**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 (2100290120027)

ANURAG KUMAR (2100290120038)

Supervised by   
**ABHISHEK GOYAL   
(Assistant 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 :- 18/02/25

## **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: 18/02/25 Supervisor**

**Dr.** Abhishek Goyal

(Assistant 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 today's rapidly evolving job market, the recruitment process faces several challenges, particularly in the initial stages of resume screening. Traditional methods of manually reviewing resumes are not only time-consuming but also inefficient, leading to human biases and subjective evaluations. With the increasing volume of applications, there is a growing need for automated systems to streamline the process. This paper presents ResuPro, a machine learning (ML)-based resume classifier system designed to automate the resume screening process. By using advanced machine learning algorithms and natural language processing (NLP), ResuPro classifies resumes into predefined job categories based on content. The system extracts critical information such as skills, experience, and company type (Startup, Midlevel, Enterprise), providing recruiters with a comprehensive and accurate analysis of applicants. ResuPro classifies resumes by job domain (e.g., IT, Finance, Marketing) and employs a scoring mechanism to evaluate resumes based on their alignment with job requirements, such as experience, skills, and educational background. This scoring system enhances decision-making by ranking resumes, highlighting top candidates, and providing detailed insights into their qualifications and suitability for the job. The resumes of candidates are classified with an accuracy of 88%. By automating the process, ResuPro reduces manual effort, speeds up the hiring process, and improves decision-making by evaluating resumes for relevance, minimizing human bias, and ensuring fairer assessment of applicants. This system offers a scalable, efficient, and objective approach to resume screening that can be easily adopted by organizations to enhance recruitment processes. This paper explores the design, methodology, and performance of ResuPro, demonstrating the potential of ML and NLP technologies to significantly improve the efficiency, fairness, and effectiveness of recruitment in today’s fast-paced job market.

| **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  1.1 Introduction to Project  1.2 Project Category  1.3 Objectives  1.4 Structure of Report  CHAPTER 2 LITERATURE REVIEW  2.1 Literature Review  2.2 Research Gaps  2.3 Problem Formulation  CHAPTER 3 PROPOSED SYSTEM  3.1 Proposed System  3.2 Unique Features of The System  CHAPTER 4 REQUIREMENT ANALYSIS AND SYSTEM SPECIFICATION  4.1 Feasibility Study (Technical, Economical, Operational)  4.2 Software Requirement Specification  4.2.1 Data Requirement  4.2.2 Functional Requirement  4.2.3 Performance Requirement  4.2.4 Maintainability Requirement  4.2.5 Security Requirement  4.3 SDLC Model Used  4.4 System Design  4.4.1 Data Flow Diagrams  4.4.2 Use Case Diagrams  4.5 Database Design  CHAPTER 5 IMPLEMENTATION  5.1 Introduction Tools and Technologies Used.  5.2 Dataset Description (IF ML based project)  CHAPTER 6 TESTING, AND MAINTENANCE  6.1 Testing Techniques and Test Cases Used  CHAPTER 7 RESULTS AND DISCUSSIONS  7.1 Presentation of Results (Charts/Graphs/Tables)  7.2 Performance Evaluation  7.3 Key Findings  CHAPTER 8 CONCLUSION AND FUTURE SCOPE  REFERENCES (Only **in IEEE FORMAT**)  Research Paper Acceptance Proof  Research Paper (If Presented/ Published)  Proof of patent publication (Screenshot of Publication) |  |
|  |  |

**LIST OF FIGURES**

| **Figure no.** | **Description** | **Page no.** |
| --- | --- | --- |
| 1. | figure representing data flow diagram | 16 |
| 2. | figure representing use case diagram | 17 |
| 3. | figure representing ER diagram | 18 |
| 4. | figure representing output of classification model | 25 |
| 5. | figure representing output of scoring model | 26 |
| 6. | figure representing accuracy of classification model | 27 |
| 7. | figure representing accuracy of scoring model | 27 |

**LIST OF TABLES**

| **Table no.** | **Description** | **Page no.** |
| --- | --- | --- |
| 1. | Table representing test cases for classification model | 23 |
| 2. | Table representing test cases for scoring model | 24 |

**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

**Goal:** Ensure inclusive and equitable quality education and promote lifelong learning opportunities for all.

#### Justification:

1. **Skill-Based Resume Building** – ResuPro helps job seekers create structured and domain-specific resumes, ensuring they effectively highlight their skills and qualifications, thereby improving employability.
2. **Personalized Resume Scoring** – The ML-based scoring system helps individuals understand their strengths and areas for improvement, guiding them to enhance their qualifications.
3. **Bridging the Education-to-Employment Gap** – By categorizing resumes based on skills, experience, and job roles, ResuPro supports fresh graduates and professionals in transitioning into the workforce efficiently.
4. **Empowering Learners with Career Insights** – Users receive data-driven feedback on their resumes, helping them align their career paths with industry requirements.

### 

### SDG 17: Partnerships for the Goals

**Goal:** Strengthen the means of implementation and revitalize the global partnership for sustainable development.

#### Justification:

1. **Collaboration with Educational Institutions** – ResuPro can be integrated with universities, career counseling services, and online learning platforms to assist students in building industry-ready resumes.
2. **Engagement with Recruiters and HR Professionals** – Companies can use the platform to streamline candidate screening, enhancing hiring efficiency and diversity.
3. **AI-Powered Job Market Insights** – By analyzing large-scale resume data, ResuPro provides trends and insights into job market demands, benefiting stakeholders in education and employment.
4. **Open-Source Contributions & API Integration** – The platform can collaborate with other career development platforms and recruitment systems through API integrations, expanding its reach and impact.
5. **Global Accessibility** – With cloud-based deployment, ResuPro ensures accessibility for users worldwide, promoting equal opportunities for career development across different regions.

**CHAPTER 1**

**INTRODUCTION**

**1.1 Introduction to Project**

In the constantly changing world of hiring, organizations often receive a large number of resumes, making the process overwhelming. The conventional methods of manually sifting through resumes to identify suitable candidates for specific roles are time-consuming and prone to human bias. To address these challenges, we present a cutting-edge solution – the Machine Learning (ML) based Resume Classifier System.

This project harnesses the power of advanced machine learning algorithms to automate the resume screening process, streamlining recruitment efforts and enhancing the efficiency of talent acquisition teams. By leveraging natural language processing (NLP) and pattern recognition, our system aims to categorize and prioritize resumes based on their relevance to specific job requirements.

The ML-based Resume Classifier System not only accelerates the recruitment process but also significantly reduces the risk of overlooking qualified candidates. This project represents a significant leap forward in the realm of HR technology, offering a scalable and adaptable solution to the ever-growing challenges associated with talent acquisition in today's dynamic job market. As we delve into the details of this system, we will explore its architecture, functionality, and the potential impact it can have on revolutionizing the way organizations identify and engage with top-tier talent

**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 wish to achieve the following major objectives:

1. To provide UI/UX to create a resume of a candidate.

2. To develop a system that automates the initial screening of resumes, reducing the time and resources spent on manual review processes.

3. To develop algorithms that can understand the context of job requirements and applicant qualifications, improving the precision of the classification process.

4. The proposed system will utilize Natural Language Processing (NLP) techniques, the system seeks to comprehend the semantic meaning of resumes, ensuring a more accurate assessment of candidates' qualifications and suitability for specific roles.

5. The work aims to enhance decision-making during the early stages of recruitment by providing data insights, ultimately improving the selection process and identifying top-tier talent

**1.4 Structure of Report**This report is organized to provide a clear, systematic understanding of the *ResuPro* project, covering the key areas of research, system design, testing, and implementation. The following chapters are included:

* **Chapter 2: Literature Review**This chapter reviews existing research and technologies related to resume classification systems, machine learning applications in recruitment, and the challenges in the HR domain. It also identifies the gaps in current solutions and sets the stage for the *ResuPro* system.
* **Chapter 3: Proposed System**This chapter introduces the *ResuPro* system, outlining its core features, including the ML-based resume classification, domain-specific resume scoring, and the automatic generation of tailored feedback. It also highlights the unique functionalities of the system that make it stand out from existing solutions.
* **Chapter 4: Requirement Analysis and System Specification**This chapter provides a comprehensive analysis of the system's requirements, including both functional and non-functional specifications. It also includes the feasibility study (technical, economical, and operational) and an overview of the SDLC model used. Additionally, system design aspects such as data flow diagrams, use case diagrams, and database design are discussed.
* **Chapter 5: Implementation**This chapter focuses on the implementation of the *ResuPro* system, detailing the tools and technologies used, including the integration of machine learning models for resume classification and scoring. It also covers the front-end and back-end development processes, as well as the implementation of core system features.
* **Chapter 6: Testing and Maintenance**The testing strategies and methodologies adopted for the *ResuPro* project are discussed in this chapter. It includes the test plan, testing techniques (functional, performance, security, etc.), and test cases used to ensure the quality of the system. It also presents the results from the various levels of testing performed, such as unit testing, integration testing, and user acceptance testing (UAT).
* **Chapter 7: Results and Discussions**This chapter presents the results obtained from the system’s implementation and testing. It includes performance evaluation, domain-specific resume scoring results, and insights into the effectiveness of the system. Key findings from the testing process, including issues discovered and their resolutions, are also discussed.
* **Chapter 8: Conclusion and Future Scope**The final chapter summarizes the key outcomes of the *ResuPro* project and discusses its potential impact on the recruitment industry. It also highlights areas for future improvements, including expanding the system’s capabilities, enhancing machine learning models, and integrating new technologies to further optimize the recruitment process.

**CHAPTER 2**

**LITERATURE REVIEW**

**1.1 Literature Review**

**1. IConSCEPT 2023 research paper, authored by Surendiran B, Harsha Vardhan Chirumamilla, Maruprolu Naga Raju Reddy**

An automated solution for resume classification using machine learning (ML) techniques such as Decision Tree, Random Forest, KNN, and Support Vector. The study aims to revolutionize the hiring process by addressing the challenges of manual resume screening, offering potential cost and time savings for hiring companies. The literature review discusses prior works in the field, citing examples that employ cosine similarity, NLP, and various ML algorithms. The proposed methodology involves a diverse dataset of 3446 resumes, with data pre-processing using NLP techniques, and the results showcase high accuracy across all ML models. Visual representations of model performance are provided in figures, and the research underscores the potential of ML techniques in automating and improving the efficiency of the resume screening process.

**2. Riza Tanaz Fareed et al: Resume Classification using Cosine Similarity**

Implemented a method using cosine similarity for resume classification.Utilized an NLP pipeline to extract words, including techniques like stop words and lemmatization.TF-IDF vectorizer used to vectorize words for the KNN model.Evaluated resumes based on job description using Cosine Similarity, achieving a 98.96% accuracy.

**3. Suhas H E and Manjunath A E: Model for Resume Suggestion**

Developed a model suggesting resumes for job roles using NLP, NER, Word Embedding, and Cosine Similarity.

Trained NER model on a data dump of technical skills, using the output for a word2vec model.

Evaluated resume-JD matches using Cosine Similarity, obtaining an accuracy of 79.8%.

**4. Sujit Amin et al: Web Application for Resume Screening**

Created a web application for screening resumes, divided into server-side, recruiter side, and job applicant side.

Employed an NLP pipeline using SpaCy for processing resumes.

Resumes were ranked on the recruiter side using a score calculator, utilizing a total of 220 resumes for training and testing.

**5. Pradeep Kumar Roy: Expense Reduction in Hiring**

Proposed a method to reduce expenses in hiring by addressing challenges of applicant selection, resume analysis, and suitability determination.

Utilized NER, NLP, and text classification with n-grams.

Employed machine learning algorithms with varying accuracies, with Linear Support Vector Machine Classifier achieving the highest accuracy of 78.53%.

**6. Bhushan Kinge et al: ML and NLP-Based Resume Screening**

Proposed a system for automated resume screening using NLP techniques and three machine learning algorithms: K-Nearest Neighbors, Random Forest, and Support Vector Machine.

Random Forest outperformed the other algorithms with an accuracy of 94.5%.

Aimed to enhance recruitment efficiency and accuracy while reducing workload and time for recruiters.

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

Developed a machine learning-based solution for ranking job candidates' resumes based on job requirements and qualifications.

Support Vector Machine (SVM) algorithm outperformed Naïve Bayes and K-Nearest Neighbors (KNN) in terms of accuracy and efficiency.

Highlighted the potential for the proposed system to save time and resources in the recruitment process.

**8. D. Jagan Mohan Reddy et al: Model for Predicting Job Candidate Acceptance**

Proposed a model to predict job candidates' likelihood of accepting job offers based on relevant attributes.

Used statistical measures and machine learning algorithms, with the Random Forest algorithm achieving an accuracy of 94.86%.

Demonstrated the potential of using machine learning in predicting job candidate behavior, potentially saving time and resources in recruitment.

**2.2 Research Gaps  
  
1. Accuracy and Effectiveness of Resume Classification Algorithms:**

Many existing resume screening systems still rely on rule-based or keyword-matching algorithms. However, these methods can be limited in terms of understanding the context or semantics of the text. There's room for improvement in applying more advanced machine learning algorithms (e.g., deep learning, NLP techniques) to better classify and prioritize resumes.

**Research Gap:** Developing and evaluating more accurate, context-aware resume classification models that consider the nuances in candidate qualifications.

**2. Domain-Specific Resume Scoring:**

While some systems provide resume scoring, many are generalized and don't account for specific industries or job roles. Scoring algorithms need to be tailored to particular domains, recognizing the importance of different skills, qualifications, and experience relevant to those domains.

**Research Gap:** Creating more refined, domain-specific scoring models that can more effectively evaluate a resume based on the specific requirements of a given job.

3. Bias in Resume Screening:

A common challenge in machine learning systems is reducing bias, especially when algorithms are trained on biased data sets. For example, historical hiring data may reflect certain biases in candidate selection, which could be carried over into automated systems.

Research Gap: Investigating and developing methods to mitigate biases in resume screening algorithms and ensuring fairness in recruitment processes.

**4. Usability and User Experience in Resume Building Tools:**

Although resume-building tools are widely available, many still fail to offer intuitive user experiences, especially for non-technical users. The design of such tools can be complex or unintuitive, which hinders adoption and usability.

**Research Gap:** Exploring better user interface (UI) and user experience (UX) designs for resume building tools that make them more accessible and user-friendly for a broader audience.

**5. Adaptability of Resume Tools to Emerging Job Market Trends:**

The job market is dynamic, and certain industries and skills evolve rapidly. Resume builders may not be able to keep up with these shifts and often lack the flexibility to include emerging skills and qualifications.

**Research Gap:** Developing adaptive systems that can dynamically update their templates, categories, and scoring algorithms to reflect new trends and job market demands.

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

Many resume-building systems operate in silos and lack seamless integration with other hiring platforms, such as job boards, LinkedIn, or Applicant Tracking Systems (ATS).

**Research Gap:** Researching better integration strategies for resume tools to work smoothly with various HR and recruitment platforms, allowing for a more comprehensive and streamlined recruitment process.

**2.3 Problem Formulation**

The popular resume websites today either give users lots of resumes or loads of scores, and it's hard to make them work together smoothly. This is tough for people, especially those who are new to job hunting. Trying to figure out your resume score while going through many resumes at the same time can be confusing, especially for beginners.

**CHAPTER 3**

**PROPOSED SYSTEM**

**3.1 Proposed System**

**A. Front-End Development**

The frontend of ResuPro is built using React.js, Vite, and Tailwind CSS, offering a seamless, responsive user interface for resume creation and uploading. As a single-page application (SPA), it ensures real-time updates without page reloads. The UI guides users through the resume creation process, providing suggestions and prompts to ensure a structured format.

**B. Back-End Development**

The back-end API is developed using Node.js and Express.js. Core functionalities include:

**- Authentication and Authorization:** Handled with JSON Web Tokens (JWT) for secure access to documents.

- **CRUD Operations:** Users can upload, update, and delete resume files based on their roles.

**C. Database Management**

MongoDB, a NoSQL database, is used to store both resume documents and critical metadata, including user data, upload dates, authors, and tags (e.g., job domain). MongoDB's horizontal scaling ensures the system handles large volumes of data effectively, providing flexibility and high availability.

**D. Machine Learning Model**

- **Data Collection:** The classification dataset includes 891 records from Kaggle, categorized into 24 job domains. A custom scoring dataset has 10,000 records, evaluating candidates based on experience, skills, and company level.

- **NLP Techniques:** Tokenization, stop-word removal, lemmatization, and TF-IDF vectorization are used to process resume text.

- **Classification:** Naive Bayes and Random Forest models are used to classify resumes into job domains.

- Scoring Mechanism: A Random Forest Regressor scores candidates based on skills, experience, and company level.

**E. Deployment**

- **Frontend:** Deployed on Netlify, benefiting from automatic deployments, a global CDN, and high availability.

- **ML Model:** Deployed on Streamlit, enabling efficient processing of resume data and providing real-time feedback to users.

**F. Results and Discussion**

Initial tests show strong results, with the NLP module achieving over 90% accuracy in extracting resume data. The machine learning models reached 88% classification accuracy across various job domains. Users report that the system significantly reduces resume screening time and helps identify top candidates more efficiently.

**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 Classification:** Using Natural Language Processing (NLP) and machine learning models, ResuPro automatically classifies resumes into relevant job domains (e.g., Software Engineering, Marketing, Finance), allowing recruiters to easily filter and sort resumes based on their industry-specific needs.
3. **Interactive Resume Builder:** The platform features an intuitive, guided resume creation process with prompts and suggestions, helping users structure their resumes effectively without the need for prior experience in resume formatting.
4. **Seamless Cloud Deployment:** The entire application, including both frontend (React with Tailwind CSS) and backend (machine learning models on Streamlit and API on Node.js), is deployed on scalable cloud infrastructure (Netlify and Streamlit), ensuring high availability, fast loading speeds, and seamless integration of both the front and back-end components.

**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 outlines the key functions the ResuPro system should perform:

**Resume Upload:** Users can upload resumes, which are processed by the system.

**Resume Creation:** An interactive form that allows users to fill out sections of their resume.

**Resume Scoring:** The system automatically scores the uploaded or created resume based on various factors such as skills, experience, and company type.

**Resume Classification:** The system classifies resumes into specific job domains using machine learning models.

**User Management:** Authentication (JWT-based), user profile creation, and editing.

**4.2.3 Performance Requirement**

The system should meet the following performance criteria:

**Response Time:** The system should process resumes and provide feedback (score, classification) within 5 seconds for each request.

**Scalability**: It should support up to 10,000 active users simultaneously, ensuring stable performance during peak usage periods.

**Uptime:** The system should have a minimum uptime of 99.5%, ensuring that the platform is available for use most of the time.

**4.2.4 Maintainability Requirement**

The system should be maintainable in the following ways:

**Modular Code Structure:** The frontend and backend should be modular and well-documented, making it easy for developers to maintain and extend the system.

**Version Control:** Git should be used for source code management, enabling easy tracking of changes and rollbacks if needed.

**Scalable Infrastructure:** The backend (hosted on platforms like Streamlit and Netlify) should be easy to scale and update as necessary to support more users and data.

**4.2.5 Security Requirement**

The security needs of the ResuPro platform should address:

**Data Encryption:** User data and resume documents must be encrypted both in transit (using SSL/TLS) and at rest.

**Authentication & Authorization:** Use of JSON Web Tokens (JWT) for secure authentication. Role-based access control to ensure that only authorized users can perform specific operations.

**Data Privacy:** Ensure that user data is protected and stored in compliance with relevant data protection regulations (e.g., GDPR)

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

**Flexibility**: Agile supports iterative development, allowing continuous improvement in features like resume scoring, classification, and user feedback.

**Collaboration**: Agile promotes collaboration between development teams, stakeholders, and end users, ensuring that features align with user needs and project goals.

**Frequent Releases:** Agile enables frequent releases of smaller features, so the development team can collect user feedback early and make adjustments.

**4.4 System Design**This section presents the high-level design of the system, showing how different components interact and function.

**4.4.1 Data Flow Diagrams (DFD)**

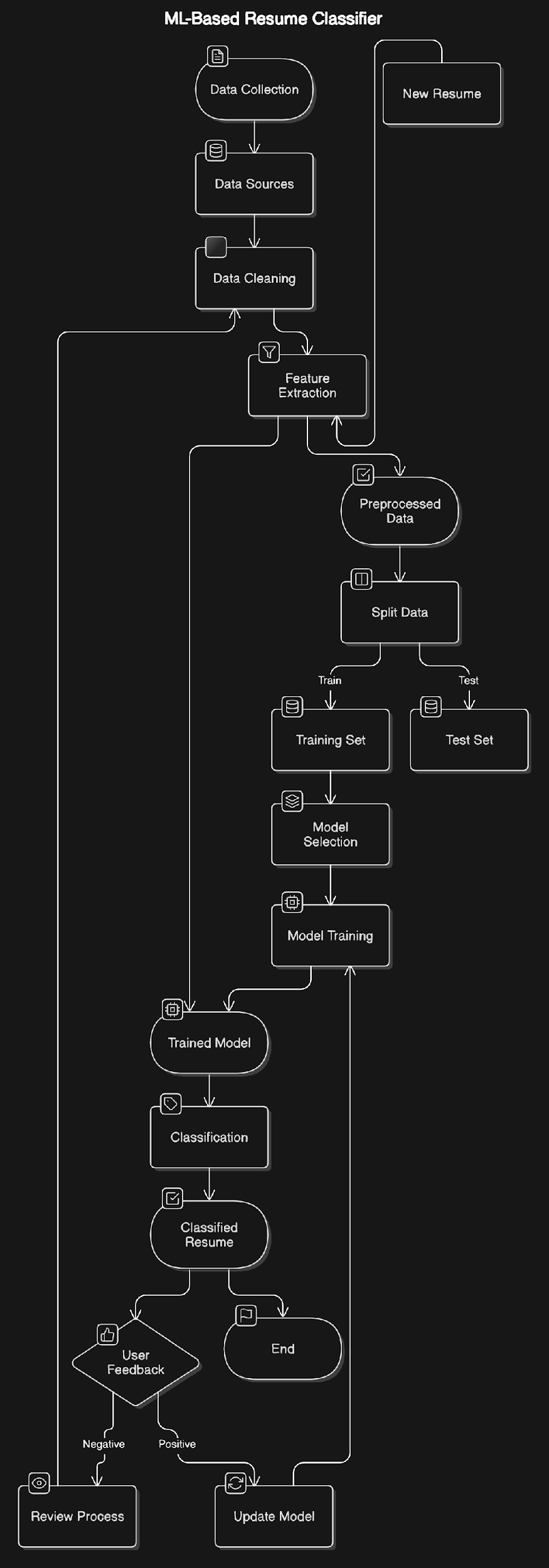


fig1: image representing data flow diagram

**4.4.2 Use Case Diagrams**

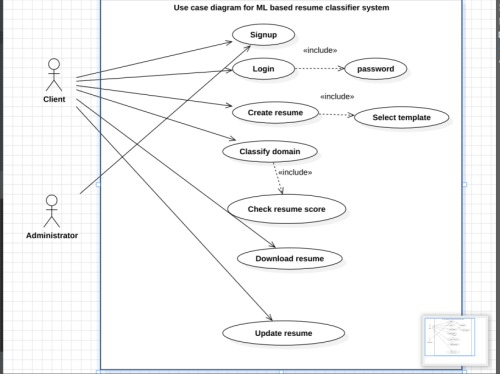


fig2: image representing use case diagram

**4.5 Database Design**

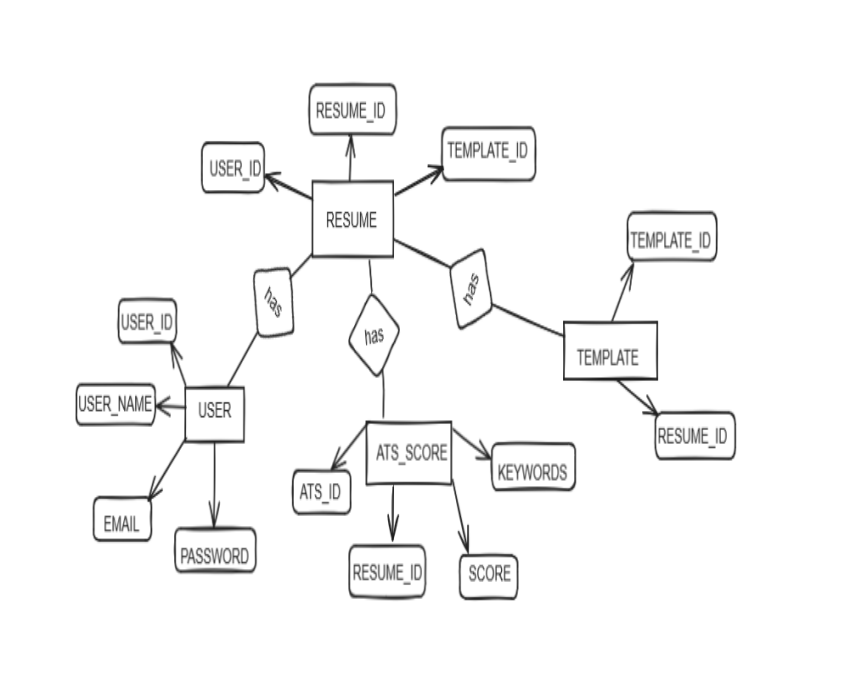


fig3: image representing ER diagram

**CHAPTER 5**

**IMPLEMENTATION**

**5.1 Introduction Tools and Technologies Used.  
1. Frontend Development:**

**1.1 Vite**: Vite is utilized as the build tool to streamline the frontend development process, ensuring rapid iteration and enhanced performance.

**1.2 React:** React.js is employed for building dynamic and interactive user interfaces, enabling seamless navigation and data handling.

**1.3 Tailwind CSS:** Tailwind CSS is utilized for styling the user interface, providing a customizable and utility-first approach to design.

**2. Backend Development:**The back-end API is developed using Node.js and Express.js. Core functionalities include:

**- Authentication and Authorization:** Handled with JSON Web Tokens (JWT) for secure access to documents.

- **CRUD Operations:** Users can upload, update, and delete resume files based on their roles.

**3. Machine Learning:**

**3.1 Natural Language Processing (NLP):** NLP techniques are employed for analyzing and processing textual data extracted from resumes. This includes tasks such as text classification, entity recognition, and semantic analysis.

**3.2 Random Forest:** Random Forest algorithms are used for ensemble learning, improving the accuracy and robustness of the resume classifier by combining multiple decision trees.

**3.3 Naive Bayes:** Naive Bayes classifiers are utilized for probabilistic classification, particularly in scenarios where feature independence assumptions hold true, such as in text classification tasks.

**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.

**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

**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:**

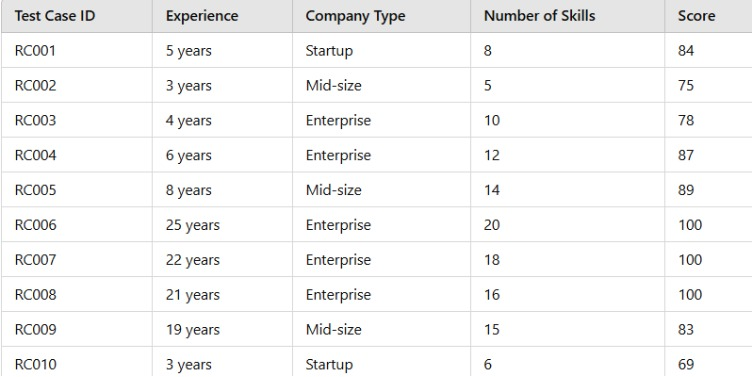


table 1: Table specifying various test cases for classification model

**B. Scoring Model:**

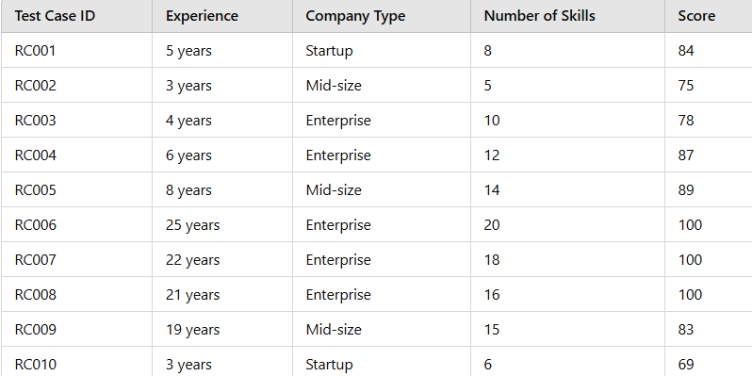
****

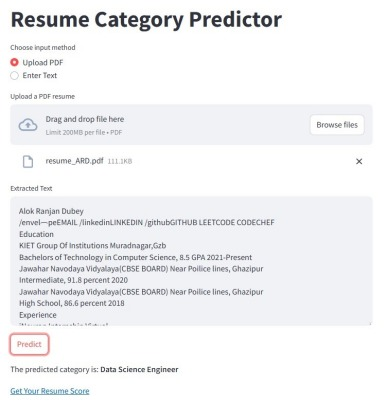
table 2: Table specifying various test cases for Scoring model

**CHAPTER 7**

**RESULTS AND DISCUSSIONS**

**7.1 Presentation of Results (Charts/Graphs/Tables)**

**7.1.1 Classification Result**

fig 4: image specifying output of classification model

**7.1.2 Scoring Result**

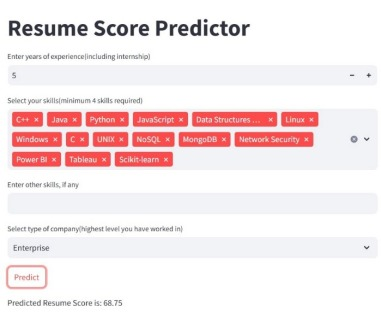


fig5: image specifying output of scoring model

**7.2 Performance Evaluation**

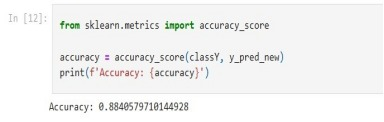
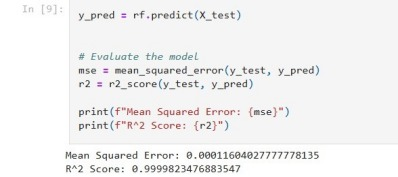


fig6: image specifying accuracy of classification model

fig7: image specifying accuracy of Scoring model

**7.3 Key Findings**

**1. Resume Scoring Accuracy:** The ResuPro system successfully evaluates resumes based on domain-specific criteria, with a high degree of accuracy in scoring resumes across various job domains. This was achieved through the integration of machine learning models that assess resume completeness, formatting, and relevance to specific job requirements.

**2. Domain Classification :** The machine learning models effectively classify resumes into job-specific domains (e.g., Software Development, Marketing, Data Science), aiding users in tailoring their resumes for particular job roles. The classification system helps ensure that the resume is aligned with the expectations of recruiters in the relevant fields.

**3. Skills Relevance:** The ResuPro platform identified key skills relevant to each job domain, which were used to evaluate the effectiveness of resumes. Users were able to see which skills were emphasized and receive recommendations on adding missing skills to make their resumes more competitive.

**CHAPTER 8**

**CONCLUSION AND FUTURE SCOPE**

The ResuPro platform has proven to be a highly effective tool for assisting users in building tailored, high-quality resumes. By leveraging machine learning to score and classify resumes, the system provides personalized feedback, helping users create resumes that stand out to recruiters in specific job domains. The combination of resume scoring, domain classification, skill analysis, and feedback integration makes ResuPro a powerful solution for job seekers who want to enhance their chances of landing interviews.

Furthermore, the ResuPro platform not only evaluates resumes but also educates users on how to make meaningful improvements to their resumes, addressing both content and presentation. The system's machine learning models have been trained on a diverse dataset, ensuring broad applicability across different industries and job roles.

**FUTURE SCOPES**

**1. Expansion of Job Domains :** The current machine learning models are trained on a set of popular job domains. Future development could include expanding the platform’s capabilities to cover more niche job categories and emerging fields (e.g., Artificial Intelligence, Blockchain, etc.).

**2. Enhanced Natural Language Processing (NLP):** Improving the NLP capabilities of ResuPro will enable the system to more effectively understand complex resume formats and extract key information, even from poorly structured or non-standard resumes. This could include more advanced parsing of unstructured data and better handling of different document formats.

**3. Integration with Job Portals:** To further streamline the job application process, ResuPro could be integrated with popular job portals like LinkedIn, Indeed, or Glassdoor. This would allow users to directly apply to jobs from within the platform, while also optimizing their resumes based on real-time job market trends.

**4. Real-time Job Market Analysis:** Future versions of ResuPro could provide real-time job market insights, offering users up-to-date information on which skills are in high demand, what recruiters are looking for, and salary expectations. This could be achieved by analyzing large-scale job listing data from various platforms.

**5. AI-Powered Resume Customization :** An advanced feature could involve AI-powered resume customization, where the system tailors resumes for specific job postings. By analyzing job descriptions and matching them with the user’s existing skills and experience, the system could generate optimized resume versions suited to particular jobs.

**6. Global Language Support:** Adding multilingual support could allow ResuPro to expand to international markets, enabling non-English-speaking users to build effective resumes in their native languages. This could be achieved through the integration of automated translation tools and region-specific resume standards.

**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 Kaganurmath, 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.