



AISSMS

COLLEGE OF ENGINEERING

ज्ञानम् सकलजनहिताय

Approved by AICTE, New Delhi, Recognized by Govt. of Maharashtra,
Affiliated to Savitribai Phule Pune University and recognized 2(f) and 12(B) by UGC
(Id.No. PU / PN/ Engg. / 093 (1992)
(Accredited by NAAC with grade A+)



A PRELIMINARY REPORT ON

RESUME EVALUATION USING MACHINE LEARNING

SUBMITTED TO THE SAVITRIBAI PHULE PUNE UNIVERSITY, PUNE IN THE
PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE AWARD OF THE
DEGREE

OF

BACHELOR OF ENGINEERING (COMPUTER ENGINEERING)

SUBMITTED BY

**SWANAND PANKAJ MODAK (20CO073)
PRASANNA DATTATRAY SHINDE (20CO124)
ANIKET SANTOSH TIWARI (20CO135)**

DEPARTMENT OF COMPUTER ENGINEERING

AISSMS COLLEGE OF ENGINEERING

PUNE 411041

**SAVITRIBAI PHULE PUNE UNIVERSITY
2022-23**



AISSMS

COLLEGE OF ENGINEERING

ज्ञानम् सकलजनहिताय

Approved by AICTE, New Delhi, Recognized by Govt. of Maharashtra,
Affiliated to Savitribai Phule Pune University and recognized 2(f) and 12(B) by UGC
(Id.No. PU / PN/ Engg. / 093 (1992)
(Accredited by NAAC with grade A+)



CERTIFICATE

This is to certify that the project report entitled

RESUME EVALUATION USING MACHINE LEARNING

Submitted by

SWANAND PANKAJ MODAK (20CO073)
PRASANNA DATTATRAY SHINDE (20CO124)
ANIKET SANTOSH TIWARI (20CO135)

is a bonafide student of this institute and the work has been carried out by him under the supervision of **Prof. S. R. Nalamwar** and it is approved for the partial fulfillment of the requirement of Savitribai Phule Pune University, for the award of the degree of **Bachelor of Engineering (Computer Engineering)**.

Prof. S. R. Nalamwar

Guide

Department of Computer Engineering

Dr. S. V. Athawale

Head

Department of Computer Engineering

Dr. D. S. Bormane

Principal

AISSMS College of Engineering, Pune

ACKNOWLEDGEMENTS

The successful completion of this project would not have been possible without the collective effort, dedication, and support of our team members, supervisors, sponsors, and various individuals who have contributed in diverse ways throughout this journey. We extend our heartfelt gratitude to each and every one of you for playing a pivotal role in making this project a reality.

First and foremost, we would like to express our deepest appreciation to our esteemed project supervisor, Prof. S. R. Nalamwar, for her unwavering guidance, support, and encouragement throughout this project. Her expertise, patience, and insightful feedback have been instrumental in shaping our understanding of the subject matter, refining our research methodology, and steering us towards the successful completion of this project.

We are also immensely grateful to our project sponsor, Mr. Sachin Supekar, for his generous financial support and invaluable guidance. His belief in our project and their willingness to invest in our endeavor has been crucial to our success. His insights and suggestions have helped us refine our project's scope and direction, and his support has been a constant source of motivation.

Finally, we would like to express our heartfelt gratitude to our families and friends for their unwavering support and encouragement throughout our academic journey. Their love, belief in us, and willingness to lend a helping hand have been a constant source of motivation and inspiration.

Swanand Modak
Prasanna Shinde
Aniket Tiwari

ABSTRACT

The traditional method of manually screening resumes is a time-consuming and inefficient process that can hinder the recruitment process and lead to missed opportunities to identify top-tier candidates. Additionally, manual screening is susceptible to human biases that can perpetuate inequities in hiring decisions. To address these challenges, this project proposes an innovative approach to resume evaluation and candidate application tracking (ATS) powered by machine learning.

The proposed system utilizes natural language processing (NLP) techniques to extract key information from resumes, including skills, education, and professional experience. This extracted information is then fed into cutting-edge machine learning algorithms that identify and classify skills, rank candidates according to their qualifications and alignment with specific job requirements, and mitigate potential biases in the decision-making process.

This intelligent ATS streamlines the recruitment process by automating the initial screening stage, allowing recruiters to dedicate their time to interacting with the most promising candidates. This leads to enhanced efficiency, reduced time-to-hire, and improved hiring outcomes.

Furthermore, the system promotes fairness and equity in hiring by mitigating potential biases in the decision-making process. This is achieved by employing techniques such as data cleansing, bias-aware algorithms, and diversity metrics. By ensuring that candidates are evaluated objectively and without prejudice, the system fosters a more inclusive and equitable hiring landscape.

The development and implementation of this intelligent ATS can significantly impact organizations by empowering them to identify the most qualified candidates, build high-performing teams, and achieve their strategic goals. By embracing this innovative approach, organizations can transform their recruitment processes, enhance their talent acquisition strategies, and reap the benefits of a more efficient, fair, and inclusive hiring environment.

Keywords: Resume evaluation, Candidate application tracking system (ATS), Machine learning, Natural language processing (NLP), Skill identification, Skill classification, Candidate ranking, Bias mitigation

TABLE OF CONTENTS

LIST OF FIGURES	vii
LIST OF TABLES	viii
1 Introduction	1
1.1 Motivation	2
1.2 Problem Statement	3
2 Literature Survey	4
3 Software Requirement Specification	5
3.1 Introduction	5
3.1.1 Project Scope	5
3.1.2 User Classes and Charecteristics	6
3.1.3 Assumptions and Dependencies	7
3.2 Functional Requirements	8
3.2.1 Resume Classification:	8
3.2.2 Candidate Tracking:	8
3.2.3 Customization and Configuration:	8
3.2.4 Automation:	9
3.2.5 Security and Compliance:	9
3.2.6 Scalability:	9
3.3 External Interface Requirements	9

3.3.1	User Interface	9
3.3.2	Software Interface	10
3.3.3	Communication Interface	10
3.4	Non-functional Requirements	11
3.4.1	Performance Requirement	11
3.4.2	Safety Requirements	12
3.4.3	Security Requirement	13
3.4.4	Software Quality Attributes	14
3.5	System Requirements	15
3.5.1	Database Requirement	15
3.5.2	Software Requirement	16
3.5.3	Hardware Requirement	17
3.6	Analysis Model: SDLC Model to be applied	18
3.7	System Implementation Plan	20
4	System Design	21
5	Other Specifications	22
5.1	Advantages	22
5.2	Limitations	23
5.3	Applications	24
6	Conclusion and Future Work	26
6.1	Conclusion	26
6.1.1	Key Findings	26

6.1.2	Implications	27
6.1.3	Recommendations	27
6.2	Future Work	27

LIST OF FIGURES

LIST OF TABLES

3.1	System Implementation Plan	20
-----	--------------------------------------	----

1

Introduction

In the dynamic realm of today's job market, recruiters face the arduous task of meticulously scrutinizing a voluminous influx of resumes to identify the most suitable candidates. Conventional manual resume screening methods are often plagued by inefficiency, time constraints, and the inherent biases of human judgment. To tackle these challenges head-on, this project delves into the development of an innovative resume evaluation and candidate application tracking system (ATS) powered by the transformative capabilities of machine learning.

This project aims to revolutionize the recruitment process by automating resume evaluation and candidate ranking through state-of-the-art machine learning techniques. The system will harness the power of natural language processing (NLP) to extract critical information from resumes, meticulously extracting relevant details such as skills, education, and professional experience. Employing cutting-edge machine learning models, the system will then identify and classify skills, categorize candidates based on their qualifications, and rank them according to their alignment with specific job requirements.

The seamless integration of machine learning into the ATS will streamline the recruitment process, dramatically reducing the time and effort expended on identifying top-tier candidates. By automating the initial screening stage, recruiters can dedicate their valuable time to engaging with the most promising applicants, leading to enhanced hiring outcomes and a more efficient recruitment process.

This project goes beyond mere efficiency; it also strives to uphold fairness and equity in the hiring process. The system will be meticulously designed to mitigate potential biases in machine learning models, ensuring that candidates are evaluated objectively and without discrimination. This commitment to fairness will foster a more inclusive and equitable hiring landscape, ensuring that the best talent is identified and

nurtured, regardless of background or identity.

The successful culmination of this project will introduce a transformative tool for recruiters, revolutionizing the recruitment process by streamlining operations, enhancing decision-making, and promoting a more equitable hiring landscape. By harnessing the power of machine learning, this innovative ATS will empower recruiters to identify the most qualified candidates, enabling organizations to build high-performing teams and achieve their strategic goals.

1.1 Motivation

The motivation for this project stems from the inherent challenges and limitations of traditional resume evaluation and candidate application tracking systems (ATS). Manual resume screening, often the standard practice, is plagued by time constraints, inefficiency, and the potential for human biases to influence decision-making. These limitations can lead to missed opportunities to identify top-tier candidates, while also perpetuating biases that hinder diversity and inclusion in the workplace.

To address these shortcomings, this project aims to leverage the power of machine learning to automate and enhance the recruitment process. By employing sophisticated NLP techniques and machine learning algorithms, the system will objectively evaluate resumes, extract key information, and rank candidates based on their qualifications and alignment with specific job requirements. This will not only streamline the recruitment process but also promote fairness and equity in hiring decisions.

Furthermore, the increasing volume of job applications and the growing demand for specialized skills necessitate a more efficient and effective approach to candidate selection. Traditional methods are often ill-equipped to handle the sheer volume of applications, leading to delays in the hiring process and the potential loss of qualified candidates. Machine learning, with its ability to process large amounts of data and identify patterns, offers a promising solution to this challenge.

By automating resume evaluation and candidate ranking, the system will enable recruiters to focus their attention on the most promising applicants, saving valuable time and resources. This will expedite the hiring process, ensuring that organizations can promptly identify and recruit the talent they need to achieve their goals.

In addition to enhancing efficiency, machine learning can also contribute to a more equitable and inclusive hiring landscape. By mitigating potential biases in the decision-making process, the system will ensure that candidates are evaluated objectively and without prejudice. This commitment to fairness will foster a more diverse and inclusive workplace, allowing organizations to tap into the full potential of their workforce.

The development of an intelligent resume evaluation and ATS powered by machine learning holds the potential to revolutionize the recruitment process, leading to enhanced efficiency, improved decision-making, and a more equitable hiring landscape. By embracing this innovative approach, organizations can empower themselves to identify the most qualified candidates, build high-performing teams, and achieve their strategic objectives.

1.2 Problem Statement

The current manual resume screening process is time-consuming, inefficient, and susceptible to human biases, leading to missed opportunities to identify top-tier candidates and perpetuating inequities in the hiring process. To address these challenges, develop an intelligent resume evaluation and candidate application tracking system (ATS) powered by machine learning that can automatically extract key information from resumes, identify and classify skills, rank candidates based on their qualifications and alignment with specific job requirements, and mitigate potential biases in the decision-making process.

2

Literature Survey

The resume classification is a broad research encompassing the studies and techniques related to automated methods for analyzing resumes to extract relevant information, categorize candidate skills, and match application to job description. Some of the key methods are as follows:

Software Requirement Specification

3.1 Introduction

In today's fast-paced and competitive job market, organizations are inundated with a massive influx of resumes from job seekers. Managing this deluge of applications efficiently, while ensuring that the right candidates are identified, is a formidable challenge. The Resume Classification and Candidate Application Tracking System project addresses this critical need by offering a powerful and automated solution that streamlines the recruitment process. This project aims to revolutionize the way organizations handle candidate applications and resumes. By integrating cutting-edge technologies from natural language processing (NLP) and machine learning, it provides an intelligent system for classifying, organizing, and tracking job applicants. Whether you are a small business looking to expand your fleet or a large enterprise processing hundreds of applications daily, this system provides a scalable and customizable solution.

3.1.1 Project Scope

The purpose of this project is to develop and implement a comprehensive system that automates the classification and tracking of job applicants and their resumes. The system is designed to streamline the recruitment process, improve the efficiency of the HR department, and enhance candidate selection.

The scope of this project encompasses the following key components:

- 1. Resume Classification:** To develop and automated system that classifies the

resumes into predefined roles or jobs using Natural Language Processing (NLP) and Machine Learning techniques.

- 2. Candidate Tracking:** To implement a candidate tracking system that enables end-to-end monitoring of candidates progress and also design a user-friendly interface for HR professionals and hiring managers to update candidate statuses
- 3. Customization and Configuration:** To provide flexibility for organizations to classify job roles and work flow and also ensures that system can adapt varying requirement processes.
- 4. Automation:** To automate resume screening and classification process which reduces manual efforts and speeds up the process for candidate selection for a job role.

3.1.2 User Classes and Charecteristics

User classes help us identify different groups of individuals who will interact with the system. Understanding these classes helps us tailor our system to meet their specific needs.

Classes and charecteristics of our project are:

1. HR Professionals:

- Lead users which are responsible for managing the hiring process.
- Creating and customizing job roles and classification criteria.
- TRacking and updating candidates application progress.
- Requires a user-friendly interface for efficient workflow management.

2. Hiring Manager:

- Collaboration with HR professionals in evaluating candidates. Has access to candidates profile and CV.
- Provides opinion and inputs to decisions.
- They need and easy user-interface to rate and comment on candidate.

3. Job Applicants:

- Followed by HR profesionals and Hiring managers, Candidates is an essential class for our project
- Candidates apply for positions externally or within the organization.
- They can submit their resume, recieve automatic status updates and track their applications.

4. System Administrator:

- System administrators manage system configuration and customization. They also ensure system security and user access control.
- They handle software updates and provide technical support for the system users.

3.1.3 Assumptions and Dependencies

Identifying assumptions and dependencies are essential so that we can ensure that essential conditions are met for successful execution of our project.

Assumptions:

- 1. Data Availability:** It is assumed that the project will have access to a sufficient volume of historical resumes and application data for training and testing the classification and tracking system.
- 2. Resource Availability:** Availability of the required resources, including personnel, hardware, software, and data storage, to implement and operate the system.
- 3. Active Participation and User Cooperation:** The successful implementation of the system assumes the cooperation and active participation of HR professionals, hiring managers, and other users in defining classification criteria and workflows.
- 4. Team Availability:** The project team members are assumed to be available as per project schedule for completing the deadlines and delivering the project on time.

Dependencies:

- 1. Data Acquisition:** This project is dependent on acquiring historical resume and application data, which is crucial for training the machine learning models and testing the system.
- 2. NLP and Machine Learning Models:** The project depends on the development and fine-tuning of NLP and machine learning models for resume classification, which may require iterative testing and refinement.
- 3. User Training:** After successful user adoption is dependent on the development of user training materials and conducting training sessions for HR professionals, hiring managers, and other system users.

-
- 4. Testing and Quality Assurance:** The project depends on thorough testing and quality assurance processes to ensure the system's accuracy, security, and usability.
 - 5. External Services:** Any external services, such as email notifications or mobile notifications may be dependent on third-party providers' availability and reliable functioning.

3.2 Functional Requirements

3.2.1 Resume Classification:

- The system automatically classifies the resumes into predefined job roles based on content analysis using NLP and machine learning models.
- HR professionals can define and customize criteria used for classification to align with specific job description.

3.2.2 Candidate Tracking:

- HR professionals can track candidates' progress from initial application through interview stages to onboarding.
- The system sends automated status updates to candidates, keeping them informed about the progress of their applications.

3.2.3 Customization and Configuration:

- Organizations can configure and customize their recruitment workflows, including job role-specific workflows.
- Access control is defined so that unauthorized user will not have the permission to access sensitive data and features.

3.2.4 Automation:

- Automating the initial screening and shortlisting of candidates based on predefined criteria.
- Streamline the interview scheduling process by integrating with calendars and sending notifications to candidates and interviewers.

3.2.5 Security and Compliance:

- Implementing role-based access control to ensure data security and compliance with privacy regulations.
- Enabling compliance officers to review and enforce legal and ethical standards in the recruitment process.

3.2.6 Scalability:

- Designing the system with scalability in mind to accommodate growing application volumes and organizational needs.

These system features aim to streamline the recruitment process, enhance efficiency, and provide valuable insights into candidate selection. Customization and a user-friendly interface ensure that the system meets the specific requirements of organizations while maintaining data security and compliance with industry regulations.

3.3 External Interface Requirements

3.3.1 User Interface

A web based user-interface which is accessible to HR professionals and hiring managers must be created. The user-interface must be responsive and should support

multiple devices so that users can access it anywhere and on any device.

3.3.2 Software Interface

1. Integration with HR systems

- Data import and export capabilities should be implemented for exchange of information with existing HR system or applicant tracking system.
- API for integration with third-party tools and services must also be offered, including email applications, for communication and scheduling

2. Calendar and Scheduling

- We can integrate with popular calendar applications eg. Google calendar or Outlook so that we can facilitate interview scheduling and calendar availability checks.
- Also to ensure that interview schedules and updates are synchronized with external calendar systems.

3. External Services

- Connecting the user system to external services, such as background checks or reference verification services, via APIs for additional candidate screening.
- The system is designed to work with external services which have periodic downtimes or maintenance windows.

4. Browser Compatibility

- The browser and version must be specified so that the system will be compatible to ensure a consistent and uninterrupted user experience.

3.3.3 Communication Interface

The system must be configured so that notifications are sent to applicant, HR and the hiring manager regarding the application statuses, interview invitation and other relevant updates. We can use the Simple Mail Transfer Protocol (SMTP) server configuration for email delivery.

3.4 Non-functional Requirements

3.4.1 Performance Requirement

To guarantee that the system functions are effectively and efficiently in real-world scenarios, performance requirements for machine learning based resume classification and candidate application tracking system are crucial. The system's capabilities and performance standards are outlined in these requirements.

Performance requirements are as follows:

1. Response time:

- The system should provide a responsive user interface with a maximum response time of 2 seconds for typical user interactions, such as resume classification, candidate status updates, and report generation.
- Resume classification should not take more than 5 seconds for each resume submitted.

2. Scalability:

- The system should be scalable enough to handle 10,000 applications per month without degradation in performance.
- The system must also support horizontal scaling to accommodate increased application volumes.

3. Throughput:

- The system must have the capability of processing at least 50 resumes per hour for automated screening and classification.
- The system must support 20 candidates per minute for interview scheduling and candidate tracking system.

4. Data Storage and Retrieval:

- The retrieval time for resume retrieval and document storage must be less than 3 seconds for individual resumes.
- The system should maintain a record of at least 5 years' worth of application and candidate data.

5. Availability and Uptime:

- The system should have a minimum uptime of 99.9% to ensure accessibility during business hours.
- The scheduled maintenance and updates should be communicated to users

in advance, and these activities should take place during non-peak hours.

6. Data Reporting:

- The system generated report must be completed within 5 minutes for standard reports and 15 minutes for complex or custom reports.
- The data analytics queries must be executed within 10 seconds for standard query and 30 seconds for complex queries.

7. Miscellaneous requirements:

- External service calls and integrations, such as email notifications and calendar updates, should have response times of less than 5 seconds.
- Mobile applications or mobile web interfaces should provide a responsive experience with a maximum response time of 3 seconds for common user interactions.

3.4.2 Safety Requirements

Establishing safety requirements is essential when using Machine Learning for resume classification to guarantee the system's dependability and efficiency.

Here are some safety requirements for our project:

1. Data Security

- The applicants data must be stored securely and in encrypted format to avoid unauthorized access and data breaching.
- Role based access control must be implemented to ensure that only authorized user can view and modify sensitive information.

2. Data Privacy

- The system must ensure that it complies with data protection laws and regulations, such as HIPAA, depending on the region and type of data being processed.
- User consent should be obtained and documented for data processing, and data retention policies must be adhered to.

3. User Authentication:

- Implementing strong user authentication mechanisms, such as two-factor authentication (2FA) or single sign-on (SSO), to prevent unauthorized access.
- Ensuring that any external services and third-party integrations used for

email, calendar, or data exchange are secure and adhere to security best practices.

4. Secure Communication:

- All data transmitted between the system and users should be encrypted using secure communication protocols (e.g., HTTPS) to prevent eavesdropping.

5. Disaster Recovery:

- Disaster recovery plan must be developed and tested to ensure data and system recovery in case of system failure, data loss or security breaches.
- An incident response time must also be developed to address and mitigate security breaches, including data breaches and unauthorized access with clear procedures and notifications.

6. Consent Tracking and Regular Security Audits:

- A consent tracking mechanism must be implemented to record and manage candidate consent for data processing, including data storage and communication.
- Conduct regular security audits, vulnerability assessments, and penetration testing to identify and address potential security weaknesses.

7. Audit Trials:

- A comprehensive audit logs must be maintained that record all user actions, system events, and data modifications. These logs should be protected from tampering.

3.4.3 Security Requirement

A security requirement is a goal set out for an application at its inception. Every application fits a need or a requirement. It is a statement of needed security functionality that ensures one of many different security properties of software is being satisfied. A software security requirement should be much like a functionality requirement; it shouldn't be vague or unattainable.

Here are some Security Requirements:

1. Model Security:

- Prevent unwanted access and manipulation with the ML models. Make use of version control and secure containers.

-
- To avoid malicious model injections, implement secure methods for version control and model updates.

2. System Security:

- To manage who has access to and can alter the ML system, put robust authorization and authentication procedures in place.
- To ensure that no unauthorized users gain access to the system, use intrusion detection systems, firewalls, and secure communication protocols.
- To identify and address security problems, like intrusion attempts, implement thorough monitoring and logging.
- Make that the system has fail-safes in place so that, in the case of a system malfunction or security compromise, it will revert to a safe state.

3. Adversarial Robustness:

- Apply strategies to protect machine learning models from adversarial attacks and make sure they continue to function even when intentional alterations are made.
- Conduct regular system vulnerability and adversarial attack tests.

3.4.4 Software Quality Attributes

Software Quality Attributes are features that facilitate the measurement of performance of a software product by Software Testing professionals.

It includes the following:

- **Usability:** The system should be user-friendly and self-explanatory. Proposed system is flexible, robust, and easily testable.
- **Reliability:** Software system reliability is an important attribute of software quality. System Reliability is hard to achieve because the complexity of the system tends to be high. It is the measure of how long a system/machine performs its intended function.
- **Robustness:** The program must be able to function well in a range of environmental factors.
- **Accuracy:** Ensure that the system accurately classifies resumes into predefined job roles and tracks candidate progress without errors.
- **Performance Efficiency:** Optimize system performance to handle high application volumes, ensuring that response times meet user expectations.

-
- **Scalability:** Design the system to be scalable, capable of handling growing numbers of resumes and users without significant degradation in performance.
 - **Security:** Implement robust security measures to protect candidate data, prevent unauthorized access, and ensure compliance with data protection regulations.
 - **Testing and Quality Assurance:** Conduct thorough testing, including unit testing, integration testing, and user acceptance testing, to identify and address issues before deployment.
 - **External Integration:** Verify the compatibility and reliability of external integrations with third-party services and tools.
 - **Scalability Testing:** Perform load testing to validate the system's scalability, ensuring it can handle peak usage without performance degradation.
 - **Data Reporting and Analytics:** Enable users to generate custom reports and analyze recruitment data effectively, providing valuable insights for decision-making.

3.5 System Requirements

3.5.1 Database Requirement

Database requirements are crucial as they dictate how data is stored, managed and accessed.

Some key database requirements are:

1. User Information:

- This includes information such as name, address, phone number, email ID, academic details, previous experience and all other information which can be extracted from the resume using a resume parser.

2. Data Schema Design:

- A well-structured and normalized data schema that accurately represents applicant and resume information. Include fields for personal details, job history, education, skills, and application status.

3. Relational Database Management System (RDBMS):

- As our project involves semi-structured data such as text based resumes, it needs to be stored and queried using a NoSQL database like MongoDB

which is flexible for storage and integrates with chosen technology stack.

4. Scalability and Partitioning:

- Design the database for scalability, allowing it to handle an increasing number of resumes and applicant records. Consider horizontal partitioning (sharding) if needed.

5. Backup and Recovery Strategy:

- The implemented database is backed up at regular basis to protect against data loss.. Test the restore process to ensure data recovery capabilities.

6. Backup Encryption:

- Encrypt database backups to ensure the security of data during backup and restore operations, preventing unauthorized access to sensitive information.

7. Database Monitoring:

- Implement continuous database monitoring to detect and respond to performance bottlenecks, anomalies, and potential issues in real-time. Monitoring should include resource usage, query performance, and security alerts.

3.5.2 Software Requirement

These requirements specify software components and technologies necessary for the project's successful development and operation

They are as follows:

1. Operating System (OS):

- The server hosting the system should run a supported and secure operating system, such as Linux (e.g., Ubuntu, CentOS) or Windows Server.
- The user interface should be accessible on various client platforms, including Windows, macOS, and mobile operating systems (iOS, Android).

2. Web Server:

- A web server software is required for serving web based user interfaces and handling HTTP requests.

3. Database Management System:

- A NoSQL database to store all the information about the candidate. It also aligns with the project's requirement

4. Programming Languages:

- Python is used for building the machine learning model and applying NLP techniques required for the project along with Streamlit, which is a web framework used to create an user interface.

3.5.3 Hardware Requirement

Hardware requirements will depend on various factors, including the expected system load, scalability needs, and performance goals.

Here are the hardware requirements:

1. Server Hardware:

- CPU: Multi-core processors to handle the computational load efficiently. The number of cores depends on expected user load and data processing requirements.
- RAM: Sufficient RAM to support concurrent users, database operations, and data processing tasks.
- Storage: Fast and reliable storage solutions, such as Solid-State Drives (SSDs) for low latency data retrieval and database performance.
- Network Interface: Gigabit Ethernet or higher to ensure fast data transfer and network responsiveness.

2. Load Balancer:

- If the application experiences high traffic, consider a load balancer to distribute requests across multiple application servers for load balancing and redundancy.

3. Database Server:

- The database server should meet the same CPU, RAM, and storage requirements as the application server.

4. Backup and Storage Devices:

- Deploy backup and storage devices for automated data backups, ensuring data safety and disaster recovery capabilities.

5. Internet Connection:

- Updates to software libraries necessitate internet connection. This flexibility enables adaptability to changing requirements in the application and tracking system.

3.6 Analysis Model: SDLC Model to be applied

The Agile Methodology is well suited for this project as it emphasizes on flexibility, collaboration and iterative development. Stages of development are as follows:

1. **Requirement Gathering:** The team collaborates with stakeholders to gather and document the system's functional and non-functional requirements. Key activities included in this phase are as follows:
 - Conducting interviews with HR professionals, hiring managers, and system users to understand their needs.
 - Reviewing industry best practices and compliance requirements, such as data privacy regulations.
 - Defining the system's features, user stories, and use cases in the product backlog and prioritizing requirements based on their impact on the project.
2. **Design the Requirements:** Once the requirements are gathered for this project, the design phase focuses on creating a detailed plan for the system's architecture, user interfaces, and data models. Activities included in this phase are as follows:
 - Designing the database schema to support data storage and retrieval efficiently.
 - Architecting the system with scalability, security, and maintainability in mind.
 - Developing a technical design document that outlines how the system will be built.
 - Creating wireframes to visualize the user interface and gather feedback from the user.
3. **Construction / Iteration:** During the construction or iteration phase, the actual development of the system takes place. The key points included in this phase are as follows:
 - The development is organized in sprints or iterations using agile methodology.
 - We begin by developing a resume parsing algorithm which will parse the resume (supported formats are .pdf and .docx) and extract the candidates information such as name, education, qualification and past experiences(if any).
 - After extracting the information, stop-words will be removed and the keywords will be stored in our NoSQL database.
 - We will develop a machine learning model using NLP techniques and train

this model on the keywords which are stored in our database.

- Performance of the model will be evaluated using the F1 Score.
- After each iteration of training the model, we will tune the hyperparameters for increasing the accuracy of our model until it reaches a saturation point.

4. **Testing/Quality Assurance:** This phase is essential to verify that the system meets the specified requirements and is free of defects. Following activities are included in this phase:

- Continuous testing of AI models, databases, and system components should be done concurrently with development
- We make sure that resume parsing is done without any error and model is trained after removing all stop-words.
- Conducting unit testing to test individual components and functions.
- Carrying out user acceptance testing (UAT) to involve stakeholders in testing and validation.
- Performing integration testing to ensure that different system modules work together seamlessly.

5. **Deployment:** This phase makes the system available for users. Key aspects of this phase includes:

- Initially we prepare the production environment, which includes servers, databases, and networking components.
- Followed by deploying the system to the production environment while ensuring minimal disruption to operations.
- Monitoring the deployment process for errors, and having rollback plans in case of issues.
- After deploying the production environment we communicate the release to stakeholders and provide the documentation as needed.
- And finally setting up ongoing maintenance procedures to address issues and updates on post-deployment in essential.

6. **Feedback:** Feedback is a continuous process that takes place throughout the project's lifecycle, but it's especially important after deployment. The activities involved in this phase are as follows:

- Collecting feedback from system users, HR professionals, and hiring managers to identify issues, improvements, and new requirements.
- Conducting regular retrospectives and post-implementation reviews to assess what went well and what could be enhanced in the development and deployment processes.
- Using feedback to inform future iterations and updates to the system.

3.7 System Implementation Plan

Planning is the first step involved in system implementation. It is a very basic function which describes effectively the very basic questions of how, where and when the objectives can be realized, or it serves as a guiding framework.

The System Implementation Plan table shows the overall schedule of tasks compilation and time duration required for each task.

Month	Task	Date of Execution
August	Problem statement, objectives, motivation finalisation	14/08/2023
September	Understanding project requirements, scope, feasibility	16/09/2023
October	Literature review, base paper finalisation and algorithm analysis	20/10/2023
November	Review paper draft and publication	15/11/2023

Table 3.1: System Implementation Plan

4

System Design

Other Specifications

5.1 Advantages

- **Efficiency:** Resume classification systems automate the screening process, significantly reducing the time and effort required by recruiters to review and shortlist candidates.
- **Cost-Effective:** By automating the initial stages of recruitment, organizations can save costs associated with manual resume screening and focus their resources on more strategic HR tasks.
- **Consistency:** Automated systems apply consistent criteria to evaluate resumes, minimizing the impact of human biases and ensuring fair evaluation of all applicants.
- **Scalability:** These systems can handle a large volume of resumes, making them ideal for organizations with high recruitment needs, especially during periods of rapid expansion.
- **Improved Matchmaking:** Advanced algorithms enhance the matching process, ensuring that job openings are filled by candidates with the most relevant skills and qualifications.
- **Data-Driven Insights:** Resume classification systems can provide valuable insights and analytics about the applicant pool, helping organizations make data-driven decisions in their hiring processes.
- **Time-saving:** Automated systems allow recruiters to focus on interacting with pre-screened, qualified candidates, enhancing the overall efficiency of the recruitment process.
- **Global Talent Pool:** Resume classification systems enable organizations to access a diverse pool of talent from around the world, expanding their reach beyond local markets.
- **Compliance Management:** These systems can be programmed to ensure that

hiring processes adhere to legal and industry-specific regulations, enhancing compliance management.

- **Reduction in Human Error:** Automated systems minimize the risk of human errors in the screening process, ensuring that candidates are evaluated consistently and objectively.
- **Time-to-Hire Reduction:** By swiftly shortlisting suitable candidates, organizations can reduce the time it takes to fill vacant positions, leading to improved productivity.
- **Enhanced Diversity and Inclusion:** With careful design, these systems can be leveraged to promote diversity and inclusion by reducing unconscious biases in the initial screening stages.
- **Enhanced Candidate Experience:** Well-designed systems provide instant feedback to candidates, acknowledging their application and keeping them engaged in the hiring process.

5.2 Limitations

- **Bias and Fairness Issues:** If not properly designed and trained, these systems can perpetuate biases present in historical data, leading to unfair advantages or disadvantages for certain groups.
- **Lack of Contextual Understanding:** Automated systems might misinterpret the context of skills or experiences mentioned in resumes, leading to mismatches between candidates and job requirements
- **Overemphasis on Keywords:** Some systems may overly rely on specific keywords, potentially overlooking candidates with diverse skill sets or alternative phrasing of qualifications
- **Inability to Assess Soft Skills:** Automated systems struggle to evaluate soft skills, interpersonal abilities, and cultural fit, which are essential for many roles.
- **Data Privacy Concerns:** Handling sensitive personal data raises concerns about data privacy and compliance with regulations like GDPR, necessitating robust security measures.
- **Dependence on Historical Data:** Systems trained on historical data might not adapt well to emerging job roles or changing skill requirements, leading to mismatches in job-candidate pairings
- **User Experience Challenges:** Poorly designed interfaces or complicated user experiences can deter candidates from interacting with these systems, impacting

application rates

- **Loss of Human Touch:** Automated systems lack the human touch, potentially leading to a less personalized candidate experience, which might deter some applicants.
- **Inflexibility:** Overly rigid algorithms might miss out on exceptional candidates who possess unique skills or experiences not accounted for in the predefined criteria.
- **Inadequate Assessment of Job Hopping:** Automated systems might misinterpret frequent job changes as instability, not accounting for the evolving nature of certain industries.
- **Security Vulnerabilities:** Storing vast amounts of sensitive candidate data poses security risks, requiring robust cybersecurity measures to prevent data breaches.
- **Dependence on Keywords:** Overemphasis on specific keywords can lead to false positives, where candidates are shortlisted based on keyword matches without considering the overall context.
- **Challenge with Unstructured Data:** Resumes often contain unstructured data, making it challenging for systems to extract relevant information accurately.

5.3 Applications

- **Recruitment and Staffing Agencies:** Resume classification systems help agencies efficiently match candidates with job opportunities across various industries and roles.
- **Corporate HR Departments:** Large organizations with numerous job openings benefit from automated systems to handle high volumes of applications, ensuring efficient candidate shortlisting.
- **Online Job Portals:** Platforms like LinkedIn, Indeed, and Monster use resume classification to facilitate job recommendations, enhancing user experience for job seekers.
- **HR Software Suites:** Integrated HR software solutions utilize resume classification to streamline recruitment processes, manage applicant databases, and improve overall hiring efficiency.
- **Government Employment Services:** Public employment services use these systems to assist job seekers in finding suitable positions and help employers identify potential candidates.
- **Customized Job Matching Platforms:** Niche job platforms catering to specific

industries or skill sets use resume classification to offer tailored job matches for candidates.

- **Research and Analysis:** Resume classification systems are used by researchers and analysts to study job market trends, skills demand, and other aspects of employment dynamics.
- **Skill Mapping:** Resume classification systems are used to map candidate skills against specific job requirements, identifying skill gaps and training needs within an organization.
- **Freelancer Platforms:** Platforms connecting freelancers with employers utilize resume classification to match freelancers with suitable projects, ensuring a good fit for both parties.
- **Talent Pool Management:** Organizations use these systems to maintain a talent pool of potential candidates for future job openings, streamlining the hiring process for recurring positions.
- **Employee Referral Programs:** Resume classification assists in evaluating employee referrals efficiently, ensuring that recommended candidates meet the necessary criteria.
- **Automated Interview Scheduling:** Integrated with scheduling tools, these systems facilitate the automated scheduling of interviews with shortlisted candidates, saving time for recruiters.
- **Performance Prediction:** Advanced systems analyze historical hiring data and employee performance metrics to predict the success of candidates based on their resumes.

6

Conclusion and Future Work

6.1 Conclusion

The development of an intelligent resume evaluation and candidate application tracking system (ATS) powered by machine learning has the potential to revolutionize the recruitment process, leading to enhanced efficiency, improved decision-making, and a more equitable hiring landscape. By embracing this innovative approach, organizations can empower themselves to identify the most qualified candidates, build high-performing teams, and achieve their strategic objectives.

6.1.1 Key Findings

The key findings of our project are as follows:

- Machine learning can effectively automate the resume screening process, saving time and resources for recruiters.
- Machine learning algorithms can accurately identify and classify skills from resumes, enabling better matching of candidates to job requirements.
- Machine learning can rank candidates based on their qualifications and alignment with specific job requirements, providing recruiters with a prioritized list of potential candidates.
- Machine learning techniques can mitigate potential biases in the decision-making process, ensuring fair and equitable evaluation of candidates.

6.1.2 Implications

The implementation of an intelligent ATS powered by machine learning can significantly impact organizations by:

- **Enhancing recruitment efficiency:** By automating the initial screening stage, recruiters can focus their time on interacting with the most promising candidates, leading to faster time-to-hire.
- **Improving hiring outcomes:** By identifying the most qualified candidates, organizations can build high-performing teams that are better equipped to achieve their strategic goals.
- **Promoting fairness and equity in hiring:** By mitigating potential biases in the decision-making process, organizations can foster a more inclusive and diverse workplace.

6.1.3 Recommendations

To maximize the benefits of an intelligent ATS, organizations should:

- Collect a diverse dataset of resumes to train the machine learning models.
- Continuously monitor and refine the models to ensure they are performing accurately and without bias.
- Provide training and support to recruiters to effectively utilize the system.
- Integrate the system with other human resource management systems to create a seamless recruitment workflow.

6.2 Future Work

The future scope of the project is to further develop the intelligent resume evaluation and candidate application tracking system (ATS) powered by machine learning to enhance its capabilities and address emerging trends in the recruitment landscape. This includes:

-
1. **Incorporating natural language understanding (NLU):** Advance the system's ability to extract deeper insights from resumes by utilizing NLU techniques to understand the context and semantics of the text. This will enable the system to identify nuances in language, grasp the overall sentiment of the resume, and uncover hidden patterns that may indicate a candidate's suitability for a particular role.
 2. **Developing predictive models for candidate success:** Utilize machine learning algorithms to develop predictive models that assess a candidate's potential for success in specific roles. These models can consider factors such as skills, experience, personality traits, and cultural fit to provide recruiters with data-driven insights that go beyond traditional resume screening.
 3. **Integrating with talent management systems:** Integrate the intelligent ATS with existing talent management systems to create a seamless workflow that spans the entire talent lifecycle, from initial sourcing and screening to onboarding, development, and retention. This integration will enable a holistic view of candidate engagement and development, allowing organizations to nurture top talent throughout their careers.
 4. **Adapting to emerging hiring trends:** Continuously monitor and adapt the system to address emerging trends in the recruitment landscape, such as the increasing demand for remote work, the rise of gig economy platforms, and the evolving skill requirements for specific industries. The system should be flexible enough to incorporate new data sources, adapt to changing job descriptions, and remain relevant in the face of evolving hiring practices.
 5. **Enhancing bias mitigation:** Continuously refine the system's bias mitigation techniques to ensure that candidates are evaluated objectively and without prejudice. This may involve developing new algorithms, incorporating diversity metrics, and implementing ongoing training and evaluation processes to identify and eliminate potential biases.
 6. **Exploring multi-modal data analysis:** Incorporate multi-modal data analysis techniques to consider additional sources of information beyond resumes, such as social media profiles, online portfolios, and video interviews. This will provide a more comprehensive view of a candidate's qualifications and potential, leading to more informed hiring decisions.
 7. **Automating interview scheduling and communication:** Automate the process of scheduling interviews, sending reminders, and managing communication with candidates. This will save recruiters time and streamline the interview scheduling process, allowing them to focus on in-depth candidate interactions and assessments.

-
8. **Providing real-time feedback to candidates:** Provide real-time feedback to candidates throughout the application process, highlighting their strengths, suggesting areas for improvement, and offering personalized career advice. This will enhance the candidate experience, foster positive interactions with the organization, and potentially attract top talent.
 9. **Developing a mobile-friendly ATS:** Create a mobile-friendly version of the ATS that allows recruiters and candidates to access the system conveniently from their smartphones or tablets. This will facilitate remote work, enable on-the-go candidate screening, and enhance the overall user experience.
 10. **Integrating with artificial intelligence (AI) chatbots:** Integrate the ATS with AI chatbots to provide automated responses to frequently asked questions, offer self-service options for candidates, and guide them through the application process. This will reduce the burden on recruiters, provide candidates with immediate assistance, and enhance the overall user experience.