DATA-DRIVEN RESUME ENHANCER: LEVERAGING ML AND NLP

Kavya Nandigam
Purdue University
Fort Wayne, Indiana
nandk02@pfw.edu

Karthik Balaji
Purdue University
Fort Wayne, Indiana
balak02@pfw.edu

Abstract

In the contemporary landscape of job application processes, Applicant Tracking Systems (ATS) wield considerable influence, predominantly relying on keyword matching to filter and shortlist resumes. However, this reliance poses a significant challenge for job seekers who grapple with optimizing their resumes with the right keywords, often resulting in diminished scores and missed opportunities. Moreover, the arduous task of crafting unique and compelling resumes tailored to each job description further exacerbates the difficulties, particularly for early-career applicants. While customizing resumes to align with specific job requirements is strongly advised, manual tailoring proves to be both time-consuming and susceptible to errors. This paper addresses these challenges by proposing an automated approach to resume tailoring. By leveraging natural language processing (NLP) techniques and machine learning algorithms, our system streamlines the process of aligning resumes with job descriptions, reducing the time and effort required while enhancing the quality and relevance of tailored resumes. Through empirical evaluation, we demonstrate the efficacy of our approach in improving job application success rates for candidates across diverse career stages.

1 Introduction

1.1 Motivation

Our motivation stems from the imperative to streamline and enhance the resume crafting process through innovative technological solutions. By leveraging the power of Natural Language Processing (NLP) and Natural Language Generation (NLG) techniques, we aim to automate the identification of relevant skills and experiences from both resumes and job descriptions. Our goal is to empower job seekers with a tool that not only expedites the resume tailoring process but also increases their chances of getting noticed by recruiters and securing interviews. By automating the identification and highlighting of key qualifications, our solution seeks to improve the effectiveness and efficiency of the job application process, ultimately providing job seekers with a competitive edge in their career pursuits.

Furthermore, our motivation extends beyond individual benefits to encompass broader societal implications. By democratizing access to advanced resume customization tools, we aim to promote inclusivity and equal opportunity in the job market, leveling the playing field for candidates from diverse backgrounds.

1.2 Problem Statement

Crafting a compelling resume is a critical task for job seekers amidst today's competitive market. Distilling years of education, projects, and experiences into a concise 1 or 2-page document while showcasing individuality poses a significant challenge. With the proliferation of candidates vying for positions, employers increasingly rely on automated Applicant Tracking Systems (ATS) and resume filtering tools to screen applicants based on job match criteria. In response, candidates must invest time and effort in tailoring their resumes to specific job postings, emphasizing relevant experiences, skills, projects, and achievements. However, this process is not only challenging but also time-consuming, as candidates grapple with identifying impactful keywords and maintaining conciseness. Although automated tools exist for analyzing and extracting information from resumes, there is currently no automated solution empowering job applicants to tailor their resumes to specific job openings. This gap underscores the need for innovative solutions that streamline the resume customization process, ultimately enhancing candidates' competitiveness in the job market.

This research offers a practical solution to the prevalent hurdles faced by job seekers, paving the way for more efficient and effective job application strategies in the digital age. This paper presents an innovative approach harnessing the capabilities of Natural Language Processing (NLP) and Natural Language Generation (NLG) techniques to streamline the resume tailoring process. Our objective is to automate the identification of pertinent skills and experiences from both resumes and job descriptions, thereby facilitating the enhancement of resumes by accentuating these critical qualifications. Through this automation, our solution aims to augment the scores of resumes within Applicant Tracking Systems (ATS), consequently elevating candidates' visibility to recruiters and increasing their prospects of securing interviews. This research offers a promising avenue for enhancing job application outcomes by leveraging cutting-edge NLP and NLG methodologies.

2 Related Work

The advancement of natural language processing and text mining techniques has led to various applications in tasks such as resume analysis, classification, and information retrieval. Some recent efforts focus on parsing and extracting information from resumes (Tallapragada et al., 2023) and performing job matching (Yi et al., 2007). For instance, researchers (Chen et al., 2018) have proposed a two-step approach for information extraction from resumes by identifying information blocks initially. However, despite these advancements, there is a notable gap in the literature concerning the automatic generation of job-aligned resumes. While there have been recent studies utilizing Large Language Models (LLMs) as instruction-following assistants for tasks like information retrieval and educational purposes, there remains untapped potential in leveraging LLMs for resume generation tailored to specific job descriptions. Although there is related work (Skondras et al., 2023), a recent study where structured and unstructured resumes were generated using ChatGPT, it primarily focused on data generation for downstream classification tasks rather than customizing resumes for specific job roles. Hence, our proposed approach aims to fill this gap by leveraging state-of-the-art LLMs for the personalized generation of job-aligned resumes.

3 Dataset

For this project, the Resume dataset is downloaded from Kaggle[1], which is a collection of resumes obtained from diverse sources. Each resume offers a comprehensive overview of an individual's professional background, skills, education, and work experience. The dataset is structured, featuring several key attributes that can facilitate a wide range of analyses and applications. The dataset includes columns *Category* and *Resume*, which contain textual descriptions of individuals' qualifications and proficiencies. These descriptions offer detailed insights into the candidates' expertise and capabilities. Given the nature of the dataset, it is particularly suited for tasks related to talent acquisition, job matching, and skills assessment. The *Resume* column serves as a rich source of information for training models to predict job suitability, match candidates with suitable positions, and extract relevant skills and qualifications. In this project, we will focus primarily on utilizing the *Resume* column for training our models. Delving deeper into the dataset metrics, we find that it consists of 962 entries, encapsulating various professional backgrounds. Among these entries, there exist 166 distinct resumes represented within the dataset. A mere 0.17% of the data aren't duplicate entries. Among the diverse array of resume entries in the dataset, we find that they are categorized into a concise set of 25 distinct categories. Despite the wide range of professional backgrounds and experiences represented within the dataset, the resumes are effectively organized into these specific categories, providing a structured framework for analysis and exploration. Further analysis reveals that the most prevalent category within the dataset is "Java Developer," appearing with a frequency of 84 instances. Conversely, the *Resume* column exhibits a wide array of textual content, ranging from technical skill descriptions to educational qualifications and project experiences. This rich textual data serves as a fertile ground for training machine learning models to predict job suitability, match candidates with appropriate positions, and extract pertinent skills and qualifications.

Another dataset, the Job Description dataset is downloaded from Kaggle[2], encompasses a diverse array of job descriptions for *Data Analyst* role

[1] https://www.kaggle.com/datasets/dhainjeamita/updatedresumedataset/data

[2] https://www.kaggle.com/datasets/lukebarousse/data-analyst-job-postings-google-search/data

sourced from various companies and job portals. With 39,714 entries, this dataset offers a comprehensive view of the requirements, responsibilities, and qualifications associated with Data Analyst positions across different industries and organizations. Within this dataset, 11,698 entries are identified as duplicates, emphasizing the need for data cleaning and deduplication processes. After eliminating duplicates, the dataset contains 28,016 unique and non-null values, ensuring a robust foundation for analysis and model training. Among the numerous columns provided, key attributes include company name, location, job posting source, salary range, and, most importantly, the job description itself. Each job description encapsulates the essential requirements, responsibilities, and qualifications associated with a particular job role. Structured for analytical purposes, the dataset comprises several informative attributes that offer valuable insights into job positions and their corresponding specifications. The key column within the dataset is *Description*, providing detailed textual descriptions of the job roles and their associated requirements. The dataset serves as a comprehensive resource for tasks related to talent acquisition, workforce planning, and job market analysis. The *Description* column, in particular, serves as a rich source of information for training models to categorize, match, and analyze job roles based on their unique characteristics and requirements. In this project, we will primarily focus on leveraging the *Description* column for model training and analysis. Given the variability in the length and complexity of job descriptions, we may explore strategies to effectively preprocess and represent the text data, ensuring compatibility with machine learning algorithms and natural language processing techniques.

4 Methodology

4.1 Exploring Resume Dataset Dynamics

The developed model exhibits a sophisticated approach to resume classification, incorporating advanced techniques from natural language processing and machine learning. By preprocessing textual data to extract meaningful features and employing state-of-the-art algorithms, the model not only accurately categorizes resumes but also enhances efficiency and scalability in the recruitment domain. Designed for analyzing both technical and non-technical resumes, the model automates initial candidate screening by classifying resumes into relevant categories. Leveraging natural language processing techniques, it preprocesses resume data by tokenizing text, removing stopwords, and encoding textual features. Subsequently, it trains machine learning algorithms, *Multinomial Naive Bayes* and *K-Nearest Neighbors*, to predict the category of resumes based on their content. Moreover, its interpretability is augmented through the visualization of a confusion matrix, offering stakeholders valuable insights into classification outcomes. This model revolutionizes traditional recruitment paradigms by introducing data-driven decision-making and talent acquisition strategies.

4.2 Baseline Resume Analysis Model

The baseline model introduced in this study represents a foundational approach towards streamlining candidate screening within the recruitment domain. Operating on the principle of *Jaccard similarity*, the model endeavors to quantify the degree of textual alignment between a given job description and candidate resumes. By employing tokenization and eliminating stopwords, the model standardizes the textual representation for subsequent analysis. Through the computation of the *Jaccard similarity coefficient*, the model provides a basic metric for assessing the extent of shared vocabulary and conceptual overlap between the job description and candidate resumes. However, it is imperative to acknowledge the inherent limitations of this approach, particularly its reliance on surface-level textual features and its inability to capture nuanced semantic relationships.

While the baseline model serves as an initial foray into automated candidate screening, its efficacy is inherently limited by its simplistic methodology. Nonetheless, its role within this research framework is pivotal, as it lays the groundwork for the subsequent introduction of a more sophisticated approach, namely *Resume Evaluator*. The primary objective of the baseline model is not to provide definitive screening outcomes but rather to underscore the necessity for more advanced techniques in resume evaluation. By delineating the comparative shortcomings of the baseline model and Resume Evaluator, this paper seeks to demonstrate the imperative for enhanced accuracy and efficiency in candidate assessment methodologies. Through this iterative refinement process, the research endeavors to advance the forefront of recruitment practices, ultimately culminating in the development of robust and reliable screening frameworks tailored to the demands of contemporary labor markets.

4.3 Resume Evaluator: Semantic Matching with Vector Databases

The model presented in this paper embodies a meticulously crafted methodology aimed at revolutionizing the efficiency and accuracy of resume-to-job matching processes. At its core lies a systematic approach comprising four distinctive steps meticulously engineered to harness the power of modern natural language processing and vector-based storage and retrieval systems.

Initiating with the preprocessing phase, the raw textual data extracted from job descriptions and resumes undergoes a rigorous cleaning and standardization process. This crucial step ensures the uniformity and integrity of the data, expunging any extraneous noise that could impede subsequent analyses, thus laying a robust foundation for the ensuing procedures. Following preprocessing, the textual information undergoes a sophisticated text summarization technique. This process distills the essence of verbose job descriptions and resumes into succinct representations, preserving critical details while significantly reducing redundancy. By condensing the textual content into concise summaries, this step facilitates more efficient processing of the data, paving the way for enhanced matching accuracy.The *Description* column in the Job Description dataset is summarized into *Technical Description* and *Non-technical Description* to assign respective weightages, thereby enriching the matching process with domain-specific relevance. Subsequently, the summarized text undergoes transformation into vector representations leveraging cutting-edge *Sentence Transformers*. This encoding process captures the nuanced semantic meanings embedded within the textual content, enabling the generation of vector embeddings that encapsulate the essence of the original text in a high-dimensional space. Both the Technical Description and Non-technical Description from the Job Description dataset, along with the *Resume* from Resume dataset, undergo encoding, ensuring that all relevant textual information is effectively represented in the vector space for comprehensive matching analysis. Finally, the encoded vectors find their home in the *Qdrant vector database*, a robust and scalable storage and retrieval system optimized for vector-based operations. Leveraging the capabilities of the Qdrant database, the model orchestrates efficient retrieval and matching of resumes to job postings based on their semantic similarity, culminating in a streamlined and scalable resume-to-job matching process. The model requires textual data extracted from both job descriptions and resumes, while the output encompasses matched resumes along with their corresponding similarity scores. Additionally, the model provides weighted scores to prioritize matches based on specific criteria tailored to the preferences of recruiters or organizations. The weightage is set at **0.7** for the Technical Description and **0.3** for the Non-technical Description, determining the relative importance of each aspect in the matching process.

This weightage scheme ensures that the model can accommodate varying preferences and priorities in the recruitment process, allowing recruiters to customize the matching criteria according to their specific needs and preferences. The weighted score calculation, as depicted by the equation, reflects this nuanced approach, where the final score is a weighted combination of technical and non-technical similarities.

In summation, the model represents a pioneering endeavor at the forefront of recruitment technology, epitomizing a fusion of advanced methodologies to redefine conventional talent acquisition paradigms. By harnessing the synergy of cutting-edge natural language processing and vector-based storage systems, this model offers a transformative approach to resume analysis and job matching, heralding a new era of efficiency and efficacy in the realm of talent acquisition.

4.4 Resume Enhancer

Our model, *Resume Enhancer*, embodies a sophisticated pipeline facilitated by Large Language Models (LLMs), aimed at the personalized generation of resumes tailored to specific job descriptions (JDs). Comprising three fundamental modules - *Resume Content Extractor*, *Job Description Extractor*, and *Resume Generator* - the system orchestrates a seamless workflow, leveraging advanced natural language processing techniques to curate bespoke resumes optimized for the targeted positions.

The initial phase of our pipeline, the Resume Content Extractor module, plays a pivotal role in transforming user-provided baseline resumes into meticulously structured JSON representations conducive to subsequent processing. Following this, the Job Description Extractor analyzes job descriptions provided as textual inputs, extracting crucial information such as job titles, qualifications, and responsibilities. This extracted data, presented in a structured JSON format, forms the bedrock for tailoring and customizing the generated resumes.

At the heart of our pipeline lies the Resume Generator component, where the essence of processing unfolds. Leveraging the structured user data and extracted job requirements, this module dynamically crafts personalized resumes finely tuned to the specifications outlined in the job descriptions. The workflow meticulously navigates through each section of the user data, prompting the designated Large Language Model (LLM) with section-specific cues and integrating the gleaned job details. Through adept curation of each resume segment in line with the job requisites, the model ensures alignment and relevance, ultimately producing modified versions of the user's resumes finely attuned to the targeted positions. For the GPT-4 models, we leverage the ChatCompletion backend to enhance the Resume Generator's capabilities. By utilizing the JSON output format, we can efficiently handle Large Language Model (LLM) outputs, simplifying the process of parsing these outputs into a Python dictionary for further processing. This integration allows us to harness the contextual understanding and language fluency provided by ChatGPT, enabling more refined and coherent resume generation tailored to specific job descriptions. Additionally, our system integrates with LaTeX to produce visually polished PDF resumes, ensuring alignment with the nuances of the job descriptions. Simultaneously, it generates structured JSON representations of the resumes, facilitating further analysis by Resume Evaluator model to evaluate for an improved match score. Figure 1 illustrates the flow of the project, depicting how the various models are interconnected and contribute to the overall process.

5 Experiments & Results

In the exploration of the dynamics within the Resume Dataset Exploration model, the observed performance is notably satisfactory. Demonstrating a commendable level of accuracy and precision (Refer Table 1 for results), the model exhibits robustness.

The distribution depicted in Figure 2 highlights the dataset's multidimensional nature, encompassing a broad spectrum of both technical and non-technical topics. This diversity underscores the complexity of the dataset and the need for comprehensive analysis techniques capable of accommodating various subject areas and their respective nuances. The substantial presence of technical data, comprising 62.4% of the dataset, suggests a strong emphasis on topics related to specialized fields such as technology, engineering, and sciences. Conversely, the non-technical category, representing 37.6% of the dataset, encompasses a diverse range of subjects beyond the realm of technical expertise. This includes areas such as sales, arts, and other interdisciplinary fields. While the proportion of non-technical data is relatively smaller compared to its technical counterpart, it nonetheless signifies the significance of diverse knowledge domains and their relevance within the dataset.

Figure 3, provides a comprehensive overview of the dataset by visually representing the count of entries within each category. It offers valuable insights into the distribution of resumes across different categories, thereby enhancing our understanding of the dataset's composition. Upon closer examination, we discern distinct clusters representing various occupational categories, each characterized by a specific frequency of entries. These categories encompass a diverse range of professions, including but not limited to Java Developer, Business Analyst, Testing, Advocate, and others. The visualization allows us to discern patterns in the distribution of resumes, identifying which categories are more prevalent within the dataset. The higher frequency of entries observed in the technical category, as indicated by Figure 3, reaffirms the trend observed in Figure 2. It underscores the dataset's emphasis on technical occupations and highlights the prominence of roles requiring specialized technical skills and expertise. Conversely, non-technical categories may exhibit comparatively lower frequencies, reflecting the diversity of roles beyond the technical domain.

Figures 4 and 5 discerns the classification of resumes into distinct categories, delineating the division between technical and non-technical domains. These pie charts serve as illustrative representations of the categorization process, wherein each resume is assigned to either the "Technical" or "Others" category based on the nature of its content and thematic orientation.

In Figure 4, we undertake a meticulous examina-

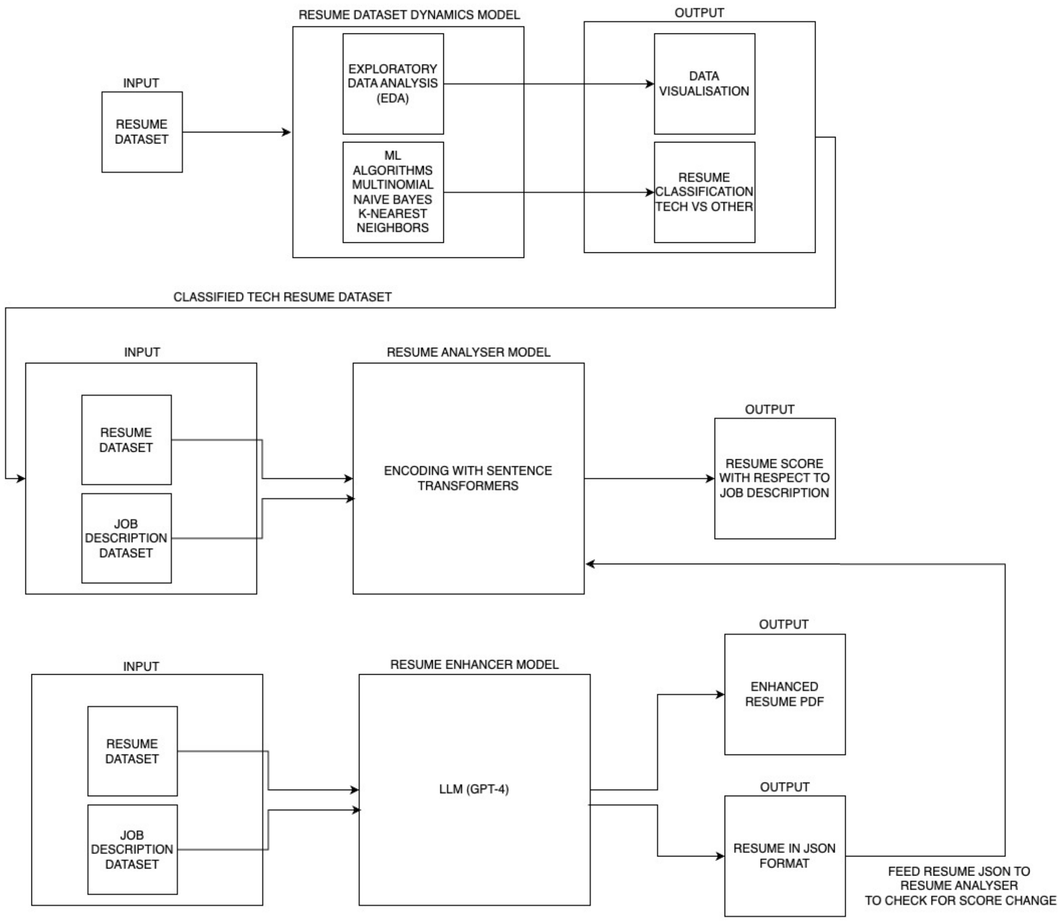


Figure 1: Overall Project Pipeline

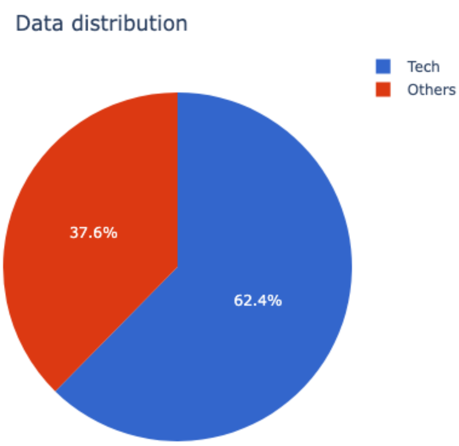


Figure 2: Resume Dataset's Data Distribtuion among Tech Vs Others

tion of the distribution within the "Others" category, which unveils a diverse array of subcategories or variations within this broader classification. This focused scrutiny affords us a nuanced understanding of the multifaceted landscape of non-technical roles encapsulated within the dataset. By dissecting the distribution within the "Others" category, we glean valuable insights into the prevalence of specific roles or themes, thus enriching our comprehension of the dataset's compositional diversity.

Shifting our focus to Figure 5, we delve into the intricacies of the distribution within the "Technical" category, providing an exhaustive portrayal of the specific areas or subjects inherent to this domain. This granular visualization equips us with a comprehensive overview of the dominant roles or topics within the technical realm, thereby furnishing invaluable guidance for subsequent analytical or modeling endeavors.

Through meticulous scrutiny of Figures 4 and 5, we attain nuanced insights into the dataset's intricate composition, discerning prevailing roles within both technical and non-technical domains. These discernments not only enrich our understanding of the dataset's diversity but also serve as catalysts for further exploration into specific categories or themes of scholarly interest.

Furthermore, the model's effectiveness is elucidated through the presentation of a confusion matrix (Refer Figure 6), affording valuable insights into the classification accuracy and potential avenues for enhancement. Collectively, this model serves as a dependable solution for the automation of resume categorization, thereby contributing significantly to the streamlining of recruitment processes and the facilitation of efficient candidate selection.

The evaluation of a chosen resume against a job description using the baseline model yielded a score of **0.0568**. This result underscores the necessity for more sophisticated models capable of providing more accurate and insightful evaluations of resume suitability for a given job description. As such, there arises a need for the development and exploration of advanced techniques that can capture deeper semantic meanings and contextual relevance, thereby enhancing the efficacy and reliability of resume screening processes in the recruitment domain.

In the context of results, it's noteworthy that a specific resume, previously evaluated utilizing the baseline model, was subjected to assessment against an identical job description using the next model, the Resume Evaluator. The outcome of this evaluation, utilizing cosine similarity as the metric, yielded a score of **0.5830**. As we proceed, the next step involves subjecting the same resume to further analysis and refinement through the application of the Resume Enhancer model. This additional step aims to iteratively improve the accuracy and relevance of the evaluation process, potentially yielding higher scores and more precise matching outcomes. The forthcoming evaluation of the regenerated resume using the Resume Enhancer model represents a crucial step towards validating the effectiveness of this enhancement approach. The outcome of this evaluation will offer valuable insights into the efficacy of the model in augmenting the quality of resume assessments and improving the accuracy of candidate screening processes.

The final model in our research endeavor, termed the Resume Enhancer, represents a pivotal advancement aimed at refining the quality of resumes in alignment with job descriptions (JDs). It operates on a multifaceted approach, meticulously curating and enhancing resumes to ensure a more precise fit with the specified job requirements. The model meticulously analyzes both the job description and the candidate's resume, identifying key skills, experiences, and qualifications essential for the role.

Subsequently, the Resume Enhancer strategically retains pertinent information from the original resume while augmenting its content to better align with the job description. This process involves leveraging natural language processing techniques to identify relevant keywords and phrases from the job description and integrating them seamlessly into the resume. By doing so, the enhanced resume not only showcases the candidate's qualifications but also highlights their suitability for the desired position.

One noteworthy aspect of the Resume Enhancer is its capability to generate resumes in both PDF and JSON formats. The PDF format ensures a visually appealing presentation, maintaining professional standards and readability. Meanwhile, the JSON format provides a structured representation of the resume's content, facilitating further analysis and processing by downstream components in our pipeline.

Upon generating the enhanced resume, the model systematically evaluates its effectiveness by

ML Algorithms	Accuracy	Precision
Multinomial Naive Bayes	98.27%	98.38%
K-Nearest Neighbors	97.23%	98.89%

Table 1: Accuracy & Precision of ML Algorithms

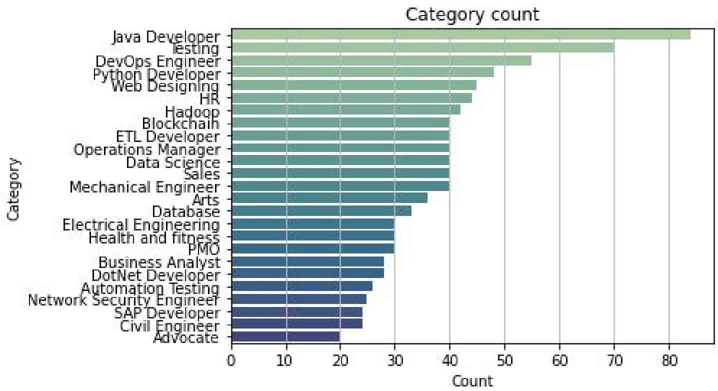


Figure 3: Count of Number of Resumes in each Category

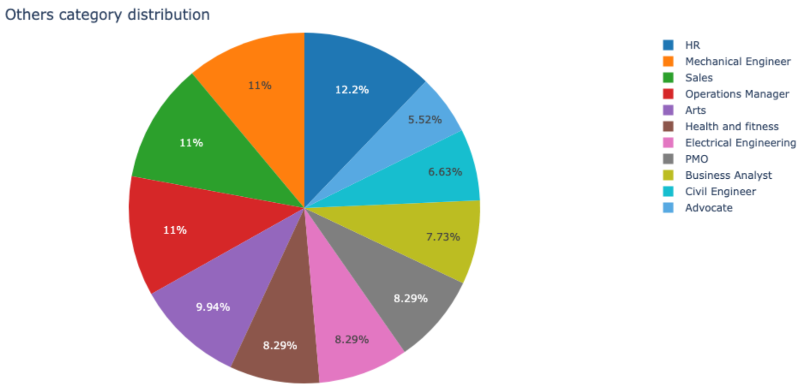


Figure 4: Distribution of Resumes in Others Category

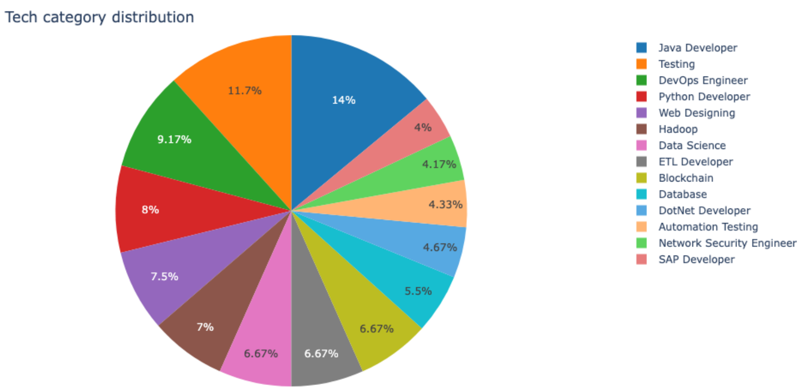


Figure 5: Distribution of Resumes in Tech Category

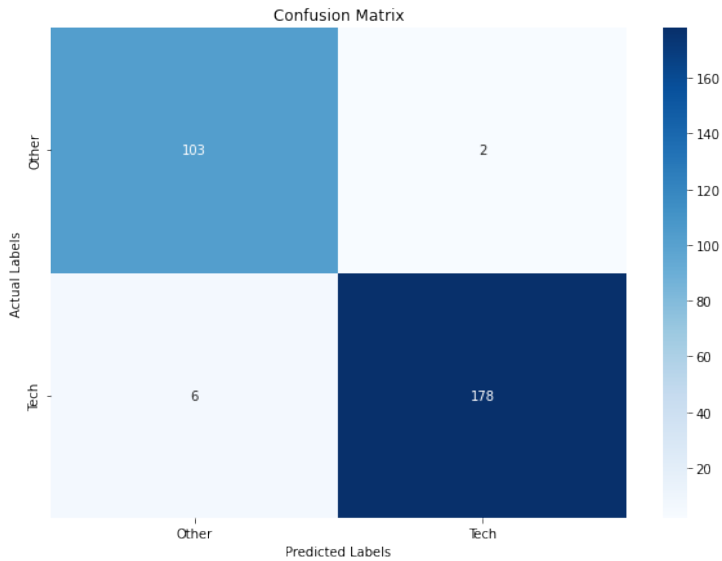


Figure 6: Resume Dataset Exploration Model's Confusion Matrix

comparing it against the original JD. This evaluation is conducted by the Resume Evaluator, which assesses the degree of alignment between the enhanced resume and the job requirements. The primary objective is to ascertain whether the enhancements result in a higher score relative to the baseline, indicating a more favorable match between the candidate and the job role.

Additionally, the enhanced resume generated a score of **0.6834** in Resume Evaluator, indicating a significant improvement in alignment with the job description. A comparative analysis of scores generated by the baseline model, Resume Evaluator before and after enhancement can be observed in the provided table (Refer Table 2), offering insights into the effectiveness of the enhancement process. Figure 7 depicts the screenshot of the generated resume in PDF format, showcasing the final output of the Resume Enhancer model.

In essence, the Resume Enhancer serves as a pivotal tool in optimizing candidate resumes, ensuring that they not only meet the specified job criteria but also stand out as compelling representations of the candidate's qualifications and suitability for the role. By integrating advanced natural language processing techniques with meticulous curation and evaluation processes, the model streamlines the resume enhancement process, ultimately enhancing recruitment efficiency and candidate selection outcomes.

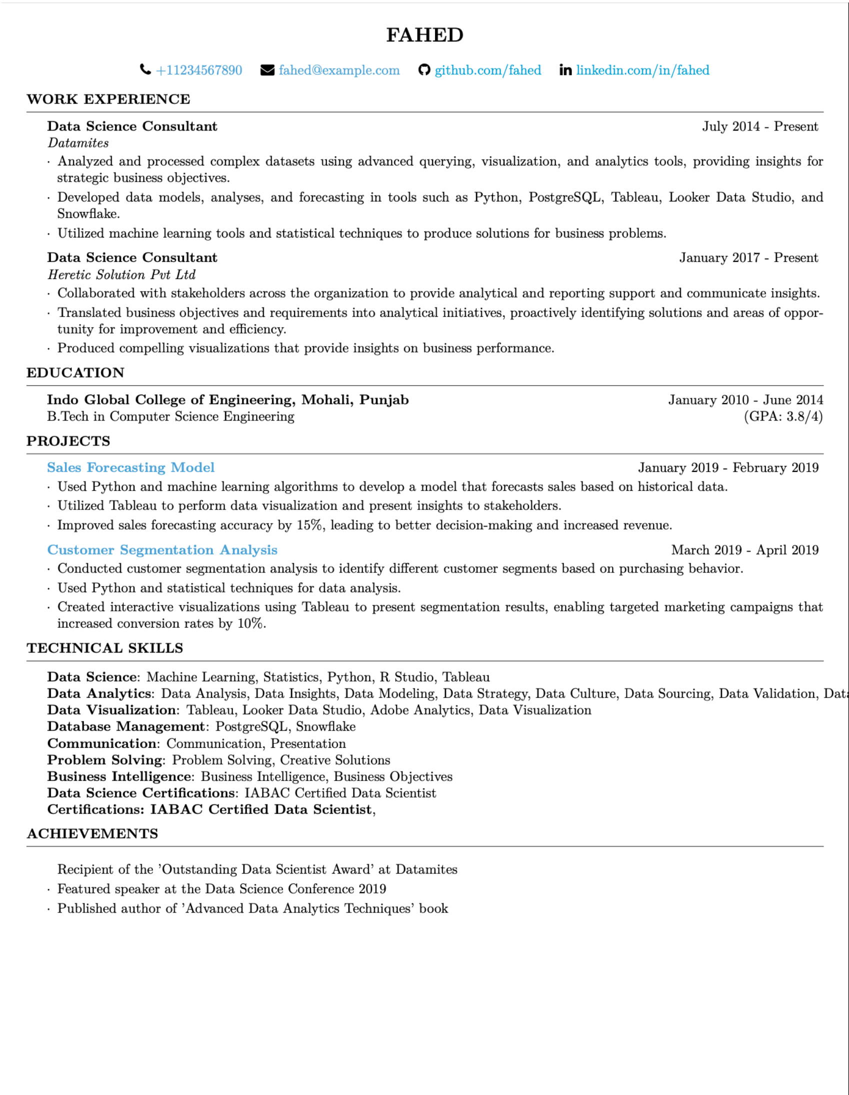
6 Additional Analysis & Future Work

One crucial aspect is to conduct a comprehensive performance comparison across all models. This comparison should include metrics such as accuracy, precision, and recall to gain a holistic understanding of their strengths and limitations. By examining how different weighting schemes and scoring mechanisms impact the overall effectiveness of resume-to-job matching, we can glean valuable insights into optimizing these models for real-world applications.

Moreover, gathering feedback from candidates who have undergone the resume enhancement process is essential. This feedback can provide valuable insights into the perceived impact of the process on their job search success and overall experience. Understanding candidates' perspectives is crucial for refining and tailoring the models to better meet their needs and expectations.

Furthermore, evaluating the applicability and effectiveness of the models across various industries and job sectors is paramount. Each industry has its unique requirements and nuances, and assessing how well the models perform in different contexts will help determine their adaptability and generalizability.

A longitudinal study could also be conducted to analyze the long-term implications of utilizing advanced resume analysis models on recruitment outcomes and organizational performance. Understanding how these models evolve and impact hiring practices over time is essential for shaping fu-



FAHED

+11234567890 fahed@example.com github.com/fahed linkedin.com/in/fahed

WORK EXPERIENCE

Data Science Consultant July 2014 - Present
Datamites
· Analyzed and processed complex datasets using advanced querying, visualization, and analytics tools, providing insights for strategic business objectives.
· Developed data models, analyses, and forecasting in tools such as Python, PostgreSQL, Tableau, Looker Data Studio, and Snowflake.
· Utilized machine learning tools and statistical techniques to produce solutions for business problems.

Data Science Consultant January 2017 - Present
Heretic Solution Pvt Ltd
· Collaborated with stakeholders across the organization to provide analytical and reporting support and communicate insights.
· Translated business objectives and requirements into analytical initiatives, proactively identifying solutions and areas of opportunity for improvement and efficiency.
· Produced compelling visualizations that provide insights on business performance.

EDUCATION

Indo Global College of Engineering, Mohali, Punjab January 2010 - June 2014
B.Tech in Computer Science Engineering (GPA: 3.8/4)

PROJECTS

Sales Forecasting Model January 2019 - February 2019
· Used Python and machine learning algorithms to develop a model that forecasts sales based on historical data.
· Utilized Tableau to perform data visualization and present insights to stakeholders.
· Improved sales forecasting accuracy by 15%, leading to better decision-making and increased revenue.

Customer Segmentation Analysis March 2019 - April 2019
· Conducted customer segmentation analysis to identify different customer segments based on purchasing behavior.
· Used Python and statistical techniques for data analysis.
· Created interactive visualizations using Tableau to present segmentation results, enabling targeted marketing campaigns that increased conversion rates by 10%.

TECHNICAL SKILLS

Data Science: Machine Learning, Statistics, Python, R Studio, Tableau
Data Analytics: Data Analysis, Data Insights, Data Modeling, Data Strategy, Data Culture, Data Sourcing, Data Validation, Dat
Data Visualization: Tableau, Looker Data Studio, Adobe Analytics, Data Visualization
Database Management: PostgreSQL, Snowflake
Communication: Communication, Presentation
Problem Solving: Problem Solving, Creative Solutions
Business Intelligence: Business Intelligence, Business Objectives
Data Science Certifications: IABAC Certified Data Scientist
Certifications: IABAC Certified Data Scientist,

ACHIEVEMENTS

Recipient of the 'Outstanding Data Scientist Award' at Datamites
· Featured speaker at the Data Science Conference 2019
· Published author of 'Advanced Data Analytics Techniques' book

Figure 7: Screenshot of Generated Resume

Model	Resume Score
Baseline Model	0.0568
Resume Analyser (In Comparison with Baseline Model Score)	0.5830
Regenerated Resume (In Comparison with Resume Analyser Score)	0.6834

Table 2: Scores of a Resume Tailored to a Specific Job Description

ture strategies and interventions.

Developing personalized matching algorithms tailored to individual candidate preferences, career goals, and organizational culture fit is another promising direction. Integrating the models seamlessly with existing Applicant Tracking Systems (ATS) and implementing real-time feedback mechanisms would also enhance their functionality and accessibility for recruiters and hiring managers.

Ethical considerations surrounding data privacy, algorithmic transparency, and potential societal implications of automated resume analysis in recruitment processes must also be addressed. Ensuring fairness and equity in candidate selection processes is paramount, and strategies to mitigate bias in resume analysis algorithms should be explored.

By continuing to explore these avenues and embracing emerging technologies and methodologies, we aim to further advance the field of recruitment optimization and contribute to more efficient and equitable hiring practices in the future.

7 Conclusion

In conclusion, this research aims to revolutionize conventional practices of resume analysis and optimization, emphasizing the empowerment of candidates to enhance their resumes and bolster their prospects in the competitive job market. Through the utilization of cutting-edge techniques in natural language processing, semantic matching, and vector embeddings, our study endeavors to equip individuals with the necessary tools and methodologies to refine and tailor their resumes for maximum impact.

Our exploration of various models, including the Tech vs. Other classifier, Baseline model, Resume Evaluator, and Resume Enhancer, has highlighted their efficacy and versatility in enhancing the efficiency and accuracy of resume-to-job matching. The results obtained from our analyses demonstrate promising performance metrics, such as accuracy, precision, and scalability, underscoring the potential of these models to revolutionize the field. Additionally, our research has identified key areas for future exploration and refinement, such as advanced semantic matching techniques, personalized algorithms, and ethical considerations surrounding algorithmic fairness and transparency.

By providing innovative tools and methodologies for resume analysis and enhancement, our work contributes to a more equitable and data-driven approach to career development and professional growth. Moving forward, continued innovation and interdisciplinary collaboration will be crucial in advancing these efforts and ensuring that individuals receive the support and resources necessary to thrive in the digital age.

8 Acknowledgements

We would like to express our sincere gratitude to *Professor Jonathan Rusert* for his invaluable guidance, encouragement, and support throughout the duration of this paper. His expertise, insights, and mentorship have been instrumental in shaping the direction of our work and contributing to its successful completion.

References

Jie Chen, Chunxia Zhang, and Zhendong Niu. 2018. A two-step resume information extraction algorithm. In *Mathematical Problems in Engineering*, pages 1–8.

Panagiotis Skondras, George Psaroudakis, Panagiotis Zervas, and Giannis Tzimas. 2023. Efficient resume classification through rapid dataset creation using chatgpt. In *2023 14th International Conference on Information, Intelligence, Systems Applications (IISA)*, pages 1–5. IEEE.

VV Satyanarayana Tallapragada, V Sushma Raj, U Deepak, P Divya Sai, and T Mallikarjuna. 2023. Improved resume parsing based on contextual meaning extraction using bert. In *2023 7th International Conference on Intelligent Computing and Control Systems (ICICCS)*, pages 1702–1708. IEEE.

Xing Yi, James Allan, and W Bruce Croft. 2007. Matching resumes and jobs based on relevance models. In *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 809–810. ACM Digital Library.