

Article

Optimizing Curriculum Vitae Concordance: A Comparative Examination of Classical Machine Learning Algorithms and Large Language Model Architectures

Mohammed Maree ^{1,*}  and Wala'a Shehada ² 

¹ Department of Information Technology, Faculty of Information Technology, Arab American University, 13 Zababdeh, Jenin P.O. Box 240, Palestine

² Department of Natural, Engineering and Technology Sciences, Ramallah Campus, Arab American University, Jenin P.O. Box 240, Palestine; w.shehada1@student.aaup.edu

* Correspondence: mohammed.maree@aaup.edu; Tel.: +970-4-2418888 (ext. 1123)

Abstract: Digital recruitment systems have revolutionized the hiring paradigm, imparting exceptional efficiencies and extending the reach for both employers and job seekers. This investigation scrutinized the efficacy of classical machine learning methodologies alongside advanced large language models (LLMs) in aligning resumes with job categories. Traditional matching techniques, such as Logistic Regression, Decision Trees, Naïve Bayes, and Support Vector Machines, are constrained by the necessity of manual feature extraction, limited feature representation, and performance degradation, particularly as dataset size escalates, rendering them less suitable for large-scale applications. Conversely, LLMs such as GPT-4, GPT-3, and LLAMA adeptly process unstructured textual content, capturing nuanced language and context with greater precision. We evaluated these methodologies utilizing two datasets comprising resumes and job descriptions to ascertain their accuracy, efficiency, and scalability. Our results revealed that while conventional models excel at processing structured data, LLMs significantly enhance the interpretation and matching of intricate textual information. This study highlights the transformative potential of LLMs in recruitment, offering insights into their application and future research avenues.



Citation: Maree, M.; Shehada, W. Optimizing Curriculum Vitae Concordance: A Comparative Examination of Classical Machine Learning Algorithms and Large Language Model Architectures. *AI* **2024**, *5*, 1377–1390. <https://doi.org/10.3390/ai5030066>

Academic Editor: Mehdi Neshat

Received: 9 July 2024

Revised: 29 July 2024

Accepted: 2 August 2024

Published: 6 August 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: digital recruitment systems; classical machine learning; large language models (LLMs); performance degradation; methodology comparison; recruitment transformation

1. Introduction

The proliferation of digital recruitment methodologies has catalyzed an increased demand for more efficacious automated solutions. By harnessing the capabilities of the internet and cutting-edge technologies, these systems have revolutionized the employment process, conferring numerous advantages over traditional recruitment practices. Historically, the recruitment landscape was dominated by newspaper advertisements and job offers, with the processing of employment applications being executed manually—a process characterized by extensive time consumption and substantial resource expenditure, alongside inherent geographical limitations. The advent of electronic recruitment systems has ameliorated these inefficiencies, rendering access to requisite competencies that is both efficient and effective [1]. The incorporation of artificial intelligence (AI), machine learning (ML), and natural language processing (NLP) technologies has further engendered transformative changes in online hiring systems. These sophisticated models possess the capability to perform comprehensive analyses of curricula vitae (CVs) and execute optimal candidate–job matching algorithms, augmented by predictive analytics. Consequently, the integration of these technologies has substantially enhanced the efficacy and efficiency of the recruitment process, providing a robust framework for future advancements in the field [2].

Despite the recent improvements in online recruitment systems, they still encounter formidable challenges, particularly concerning data privacy and AI biases. The sensitive personal data collected must be securely stored to avert breaches, which can result in identity theft and significant reputational damage. Implementing robust cybersecurity measures and clear data retention policies is imperative. Moreover, cross-border data transfers necessitate compliance with diverse regulatory frameworks. AI algorithms employed for resume screening and candidate matching can perpetuate historical biases if trained on biased datasets, leading to inequitable hiring practices. The utilization of diverse and representative training data, meticulous algorithm design, and stringent human oversight are essential to mitigate these biases. Additionally, Blockchain technology can bolster data security and accuracy within recruitment processes.

The primary objective of this article is to furnish a comprehensive review of the historical evolution and current state of online recruitment platforms. This study examined the technological advancements driving innovation in this sector, including AI, machine learning, and natural language processing. Furthermore, the article aims to explore the various models and frameworks employed in online recruitment systems, with a particular emphasis on recent developments such as large language models. By synthesizing extant data and identifying gaps, this article seeks to provide valuable insights to developers, researchers, and professionals engaged in the development and implementation of future recruitment systems. Ultimately, this study aspired to guide future research and development endeavors by proposing strategies to address current challenges and capitalize on emerging trends, thereby enhancing the efficacy and equity of electronic recruitment methodologies. Our motivation in this context originated from the desire to improve the accuracy and efficiency of resume analysis throughout the recruitment process by utilizing advanced machine learning and large language model (LLM) technology. The issue being addressed is the inherent subjectivity and probable bias in traditional human-based resume screening, which can result in inconsistent hiring decisions. By automating resume analysis with sophisticated algorithms, the study aimed to ensure a more objective, fair, and consistent manner for evaluating candidates, thereby increasing the recruiting process's reliability and effectiveness. Accordingly, this study aimed to discover the best models and approaches for effectively categorizing job applicants based on their resumes, thereby streamlining the recruiting process for firms.

The remainder of this manuscript is structured as follows: Section 2 presents a comprehensive literature review that examines recent advancements in AI and ML, with a focus on the influence of large language models (LLMs). It also outlines the methodology employed in this study. Section 3 presents the findings obtained from the experiments conducted during the evaluation of the various ML and LLM models employed in this work. In Section 4, future directions are explored, including the integration of Blockchain for secure data management, algorithm optimization strategies, and the potential role of virtual reality in candidate assessment. Finally, Section 5 provides concluding remarks for this article.

2. Materials and Methods

The integration of AI, ML, and LLMs into online recruitment platforms has revolutionized the employment landscape. AI and ML algorithms enhance candidate screening and matching capabilities by analyzing extensive datasets to identify optimal candidates based on skill sets, professional experience, and cultural alignment [2]. LLMs such as GPT, LLAMA, and BERT contribute to streamlining recruitment workflows by automating communication, generating interview queries, and providing immediate feedback during candidate interactions [3]. Collaboratively, these technologies enhance operational efficiency, mitigate bias, and enhance the overall candidate experience throughout the recruitment process.

2.1. Enhancements in Online Recruitment Systems through AI and ML

Recent research has extensively examined the integration of AI and machine learning in online recruitment platforms. Several models and methodologies have been examined in the literature, emphasizing the potential of AI and machine learning to improve various parts of recruiting procedures. For instance, in [2], the authors focused on the usage of machine learning and artificial intelligence to improve the staff selection process. The work used latent semantic analysis along with bidirectional encoder representations from transformers (BERT) to detect and comprehend hidden patterns in textual resume data. Support vector machines (SVM) were then used to create and improve the screening model. The study showed that LSA and BERT could effectively retrieve essential themes, but SVM enhanced the model's efficiency through cross-validations and variable selection procedures. This approach offers HR practitioners useful insights for building and enhancing recruitment processes, and it also offers more interpretable outcomes than current machine learning-based resume screening systems. In [4], the researchers investigated the creation and implementation of an interview bot intended to improve the recruitment process. The project utilized NLP and AI technologies to develop a system for conducting initial interviews with job candidates. The bot evaluated candidate responses using complex NLP algorithms to determine their eligibility for the role. This automated technique intended to improve the recruitment process, decrease human bias, and increase productivity by swiftly filtering out unsuitable candidates, freeing up the human recruiters to focus on the most promising individuals. The research illustrated the potential of AI-driven technologies to change traditional human resource procedures, delivering a more objective and scalable solution for talent acquisition. The authors of [5] explored the advancements in recruitment technology for developing smart recruitment systems (SRS), which use deep learning and natural language processing (NLP) to classify and rank resumes. These systems standardize and parse resume information, transforming diverse forms into English and extracting relevant data. By including executive requirements in the scoring process, the algorithms ensure that resume rankings meet the unique objectives of recruiters. This strategy strives to supply high-quality applicants by evaluating resumes based on technical capabilities and other important information, increasing the hiring process's efficacy and efficiency.

Previous research has emphasized the challenges organizations face when manually sifting, analyzing, recruiting, and alerting job applicants. For example, the authors of [6] proposed a system to automate the recruitment procedure utilizing the waterfall model, which is simple and sequential in nature. This intelligent job application screening and recruitment system, developed with PHP and MySQL, made use of artificial intelligence techniques to improve human resource management. The approach made the recruitment process faster, more efficient, and inexpensive. According to evaluations, this strategy improves the accuracy with which individuals are matched to appropriate occupations. In a similar line of research [7], the researchers proposed an analytics-based method to improve the competitiveness of the HR recruitment process. The authors used existing tools to extract features from job postings and match them with potential candidates' resumes in a database. The finest suitable candidates were chosen using similarity analysis, taking into account the desired traits and their resumes. In one case, a structure learned on the profiles of 1029 job candidates at an IT company reduced screening manual efforts by 80%. This decrease is predicted to save significantly on time and operational costs. Although the study was limited to the situation of a technology company in India, the suggested artificial intelligence-based approach can be used across many industries. The study in [8] provides a complete review of current e-recruitment systems, classifying them based on a variety of evaluation criteria. The suggested approach in this work differed from typical systems in that it used several semantic resources, such as WordNet and YAGO3, along with NLP, extracting features, and skill-relatedness algorithms, to discover the latent semantic dimensions in resumes and job postings. Unlike prior algorithms that considered the entire text of documents, this approach matched specific sections of applications to the relevant parts of job postings. To address missing background knowledge, the HS dataset

was used to augment job postings with semantically relevant ideas. The initial studies with a real-world dataset yielded high-precision results, demonstrating the approaches' usefulness in giving relevance scores to candidate resumes and job offers. Also, in [9], the research presented an automated online recruitment method that combined numerous semantic resources and statistical concept-relatedness measurements. The system started by using NLP techniques to discover and extract applicant concepts from job postings and resumes. The retrieved concepts were then refined using statistical concept-relatedness measures. As a result, various semantic resources were used to determine the semantic elements of resumes and job postings. To compensate for the restricted domain coverage of these semantic resources, the HS dataset was used to supplement the job postings with new and semantically relevant ideas. The initial testing with different resumes and job postings showed encouraging and precise results, verifying the effectiveness of the suggested approach.

In [10], a decision-making model was offered as a critical requirement for the use of AI technology in corporate employment interviews. The Analytic Hierarchy Process (AHP) approach was used to create the model, which included eighteen primary criteria defined by prior research using the Unification Model of Usage and Adoption of Technology. The factors were grouped into the following four major categories: expectations of performance, expectations of effort, societal impact, and facilitating environment. A survey of 40 professionals who were either users or producers of AI job interview solutions discovered that the simplicity of use, job fitness, perceived utility, and perceived consistency were the most important criteria influencing the adoption of an AI-based job interview system. The results indicated that greater attention should be focused on integrating AI job interview systems into an organization's internal recruitment procedures rather than focusing on external surroundings or situations.

In the current job market, organizations must hire and retain people who are the best match for the position. Research shows that employees who feel their jobs are important and enjoyable are more effective and less inclined to quit. Throughout the recruitment procedure, in order to select the most qualified candidates, large numbers of applicants are considered. However, due to the large number of candidates, rigorous screening and interviews with each candidate are frequently impractical. To address this, researchers have created systems such as JobFit, which uses recommender systems, machine learning algorithms, and historical data to forecast the best candidates for a job [11]. JobFit examines job requirements and the profiles of applicants to generate a JobFit score that rates candidates based on their compatibility. This approach assists HR professionals by selecting a small group of top candidates for additional screening and interviews, making sure the best applicants are not ignored.

2.2. Enhancements in Online Recruitment Systems through LLMs

Previous research has shown how large language models (LLMs) have transformed the processing of natural language jobs across numerous domains due to their exceptional capabilities. However, the possibility of graph semantic processing in employment suggestions has not been thoroughly examined [3]. This work aimed to demonstrate the potential of large language models to understand behavioral graphs and utilize their knowledge to improve recommendations for online recruitment, specifically, the advertising of OOD applications. They provide a novel technique for studying behavioral graphs and detecting underlying patterns and correlations by utilizing the extensive contextual knowledge and semantic descriptions supplied by large language models. They suggest a meta-path-prompting constructor to help LLM recommenders grasp the syntax for behavioral graphs for an initial time, as well as a path supplementation component to reduce the prompt bias resulting from the path-based sequencing input. This design allows for individualized and accurate employment recommendations for specific individuals. This method was evaluated on large, actual-world data sets, demonstrating its capability to increase the importance and value of proposed outcomes. This study not only demonstrated the hidden

strengths of large language models but also provided useful insights for constructing advanced recommendation systems for the field of employment. The findings improved the area of processing natural languages while having clear implications for improving job search processes. Large language models (LLMs) are growing in popularity as industry utilize their extensive knowledge and sophisticated reasoning capabilities as a potential strategy for improving resume completeness for more precise job suggestions. However, directly employing LLMs can lead to concerns like fabricated generation and the few-shot issue, which can reduce resume quality. To overcome these problems, ref. [12] provided a revolutionary LLM-based GANs Interactive Recommendation (LGIR) technique. The researchers improved resume completeness by extracting explicit user features (such as skills and interests) from self-descriptions and inferring implicit qualities from behaviors. To address the few-shot problem, which occurs when limited interaction records prevent high-quality resume production, they used Generative Adversarial Networks (GANs) to align low-quality resumes with high-quality created equivalents. Extensive trials using three large real-world recruiting datasets have demonstrated the effectiveness of their proposed strategy. Current job recommendation systems primarily rely on shared filtering or person–job matching algorithms. Such models frequently function as “black-box” structures and are without the transparency required to deliver clear information to job seekers. Furthermore, traditional matching-based systems can only rank and retrieve current vacancies from a database, limiting their utility as comprehensive career advisors. To address these limitations, the authors of [13] introduced GIRL (Generative Job Recommendation Based on Large Language Models), a new approach that leverages recent breakthroughs in LLM. GIRL employs the Supervising Fine-Tuning (SFT) approach to train an LLM-based generation to generate appropriate job descriptions (JDs) depending on a job seeker’s curriculum vitae. To improve the generator’s performance, they trained and fine-tuned a model that assesses the match between CVs and JDs utilizing a Proximal Policies Optimizing (PPO)-based reinforced learning (RL) technique. This connection with recruiter feedback ensures that the resulting outputs better reflect employer preferences. GIRL is a job seeker-centric generative algorithm that provides employment ideas without the necessity for a pre-defined candidate list, thereby boosting the capabilities of existing job recommendation models via produced content. Extensive experiments on an extensive real-world dataset have confirmed this technique’s great success, confirming GIRL as a paradigm-shifting solution in job system recommendation that promotes a more personal and detailed job-seeking experience. Table 1 shows a summary of the reviewed systems.

Table 1. A summary of the reviewed systems.

Reference	Methodology	Summary
[2]	Latent Semantic Analysis (LSA) and BERT with Support Vector Machines (SVM)	Improves resume screening by detecting patterns in textual data and enhancing model efficiency through cross-validation
[4]	Natural Language Processing (NLP) and AI	Development of an interview bot that evaluates candidate responses to improve recruitment efficiency and reduce bias
[5]	Deep Learning and NLP (SRS)	Smart Recruitment Systems (SRS) classify and rank resumes by standardizing and parsing information, improving candidate selection
[11]	(JobFit) Recommender Systems, Machine Learning, and Historical Data	JobFit predicts the best candidates for a job by generating a compatibility score, aiding HR professionals in candidate selection
[6]	Waterfall Model, PHP, and MySQL	An automated recruitment system that uses AI to improve efficiency and accuracy in matching candidates to jobs
[7]	Analytics-Based Recruitment	An analytics-based method to match job postings with resumes, significantly reducing manual screening efforts

Table 1. Cont.

Reference	Methodology	Summary
[10]	Analytic Hierarchy Process (AHP)	A decision-making model for AI recruitment systems focusing on criteria like ease of use, job fit, and perceived utility
[3]	Behavioral Graph Analysis (BGA) with LLMs	Enhances job recommendations by understanding and utilizing behavioral graphs to improve recruitment outcomes
[12]	LLM-Based GANs Interactive Recommendation (LGIR)	Improves resume completeness by extracting user features and using GANs to align low-quality resumes with high-quality generated equivalents
[13]	GIRL Generative Job Recommendation with LLMs and Reinforced Learning	Generates job descriptions based on CVs, providing personalized job suggestions and enhancing the job-seeking experience
[8]	Semantic Resources (WordNet and YAGO3) and NLP	An e-recruitment system that match resumes with job postings by extracting features and using skill-relatedness algorithms to improve candidate relevance

In our proposed research work, we used a dataset of resumes from various industries to extract features and train multiple machine learning models (Naive Bayes, SVM, Logistic Regression, Decision Trees, and XGBoost) and large language models (ChatGPT and LLaMA) for job classification. The performance of the models in categorizing the resumes was tested using accuracy measures. Figure 1 depicts the outline of the methodology, followed by a detailed explanation.

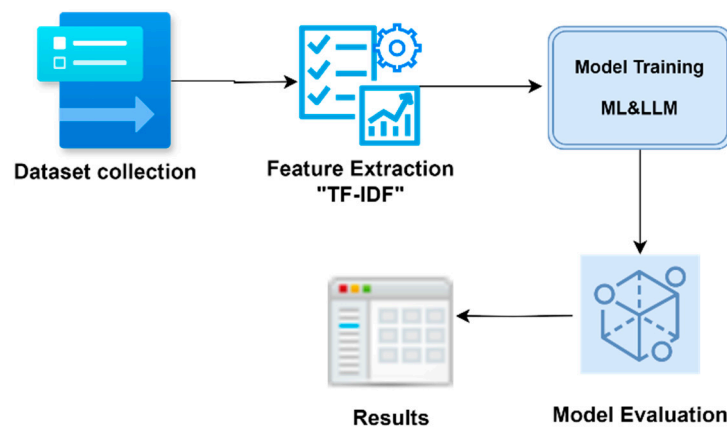


Figure 1. Online recruitment architecture for comparing the exploitation of ML against LLMs.

- **Dataset Description:**

This study’s dataset consisted of resumes and job categories in multiple key industries like finance, technology, and healthcare. Our dataset included 962 resumes that were obtained from a publicly available resource, and it meant to provide a wide representation of job categories to enable the training and evaluation of the classification algorithms [14].

The output labels were the job categories extracted from the resumes, representing the different job titles that the classification models aimed to predict. There was a total of 25 unique classes allocated, corresponding to distinct job categories. These categories, along with the number of resumes that fell under each category, are presented in Table 2.

Table 2. Distribution job categories (classes) in the original dataset.

Job Category	Number of Resumes
Data Science	40
HR	43
Advocate	19
Arts	35
Web Designing	44
Mechanical Engineer	39
Sales	39
Health and Fitness	29
Civil Engineer	23
Java Developer	83
Business Analyst	27
SAP Developer	23
Automation Testing	25
Electrical Engineering	53
Operation Manager	39
Python Developer	47
DevOps Engineer	54
Network Security Engineer	24
PMO	29
Database	32
Hadoop	41
ETL Developer	39
DotNet Developer	27
Blockchain	39
Testing	69

- **TF-IDF Vectorization:**

Term Frequency-Inverse Document Frequency (TF-IDF) was used to translate the textual information contained in the resumes into numerical features [15]. This model emphasizes the value of a word in a document to the total corpus. The Scikit-Learn library's TF-IDF vectorizer was used, with its parameters set to eliminate English stop words, and we set a maximum document frequency threshold of 85% to filter out extremely common phrases. As a result, each resume was converted into a feature vector that could be used as an input by a machine learning model.

- **Machine Learning Models:**

Multiple machine learning models were trained using the TF-IDF features extracted from the resumes. The models included in this experiment were the following:

- **Multinomial Naïve Bayes:**

This is a probabilistic framework that uses Bayes' theorem and a multinomial distribution. It works particularly well for text categorization tasks. The model computes the probability of each job category using the features retrieved from the resumes and assigns the category with the highest probability [15]. In our research, the Multinomial Naive Bayes model utilized an alpha parameter set to 1.0 for smoothing, which was crucial for handling the feature independence assumptions.

- **Support Vector Machine (SVM):**

This is a classifier that locates the optimal hyperplane in a high-dimensional space that most effectively separates classes. A linear kernel was utilized in this study to efficiently handle the text data, and a regularization parameter (C) was set to 1.0 to manage overfitting [16].

- **Logistic Regression:**

This is a linear model that has been adapted to multi-class classification using the one vs rest method. The model calculates the probabilities of each job category and chooses

the one with the greatest probability. The model was set with a maximum iteration limit of 1000 to ensure convergence, and a regularization parameter (C) was also set to 1.0 for robustness [17].

- **Decision Trees:**

This is a non-linear model that divides data into subsets according to feature values. A tree is built by recursively separating the data at the feature that provide the most information gain, and we fine-tuned this with a maximum depth constraint of 40 to prevent overfitting while leveraging a random state set to 42 for reproducibility [17].

- **XGBoost:**

This is a gradient-boosted decision tree implementation that has been tuned for performance and speed. XGBoost develops an ensemble of trees sequentially, with each tree correcting the mistakes of the preceding ones. We employed a SoftMax objective for the multi-class, and the required label encoding was set to false for handling the categorical variables.

- **Large Language Models:**

A large language model (LLM) is an artificial intelligence system that understands, generates, and manipulates human language with high efficiency. These models are trained on massive volumes of text data and employ sophisticated algorithms to understand language patterns, context, and nuances. The models included in this experiment were ChatGPT [18] and LLaMA [19].

- **ChatGPT:**

Powered by OpenAI's GPT-3.5-turbo, this model was used to map resumes to their corresponding job categories. The resumes were submitted as inputs to the model, along with a system prompt that presented a list of potential job categories. The program identified the most relevant job category for each resume.

- **Meta's LLaMA:**

The LLaMA model was also applied for resume classification. Similar to ChatGPT, the model accepted resumes as inputs, along with a system prompt showing job categories. The model's responses were processed to retrieve the expected job categories.

- **Evaluation Metrics:**

To assess the models' performance, we utilized the accuracy metric, which is the ratio of accurately predicted job categories to the total number of predictions [20]. It is computed as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

where TP (true positive) is the number of correctly predicted positive cases, TN (true negative) is the number of correctly predicted negative cases, FP (false positive) is the number of falsely predicted positive cases, and FN (false negative) is the number of falsely predicted negative cases.

3. Results

In this experiment, we used a dataset of resumes and their corresponding job categories, which we divided into training and testing sets to simplify model training and evaluation. The dataset was divided so that 80% of the data were allocated to the training set and 20% to the testing set. To ensure that the data-split could be reproduced, a random seed of 42 was employed. Therefore, our code read the input dataset, randomly selected 80% of the records for the training set, and allocated the remaining 20% to the testing set. The split data were recorded in two different .csv files named "_dataset.csv" and "_predict.csv".

To better support our results, we created a new dataset (called the UPDATED DATASET) from the original dataset with semantic adjustments and reduced resume lengths. The updated dataset had the same size and distribution of classes as the original dataset, and we only made semantic modifications to the resumes. This provided a unique opportunity to evaluate the robustness and adaptability of both machine learning models and LLMs.

3.1. Category Mapping

When applying the large language models, we noticed that the models could answer the job classification using semantically related terms that carried the same meaning. For example, the “data science” category was classified, according to the large language models, into “Data Scientist”. To obtain more precise results, the code referred to a JSON file named “data_map.json”. This file contained concept mappings from the raw job title labels contained in the “Category” column that were converted to a more standardized, semantically overt form. This mapping could assist in enhancing the accuracy of the machine learning models by decreasing the ambiguity and inconsistency in the labels. Figure 2 illustrates an example of how the concept mapping could have been organized in the “data_map.json” file.

```
{
  "Data Science": [
    "Data Scientist",
    "Data Science Consultant"
  ],
  "HR": [
    "Human Resources Manager",
    "HR Manager"
  ]
}
```

Figure 2. Concept mapping captured in the “data_map.json” file.

3.2. API Integration

To use large language models (LLMs) such as ChatGPT and LLaMA, the code communicates with the application programming interfaces (APIs) of ChatGPT and LLaMA to incorporate them into a job title prediction process. These APIs serve as gates to the LLMs’ functionality. The code first generates a message, or a prompt, instructing the LLM to behave as a recruitment expert.

For the used ChatGPT OpenAI GPT-3.5-turbo model, the request included the resume’s content and a system prompt to select the most suitable job title from a predefined list, and the response was parsed to extract the job title suggested by ChatGPT.

For Meta’s LLaMA model, similar to ChatGPT, the request comprised the resume text and a system prompt. The answer was retrieved in the following two steps: first, fetching the prediction URL, and second, retrieving the final job title from the prediction URL.

Once the prompt was created, the code used libraries such as requests to send it to the ChatGPT and LLaMA APIs. Finally, the code extracted the anticipated job title from each LLM’s response, using their insights alongside the predictions from the machine learning models.

Machine learning models were employed to predict job categories for resumes in the prediction data. For ChatGPT and LLaMA, the functions getGPT and getLLAMA were used to obtain predictions for the resumes. The results were saved into a CSV file. Overall, the evaluation and comparative analysis process was critical for determining the effectiveness of the various methodologies utilized for the resume-to-job classification tasks.

3.3. Comparative Analysis

In this section, we compare the performance of several models on the original and updated datasets. The models tested were GPT, LLAMA, SVM, Multinomial Naive Bayes, Logistic Regression, Decision Trees, and XGBoost. The datasets were divided into 80% for training and 20% for testing (see Table 3).

Table 3. Accuracy performance of the machine learning algorithms and large language models.

Model	Original Dataset	Updated Dataset
GPT	90.6%	84.9%
LLAMA	72.5%	81.3%
SVM	100%	100%
Multinomial Naive Bayes	91.7%	83.4%
Logistic Regression	97.9%	97.9%
Decision Trees	98.9%	91.7%
XGBoost	99.0%	96.8%

As demonstrated in Table 3, our findings highlighted the impacts of data changes on model accuracy, revealing important implications for model selection and application in resume screening. Summaries of the performance are as follows:

1. **GPT-3:** The accuracy of GPT-3 dropped from 0.906 to 0.849, illustrating its susceptibility to changes in a dataset. This suggested that while GPT-3 was highly accurate with the original data, its performance significantly declined with the new data, emphasizing the need for regular retraining or fine-tuning.
2. **LLAMA:** In contrast, LLAMA's accuracy increased substantially from 0.725 to 0.813 with the updated dataset. This improvement indicated that LLAMA can adapt well to new features and distinctions present in updated data, making it a suitable choice for dynamic environments where data frequently change.
3. **SVM:** Support Vector Machine (SVM) achieved perfect accuracy (100%) on both datasets, demonstrating exceptional robustness. This consistency suggested that SVM is highly reliable and less sensitive to variations in data, making it an excellent choice due to its stable and consistent performance.
4. **Multinomial Naive Bayes:** The accuracy of Multinomial Naive Bayes decreased from 0.917 to 0.834, likely due to the changes in the word distributions. This model's performance indicated that it may be more sensitive to variations in text data, necessitating careful monitoring and possible adjustments when data characteristics change.
5. **Logistic Regression:** Logistic Regression maintained a high accuracy of 0.979 across both datasets, demonstrating stability. This model's consistent performance suggested that it can reliably handle data variations without significant losses in accuracy, making it a strong candidate for applications requiring steady performance.
6. **Decision Trees:** The accuracy of Decision Trees was reduced from 0.989 to 0.917, indicating possible overfitting on the original dataset. This drop highlighted the model's potential vulnerability to overfitting and suggested that it may require pruning or other regularization techniques to improve generalization to new data.
7. **XGBoost:** XGBoost's accuracy dropped slightly from 0.99 to 0.968, showing some sensitivity to data changes but still performing well overall. This model's high performance, despite the slight decline, underscored its resilience and adaptability, making it a reliable choice for varied data circumstances.

These findings emphasize the importance of assessing model resilience and flexibility to ensure a consistent performance across different data conditions. In future work, we aim to explore these aspects further, providing more in-depth analyses and strategies for optimizing model selection and application in resume screening.

4. Discussion

Large language models are often trained on vast amounts of text data and are capable of generating human-like text. However, there can be a mismatch between the general capabilities of these models and the specific requirements of the tasks or applications. For example, while LLMs excel at generating coherent text, they may struggle with tasks requiring deep domain knowledge or specific contextual understanding. In the context of our work, domain knowledge about occupational categories was not precisely captured by the exploited LLMs. For instance, as depicted in Figure 3, for the resume of an “HR Manager”, the result (“Human Resources Coordinator”) was retrieved when consulting ChatGPT about the most appropriate job title for the given resume. As such, mismatches and ambiguity can undermine the accuracy and reliability of an LLM’s outputs. In applications requiring precise information or sensitive content (e.g., medical advice or legal documents), errors or misinterpretations due to mismatched capabilities or ambiguous outputs can have serious consequences.

```

1 {
2   "model": "gpt-3.5-turbo-0125",
3   "prompt": "what is the first computer ",
4   "max_tokens": 100,
5   "response_format": {
6     "type": "json_object"
7   },
8   "messages": [
9     {
10      "role": "system",
11      "content": "as an online recruitment expert, what is the most appropriate job for the following resume? return one job title only"
12    },
13    {
14      "role": "user",
15      "content": "TECHNICAL SKILLS Typewriting & TORA SPSEducation Details January 2017 MBA Chidambaram, Tamil Nadu SNS College of Engineering January 2014 HSC at SAV Higher Secondary School MBA SNS College of Engineering SSLC Finance at KamaraJ Matriculation School HR Skill Details Human resource, Finance-Exprience - Less than 1 year monthsCompany Details company description --"
16    }
17  ]
18 }
19
20 {
21   "id": "chatcmpl-91m2e5d8dfDLHmJY4kN01drP40jig",
22   "object": "chat.completion",
23   "created": 1729458132,
24   "model": "gpt-3.5-turbo-0125",
25   "choices": [
26     {
27       "index": 0,
28       "message": {
29         "role": "assistant",
30         "content": "Human Resources Coordinator"
31       },
32       "logprobs": null,
33       "finish_reason": "stop"
34     }
35   ],
36   "usage": {
37     "prompt_tokens": 142,
38     "completion_tokens": 3,
39     "total_tokens": 145
40   },
41   "system_fingerprint": null
42 }

```

Figure 3. Domain concept mismatch between an actual job title and the LLM’s suggestion.

To provide a clearer understanding of the data and the models’ performances, we present two example resumes, along with their corresponding classes and model predictions (see Figure 4). For instance, a resume for Web Designing with “Bachelor of Engineer in Information Technology with a background in web and graphics design, proficient in Bootstrap, HTML5, JavaScript, jQuery, Corel Draw, Photoshop, and Illustrator. Experienced in website design and development, participating in project competitions and national-level paper presentations. Skilled in analyzing user requirements, designing modules, and fixing bugs. Undertaken projects utilizing various technologies, such as HTML5, CSS3, Bootstrap, and JavaScript. Noteworthy projects include developing a disease diagnosis expert system and a shopping management system. Engaged in extracurricular activities as a coordinator for project competitions.” was correctly classified by most models, including Naive Bayes, SVM, Logistic Regression, XGBoost, and GPT-4, but Decision Trees misclassified it as a Blockchain and LLaMA classified it as a Web Designer or Frontend Developer. Although this was correct, it was considered a wrong answer due to the different context. Another example resume (a PMO “With over 10 years of experience at global brands like Wipro, this resume showcases expertise in PMO, ITIL management, process improvements, project audits, risk management, and workforce resource management. Proficient in project management tools including CA Clarity and Visio, as well as technical skills in SAP and Microsoft Office. Holds an MBA in HR and Finance and has held roles such as Senior Executive PMO Consultant at Ensono LLP. Responsibilities include creating structured reports, SLA management, invoicing, maintaining project data, and driving project progress. Strong background in operations, staffing, HR, and PMO, with experience across various delivery models and project management tasks. Proficient in communication, stakeholder man-

agement, and reporting, with a diverse skill set encompassing project planning, resource utilization, and compliance.”) was accurately classified by all machine learning models, though GPT predicted “PMO Manager”. Although this was correct, it was considered incorrect as it was not included in the data_map.json file. For this same resume, LLaMA suggested “PMO Consultant”, which was considered correct (“yes”) after being included in the data_map.json file.

Category	Resume	Multinomial Naive Bayes	SVM	Logistic Regression	Decision Tree	XGBoost	gpt	is_gpt_correct	lama	is_lama_correct
Web Designing	Bachelor of Engin	Web Designing	Web Designing	Web Designing	Blockchain	Web Designing	Web Designing	yes	Based on the information provided in the resume, the most appropriate job title for the candidate would be: Web Designer or Frontend Developer. Based on the information provided in the resume, the most appropriate job title for this candidate would be: * PMO Consultant	no
PMO	With over 10 yea	PMO	PMO	PMO	PMO	PMO	PMO Manager	no	This	yes

Figure 4. Two example resumes, along with their corresponding classes and model predictions.

In subsequent sections, we aim to delineate the significant avenues that are pertinent to resume screening and classification. These directions draw inspiration from recent advancements in Blockchain and virtual reality (VR) technologies, which, to our understanding, have yet to be explored within the scope of this research endeavor.

4.1. Future Direction in Online Recruitment Systems

Integrating Blockchain technology in online recruiting platforms improves data security [21–23] by retaining information about candidates in a decentralized and secure manner. This ensures strong protection against unauthorized access and fraud. Furthermore, Blockchain simplifies the verification of credentials, making background checks more efficient and reliable [24].

4.2. The Role of Virtual Reality in Candidate Assessment

Virtual reality (VR) is changing candidate assessments by providing immersive job simulations that enable realistic and objective evaluations of candidates’ abilities and performance, removing biases and providing a more accurate assessment than traditional approaches.

5. Conclusions

Technological advancements have significantly transformed the landscape of online recruitment, enhancing efficiency in hiring processes and expanding opportunities for both employers and job seekers. This study undertook a comparative analysis of traditional machine learning models and large language models (LLMs) in the context of resume-to-job category-matching across two distinct datasets. Traditional models such as Logistic Regression, Decision Trees, Naive Bayes, and SVM demonstrated robust performances when applied to structured data. Conversely, LLMs such as GPT-3.5-turbo and LLAMA exhibited exceptional proficiency in interpreting and matching complex, unstructured textual information. Our findings underscore the substantial potential of LLMs to revolutionize recruitment practices by effectively capturing language nuances and contexts, thereby enhancing the accuracy and scalability of online recruitment systems.

However, we also identified a critical issue that warrants further attention in this domain—specifically, the challenge of concept mismatching and ambiguity. This issue becomes particularly pronounced when LLMs are tasked with comprehending domain-specific knowledge or assuming an expert role within specific domains, such as the occupational categories in our current study. Addressing this challenge necessitates additional research efforts aimed at aligning LLM outputs with expert domain knowledge encoded in ontological repositories. Such alignment would enable LLMs to more accurately discern

semantic orientations and domain concepts, thereby yielding more precise and contextually relevant outcomes.

In summary, while LLMs offer significant advancements in recruitment technology, mitigating the effects of concept mismatching and ambiguity through enhanced semantic alignment remains a critical area for further investigation and development.

Author Contributions: Conceptualization, M.M. and W.S.; methodology, M.M. and W.S.; software, W.S.; validation, M.M. and W.S.; formal analysis, M.M.; investigation, M.M. and W.S.; resources, M.M.; data curation, M.M. and W.S.; writing—original draft preparation, W.S.; writing—review and editing, M.M.; visualization, W.S.; supervision, M.M.; project administration, M.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data are contained within the article.

Conflicts of Interest: The authors declare no conflicts of interest.

References

- Arman, M. The Advantages of Online Recruitment and Selection: A Systematic Review of Cost and Time Efficiency. *Bus. Manag. Strategy* **2023**, *14*, 220–240. [CrossRef]
- Tian, X.; Pavur, R.; Han, H.; Zhang, L. A machine learning-based human resources recruitment system for business process management: Using LSA, BERT and SVM. *Bus. Process Manag. J.* **2023**, *29*, 202–222. [CrossRef]
- Wu, L.; Qiu, Z.; Zheng, Z.; Zhu, H.; Chen, E. Exploring large language model for graph data understanding in online job recommendations. In Proceedings of the AAAI Conference on Artificial Intelligence, Vancouver, BC, Canada, 20–28 February 2024; Volume 38, pp. 9178–9186.
- Josu, A.; Brinner, J.; Athira, B. Improving job recruitment with an interview bot: A study on NLP techniques and AI technologies. In *Advances in Networks, Intelligence and Computing*; CRC Press: Boca Raton, FL, USA, 2024; pp. 579–586.
- Ponnaboyina, R.; Makala, R.; Venkateswara Reddy, E. Smart recruitment system using deep learning with natural language processing. In *Intelligent Systems and Sustainable Computing: Proceedings of ICISSC 2021*; Springer Nature: Singapore, 2022; pp. 647–655.
- Ele, B.I.; Ele, A.A.; Agaba, F. Intelligent-based job applicants' assessment and recruitment system. *Technology* **2024**, *6*, 25–46.
- Ujlayan, A.; Bhattacharya, S.; Sonakshi, A. Machine Learning-Based AI Framework to Optimize the Recruitment Screening Process. *JGBC* **2023**, *18* (Suppl. S1), 38–53. [CrossRef]
- Maree, M.; Kmail, A.B.; Belkhatir, M. Analysis and shortcomings of e-recruitment systems: Towards a semantics-based approach addressing knowledge incompleteness and limited domain coverage. *J. Inf. Sci.* **2019**, *45*, 713–735. [CrossRef]
- Kmail, A.B.; Maree, M.; Belkhatir, M.; Alhashmi, S.M. An Automatic Online Recruitment System based on Multiple Semantic Resources and Concept-relatedness measures. In Proceedings of the 23rd IEEE International Conference on Tools with Artificial Intelligence (ICTAI'15), Vietri sul Mare, Italy, 9–11 November 2015.
- Lee, B.C.; Kim, B.Y. A decision-making model for adopting an ai-generated recruitment interview system. *Management (IJM)* **2021**, *12*, 548–560.
- Appadoo, K.; Soonnoo, M.B.; Mungloo-Dilmohamud, Z. Job recommendation system, machine learning, regression, classification, natural language processing. In Proceedings of the 2020 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE), Gold Coast, Australia, 16–18 December 2020; pp. 1–6.
- Du, Y.; Luo, D.; Yan, R.; Wang, X.; Liu, H.; Zhu, H.; Song, Y.; Zhang, J. Enhancing job recommendation through llm-based generative adversarial networks. In Proceedings of the AAAI Conference on Artificial Intelligence, Vancouver, BC, Canada, 20–28 February 2024; Volume 38, pp. 8363–8371.
- Zheng, Z.; Qiu, Z.; Hu, X.; Wu, L.; Zhu, H.; Xiong, H. Generative job recommendations with large language model. *arXiv* **2023**, arXiv:2307.02157.
- Available online: <https://www.kaggle.com/datasets/jillanisofttech/updated-resume-dataset> (accessed on 15 June 2024).
- Wendland, A.; Zenere, M.; Niemann, J. Introduction to text classification: Impact of stemming and comparing TF-IDF and count vectorization as feature extraction technique. In *Systems, Software and Services Process Improvement: 28th European Conference, EuroSPI 2021, Krams, Austria, 1–3 September 2021, Proceedings 28*; Springer International Publishing: Berlin/Heidelberg, Germany, 2021; pp. 289–300.
- Xu, S.; Li, Y.; Wang, Z. Bayesian multinomial Naïve Bayes classifier to text classification. In *Advanced Multimedia and Ubiquitous Engineering: MUE/FutureTech 2017*; Springer: Singapore, 2017; pp. 347–352.

17. Pisner, D.A.; Schnyer, D.M. Support vector machine. In *Machine Learning*; Academic Press: Cambridge, MA, USA, 2020; pp. 101–121.
18. Qi, Z. The text classification of theft crime based on TF-IDF and XGBoost model. In Proceedings of the 2020 IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA), Dalian, China, 27–29 June 2020; pp. 1241–1246.
19. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. In *Advances in Neural Information Processing Systems*; MIT Press: Cambridge, MA, USA, 2017; Volume 30.
20. Kmail, A.B.; Maree, M.; Belkhatir, M. MatchingSem: Online Recruitment System based on Multiple Semantic Resources. In Proceedings of the 12th IEEE International Conference on Fuzzy Systems and Knowledge Discovery (FSKD'15), Zhangjiajie, China, 15–17 August 2015.
21. Available online: <https://www.mckinsey.com/capabilities/risk-and-resilience/our-insights> (accessed on 15 June 2024).
22. Available online: <https://www.emptor.io/blog/data-security-in-hiring-best-practices-for-protecting-sensitive-information/> (accessed on 15 June 2024).
23. GDPR. General Data Protection Regulation. 2018. Available online: <https://gdpr-info.eu/> (accessed on 21 November 2020).
24. Dhanala, N.S.; Radha, D. Implementation and Testing of a Blockchain based Recruitment Management System. In Proceedings of the 2020 5th International Conference on Communication and Electronics Systems (ICCES), Coimbatore, India, 10–12 June 2020; pp. 583–588. [CrossRef]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.