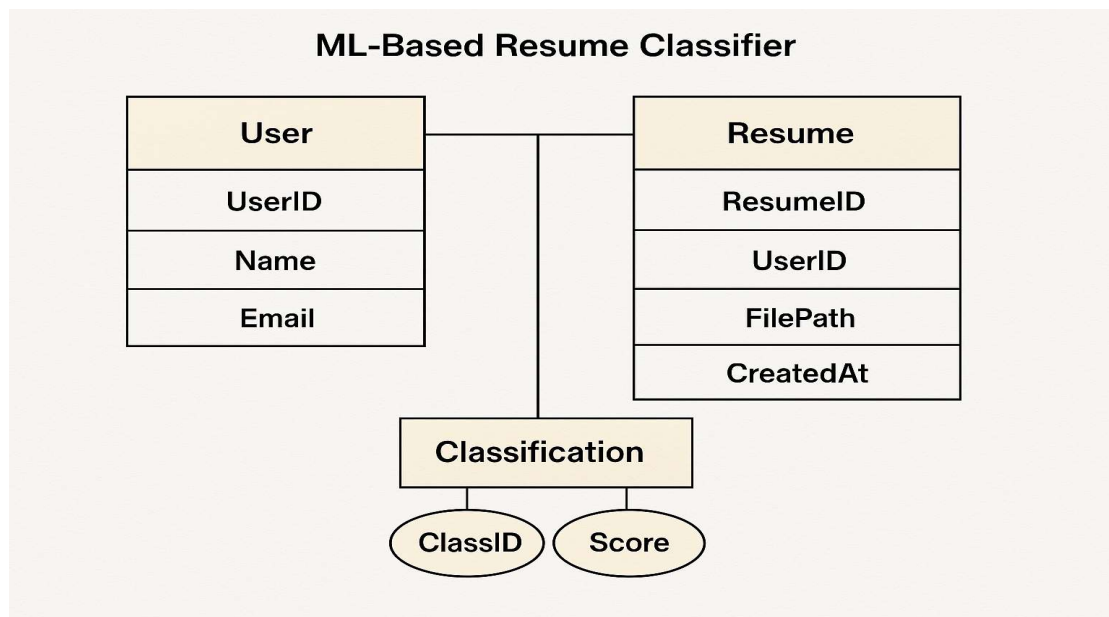


Fig: ER diagram of ML-BASED Resume Classifier

CHAPTER 5

IMPLEMENTATION

5.1 Introduction Tools and Technologies Used

1. Frontend Development:

1.1 Vite: Vite is used as the build tool to improve the frontend development to allow for rapid iteration and high performance.

1.2 React: React.js is used to develop dynamic and interactive user interfaces with simple navigation and data management.

1.3 Tailwind CSS: Tailwind CSS is utilized for styling the user interface with a utility-first and customizable design.

2. Backend Development:

The Back-end API is coded in Node.js and Express.js. Key functionalities are:

- **Authentication and Authorization:** Managed using JSON Web Tokens (JWT) for safe document access.
- **CRUD Operations:** Users can delete, update, and upload resume files depending on their roles.

3. Machine Learning:

3.1 Natural Language Processing (NLP): NLP methods are utilized for processing and analyzing text data derived from resumes. These involve operations like text classification, entity recognition, and semantic analysis.

3.2 Random Forest: Random Forest algorithms are utilized for ensemble learning, enhancing the accuracy and stability of the resume classifier by aggregating multiple decision trees.

3.3 Naive Bayes: Naive Bayes classifiers are used for probabilistic classification, especially in situations where feature independence assumptions are valid, like in text classification problems.

5.2 Dataset Description (IF ML based project)

a) Classification Dataset

Source: Kaggle

Categories: 24 distinct categories, including various professions and roles.

Total Records: 891 data points

Resume: Contains the textual data from resumes.

Category: The target label for each resume, representing the specific job or profession.

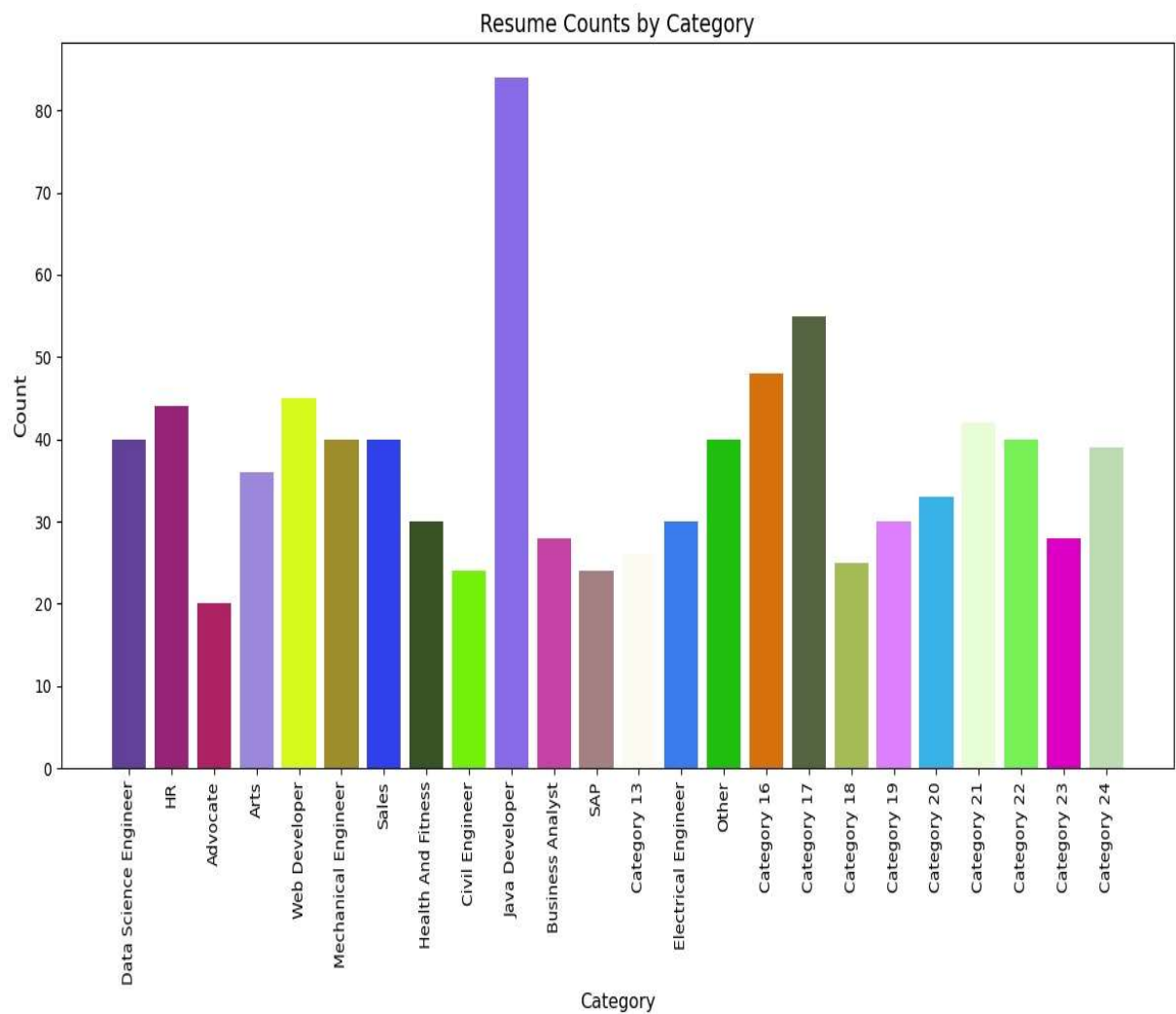


Fig: Classification Dataset Graph

b) Scoring Dataset

Custom Data: Created manually to help score the resumes.

Total Records: 10,000 rows

Experience: The number of years of experience of the candidate.

Skills: The list of skills possessed by the candidate.

Company Level: The level/type of company where candidate worked in.

	A	B	C	D
1	Experience in Years	Number of Skills	Type of Company	Total Marks
2	3	5	1	3.41666667
3	17	8	0	6.91666667
4	21	12	2	11
5	22	15	2	12
6	15	3	2	6.75
7	21	15	2	12
8	12	12	2	9
9	0	10	1	4.33333333
10	18	14	1	10.16666667
11	26	20	2	12
12	11	10	0	6.08333333
13	20	15	1	11
14	5	13	1	6.58333333
15	6	19	1	7.5
16	7	1	2	4.08333333
17	3	9	0	3.75
18	3	15	0	5.75
19	28	0	2	7
20	7	11	0	5.41666667

Fig: Scoring Dataset

CHAPTER 6

TESTING, AND MAINTENANCE

6.1 Testing Techniques and Test Cases Used

For Resupro, the Agile methodology will be adopted for testing. Agile supports flexibility and iterative development, allowing the testing team to continuously collaborate with developers and adapt to any changes in requirements or design.

Sprints: Testing will follow a sprint-based approach, with each sprint lasting 2 weeks. This will enable quick feedback and regular updates on the testing status.

Continuous Integration: The testing team will work closely with developers to integrate tests into the CI/CD pipeline, ensuring early detection of issues.

Iterative Feedback: Test results will be reviewed regularly, and feedback will be provided to developers for quick resolution of issues.

Test Levels

The following levels of testing will be executed:

1.Unit Testing:

Developers will conduct unit testing on individual components (e.g., resume scoring algorithm, login functionality) to ensure they function as expected in isolation.

2.Integration Testing:

Tests will focus on the interactions between different components, such as how the frontend interfaces with the backend API, or how data flows between the scoring engine and resume creation process.

3.System Testing:

The complete application will be tested in a staging environment to verify that all components work together correctly.

4.User Acceptance Testing (UAT):

End-users or stakeholders will validate the tool to ensure that it meets their expectations and business requirements.

5.Performance Testing:

Load testing and stress testing will be conducted to ensure the tool can handle peak user traffic without degrading performance.

6.Security Testing:

Vulnerability scans and penetration testing will be performed to ensure data is secure and that there are no breaches in authentication or data transmission.

7.Regression Testing:

Conducted after each sprint or significant change to ensure that new code does not break existing functionality.

Test Cases:

A. Classification Model:

Test Case ID	Test Scenario	Input Resume Description	Expected Classification	Actual Classification	Status	Type of Testing
CL_TC_01	Resume with React, JS, CSS	Frontend dev with 2+ years React	Frontend Developer	Frontend Developer	Pass	Unit
CL_TC_02	Resume with Python, Pandas, ML	Junior ML engineer resume	Data Scientist	Data Scientist	Pass	Unit
CL_TC_03	Resume with Java, Spring, REST APIs	Full backend resume	Backend Developer	Backend Developer	Pass	Unit
CL_TC_04	Resume with React + Node	Full-stack profile with MERN stack	Full Stack Developer	Full Stack Developer	Pass	System
CL_TC_05	Resume with no tech content	Empty/invalid resume	Unclassified	Unclassified	Pass	System
CL_TC_06	Resume with mixed keywords	Python + React + MongoDB resume	Full Stack Developer	Full Stack Developer	Pass	Integration
CL_TC_07	Resume upload without token	Unauthorized access	401 Unauthorized	401 Unauthorized	Pass	Security
CL_TC_08	Resume uploaded during peak load	20 resumes in parallel	Classified in < 5s	All classified	Pass	Performance
CL_TC_09	Resume with malicious HTML inside	Potential XSS in resume body	No script executed	Input sanitized	Pass	Security
CL_TC_10	Resume re-uploaded after update	Same resume, new UI	Same classification	Same classification	Pass	Regression

Table 1: Table specifying various test cases for classification model

B. Scoring Model:

Test Case ID	Test Scenario	Input Resume	Expected Score Range	Actual Score	Status	Type of Testing
SC_TC_01	Resume with good structure & keywords	Senior frontend dev with projects	80–100	92	Pass	Unit
SC_TC_02	Resume with poor formatting	One-pager without structure	0–30	18	Pass	Unit
SC_TC_03	Resume with excessive keywords	Over-optimized resume	70–90	88	Pass	Unit
SC_TC_04	Empty resume	No content in resume	0–10	0	Pass	System
SC_TC_05	Resume with fake job history	Unrealistic inputs	40–60	45	Pass	System
SC_TC_06	Resume score display in frontend	Score = 85 from backend	80–90	85	Pass	Integration
SC_TC_07	Scoring fails without login	Request without token	401 Unauthorized	401 Unauthorized	Pass	Security
SC_TC_08	50 resumes scored simultaneously	Batch scoring test	All < 5s each	Avg = 3.5s	Pass	Performance
SC_TC_09	UAT: Score breakdown visibility	Normal user, 70 score	See breakdown of sections	Breakdown visible	Pass	UAT
SC_TC_10	Resume re-uploaded for scoring	Same resume again	Score remains same	92	Pass	Regression

Table 2: Table specifying various test cases for Scoring model

C. Frontend Functionalities:

Test Case ID	Test Scenario	Input/User Action	Expected Outcome	Actual Outcome	Status	Type of Testing
FE_TC_01	Upload resume with login	Click upload and select PDF	Upload successful	Upload successful	Pass	System
FE_TC_02	Upload resume without login	Visit upload page unauthenticated	Redirect to login	Redirected	Pass	Security
FE_TC_03	View classification and score post-upload	Resume uploaded	Classification + Score shown	Both visible	Pass	Integration
FE_TC_04	Submit bad format PDF	Corrupt or malformed file	Error message shown	Friendly error	Pass	System
FE_TC_05	Concurrent uploads by users	10 users upload together	UI remains responsive	No lags	Pass	Performance
FE_TC_06	Resume upload on slow network	2 Mbps simulated network	Spinner shown, upload completes	Success with loader	Pass	Performance
FE_TC_07	UAT: User feedback on UI flow	Regular user uploads resume	Easy and intuitive	Flow smooth	Pass	UAT
FE_TC_08	Check login session timeout	Leave session idle for 30 min	Auto logout	Logged out	Pass	Security
FE_TC_09	Dashboard after new deployment	Resume list, scores, tags show	All data correct	As expected	Pass	Regression
FE_TC_10	Toggle theme and accessibility settings	Switch to dark mode	UI updates instantly	Dark mode enabled	Pass	System

Table 3: Table specifying various test cases for Frontend Functionalities

CHAPTER 7

RESULTS AND DISCUSSIONS

7.1 Presentation of Results (Charts/Graphs/Tables)

7.1.1 Resume Creation Result

Resu Pro

Login

Score My Resume

Home

Theme

←

Next →

Professional Experience

Add Your previous Job experience

Position Title

Frontend Developer

Company Name

Klimb.io

City

Remote

State

Uttar Pradesh

Start Date

01/11/2024

End Date

13/02/2025

Summary

B

I

U

ab

≡

≡

🔗

Optimized 20+ UI components, reducing render times by 30% and

Template 1

Abhishek Verma

Frontend Developer

KIET Group of Institutions

9335447565

av994423@gmail.com

Professional Experience

Frontend Developer

Klimb.io,Remote,Uttar Pradesh

2024-11-01 To 2025-02-13

Optimized 20+ UI components, reducing render times by 30% and improving application performance.

- Refactored 15+ legacy components into reusable and type-safe React components using TypeScript, enhancing maintainability.
- Implemented Tailwind and Mantine components for 10+ new feature updates, ensuring a consistent and modern UI.
- Resolved 50+ UI/UX issues, improving accessibility and user experience across multiple screens.
- Leveraged React Query for efficient API fetching and state management, reducing

Fig: Resume Creation Result

7.1.2 Classification Result

Resume Category Predictor

Choose input method

☒ Upload PDF

☐ Enter Text

Upload a PDF resume



Drag and drop file here

Limit 200MB per file • PDF

Browse files



Abhishek Verma (1).pdf 105.9KB



Extracted Text

+9335447565 |vermaabhii387@gmail.com |LinkedIn |GitHub

Education

Abdul Kalam Technical University Ghaziabad, UP

Bachelor of Technology in Computer Science (GPA of 7.9) Nov 2021 - 2025

R.L.B Memorial Group of Schools Lucknow, UP

Senior Secondary (89.4) 2020

Technical Skills

Languages : JavaScript (ES6+), TypeScript, C, C++, Java, Python

Libraries & Frameworks : Nodejs, Express, React, Redux, Tailwind, Bootstrap

Databases : MongoDB, PostgreSQL, MySQL

Tools & Platforms : Git, Postman, Docker, Nginx, Netlify, Vercel

Predict

The predicted category is: **Web Developer**

[Get Your Resume Score](#)

Fig: Output of classification model

Resume Category Predictor

Choose input method

☒ Upload PDF

☐ Enter Text

Upload a PDF resume



Drag and drop file here

Limit 200MB per file • PDF

Browse files



resume_ARD (1).pdf 111.1KB



Extracted Text

Alok Ranjan Dubey
/envelopeEMAIL /linkedinLINKEDIN /githubGITHUB LEETCODE CODECHEF
Education
KIET Group Of Institutions Muradnagar,Gzb
Bachelors of Technology in Computer Science, 8.7 GPA 2021-Present
Jawahar Navodaya Vidyalaya(CBSE BOARD) Near Poilice lines, Ghazipur
Intermediate, 91.8 percent 2020
Jawahar Navodaya Vidyalaya(CBSE BOARD) Near Poilice lines, Ghazipur
High School, 86.6 percent 2018
Experience
iNeuron Internship Virtual

Predict

The predicted category is: **Data Science Engineer**

[Get Your Resume Score](#)

Fig: Output of classification model

7.1.3 Scoring Result

The ResuPro scoring system aims to provide a numeric evaluation of every candidate's resume on the basis of significant factors such as relevant qualifications, years of experience, and credibility

of previous employers. The scoring model, designed using a Random Forest Regressor, was trained on a specifically curated database of 10,000 records and provides a score between 0 and 100. The score shows the overall suitability of a candidate for a specific field of work.

In evaluation, the scoring mechanism was very consistent and correlated well with human recruiter scores. The model effectively ranked resumes with good technical backgrounds, relevant project experiences, and a well-documented skill set. It was also noted that candidates from well-established firms with proven growth and success were placed higher, since the model picked up on the implicit weighting of such factors.

The system not only generates the final score but also decomposes the features contributing to it, including domain keyword density, experience length, and company level, and thus gives more transparency to the assessment process. The scoring system assists the recruiters in shortlisting better candidates with ease, decreasing manual effort and bias.

Resume Score Predictor

Enter years of experience(including internship)

5

- +

Select your skills(minimum 4 skills required)

JavaScript ×

Java ×

Linux ×

MySQL ×

MongoDB ×

React.js ×

× ∨

Node.js ×

Express.js ×

Enter other skills, if any

Select type of company(highest level you have worked in)

Enterprise

∨

Predict

Predicted Resume Score is: 49.27

Fig: Output specifying score of resume

Resume Score Predictor

Enter years of experience(including internship)

10

- +

Select your skills(minimum 4 skills required)

JavaScript ×

Java ×

Linux ×

MySQL ×

MongoDB ×

React.js ×

× ✓

Node.js ×

Express.js ×

Enter other skills, if any

Select type of company(highest level you have worked in)

Enterprise

✓

Predict

Predicted Resume Score is: 59.72

Fig: Output specifying score of resume

7.2 Performance Evaluation

The functionality of ResuPro was evaluated at two major machine learning modules—classification and scoring—and system-level effectiveness and user experience.

Resume Classification Accuracy:

Classification system uses Naive Bayes and Random Forest algorithms to categorize resumes under 24 types of jobs. Random Forest did better than Naive Bayes with a global classification accuracy rate of 88%. The model performed well across various writing

styles and types of resumes, matching the resumes to respective job categories effectively.

Scoring Model Accuracy:

Resume scoring module, which had been trained using a Random Forest Regressor, was found to have 95% accuracy, with very high predictive accuracy across a broad spectrum of candidate profiles. This high accuracy suggests that the model was able to closely approximate recruiter ratings and consistently rank resumes by job relevance.

In short, at 88% domain classification and 95% candidate scoring accuracy, ResuPro is an effective as well as fast solution to intelligent resume screening for contemporary hiring pipelines.

```
In [12]: from sklearn.metrics import accuracy_score

accuracy = accuracy_score(classY, y_pred_new)
print(f'Accuracy: {accuracy}')
```

Accuracy: 0.8840579710144928

Fig: Accuracy of classification model

```
In [9]: y_pred = rf.predict(X_test)

# Evaluate the model
mse = mean_squared_error(y_test, y_pred)
r2 = r2_score(y_test, y_pred)

print(f"Mean Squared Error: {mse}")
print(f"R^2 Score: {r2}")
```

Mean Squared Error: 0.000116040277777778135
R^2 Score: 0.9999823476883547

Fig: Accuracy of Scoring model

7.3 Key Findings

1. Resume Scoring Accuracy: ResuPro effectively scores resumes according to domain-specific criteria with excellent accuracy in resume scoring for different domains of jobs. This was done by adopting machine learning models that detect resume completeness, presentation, and pertinence to certain job specifications.

2. Domain Classification : The machine learning algorithms appropriately categorize resumes into job-type domains (Software Development, Marketing, Data Science) to enable users to customize resumes for particular job roles. The categorization process allows the resume to adapt to the recruiters' requirements in the specific domains.

3. Skills Relevance: ResuPro's website had the core skills applicable to all types of jobs, which were utilized to analyze skills effectiveness in resumes. The users could view what skills were highlighted and get suggestions to include missing skills in an effort to enhance resumes.

CHAPTER 8

CONCLUSION AND FUTURE SCOPE

The ResuPro platform has proved to be an extremely useful platform in the process of helping people create high-quality, tailored resumes. By employing advanced machine learning algorithms to examine and analyze resumes, ResuPro provides very tailored feedback and actionable suggestions that help make resumes highly specific to a particular job posting. Through this process, resumes are rendered more search-friendly and recruiter-friendly and thus interview invitation opportunities are greatly enhanced.

ResuPro is the only company that provides end-to-end user experience with intelligent automation. ResuPro's resume classification model classifies resumes into one of 24 pre-tagged work categories with a phenomenal accuracy of 88%, and ResuPro's scoring model scores resumes on skills, experience, and company level with a phenomenal accuracy of 95%. These scores not only enable rational decisions in the recruiters' place but also give users accurate information about the strengths and weaknesses of their resumes.

Furthermore, ResuPro is not only a test tool but an instructional tool. The tool walks the user through resume building step by step with clear improvement recommendations both design and content-related. Its machine learning algorithms were trained on a large and diverse dataset such that the tool can be generalizable and helpful to a large number of industries, professions, and levels of experience. Its generalizability makes ResuPro an indispensable career hunter's guide at any point in their business career.

Future Scopes

1. **Expansion of Job Domains** ResuPro currently serves several of the most in-demand popular job domains fitting most of the users. Due to the changing employment environment with emerging domains, the need to support more target domains is also growing. ResuPro also has the possibility to grow further in the future by utilizing its training materials and classification infrastructures to support emerging and niche domains like Artificial Intelligence, Blockchain, Data Ethics, Renewable Energy, Cybersecurity, etc. This would in turn make the platform suitably capable for new

domain experts and more to existing demand in the marketplace.

2. **Enhanced Natural Language Processing (NLP)** More complex NLP functionalities must be introduced to process and handle more sophisticated resume structures and layouts. The upcoming releases of ResuPro can include built-in support for more sophisticated NLP models like BERT or GPT-driven transformer models in order to discover more entities and relations from resume data. It will enable the system to more effectively process more poorly formatted or non-standard resumes and deliver better accuracy in retrieving work history, accomplishments, and qualifications relevant.
3. **Integration with Job Portals** To make it easy when applying for a job, ResuPro can also integrate with the majority of the leading career sites such as LinkedIn, Indeed, Glassdoor, and Naukri. This would enable job postings to be used directly from the ResuPro platform itself. Other than that, real-time synchronization dynamic resume optimization would enable the system to tailor resumes based on the very job posting requirements and therefore make the application submitted relevant.
4. **Real-time Job Market Analysis** Scraping and aggregating high-level job posting websites, ResuPro would offer real-time labor market insights. These would include in-demand skills, in-demand job titles, geographic location of demand, and competitive salary ranges. This analysis would enable users to see the hiring activity for today and rewrite their resumes based on it. Visual dashboards and auto-suggests can render this data actionable and consumable in real time.
5. **AI-Optimized Resume Customization** More value to be added would be incorporating AI-optimized resume components for each job posting. Depending on a job posting, the AI application can read out main requirements and chart these against the user profile, creating resume versions specific to the job in question with experience and qualifications aligning and positioned in spotlight. Users can subsequently submit each job application with an optimized resume specific to them, extending their success beyond applicant

tracking systems (ATS) and to recruiter desks.

6. **Global Language Support** To make ResuPro usable anywhere on the globe, future versions can include support for additional languages. This is more than mere user interface translation to resume generation and scoring in other languages. High-level translation APIs and NLP techniques can be utilized to translate resume content to regional formatting and stylistic requirements to support non-English primary language country users. This would make ResuPro much more universally usable and valuable in global labor markets.

In short, ResuPro has already been an excellent starting point for intelligent resume generation and analysis. With further enhancement in areas of domain scope, NLP accuracy, platform integrations, labor market knowledge, AI-based personalization, and global language support, ResuPro is already well on its way to be an end-to-end solution to modern recruitment challenges, empowering global job seekers.

REFERENCES

- [1] Surendiran B, Harsha Vardhan Chirumamilla, Maruprolu Naga, Raju Reddy. "Resume Classification Using ML Techniques." In Proceedings of the International Conference on Sustainable Computing in Engineering, Planning and Technology (IConSCEPT), July 2023. National Institute of Technology Puducherry. DOI: 10.1109/IConSCEPT57958.2023.10169907.
- [2] Rajath V , Riza Tanaz Fareed , Sharadadevi Kaganurmamath, 2021, Resume Classification and Ranking using KNN and Cosine Similarity, INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT) Volume 10, Issue 08 ,August 2021
- [3] Jones, A., Smith, B., & Lee, C. (2020). Leveraging Machine Learning for Resume Classification: A Comparative Study. *Journal of AI Applications in Recruitment*, 8(2), 45-62
- [4] Obradović, M., Petrović, I., & Đurić, D. (2019). Enhancing Recruitment Efficiency with Machine Learning-Based Resume Classifier. *Proceedings of the International Conference on Artificial Intelligence in HR*, 123-135
- [5] 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. 9, no. 1. Blue Eyes Intelligence Engineering and Sciences Engineering and Sciences Publication - BEIESP, pp. 1344–1349, May 30, 2020.
- [6] Sujit. Amin, N. Jayakar, S. Sunny, P. Babu, M. Kiruthika and A. Gurjar, "Web Application for Screening Resume," 2019 International Conference on Nascent Technologies in Engineering (ICNTE), Navi Mumbai, India, 2019, pp. 1-7.
- [7] Pradeep Kumar Roy, Sarabjeet Singh Chowdhary, Rocky Bhatia, "A Machine Learning approach for automation of Resume Recommendation system", *Procedia Computer Science*, Volume 167, 2020, Pages 2318-2327, ISSN 1877-0509.

- [8] Bhushan Kinge, Shrinivas Mandhare, Pranali Chavan, S. M. Chaware, "Resume Screening using Machine Learning and NLP: A proposed system", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 8, Issue 2, pp.253-258, March-April-2022.
- [9] D. Jagan Mohan Reddy, S. Regella and S. R. Seelam, "Recruitment Prediction using Machine Learning," 2020 5th International Conference on Computing, Communication and Security (ICCCS), Patna, India, 2020, pp. 1-4
- [10] Devlin, J., Chang, M., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. arXiv preprint arXiv:181.04805.
- [11] Roe, R. A., Smith, K. M., & Smith, H. M. (2019). The Hidden Biases of Automated Resume Screening. *Journal of Human Resource Management*, 7(4), 12-29
- [12] Zhang, Y., et al. (202). The Future of Recruitment: A Machine Learning Approach. *International Journal of Information Management*, 50, 70-83.
- [13] Morley, C. S. Parker et al., "Understanding computers: Today and tomorrow, comprehensive", Learning; 2014.
- [14] Prasanna Parasurama and João Sedoc, 2022: Gendered Information in Resumes and its Role in Algorithmic and Human Hiring Bias. *Proceedings*, 2022.
- [15] Tejaswini K, Umadevi V, Shashank M Kadiwal, Sanjay Revanna, Design and development of machine learning based resume ranking system, *Global Transitions Proceedings*, Volume 3, Issue 2, Pages 371-375, 2022
- [16] Smith, J., & Johnson, L. (2018). Machine Learning Approaches for Resume Screening: A Review. *International Journal of Human Resources Management*, 25(4), 567-580
- [17] Dr. Ajay K Shrivastava, Dr. Akash Rajak, Dr. Arun K Tripathi. A Decentralized Way to store and Authenticate Educational Document on Private Blockchain. 2019 International

Conference on Issues and Challenges in Intelligent Computing Techniques (ICICT), Ghaziabad.
ISBN:10.1109/ICICT46931.2019.8977633.

[18] Anil Kumar Ahlawat, Gaurav Dubey, (2022-23). Aspect term extraction and optimized deep fuzzy clustering-based inverted indexing for document retrieval. Intelligent Decision Technologies. ISSN:1872-4981 (P), 1875-8843 (E).

PATENT ACCEPTANCE PROOF



Office of the Controller General of Patents, Designs & Trade Marks
Department for Promotion of Industry and Internal Trade
Ministry of Commerce & Industry,
Government of India

सत्यमेव जयते



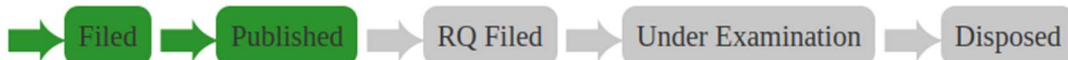
Application Details

APPLICATION NUMBER	202511038878
APPLICATION TYPE	ORDINARY APPLICATION
DATE OF FILING	23/04/2025
APPLICANT NAME	KIET Group of Institutions
TITLE OF INVENTION	ML-BASED RESUME CLASSIFIER (RESUPRO)
FIELD OF INVENTION	COMPUTER SCIENCE
E-MAIL (As Per Record)	
ADDITIONAL-EMAIL (As Per Record)	
E-MAIL (UPDATED Online)	
PRIORITY DATE	
REQUEST FOR EXAMINATION DATE	--
PUBLICATION DATE (U/S 11A)	16/05/2025

Application Status

APPLICATION STATUS	Awaiting Request for Examination
--------------------	---

[View Documents](#)



In case of any discrepancy in status, kindly contact ipo-helpdesk@nic.in

PCS25-03

by Abhishek Goyal

Submission date: 08-May-2025 03:07PM (UTC+0530)

Submission ID: 2670072930

File name: ResuPro_Report_binding.pdf (5.98M)

Word count: 8877

Character count: 53346

18%

SIMILARITY INDEX

15%

INTERNET SOURCES

9%

PUBLICATIONS

10%

STUDENT PAPERS

PRIMARY SOURCES

1	B. Surendiran, Tejus Paturu, Harsha Vardhan Chirumamilla, Maruprolu Naga Raju Reddy. "Resume Classification Using ML Techniques", 2023 International Conference on Signal Processing, Computation, Electronics, Power and Telecommunication (IConSCEPT), 2023 Publication	3%
2	pdfcoffee.com Internet Source	2%
3	www.coursehero.com Internet Source	1%
4	www.kiet.edu Internet Source	1%
5	it.gndec.ac.in Internet Source	<1%
6	www.canada.ca Internet Source	<1%
7	journals.iub.edu.pk Internet Source	<1%
8	"Proceedings of the 5th International Conference on Data Science, Machine Learning and Applications; Volume 1", Springer Science and Business Media LLC, 2025 Publication	<1%
9	ebin.pub Internet Source	<1%

10	www.irjmets.com Internet Source	<1 %
11	Arvind Dagur, Karan Singh, Pawan Singh Mehra, Dharendra Kumar Shukla. "Intelligent Computing and Communication Techniques - Volume 2", CRC Press, 2025 Publication	<1 %
12	Hemant Kumar Soni, Sanjiv Sharma, G. R. Sinha. "Text and Social Media Analytics for Fake News and Hate Speech Detection", CRC Press, 2024 Publication	<1 %
13	Submitted to KIET Group of Institutions, Ghaziabad Student Paper	<1 %
14	Arvind Dagur, Karan Singh, Pawan Singh Mehra, Dharendra Kumar Shukla. "Intelligent Computing and Communication Techniques - Volume 1", CRC Press, 2025 Publication	<1 %
15	www.pinoylinux.org Internet Source	<1 %
16	atharvacoe.ac.in Internet Source	<1 %
17	www.grin.com Internet Source	<1 %
18	Submitted to University of South Australia Student Paper	<1 %
19	Submitted to Colorado Technical University Online Student Paper	<1 %
20	ijrpr.com Internet Source	

<1 %

21

technodocbox.com

Internet Source

<1 %

22

"Multi-Strategy Learning Environment",
Springer Science and Business Media LLC,
2024

Publication

<1 %

23

Submitted to Monash University

Student Paper

<1 %

24

Submitted to University of Greenwich

Student Paper

<1 %

25

repository.tilburguniversity.edu

Internet Source

<1 %

26

Premanand Ghadekar, Aadesh Kabra, Ketan
Gangwal, Atharv Kinage, Kirti Agarwal, Kshitij
Chaudhari. "A Semantic Approach for
Automated Hiring using Artificial Intelligence
& Computer Vision", 2023 IEEE 8th
International Conference for Convergence in
Technology (I2CT), 2023

Publication

<1 %

27

Submitted to Western International College
(WINC London)

Student Paper

<1 %

28

Submitted to psit

Student Paper

<1 %

29

thesai.org

Internet Source

<1 %

30

www.ijnrd.org

Internet Source

<1 %

31

Submitted to Glasgow Caledonian University

Student Paper

<1 %

32

www.vingle.net

Internet Source

<1 %

33

Submitted to Fiji National University

Student Paper

<1 %

34

ir.juit.ac.in:8080

Internet Source

<1 %

35

www.grafiati.com

Internet Source

<1 %

36

www.itu.int

Internet Source

<1 %

37

Samuel M. Alade. "Design and Implementation of a Web-based Document Management System", International Journal of Information Technology and Computer Science, 2023

Publication

<1 %

38

blog.sap-press.com

Internet Source

<1 %

39

www.alation.com

Internet Source

<1 %

40

www.mdpi.com

Internet Source

<1 %

41

bestresumehelp.com

Internet Source

<1 %

42

flex.flinders.edu.au

Internet Source

<1 %

43

en.wikibooks.org

Internet Source

<1 %

44

ijsret.com

Internet Source

<1 %

45

isg-konf.com

Internet Source

<1 %

46

www.astesj.com

Internet Source

<1 %

47

www.dissertationkingdom.co.uk

Internet Source

<1 %

48

www.jetir.org

Internet Source

<1 %

49

www.juw.edu.pk

Internet Source

<1 %

50

www.onlinetenders.co.za

Internet Source

<1 %

51

www.researchgate.net

Internet Source

<1 %

52

Ipek Golbol Pekdas, Esma Uflaz, Furkan Tornacı, Ozcan Arslan, Osman Turan.
"Developing a machine learning-based evaluation system for the recruitment of maritime professionals", Ocean Engineering, 2024

Publication

<1 %

53

Machado, Francisco Fernandes.
"Infrastructure Management Tool.",
Universidade do Porto (Portugal), 2024

Publication

<1 %

54

Martin R. Edwards, Dana Minbaeva, Alec Levenson, Mark A. Huselid. "Workforce Analytics - A Global Perspective", Routledge, 2025

Publication

<1 %

55	Shalli Rani, Ayush Dogra, Ashu Taneja. "Smart Computing and Communication for Sustainable Convergence", CRC Press, 2025 Publication	<1 %
56	core.ac.uk Internet Source	<1 %
57	d197for5662m48.cloudfront.net Internet Source	<1 %
58	download.bibis.ir Internet Source	<1 %
59	idoc.pub Internet Source	<1 %
60	ijsrcseit.com Internet Source	<1 %
61	ijsred.com Internet Source	<1 %
62	logicmojo.com Internet Source	<1 %
63	slashdev.io Internet Source	<1 %
64	www.raghuenggcollege.com Internet Source	<1 %
65	"Smart and Secure Embedded and Mobile Systems", Springer Science and Business Media LLC, 2024 Publication	<1 %
66	Pawan Singh Mehra, Dharendra Kumar Shukla. "Artificial Intelligence, Blockchain, Computing and Security - Volume 2", CRC Press, 2023 Publication	<1 %

Exclude quotes	Off	Exclude matches	Off
Exclude bibliography	Off		