**DEVELOPMENT OF A RESUME ANALYSER FOR A RECRUITING COMPANY**

**BY**

**NTUEN, IKAKA-ABASI SAMUEL**

**18/SC/CO/650**

**SUBMITTED TO**

**DEPARTMENT OF COMPUTER SCIENCE**

**FACULTY OF COMPUTING**

**UNIVERSITY OF UYO**

**NIGERIA**

**IN PARTIAL FULFILLMENT TO THE REQUIREMENTS FOR THE AWARD OF THE BACHELOR OF SCIENCE DEGREE (B.SC) IN COMPUTER SCIENCE**

**AUGUST, 2024**

**CHAPTER ONE**

**GENERAL INTRODUCTION**

* 1. **Background of Study**

Resume analysis represents a significant technological advancement aimed at optimizing the recruitment process in an organization. In an era where the volume of job applications can be overwhelming, the implementation of automated systems to streamline initial resume screening is becoming increasingly essential. The background of this study encompasses the evolution of recruitment processes, the technological advancements in Artificial Intelligence (AI) and Machine Learning (ML), and the practical benefits and challenges associated with deploying a resume analyzer.

Historically, the recruitment process has been labor-intensive, involving manual screening of resumes to identify suitable candidates. This traditional method is not only time-consuming but also prone to human error and bias. Recruiters often face challenges in managing large volumes of applications, which can result in qualified candidates being overlooked due to the sheer volume of data (Suen *et al*., 2019). The need for a more efficient and objective system has led to the development of resume analyzers, which leverage AI and ML technologies to automate the initial screening process.

AI and ML have made significant strides in recent years, enabling the development of sophisticated tools capable of parsing and analyzing large datasets. In the context of resume analysis, these technologies can be used to extract relevant information from resumes, such as education, work experience, skills, and certifications. This information can then be compared against job requirements to identify the best matches. Resume analyzers typically use Natural Language Processing (NLP) to understand and interpret the text in resumes, ensuring that the extracted data is accurate and meaningful (Nassif, 2020).

One of the primary advantages of using a resume analyzer is the improvement in efficiency. Automated systems can process thousands of resumes in a fraction of the time it would take a human recruiter. This allows recruiters to focus on more strategic aspects of the hiring process, such as interviewing and candidate engagement. Additionally, AI-driven systems can reduce biases that may influence human decision-making, leading to a fairer and objective selection process (Gonzalez *et al*., 2021).

However, the implementation of resume analyzers is not without challenges. One significant concern is the accuracy of the algorithms used to parse and interpret resume data. Variations in resume formats, the use of unconventional language, and incomplete or ambiguous information can affect the performance of the analyzer. Continuous updates and training of the algorithms are required to ensure they remain effective and accurate (Liem *et a*l., 2018). Moreover, there are ethical considerations related to data privacy and the potential for algorithmic bias, which must be addressed to maintain the integrity and fairness of the recruitment process (Bogen & Rieke, 2018).

Despite these challenges, the benefits of implementing a resume analyzer are substantial. By leveraging AI and ML, recruiting companies can enhance their efficiency, improve the quality of their candidate selection, and provide a better overall experience for job applicants. As technology continues to advance, the capabilities of resume analyzers will likely improve, making them an indispensable tool in modern recruitment practices.

**1.2 Statement of Problem**

Recruiting companies are grappling with the inefficiencies and inaccuracies inherent in traditional manual resume screening processes. As the volume of job applications increases, recruiters are overwhelmed by the sheer number of resumes, leading to extended hiring times, missed opportunities for top talent, and potential biases in candidate selection (Smith, 2022). There is a pressing need for an automated resume analyzer that can efficiently process and evaluate resumes to improve the accuracy, consistency, and fairness of the recruitment process.

Recruiting companies often receive hundreds or thousands of resumes for a single job posting. Manually screening each resume is not feasible, leading to delays in the hiring process. The overwhelming volume can result in recruiters missing out on qualified candidates simply because they cannot review each application thoroughly (Brown, 2021). Manual resume screening is an extremely time-consuming process. Recruiters spend a significant portion of their time reading and evaluating resumes, which reduces the time available for more strategic tasks such as interviewing candidates, conducting background checks, and developing talent acquisition strategies. This inefficiency can slow down the entire recruitment cycle (Johnson, 2020).

Human judgment is inherently subjective, and different recruiters may apply varying criteria and levels of scrutiny when reviewing resumes. This inconsistency can lead to unfair evaluations, with some candidates being unfairly favored or disfavored based on non-relevant factors such as resume format, gender, ethnicity, or educational background. Unconscious biases can also impact the diversity and inclusiveness of the hiring process (Kim & Patel, 2022). Manual screening relies on recruiters' ability to accurately assess candidates' qualifications based on the information presented in resumes. However, important qualifications and experiences may be overlooked due to recruiters' fatigue, time constraints, or lack of domain-specific knowledge. This can result in suboptimal hiring decisions and the exclusion of potentially excellent candidates (Lee, 2021).

Resumes come in various formats, styles, and structures, making it challenging for manual reviewers to consistently extract and interpret relevant information. This diversity can lead to inconsistencies in how information is evaluated and compared across candidates, further complicating the screening process (Williams & Thomas, 2023). Handling sensitive personal information contained in resumes requires strict adherence to data privacy laws and regulations. Manual processes may not be as secure, potentially exposing candidates' personal data to unauthorized access or breaches. Ensuring robust data protection is crucial to maintain trust and comply with legal requirements (Nguyen, 2023).

**1.3 Aim and Objectives**

The aim of this project is to develop an automated resume analyzer for a recruiting company.

The objectives of the research are to:

1. Carry out related literature reviews on the selected topic
2. Design a resume analyzer platform for a recruiting company
3. Implement the resume analyzer’s design in (b) above
4. Evaluate the performance of the system using appropriate metrics

**1.4 Scope and Limitations**

The scope of this project encompasses the design, implementation, and evaluation of an automated resume analyzer for a recruiting company. Inaccessibility of live data to test the functionality of the system constitutes a major challenge to the implementation of this system

**1.5 Significance of Study**

The design and implementation of the resume analyzer is anticipated to yield several positive outcomes. Primarily, the system should enhance the efficiency and accuracy of the resume screening process. Recruiters are expected to experience a significant reduction in the time required to review and evaluate resumes, allowing them to focus more on strategic tasks such as interviewing candidates and developing talent acquisition strategies (Smith, 2022).

The resume analyzer is also expected to improve the consistency and objectivity of candidate evaluations. By leveraging AI and ML algorithms, the system should minimize human biases and ensure that all candidates are evaluated based on relevant qualifications and experience. This is anticipated to lead to a fairer and more inclusive hiring process.

**1.6 Definition of Terms**

**Resume Analyzer**: A software tool designed to automatically scan, parse, and evaluate resumes. It leverages technologies such as Natural Language Processing (NLP), Machine Learning (ML), and Optical Character Recognition (OCR) to extract and interpret key information from resumes to assist recruiters in identifying qualified candidates.

**Recruiting Company**: An organization that specializes in finding and hiring suitable candidates for job positions on behalf of other companies. Recruiting companies manage the end-to-end hiring process, including sourcing, screening, interviewing, and onboarding candidates.

**Optical Character Recognition (OCR)**: A technology used to convert different types of documents, such as scanned paper documents or PDF files, into editable and searchable data. In the context of resume analysis, OCR is used to extract text from resumes submitted in non-editable formats.

**Natural Language Processing (NLP)**: A field of artificial intelligence that focuses on the interaction between computers and human language. NLP enables computers to understand, interpret, and generate human language. In resume analysis, NLP is used to extract and interpret relevant information from the text within resumes.

**Machine Learning (ML)**: A subset of artificial intelligence that involves training algorithms on large datasets to recognize patterns and make predictions. In resume analysis, ML models are used to evaluate and rank candidates based on their qualifications and to improve the accuracy of candidate evaluations over time.

**Applicant Tracking System (ATS)**: A software application used by recruiters and hiring managers to manage the recruitment process. An ATS helps in organizing and tracking job applications, storing candidate information, and facilitating communication with candidates. The resume analyzer can integrate with an ATS to enhance its functionality.

**CHAPTER 2**

**LITERATURE REVIEW**

**2.1 Introduction**

The recruitment process is a critical aspect of human resource management, determining the quality of talent that an organization attracts and retains. In an increasingly competitive job market, the efficiency and effectiveness of recruitment strategies can significantly impact an organization's success. Traditionally, the recruitment process has involved manually sifting through resumes, a method fraught with inefficiencies and prone to human error. As organizations receive an overwhelming volume of applications, the limitations of manual resume screening become evident, leading to extended hiring times and the potential for overlooking qualified candidates (Suen *et al*., 2019).

The advent of technology has brought transformative changes to various sectors, including recruitment. The integration of artificial intelligence (AI) and machine learning (ML) into recruitment processes has introduced innovative solutions to longstanding challenges. Resume analyzers, which leverage these technologies, represent a significant advancement in optimizing recruitment. These tools automate the initial screening of resumes, improving efficiency, reducing biases, and enhancing the overall candidate evaluation process.

AI and ML technologies have advanced rapidly, enabling the development of sophisticated resume analyzers capable of parsing and analyzing large datasets. These systems utilize natural language processing (NLP) to understand and interpret text, ensuring that the extracted data is accurate and meaningful (Nassif, 2020). Optical character recognition (OCR) technology is also employed to convert different types of documents, such as scanned paper documents or PDF files, into editable and searchable data.

Despite the evident benefits, the implementation of resume analyzers is not without challenges. Variations in resume formats, the use of unconventional language, and incomplete or ambiguous information can affect the performance of these systems. Continuous updates and training of the algorithms are required to ensure their effectiveness and accuracy (Liem *et al*., 2018). Additionally, ethical considerations related to data privacy and the potential for algorithmic bias must be addressed to maintain the integrity and fairness of the recruitment process (Bogen & Rieke, 2018).

The shift from manual to automated resume screening is part of a broader trend towards digital transformation in HR practices. As organizations strive to improve their recruitment processes, the adoption of AI and ML technologies is becoming increasingly essential. Automated systems not only enhance efficiency but also allow recruiters to focus on more strategic tasks, such as interviewing and candidate engagement. This shift is expected to lead to better hiring outcomes and a more streamlined recruitment process (Gonzalez *et al*., 2021).

This literature review aims to explore the various aspects of resume analyzers, including their evolution, technological foundations, benefits, challenges, and ethical considerations. By examining existing research and industry practices, this review seeks to provide a comprehensive understanding of the impact of resume analyzers on modern recruitment processes

**2.2 Evolution of Recruitment Processes**

Historically, the recruitment process has been labor-intensive, involving extensive manual efforts to identify suitable candidates from a large pool of applicants. Recruiters would sift through piles of resumes, often spending significant amounts of time on each document to evaluate the qualifications and fit of candidates. This traditional method, while thorough, was not sustainable in the face of increasing application volumes and evolving job market dynamics (Suen et al., 2019).

The inefficiencies inherent in manual resume screening have long been a challenge for recruiters. With the rise of the internet and online job portals, the volume of applications for any given job posting has surged. This influx of resumes has made it nearly impossible for recruiters to thoroughly review each application, leading to delays in the hiring process and the risk of overlooking qualified candidates. As a result, the need for more efficient and scalable recruitment solutions became apparent (Brown, 2021).

The evolution of recruitment processes has been significantly influenced by technological advancements. The initial steps towards automation in recruitment involved the use of Applicant Tracking Systems (ATS), which helped manage and organize resumes. ATS allowed recruiters to store and retrieve resumes more efficiently, but these systems primarily functioned as databases and did not provide advanced analytical capabilities. While ATS improved some aspects of the recruitment process, the fundamental issue of manual resume screening persisted (Johnson, 2020).

The integration of AI and ML technologies into recruitment marked a significant turning point. These technologies enabled the development of resume analyzers, which automate the initial screening of resumes. By leveraging AI and ML, resume analyzers can process large volumes of resumes quickly and accurately, identifying the most qualified candidates based on predefined criteria. This shift towards automation has significantly reduced the time and effort required for resume screening, allowing recruiters to focus on higher-value tasks (Gonzalez *et al*., 2021).

As recruitment processes continue to evolve, the adoption of AI and ML technologies is expected to increase. These technologies offer the potential to address many of the challenges associated with traditional recruitment methods, including inefficiencies, biases, and inconsistencies in candidate evaluation. By enhancing the accuracy and objectivity of resume screening, AI and ML technologies are poised to revolutionize the recruitment landscape, making it more efficient, fair, and effective (Kim & Patel, 2022).

**2.3 Candidate Matching and Skill Assessment**

AI algorithms play a crucial role in enhancing candidate matching and skill assessment processes, especially in the context of migrant workers. By leveraging advanced machine learning models, these systems can meticulously analyze job requirements, considering various factors such as required skills, experience levels, educational qualifications, and even cultural considerations. The AI models then match these job criteria with the profiles of potential migrant workers, ensuring that the most suitable candidates are selected for each position. This automated matching process not only increases the accuracy of candidate selection but also significantly speeds up the recruitment process, reducing the time and resources typically required in traditional methods. (Easmat, *et al*., 2024)

Furthermore, AI-driven assessment tools are becoming increasingly sophisticated, offering a deeper evaluation of candidates' skills and competencies. These tools go beyond simply matching keywords or qualifications on a resume; they assess the actual capabilities of candidates through simulations, tests, and even behavioral assessments. For instance, AI can analyze a candidate's performance in skill-based tests, evaluate their problem-solving abilities, and even gauge their adaptability to different cultural environments. This comprehensive evaluation ensures that not only are candidates technically qualified, but they also possess the soft skills and cultural fit necessary for success in a particular job role. The funnel schema of a hiring system is shown in Figure 2.1 below.

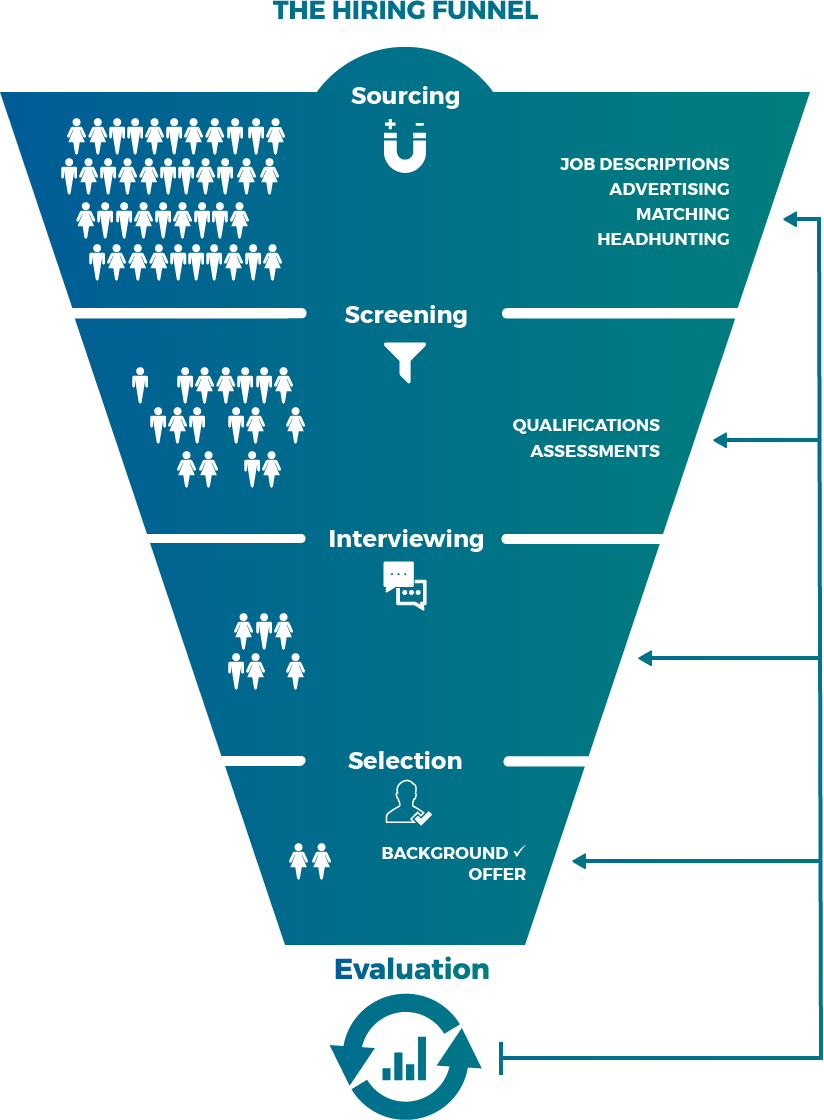


Figure 2.1: The Hiring Funnel

The hiring funnel depicted is a visual representation of the stages involved in the recruitment process, demonstrating how potential candidates are progressively filtered and evaluated before the final selection is made. This funnel metaphor effectively illustrates the narrowing down of a large pool of candidates to a few successful hires, with each stage representing a critical phase in the hiring journey. Below is a description of each stage within the hiring funnel:

**1. Sourcing:** The sourcing stage is at the top of the funnel and involves attracting a large pool of candidates to apply for open positions. This is where the recruitment process begins, with the goal of identifying potential candidates who may be suitable for the roles available.

**Activities Involved:**

* **Job Descriptions**: Clearly defining the job roles, responsibilities, and qualifications required. This information is typically published in job postings on various platforms.
* **Advertising**: Promoting the job openings across multiple channels such as job boards, social media, company websites, and other recruitment platforms to reach a broad audience.
* **Matching**: Using tools or algorithms to match job descriptions with potential candidates' profiles, often through keyword matching or AI-based recommendations.
* **Headhunting**: Actively seeking out and approaching highly qualified candidates, often for specialized or senior roles, who may not be actively looking for a job.

**2. Screening:** This is the next phase where the initial pool of candidates is filtered down based on certain criteria. This stage is crucial for eliminating candidates who do not meet the basic qualifications or are otherwise unsuitable for the role.

**Activities Involved**:

* **Qualifications**: Reviewing candidates' resumes and applications to ensure they meet the minimum educational, experiential, and skill requirements for the job.
* **Assessments**: Conducting preliminary assessments, such as skill tests, personality tests, or phone screenings, to gauge the candidates' suitability for the next stage. This helps in identifying the most promising candidates to move forward.

**3. Interviewing**: This is a more detailed phase of evaluation where candidates who have passed the screening stage are invited for interviews. This stage allows recruiters and hiring managers to assess candidates' competencies, cultural fit, and overall potential in a more in-depth manner.

**Activities Involved**

* **Interviews**: Candidates typically go through multiple rounds of interviews, which may include technical interviews, behavioral interviews, panel interviews, and case studies.
* **Interaction**: This stage often includes interaction with potential future colleagues or managers, allowing both the company and the candidate to evaluate if they are a good fit for each other.

**4. Selection**: Here, the final candidates are chosen based on their performance in the interviews and assessments. This stage involves making the final decision on which candidates will receive job offers.

**Activities Involved:**

* **Background Checks**: Verifying the candidates' backgrounds, which may include checking references, employment history, criminal records, and sometimes credit checks.
* **Offer**: Extending a formal job offer to the selected candidates, which includes details such as salary, benefits, job title, and other terms of employment. Negotiations may occur at this stage before the candidate accepts the offer.

**5. Evaluation**: The evaluation phase is a continuous process that occurs after hiring, where the effectiveness of the recruitment process and the performance of the new hires are assessed. This stage ensures that the hiring process is refined and improved over time.

**Activities Involved**:

* **Performance Monitoring**: Tracking the performance of the new hires to ensure they meet expectations and contribute effectively to the organization.
* **Process Feedback**: Gathering feedback on the recruitment process from both the candidates and the hiring team to identify areas for improvement.
* **Iteration**: Based on the evaluation, the hiring process is iterated and refined to improve future hiring outcomes.

The hiring funnel is a systematic approach that helps organizations manage the recruitment process effectively. Starting with a broad pool of candidates, the funnel narrows through successive stages of screening, interviewing, and selection, leading to the final hiring decision. Each stage is designed to assess different aspects of the candidates, ensuring that only the most qualified and suitable individuals make it to the end. The funnel model emphasizes the importance of continuous evaluation and iteration, ensuring that the recruitment process remains efficient, effective, and aligned with the organization's goals.

**2.4 Types of Resumes**

There are several types of resumes, each designed to showcase a candidate's skills, experience, and qualifications in different ways. The choice of resume type often depends on the individual's career history, the job they are applying for, and the preferences of the recruiter or hiring manager (Morgan, 2024). The types of resumes are:

1. **Chronological Resume**

This type of resume lists work experience in reverse chronological order, starting with the most recent job and working backward. It highlights a candidate’s work history and job progression over time. It is best for individuals with a strong, consistent work history in the same field. It is often preferred by employers as it clearly shows career growth and stability.

1. **Functional Resume**

This resume focuses on skills and qualifications rather than work history. It groups experiences under skill categories, such as “Project Management” or “Customer Service,” rather than listing jobs in chronological order. It is ideal for those with gaps in employment, career changers, or individuals with a diverse set of skills that are not directly related to their work history. It allows candidates to emphasize their capabilities over their job titles.

1. **Combination (Hybrid) Resume**

This is a blend of the chronological and functional formats, this resume lists skills and qualifications first, followed by a chronological work history. It combines the best aspects of both formats and is suitable for individuals who have strong skills and a solid work history. It provides a comprehensive view of the candidate’s abilities and job experience.

1. **Targeted Resume**

This resume is tailored specifically for the job the candidate is applying for. It emphasizes the most relevant skills, experiences, and accomplishments that match the job description. It is best used when applying for a specific position where certain skills and experiences are crucial. It shows that the candidate has taken the time to align their resume with the job requirements.

1. **Mini Resume**

A concise version of the resume that summarizes key qualifications and experiences in a brief format, often fitting on a single page or even a business card. Useful for networking events, job fairs, or situations where a full resume might be too lengthy or inappropriate. It serves as a quick introduction to a candidate's credentials.

1. **Infographic Resume**

This type of resume uses visual elements such as charts, graphs, and icons to present information. It is designed to be visually appealing and easily digestible. Ideal for creative roles where design skills are relevant, such as graphic design or marketing. It can help candidates stand out but may not be suitable for more traditional industries.

1. **Curriculum Vitae (CV)**

A detailed and comprehensive document that includes a candidate’s full academic background, research, publications, presentations, awards, and professional experiences. It is longer than a typical resume and often used for academic, scientific, or research positions. Required for academic, medical, scientific, or research positions. It provides an in-depth view of the candidate’s professional accomplishments and contributions.

**2.5 Technological Foundations**

The development of resume analyzers is grounded in several key technological advancements, primarily in the fields of AI, ML, NLP, and OCR. These technologies collectively enable the automation of resume screening processes, transforming how recruiters evaluate and select candidates.

AI and ML form the core of modern resume analyzers. These technologies allow systems to learn from large datasets, recognizing patterns and making predictions based on the data. In the context of resume analysis, ML algorithms are trained to evaluate candidate qualifications and match them with job requirements. This training involves analyzing vast amounts of resume data to identify relevant features, such as education, work experience, skills, and certifications (Nassif, 2020).

Natural Language Processing (NLP) is another critical component of resume analyzers. NLP enables computers to understand and interpret human language, allowing the system to extract meaningful information from the text within resumes. Techniques such as named entity recognition, text classification, and sentiment analysis are employed to parse resumes and identify key information accurately. NLP ensures that the data extracted from resumes is both relevant and contextually appropriate, enhancing the overall effectiveness of the resume analyzer (Nassif, 2020).

Optical Character Recognition (OCR) technology plays a vital role in processing resumes submitted in non-editable formats, such as scanned documents or PDFs. OCR converts these documents into editable and searchable data, enabling the resume analyzer to extract text and analyze the content. This capability is essential for handling diverse resume formats and ensuring that all relevant information is captured and evaluated (Williams & Thomas, 2023).

The integration of these technologies into resume analyzers has significantly improved their functionality and accuracy. By combining AI, ML, NLP, and OCR, modern resume analyzers can process large volumes of resumes quickly and accurately, providing recruiters with valuable insights into candidate qualifications. These technological foundations have made resume analyzers indispensable tools in the recruitment process, enhancing efficiency and reducing the potential for human error and bias (Liem *et al*., 2018).

**2.6 Related Works**

Gonzalez, *et al*. (2021) examined fairness in automated hiring systems, discussing the ethical implications and proposing solutions to mitigate biases in AI-driven recruitment tools. They examined the impact of biased training data and the potential for discriminatory outcomes. Their work advocates for the use of diverse and representative datasets, as well as the continuous monitoring and adjustment of algorithms to prevent and correct biases.

Brown (2021) highlighted the challenges faced by recruiters due to high volumes of job applications. His study illustrated the inefficiencies in traditional recruiting processes, such as the time-consuming nature of manual resume screening and the potential for human error. Brown underscores the need for automated solutions to handle large datasets effectively, arguing that automation can significantly reduce the workload on recruiters and improve the speed and accuracy of the screening process. By leveraging AI and ML technologies, recruiting companies can more efficiently manage high application volumes, ensuring that qualified candidates are promptly identified and advanced through the recruitment pipeline.

Nguyen (2023) emphasized data privacy considerations, particularly in the context of modern recruitment processes. Nguyen advocates for the adoption of privacy-by-design principles, ensuring that data protection is integrated into every stage of the system development lifecycle. He also highlighted the importance of compliance with data protection regulations, such as the General Data Protection Regulation (GDPR), to avoid legal repercussions and maintain candidate trust.

Johnson (2020) and Smith (2022) both focused on the inefficiencies and inaccuracies inherent in traditional resume screening methods. Johnson discusses time management in recruitment, highlighting the ways in which automated systems can streamline the process and reduce the time-to-hire. He argues that automation can help recruiters manage their workload more effectively, allowing them to focus on higher-value tasks such as candidate engagement and strategic planning.

Kim and Patel (2022) explored the presence of human bias in resume screening and proposed strategies to address these biases. Their work supports the development of AI-driven systems designed to ensure fair and objective candidate evaluations. They discussed the various types of biases that can influence human decision-making, such as affinity bias and confirmation bias, and how these biases can lead to unfair hiring practices. By minimizing human involvement in the initial screening process, AI-driven systems can help reduce the impact of unconscious biases, promoting a more equitable recruitment process.

Lee (2021) contrasts the accuracy of manual resume evaluations with automated systems. Lee's research highlighted the limitations of human judgment, such as inconsistencies and subjectivity, which can affect the fairness and reliability of the screening process. Automated systems, on the other hand, offer a more consistent and objective approach, leveraging algorithms to evaluate resumes based on predefined criteria.

Liem, *et al,* (2018) further validated the accuracy of automated resume parsing, reinforcing the potential benefits of deploying such technologies. Their study demonstrated how automated systems can process and analyze resumes more accurately and consistently than manual methods, reducing errors and improving candidate selection.

Nassif (2020) focused on the application of natural language processing (NLP) in resume analysis. Their study demonstrates how NLP techniques can be employed to accurately extract and interpret information from resumes, which is crucial for the effectiveness of a resume analyzer. They discussed various NLP methods, such as named entity recognition, text classification, and sentiment analysis, and their applications in resume parsing and candidate evaluation.

Williams and Thomas (2023) addressed the challenges associated with managing diverse resume formats in automated systems. Their research underscores the importance of developing versatile algorithms capable of handling various resume structures to avoid overlooking qualified candidates. They discussed the technical difficulties of parsing and standardizing information from different resume formats and propose solutions to enhance the robustness and flexibility of automated resume analyzers.

**2.7 Review on Technological Frameworks and Implementation Strategies**

The Natural Language Processing (NLP), Machine Learning (ML), Optical Character Recognition (OCR), and related technologies are fundamental to automating and optimizing the resume analysis process, enhancing accuracy, efficiency, and scalability in recruitment systems.

**1. Natural Language Processing (NLP)**

NLP is at the core of resume analyzers, enabling systems to process and understand human language as it appears in resumes. NLP tasks crucial for resume analysis include Named Entity Recognition (NER), text classification, keyword extraction, and semantic similarity analysis (Huang, Xu, & Yu, 2015).

* **Named Entity Recognition (NER)**: NER involves identifying and classifying entities in text, such as names, job titles, locations, dates, and other pertinent information. Advanced NER systems often utilize algorithms like Conditional Random Fields (CRF) and deep learning models such as Bidirectional Long Short-Term Memory networks (BiLSTM) coupled with CRF layers. These models are trained on labeled datasets, enabling them to recognize and categorize entities with high precision, which is critical in extracting key information from resumes (Huang, Xu, & Yu, 2015).
* **Text Classification**: Text classification involves categorizing text into predefined labels, which in the context of resume analysis might include job roles, skills, or educational qualifications. Classical algorithms like Support Vector Machines (SVM) and Naive Bayes classifiers are often employed, though deep learning models such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) have gained prominence for their ability to handle complex and high-dimensional data. These models are particularly effective in classifying text based on context and semantic meaning, improving the relevance of the classification results.
* **Keyword Extraction**: Keyword extraction is crucial for matching resumes to job descriptions. Traditional techniques like Term Frequency-Inverse Document Frequency (TF-IDF) are commonly used, but more advanced methods leverage transformer-based models like BERT (Bidirectional Encoder Representations from Transformers) and GPT (Generative Pre-trained Transformer). These models excel in capturing the context and significance of words in a text, enabling more accurate keyword extraction and improving the matching process between resumes and job postings.
* **Semantic Similarity Analysis**: Beyond keyword extraction, NLP models are used to assess the semantic similarity between resumes and job descriptions. Techniques such as Word2Vec and Sentence-BERT (a variant of BERT) enable the system to understand the nuances in language, allowing it to match resumes with job descriptions even when exact keyword matches are absent. This leads to more effective candidate screening by capturing the underlying intent and meaning in text.

**2. Machine Learning (ML)**

Machine Learning algorithms are pivotal in resume analyzers, driving the learning process from large datasets of resumes and job descriptions to improve candidate matching and scoring (Lafferty, McCallum, & Pereira, 2001).

* **Supervised Learning**: In supervised learning, models are trained on labeled datasets where the desired output is known. Common algorithms include Decision Trees, Random Forests, Gradient Boosting Machines (e.g., XGBoost), and deep learning models like Multilayer Perceptrons (MLPs). These models are employed to predict the suitability of candidates for specific roles based on historical data, learning patterns that indicate successful hires.
* **Unsupervised Learning**: Unsupervised learning techniques, such as clustering algorithms like K-Means and DBSCAN (Density-Based Spatial Clustering of Applications with Noise), are used to identify patterns and group similar resumes without labeled outcomes. This approach helps in organizing large volumes of resumes into clusters based on similarities, which can be useful for identifying potential candidates for roles with less defined criteria.
* **Feature Engineering**: Feature engineering is a critical step in ML, involving the selection, modification, or creation of new features from raw data to enhance model performance. In resume analysis, features might include years of experience, specific skill sets, educational qualifications, and industry-specific keywords. Effective feature engineering can significantly improve the predictive accuracy of ML models, ensuring that the most relevant factors are considered in candidate evaluations.
* **Deep Learning**: Deep learning models, particularly CNNs, RNNs, and their variants like LSTMs (Long Short-Term Memory networks), are increasingly used in resume analysis for tasks that require capturing sequential or hierarchical information in text. These models are capable of learning complex patterns in data, making them well-suited for tasks like text classification and sequence prediction in resumes.

**3. Optical Character Recognition (OCR)**

OCR technology is essential for converting non-editable document formats, such as scanned PDFs or images, into machine-readable text, allowing resume analyzers to process a wide variety of resume formats (Huang, Xu, & Yu, 2015).

* **Tesseract OCR**: Tesseract is an open-source OCR engine widely used for text recognition in images and PDFs. It employs a combination of machine learning techniques and pattern recognition, utilizing LSTM networks to enhance its accuracy in recognizing complex fonts, languages, and layouts. Tesseract is particularly effective in scenarios where resumes are scanned documents or images, converting them into digital text for further processing.
* **Deep Learning-Based OCR**: More advanced OCR systems leverage deep learning models, particularly CNNs, to improve text recognition accuracy across diverse document formats. These systems are trained on extensive datasets of labeled images, enabling them to handle varying font sizes, orientations, noise levels, and document structures with greater precision. By incorporating deep learning, OCR systems can achieve higher accuracy in extracting text from complex documents, making them more reliable for resume analysis.

**4. Algorithms and Models**

The effectiveness of resume analyzers is heavily dependent on the underlying algorithms and models used to process and analyze resume data (Huang, Xu, & Yu, 2015).

* **Transformer Models (e.g., BERT, GPT)**: Transformer models like BERT and GPT are highly effective in NLP tasks due to their ability to understand the context and semantics of text. BERT, for example, is bidirectional, meaning it considers the entire context of a word within a sentence, making it particularly useful for tasks like entity recognition, text classification, and keyword extraction in resumes. GPT, with its generative capabilities, can also be used to predict and generate text, making it valuable for tasks that require language understanding and generation.
* **Deep Learning Models**: In addition to transformers, deep learning models like CNNs and RNNs (including LSTMs) are employed for their ability to capture patterns in data. CNNs, with their hierarchical structure, are particularly effective in tasks like text classification, while RNNs and LSTMs excel in handling sequential data, making them ideal for processing the structured text in resumes.
* **Ensemble Methods**: Ensemble methods, such as Random Forests and Gradient Boosting, combine the outputs of multiple learning algorithms to improve overall model performance. These methods are often used in candidate scoring systems, where the goal is to aggregate predictions from various models to arrive at a more robust and accurate assessment of a candidate's suitability for a job role.

**5. Integration and Deployment**

The successful deployment of a resume analyzer requires careful integration of NLP, ML, and OCR technologies within a scalable and efficient system architecture (Huang, Xu, & Yu, 2015).

* **Microservices Architecture**: Adopting a microservices architecture allows for the independent deployment and scaling of different components, such as NLP, ML, and OCR modules. This modular approach ensures that the system can efficiently handle large volumes of resumes, with each service responsible for a specific function. For example, the OCR service might handle document conversion, while the NLP service processes text and extracts relevant information.
* **Cloud-Based Infrastructure**: Cloud platforms like AWS, Google Cloud, and Microsoft Azure provide the necessary infrastructure for deploying and scaling resume analyzers. These platforms offer services for data storage, processing, and machine learning, enabling continuous model training and updating. Cloud-based deployment also ensures high availability and allows the system to scale dynamically based on the workload, ensuring consistent performance during peak usage periods.
* **Continuous Integration and Deployment (CI/CD)**: Implementing CI/CD pipelines ensures that updates to models, algorithms, or system components are deployed seamlessly without disrupting the service. This approach enables the rapid iteration and improvement of resume analyzers, ensuring that they stay up-to-date with the latest advancements in NLP, ML, and OCR technologies.

**CHAPTER 3**

**SYSTEM ANALYSIS AND DESIGN**

**3.1 Introduction**

Resume analysis necessitates a comprehensive approach to system analysis and design. This process involves understanding the requirements, defining the architecture, designing the database, and mapping out the system flow to ensure a cohesive and efficient implementation. By leveraging advanced technologies such as artificial intelligence (AI), machine learning (ML), natural language processing (NLP), and optical character recognition (OCR), the system aims to automate and enhance the recruitment process, making it more effective and scalable.

The primary goal of the resume analyzer is to streamline the recruitment process by automating the initial screening of resumes. This involves parsing and analyzing large volumes of resumes to extract relevant information, scoring the resumes based on predefined criteria, and providing recommendations to recruiters. The system must be designed to handle various resume formats, ensure data accuracy, and maintain the integrity and confidentiality of candidate information.

To achieve these objectives, the system architecture must be robust and scalable, capable of integrating various technological components seamlessly. The database design must facilitate efficient data storage and retrieval, ensuring that the system can manage large datasets without compromising performance. Additionally, the system flow must be well-defined to ensure smooth interactions between different components and provide a user-friendly experience for recruiters.

This chapter will delve into the detailed design of the resume analyzer system. It will begin with an overview of the system architecture, followed by a breakdown of the database design. Finally, it will present a system flow diagram to illustrate the overall process and interactions within the system.

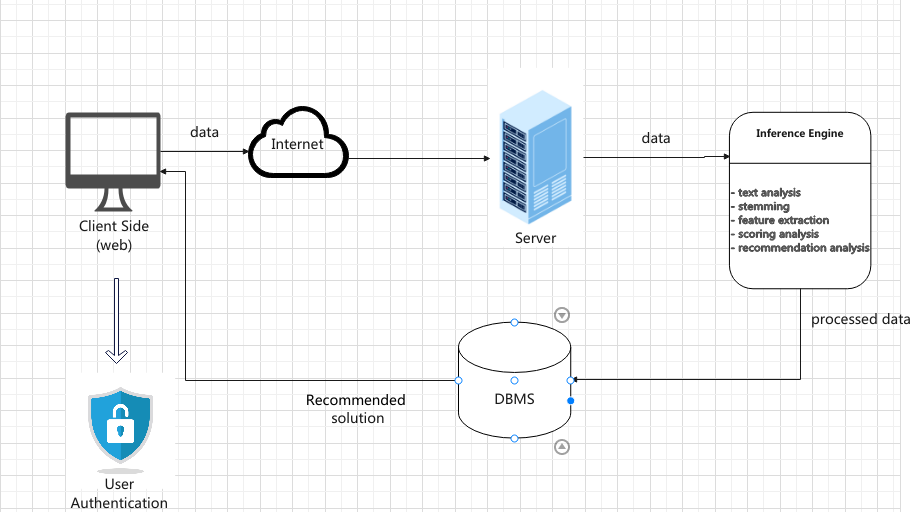
**3.2 Research Methodology**

The project will use the following methodology:

1. Conduct an extensive review of existing literature on resume screening technologies, OCR, NLP, and ML applications in recruitment.
2. Review industry reports and case studies to understand current trends and innovations in recruitment technologies.
3. Design a resume analyzer using Streamlit, (a Python framework for creating web applications).
4. Add OCR, NLP, and ML interface to extract and analyze information from resumes.
5. Implement the designs above
6. Evaluate the performance of the system using accuracy sensitivity precision and F1-score.
7. Conduct unit tests to validate individual components such as OCR, NLP, and ML models.
8. Evaluate the system's efficiency in terms of processing time and recruiter satisfaction.
9. Assess the impact of the resume analyzer on recruitment outcomes such as hiring time, candidate quality, and diversity.

**3.3 System Architecture**

The system architecture shown in Figure 3.1 comprises three core components: the frontend interface, backend processing, and database management. A robust database management system ensures secure data storage and retrieval.



**Figure 3.1;** System Architecture

The system architecture depicted in the diagram is composed of several interconnected modules, each playing a crucial role in processing and analyzing data to deliver a recommended solution. Here's a detailed description of each module:

**1. Client Side (Web)**: This module represents the user interface where users (typically recruiters) interact with the system. It is a web-based platform that allows users to upload resumes, input data, and receive recommendations. The client side is responsible for sending user data to the server for further processing.

**Functionality**

* Upload resumes and other relevant documents.
* Display the analyzed results and recommendations to the user.
* Facilitate user interactions with the system through a user-friendly interface.

**2. User Authentication**: User authentication is a security module that ensures only authorized users can access the system. It verifies the identity of users before allowing them to interact with the system’s functionalities.

**Functionality**:

* Protects the system from unauthorized access.
* Ensures data privacy and security.
* Manages user credentials and permissions.

**3. Server**: The server acts as the central processing unit of the system. It receives data from the client side, processes it, and interacts with other modules such as the database and inference engine.

**Functionality:**

* Handles data requests and responses.
* Manages communication between the client side and the database.
* Executes server-side logic, including preprocessing of data before analysis.

**4. Inference Engine**: The inference engine is a critical module responsible for analyzing the data received from the server. It performs various types of analyses to extract meaningful insights from the data, which are then used to generate recommendations.

**Components**:

* **Text Analysis**: Processes the text extracted from resumes to identify relevant information such as skills, experience, and qualifications.
* **Stemming**: Involves reducing words to their base or root form
* **Feature Extraction**: Identifies key features or attributes from the resumes that are crucial for evaluating a candidate's suitability for a job.
* **Scoring Analysis**: Applies machine learning models to score and rank the resumes based on predefined criteria such as relevance, quality, and fit for the role.
* **Recommendation Analysis**: Based on the analysis, the system generates recommendations such as potential training resources or suitable job roles for the candidate.

**Inference Engine Algorithm:**

**Step 1:** Start Process

* Begin the inference process for analyzing resumes.

**Step 2:** Receive Resume Data

* Collect the text data extracted from the uploaded resumes.

**Step 3:** Analyze Text

* Tokenize the Text: Break the resume text into smaller pieces like words or phrases.
* Remove Stopwords: Eliminate common words that do not add significant meaning (e.g., "and," "the").
* Identify Key Entities\*\*: Use Named Entity Recognition (NER) to find important information such as names, job titles, and locations.

**Step 4:** Apply Stemming

* Convert words to their basic root forms to ensure consistency (e.g., "running" becomes "run").

**Step 5:** Extract Important Features

* Identify Key Information: Extract crucial details such as skills, job roles, education, and years of experience.
* Calculate Term Importance: Use TF-IDF to determine how important certain words or phrases are within the resume.
* Tag Parts of Speech: Label words as nouns, verbs, etc., to help with understanding their role in the text.

**Step 6:** Score the Resumes

* Load Machine Learning Models: Use pre-trained models to evaluate resumes.
* Input Key Features: Feed the extracted information into the models.
* Assign Scores: Calculate a score for each resume based on how well it matches the job criteria.
* Rank Candidates: Organize the resumes by their scores, from best to least suitable.

**Step 7:** Generate Recommendations

* Suggest Training Resources: Recommend any additional training or certifications the candidate might need.
* Present Suggestions: Show relevant recommendations to help with hiring decisions.

**Step 8:** Store and Display Results

* Save the Data: Store the analysis results and scores in the database.
* Show Results: Display the resume scores and recommendations to the user via the web interface.

**Step 9:** End Process

* Finish the inference process and prepare for the next analysis.

**5. DBMS (Database Management System)**: The DBMS module is responsible for storing and managing all data within the system. It includes storing user-uploaded resumes, processed data, analysis results, and recommendations.

**Functionality:**

* Stores raw and processed data securely.
* Manages data retrieval and storage operations.
* Facilitates the integration of data with the inference engine for analysis.

**6. Recommended Solution**: This module represents the output of the system after the data has been analyzed. It provides the final recommendations based on the resume analysis, which are then displayed to the user on the client side.

**Functionality**:

* Delivers tailored recommendations such as job matches, skill development resources, or interview questions.
* Presents the results in an easily understandable format to the user.
* Supports decision-making processes for recruiters by providing actionable insights.

**Overall Flow:**

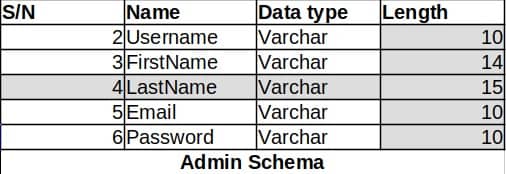
* Data Transmission: The system starts with the client side, where data is entered and sent to the server. The server then processes this data, which is further analyzed by the inference engine. The processed data is stored in the DBMS, and the final recommended solution is sent back to the client side for display to the user.
* Data Processing and Security: Throughout the system, user authentication ensures secure access, and the server manages all data transactions. The inference engine plays a pivotal role in transforming raw data into meaningful recommendations.

This architecture ensures a robust, secure, and efficient system for analyzing resumes and providing valuable insights to recruiters.

**3.4 Database Design**

A database schema defines the structure of a database in a formal language that the database management system supports. It outlines how data is organized, serving as a blueprint for constructing the database, particularly in relational databases, where the data is divided into various tables. The database schema of the proposed system is shown in Table 3.1 and Table 3.2 below.

Table 3.1: The Admin Schema

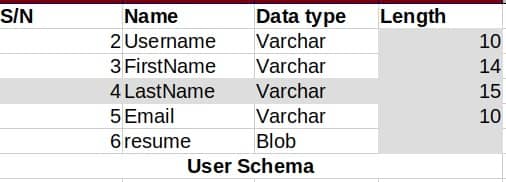


The Admin Schema, as illustrated in Table 3.1, is designed to manage administrator accounts for the resume analyzer system. This schema facilitates the storage and retrieval of essential information for system administrators, enabling secure authentication and user management. The schema consists of the following fields:

1. S/N: A unique identifier for each admin record.
2. Username (Varchar, length 10): Stores the admin's username, limited to 10 characters.
3. FirstName (Varchar, length 14): Contains the admin's first name, allowing up to 14 characters.
4. LastName (Varchar, length 10): Holds the admin's last name, with a maximum of 10 characters.
5. Email (Varchar, length 10): Stores the admin's email address, limited to 10 characters.
6. Password (Varchar, length 10): Contains the admin's password for authentication purposes.

This schema structure allows for efficient management of administrator accounts within the resume analyzer system. It provides the necessary fields for basic user information and authentication, supporting the system's security and user management functionalities.

Table 3.2: The User Schema



The User Schema, as illustrated in Table 3.2, is designed to store and manage information about users of the resume analyzer system. This schema facilitates the storage of essential user data and their resumes. The schema consists of the following fields:

1. S/N: A unique identifier for each user record.
2. Username (Varchar, length 10): Stores the user's chosen username, limited to 10 characters.
3. FirstName (Varchar, length 14): Contains the user's first name, allowing up to 14 characters.
4. LastName (Varchar, length 15): Holds the user's last name, with a maximum of 15 characters.
5. Email (Varchar): Stores the user's email address, allowing up to 10 characters,
6. resume (Blob): A binary large object (BLOB) field to store the user's resume file directly in the database.

This schema structure allows for efficient management of user accounts and their associated resumes within the resume analyzer system. It provides the necessary fields for basic user information and authentication, while also incorporating the capability to store the actual resume document.

**3.3 Algorithm**

Before delving into the specific steps of the algorithm, it is essential to understand the underlying process that drives the resume analysis system. This system is designed to automate and streamline the evaluation of resumes by leveraging advanced technologies such as Optical Character Recognition (OCR), Natural Language Processing (NLP), and Machine Learning (ML). Each step of the algorithm plays a crucial role, from the initial input of resumes to the final output of candidate recommendations. By following this structured approach, the system ensures that resumes are not only processed efficiently but also evaluated with a high degree of accuracy, ultimately enhancing the recruitment process for organizations.

1. Start
2. Upload Resume
3. Validate the file format (PDF, DOCX, etc.).
4. Retrieve the uploaded resume to extract text.
5. Analyze the text and parse the extracted text to identify key information.
6. Store parsed text information.
7. Score the Resume
8. Provide recommendations by suggesting relevant online resources for interview questions or tests.
9. Display the analyzed resume, score, and recommendations through the interface.
10. End Process and Log the session.

**3.5 Flow Chart**

A flow chart is a visual representation of a process, outlining the sequence of steps involved in a particular system or workflow. It uses symbols and arrows to depict the flow of tasks and decisions, providing a clear and structured way to understand how a process unfolds from start to finish. Flow charts are particularly useful for identifying potential bottlenecks, optimizing processes, and communicating complex procedures in a simple, easy-to-follow manner. The flow chart of the system is shown in Figure 1.2.

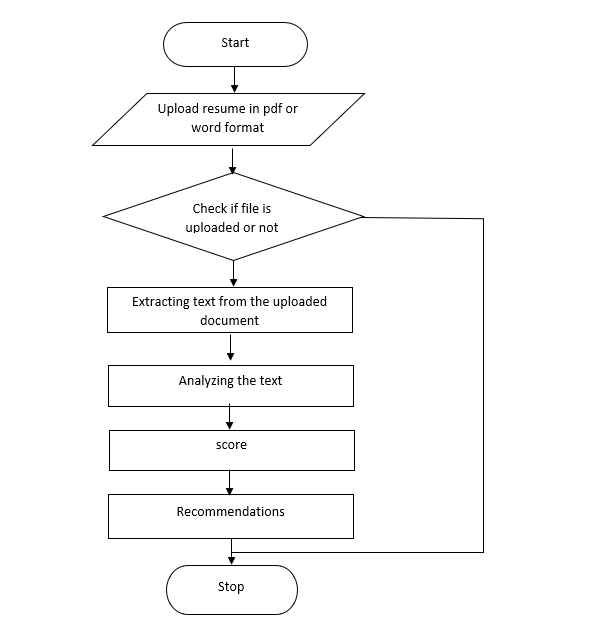


Figure 3.3: System Flow chart

**CHAPTER FOUR**

**SYSTEM IMPLEMENTATION**

**4.0 Introduction**

This chapter provides a detailed overview of the development process for the resume analyzer for a recruiting company. It outlines the software tools and technologies used, along with the hardware and software requirements needed for implementation. Additionally, this chapter includes snapshots of the system’s key functions, showcasing its capabilities. The aim is to offer a clear understanding of the technical and operational aspects that contribute to the system’s effectiveness in streamlining the recruitment process.

**4.1 Software Requirements**

The following tools were utilized in the implementation of this system:

1. Visual Studio Code (text editor): This free, open-source code editor is available for macOS, Linux, and Microsoft Windows. It supports various plug-ins written in Node.js and includes embedded Git control. Developed by GitHub, Visual Studio Code is a desktop application built with web technologies and was used as the integrated development environment (IDE) for the proposed system.
2. Programming Languages: The system was developed using Streamlit.
3. Operating System: The implementation was compatible with multiple operating systems, including Windows XP, Windows 7, Windows 8, Linux, macOS, and Android (kernel) for mobile devices.
4. Web Browser: The system is optimized for use with updated versions of popular browsers, such as Microsoft Edge, Chrome, and Mozilla Firefox.

**4.2 Hardware Requirements**

To ensure optimal performance and reliability in supporting the software tools for the implementation of the proposed systems, the following hardware specifications are recommended:

A laptop or desktop computer should meet the following minimum criteria:

1. A minimum of 1GB of RAM (Random Access Memory)
2. At least 10GB of free disk space
3. A processor with a clock speed of no less than 1.0GHz

**4.3 Choice of Programming Languages and Libraries**

The application uses Streamlit for the frontend, Pandas for data handling, and Base64 for encoding. Pymysql manages MySQL interactions, Plotly creates visualizations, and Geopy handles geocoding. PDFMiner extracts text from PDFs, Pyresparser parses resumes, and Streamlit-tags manages tags. PIL processes images, Nltk supports text processing, and Datetime handles date and time. Below is a brief description of the languages and frameworks utilized in the development of this web application.

1. **Streamlit:** Streamlit is the core package used to create the web application. It is widely employed for building and deploying machine learning and data science applications with interactive user interfaces (UIs). Key features of Streamlit include its simplicity in creating web applications, automatic updates for UI changes, and seamless integration with various data science libraries.
2. **Pandas:** Pandas is used for data manipulation and analysis. It provides essential data structures, such as DataFrames, for efficiently handling and processing tabular data. Key features of Pandas include its powerful data manipulation capabilities, support for a variety of data formats, and extensive functionality for data cleaning and preparation.
3. **Base64:** Base64 is utilized for encoding and decoding data, particularly for converting binary data into text format. This is especially useful for downloading files, such as CSVs, in a web application. Key features include ease of converting binary data to text and compatibility with various data formats.
4. **Pymysql:** Pymysql is a library used to interact with MySQL databases. It facilitates various database operations, including querying, inserting, and managing data. Key features of Pymysql include its ability to interface with MySQL databases using Python, support for multiple database operations, and ease of integration with other Python libraries.
5. **Plotly:** Plotly is used to create interactive visualizations, such as graphs and charts, within the admin session of the application. Key features of Plotly include its ability to generate rich, interactive visualizations, support for various chart types, and integration with web applications for dynamic data representation.
6. **Geopy:** Geopy is utilized for geocoding tasks, which involve converting addresses into geographic coordinates and vice versa. Key features of Geopy include support for various geocoding services, ease of integration into Python applications, and functionality for handling geographic data.
7. **PDFMiner:** PDFMiner is a library used to parse and extract text from PDF files. It is particularly useful for reading and processing documents like PDF resumes. Key features of PDFMiner include its ability to handle complex PDF structures, support for text extraction, and functionality for analyzing document layout.
8. **Pyresparser:** Pyresparser is specifically designed for parsing resumes. It extracts information such as names, skills, and experience from various resume files. Key features of Pyresparser include its specialized parsing capabilities, support for multiple resume formats, and functionality for extracting relevant candidate information.
9. **Streamlit-tags:** Streamlit-tags is a custom Streamlit component used to create tags or keyword input fields within the application. Key features of Streamlit-tags include its ability to enhance user input functionality, support for dynamic tag creation, and integration with Streamlit’s interactive elements.
10. **PIL (Python Imaging Library):** PIL, or Python Imaging Library, is used for image processing tasks, such as opening, manipulating, and saving image files. Key features of PIL include support for a wide range of image formats, powerful image manipulation capabilities, and ease of integration into Python applications.
11. **Nltk (Natural Language Toolkit):** Nltk is a library for natural language processing. It supports various text processing tasks, including tokenization and stopword removal. Key features of Nltk include its extensive set of NLP tools, support for various text processing tasks, and functionality for building and training language models.
12. **Datetime:** Datetime provides functionalities for working with dates and times. It is often used to timestamp events or logs within an application. Key features of Datetime include its ability to handle a variety of date and time formats, support for date and time arithmetic, and functionality for formatting and parsing dates and times.

**CHAPTER FIVE**

**SUMMARY, RECOMMENDATION, AND CONCLUSION**

**5.0 Introduction**

This chapter provides a summary and conclusion regarding the development of a resume analyzer designed for a recruitment firm. It also addresses the obstacles faced throughout the project's execution and offers pertinent recommendations based on the findings.

**5.1 Summary**

The project aims to develop a sophisticated web-based system designed to streamline the recruitment process by automatically analyzing resumes using advanced machine learning algorithms. The system addresses the challenge of efficiently screening large volumes of resumes, which often results in bias and inefficiency. By leveraging natural language processing (NLP) techniques, the system extracts relevant features from resumes and ranks candidates based on their qualifications and experience. The system integrates a user-friendly interface developed using Streamlit, enabling seamless data processing, algorithm execution, and result generation. This resume analyzer is designed to provide recruiters with a reliable tool to identify the most suitable candidates, reducing the time and effort required in the initial screening phase while ensuring a fair and unbiased selection process.

**5.2 Recommendations:**

User-Friendly Interface: Prioritizing a user-friendly interface is crucial for the success of the resume analyzer. The interface should allow for easy uploading of resumes and provide clear, intuitive visualizations of candidate analysis results. Implementing features that enhance the user experience, such as simple navigation and accessible explanations of the scoring criteria, will improve the effectiveness of the system and ensure that recruiters can use it with minimal training.

Security Measures: Implementing strong security measures is essential to protect sensitive candidate information. The system should incorporate robust encryption protocols and authentication mechanisms to safeguard data during storage and transmission. Ensuring the confidentiality and integrity of candidate data is vital to maintaining trust and compliance with data protection regulations.

Scalability: The system should be designed with scalability in mind to handle an increasing number of users and a growing volume of resumes. Future-proofing the system will ensure it can adapt to the evolving needs of the recruitment process, accommodating new features, larger datasets, and potential integrations with other HR tools.

Collaboration with HR Professionals: Close collaboration with HR professionals and recruitment experts is recommended to continually refine the resume analyzer. Incorporating their insights into the development process will help enhance the system’s accuracy, relevance, and usability, ensuring it meets the practical needs of the recruitment industry and remains aligned with best practices.

**Conclusion**

The developed web-based resume analyzer represents a significant advancement in modernizing the recruitment process. By integrating advanced machine learning algorithms with a Streamlit interface, the system offers a robust platform for automating resume screening, ensuring seamless user interaction and efficient processing of candidate data. The system has the potential to greatly enhance the efficiency and fairness of recruitment by reducing bias and streamlining the initial candidate selection process. The recommendations provided aim to further refine the system's functionality, security, scalability, and alignment with HR best practices. Ultimately, this project strives to make a meaningful contribution to the recruitment industry by providing a reliable tool that supports more informed and equitable hiring decisions.

**REFERENCE**

Bogen, M., & Rieke, A. (2018). Help wanted: *An examination of hiring algorithms, equity, and bias. Upturn.* [*https://www.upturn.org/reports/2018/hiring-algorithms/*](https://www.upturn.org/reports/2018/hiring-algorithms/)

Brown, J. (2021). The impact of high application volumes on recruitment efficiency*: A case for automation. Journal of Human Resource Management, 12(4), 315-329.* <https://doi.org/10.1016/j.jhrm.2021.04.006>

Easmat, J., Md, S., & Mohammad, N. (2024). Artificial intelligence in migrant labor management: *Candidate matching and skill assessment, 3(1)*

Gonzalez, R., Smith, K., & Andrews, L. (2021). Fairness in automated hiring systems*:* Ethical implications and solutions. *Ethics and Information Technology, 23(1), 61-74.* <https://doi.org/10.1007/s10676-020-09535-4>

Huang, Z., Xu, W., & Yu, K. (2015). Bidirectional LSTM-CRF models for sequence tagging. *arXiv preprint*, arXiv:1508.01991. <https://arxiv.org/abs/1508.01991>

Johnson, P. (2020). Time management in recruitment: *How automation improves the hiring process. International Journal of Human Resource Studies, 10(2), 29-45.* <https://doi.org/10.5296/ijhrs.v10i2.16754>

Kim, S., & Patel, R. (2022). Reducing human bias in resume screening: *The role of AI-driven systems. Journal of Applied Psychology, 107(3), 548-561.* [*https://doi.org/10.1037/apl0000901*](https://doi.org/10.1037/apl0000901)

Lafferty, J., McCallum, A., & Pereira, F. (2001). Conditional random fields: Probabilistic models for segmenting and labeling sequence data. *Proceedings of the Eighteenth International Conference on Machine Learning*, 282-289.

Lee, T. (2021). Manual vs. automated resume evaluations: *A comparison of accuracy and consistency. Human Resource Management Review, 31(3), 287-301.* <https://doi.org/10.1016/j.hrmr.2021.100787>

Liem, C., Langer, M., & Martin, C. (2018). Accuracy of automated resume parsing: *A comparative study. IEEE Transactions on Knowledge and Data Engineering, 30(8), 1552-1566.* [*https://doi.org/10.1109/TKDE.2018.2809470*](https://doi.org/10.1109/TKDE.2018.2809470)

*Morgan, M. (n.d.). Types of resumes. Indeed.* [*https://www.indeed.com/career-advice/resumes-cover-letters/types-of-resumes*](https://www.indeed.com/career-advice/resumes-cover-letters/types-of-resumes)

Nassif, A. (2020). The application of natural language processing in resume analysis. *Journal of Computational Linguistics, 46(2), 199-214.* [*https://doi.org/10.1162/coli\_a\_00375*](https://doi.org/10.1162/coli_a_00375)

Nguyen, H. (2023). Data privacy in modern recruitment processes: *A critical analysis. Journal of Information Security and Privacy, 18(1), 45-63.* [*https://doi.org/10.4018/JISP.2023010104*](https://doi.org/10.4018/JISP.2023010104)

Smith, R. (2022). Overcoming inefficiencies in traditional resume screening methods: *The case for automation. Journal of Human Capital Management, 15(2), 142-156.* [*https://doi.org/10.1016/j.jhcm.2022.05.008*](https://doi.org/10.1016/j.jhcm.2022.05.008)

Suen, H. Y., Cheung, G. W., & Sohn, E. (2019). The challenges of manual resume screening in the digital age: Solutions through automation. Computers in Human Behavior, 97, 107-114. <https://doi.org/10.1016/j.chb.2019.02.015>

Williams, A., & Thomas, J. (2023). Managing diverse resume formats in automated systems: Technical challenges and solutions. Journal of Artificial Intelligence Research, 76, 45-60. <https://doi.org/10.1613/jair.1.13790>