future internet

MDPI

Article

Generating Synthetic Resume Data with Large Language Models for Enhanced Job Description Classification [†]

Panagiotis Skondras, Panagiotis Zervas * and Giannis Tzimas

Data and Media Laboratory, Department of Electrical and Computer Engineering, University of Peloponnese, 22100 Tripoli, Greece; eced2024@go.uop.gr (P.S.); tzimas@uop.gr (G.T.)

* Correspondence: p.zervas@uop.gr

[†] This paper is an extended version of our paper published in the Fourteenth International Conference on Information, Intelligence, Systems and Applications (IISA 2023), Volos, Greece, 10–12 July 2023.

Abstract: In this article, we investigate the potential of synthetic resumes as a means for the rapid generation of training data and their effectiveness in data augmentation, especially in categories marked by sparse samples. The widespread implementation of machine learning algorithms in natural language processing (NLP) has notably streamlined the resume classification process, delivering time and cost efficiencies for hiring organizations. However, the performance of these algorithms depends on the abundance of training data. While selecting the right model architecture is essential, it is also crucial to ensure the availability of a robust, well-curated dataset. For many categories in the job market, data sparsity remains a challenge. To deal with this challenge, we employed the OpenAI API to generate both structured and unstructured resumes tailored to specific criteria. These synthetically generated resumes were cleaned, preprocessed and then utilized to train two distinct models: a transformer model (BERT) and a feedforward neural network (FFNN) that incorporated Universal Sentence Encoder 4 (USE4) embeddings. While both models were evaluated on the multiclass classification task of resumes, when trained on an augmented dataset containing 60 percent real data (from Indeed website) and 40 percent synthetic data from ChatGPT, the transformer model presented exceptional accuracy. The FFNN, albeit predictably, achieved lower accuracy. These findings highlight the value of augmented real-world data with ChatGPT-generated synthetic resumes, especially in the context of limited training data. The suitability of the BERT model for such classification tasks further reinforces this narrative.

Keywords: metadata extraction; resumes; CV; big data; multiclass classification; ChatGPT; large language models; deep learning; embeddings; labor market analysis

check for updates

Citation: Skondras, P.; Zervas, P.; Tzimas, G. Generating Synthetic Resume Data with Large Language Models for Enhanced Job Description Classification. *Future Internet* **2023**, *15*, 363. https://doi.org/10.3390/fi15110363

Academic Editors: Francesca Fallucchi and Ernesto William De Luca

Received: 30 September 2023
Revised: 1 November 2023
Accepted: 7 November 2023
Published: 9 November 2023

1. Introduction

In the fast-paced world of recruitment, where companies refine numerous resumes to identify the ideal candidate, the demand for automated solutions has never been more pressing. Recent advancements in machine learning algorithms now allow for the extraction of information from resumes with increased accuracy and speed. This extraction process involves parsing the resume to identify specific sections, such as the candidate's work experience, education, skills and other relevant details. Once extracted, this information can be employed to classify resumes based on set criteria, streamlining the selection process. With the surge of online resume platforms and the increasing dominance of digital recruitment (e-recruitment), there has been an overwhelming flood of data, intensifying the challenge of extracting valuable insights and metadata from resumes. This article delves into the extraction of metadata from digital resumes characterized by big data attributes, necessitating regular updates and ongoing monitoring.

The efficacy of machine learning algorithms in resume classification is deeply rooted in the quality and size of the training data. Moreover, the selection of the appropriate model architecture is crucial for optimal performance. While large language models (LLMs) like GPT-4 [1] have demonstrated remarkable results in various natural language processing tasks, they might not be the best fit for tasks demanding a deep understanding of specific language structures, such as resume classification. This is attributed to LLMs' [2] constrained capacity to encode and comprehend syntax, semantics and other linguistic nuances vital for resume classification.

In machine learning, possessing a curated dataset is essential for training powerful models. This holds particularly true for classification tasks where precise tags or labels are important for achieving a noteworthy performance. The choice of datasets for training and evaluation is crucial. Our approach is multifaceted: we initially deployed web crawlers adept at aggregating resumes from online sources like Indeed.com. Following this, we employed natural language processing (NLP) techniques for data sanitization and preprocessing, ensuring the uniformity and quality of the collected resumes. A significant difficulty is the lack of labeled data, which is vital for machine learning algorithms' efficiency. To address this, we utilized the OPEN AI API [3], and through prompt engineering [4–9], we generated labeled resumes [10–15], including candidate categorization, skill identification and experience evaluation. This strategy underlines the strengths of LLMs, guaranteeing the creation of dependable and precise labeled datasets for efficient metadata extraction.

Our methodology emphasizes the use of synthetic resumes for rapid training data creation and their utilization for data augmentation [16–19], especially in categories with few samples. Not only did we rely on the ChatGPT resume dataset, but we also incorporated a new dataset sourced from Indeed.com using crawlers. This hybrid approach ensured our model's versatility, enabling it to perform efficiently on both generated and real-world data. Aiming to enhance the metadata extraction process, this research couples the potential of embeddings and deep learning techniques. We trained deep learning models on data labeled using big language models (like GPT-3) to capture complex patterns and learn the detailed meanings of words based on context. This strategy enables precise and rapid metadata extraction and leverages deep learning algorithms' inherent capabilities and the rich semantic information within the chosen embeddings.

Our research's primary objective was to extract metadata related to job resume classification across diverse occupational categories, especially those marked by sparse samples. To realize this goal, the multiclass classification task was executed using two distinct models: a transformer model, specifically BERT [20], and a feedforward neural network (FFNN). A comparative analysis of the results produced using these models offers useful insights into their performance. The experiments underscore that blending synthetic data with real data produces great results, affirming our approach's efficacy in generating precise metadata and paving the way for enhanced candidate matching and e-recruitment decisions.

The article begins by reviewing related works in the field of resume classification, especially those that leverage machine learning algorithms. We then proceed to the methodology, which also embodies the creation and preprocessing of the datasets, the models we adopted and their configuration for the training and evaluation for the multiclass classification of resumes. Afterward, we present and analyze the results from our experiments. The article concludes with insights and suggestions for potential future research in this area.

2. Previous Work

Resume classification for e-recruiting has garnered significant attention in recent years. Malinowski, J., Keim, T., Wendt, O. and Weitzel, T. [21] proposed a recommender system to align candidates' skills with recruiter requirements. Yi, X., Allan, J. and Croft, W. B. [22] introduced a structured relevance model that leverages labeled resumes to retrieve analogous ones. Tallapragada, V. V. S., Raj, V. S., Deepak, U., Sai, P. D. and Mallikarjuna, T. [23], exploited the concept of understanding the document contextually with the BERT model. They proposed the use of bidirectional encoder representations from transformer (BERT) vectorization to identify the text contextually. The presented results showed that the algorithm they constructed parsed the resume better than existing techniques. The BERT vectorization seemed to figure out what the text meant in a specific context. The resumes for the study were chosen randomly from a database where all persisted resumes were randomly generated for the specific study. Although the results showed a promising solution, the data that was used to perform the experiments were only 100 resumes.

Convolutional neural networks (CNNs) have been employed to enhance the accuracy of resume classification for e-recruiting. Several deep learning models for resume classification have been proposed, such as skill prediction based on multilabel resume classification using CNNs, which achieved state-of-the-art results [24]. Skills are categorized into high and low levels. High-level skills represent specialties (e.g., web developer and front-end developer), while low-level skills (e.g., CSS, HTML and PHP) denote technical expertise. The underlying assumption is that high-level skills are defined through a collection of low-level ones. Thus, if low-level skills are accurately identified, high-level ones can be predicted. The model achieved a recall of 98.79% and a precision of 91.34%, surpassing a 99% accuracy for specific competencies. Transformer-based models like BERT and GPT-3 have also been applied to resume classification tasks. Li, X., Shu, H., Zhai, Y. and Lin, Z. [25] proposed an information extraction method grounded in named entity recognition. Skills described in a resume are pivotal information to extract.

An approach to extract information from resumes written in five different languages using the transformer model (BERT) was successfully implemented [26]. This model extracts and classifies relevant document sections (personal information, education, past employment and skills) and, at a more granular level, extracts and classifies specific details, such as names, dates, organizations, positions, university degrees, individual skills and their self-assessed competence levels. Four labels representing skill competence were defined: excellent (high proficiency/advanced level), good (intermediate level), bad (beginner level) and null (missing self-assessment).

While these models show potential in enhancing classification accuracy, acquiring large labeled datasets remains a challenge. Furthermore, a meticulous evaluation is essential to avoid biases towards particular candidate groups or job postings. The creation of effective strategies for collecting and labeling high-quality resume data is crucial to support the development of precise and impartial resume classification models for e-recruitment.

3. Methodology

3.1. Overview

The objective of our study was to explore the feasibility and benefits of augmenting real-world job resume data crawled from Indeed.com with synthetic resumes generated using ChatGPT and to perform the multiclass classification task. Our investigation spanned across 15 distinct job categories, ranging from technical roles like software developers to professional roles such as lawyers and various others in between. This section provides a detailed methodology encompassing data collection, data synthesis, machine learning methods and the experimental setup. Figure 1 presents the overall schema of our study; each step was described in the following sections.

3.2. Data Collection from Online Sources and Processing

Resumes typically fall into two categories: structured and unstructured (freestyle). Structured resumes present a candidate's details in distinct sections, such as education and work experience, usually in reverse chronological order. Unstructured resumes, however, follow a narrative approach without specific divisions. Indeed.com predominantly showcases structured resumes, but variations within this format can complicate metadata extraction. To enhance labor market insights, it is crucial to utilize data from platforms like Indeed.com that mirror industry needs and offer insights into a diverse range of job categories, enabling a thorough analysis of labor market trends, skill prerequisites and other employment-influencing factors. We delved into the construction of the dataset from indeed.com (Indeed_Dataset) in the subsequent sections.

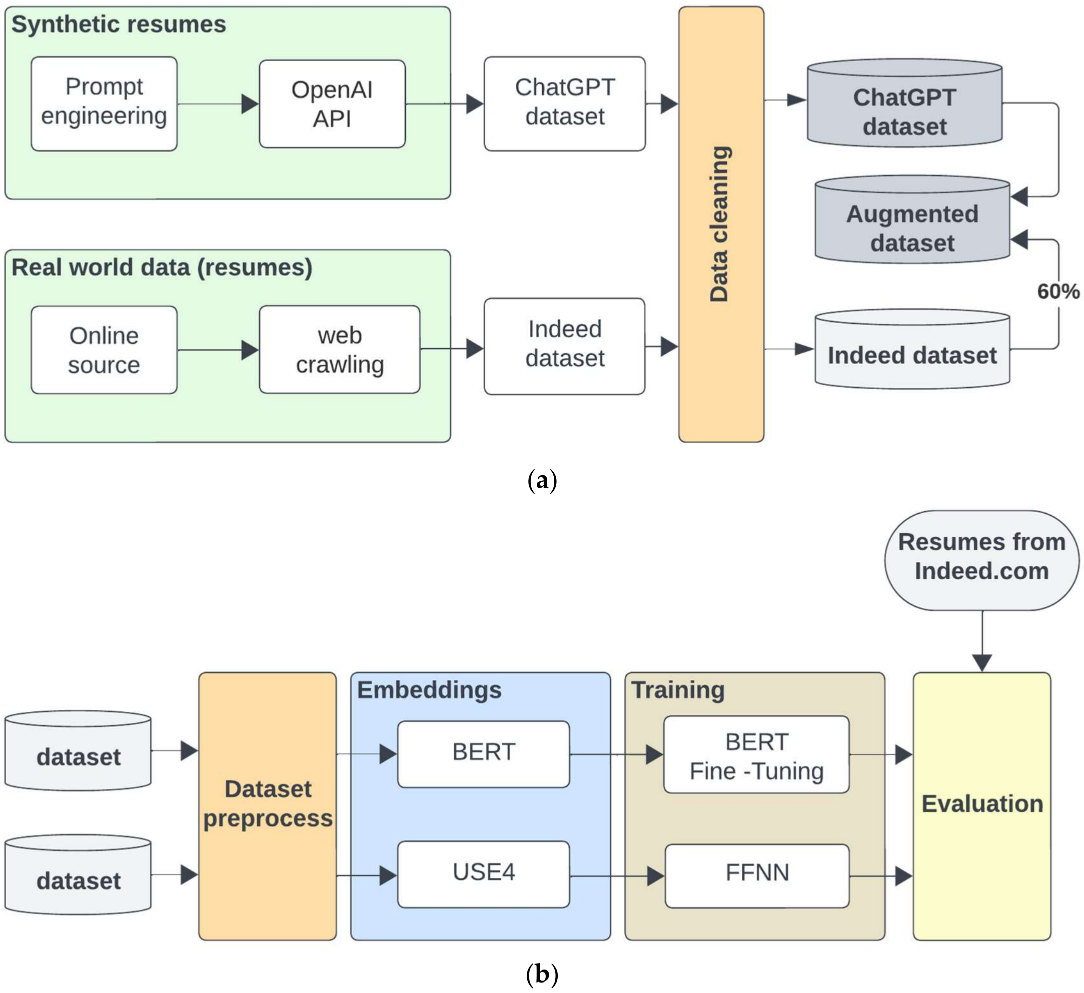

Figure 1. (**a**) Data collection and preprocess pipeline; (**b**) training and evaluation pipeline.

Online sources: The Indeed_Dataset was composed of resumes sourced exclusively from the Indeed website.

Data acquisition and processing: Using web crawlers, resumes from Indeed were methodically acquired. Tailored search queries were input into the search system producing search outcomes. Each result was systematically cataloged in a database and linked with its corresponding occupation label. This method was employed iteratively until the dataset reached its desired volume for the analysis.

The search results were algorithmically generated, emphasizing the most relevant outcomes. This indicated that the search engine performed internal calculations using search algorithms and presented the most relevant resumes. Furthermore, the search engine provided customization options through various filters, such as location, time and category. When applied, these filters produced unique search results tailored to specific query parameters.

Data cleaning: Throughout our data collection and processing, emphasis was placed on maintaining the consistency and accuracy of the data. This step involved:

- Correcting typographical errors.
- Removing duplicate records.
- Stripping special characters like &NBSP (HTML element), \r and _.
- Managing empty lines within a text block.
- Using regular expressions to identify and exclude URLs embedded in job descriptions.
- Deleting site-specific prefix keywords.

The executed procedures guaranteed the creation of a dataset that accurately represented real-world information, while minimizing any errors or unwanted inconsistencies. As a result of these processes, a single dataset was created (the Indeed dataset), containing authentic data extracted from an online source (Indeed.com) and encapsulating 15 diverse job categories.

3.3. Utilization of ChatGPT for Resume Generation

Our aim was to simplify the data collection process and employ augmentation techniques (through enhancement) to boost the performance of models trained for resume classification. We generated a substantial corpus of resumes using ChatGPT, guided by customized prompts. We crafted specific prompts to obtain resumes for the 15 job categories of interest. To ensure a broad range of content, we leveraged the O*NET [27,28] web service to create dictionaries encompassing synonyms for the 15 professions, their qualifications and other related terms. This comprehensive approach allowed us to generate prompts that would obtain a broad spectrum of responses from ChatGPT, capturing the essence of each occupation. To further diversify the content, we requested resumes in various writing styles, such as scientific, literary, colloquial and others. Additionally, we aimed for different levels of English proficiency in the generated content, ensuring a wide range of linguistic nuances.

To streamline this process, we utilized the OpenAI API [1,29] through a Python script that managed all fundamental tasks. The uniqueness of each prompt ensured that the resulting resumes displayed a spectrum of nuances, capturing variations in content depth, writing style and job titles. To enhance the dataset's richness, we integrated WordNet [30], a lexical database, to facilitate word substitutions with their synonyms. This strategy aimed to further diversify the dataset while preserving the core essence of each resume.

In the preprocessing phase, we emphasized data integrity and uniformity. This involved eliminating redundant or irrelevant details and standardizing data formats, preparing the dataset for the classification task. Due to the dynamic parameterization of our prompts, there were occasions where ChatGPT declined to generate a resume. For example, when tasked with creating a resume for a lawyer with only a high school education, the model tactfully refused, stating that it is impossible for a lawyer to hold only formal education for being a lawyer.

Upon completing the resume generation, we undertook a manual review to rectify any inconsistencies. This hands-on check ensured that the data were both consistent and valid. As a result, the 'ChatGPT_dataset' was created, which comprised synthetic data produced with ChatGPT, spanning the required job categories.

3.4. Model Architectures and Training

In the context of our study, we implemented two learning architectures: a baseline FFNN and a version of the BERT model. The FFNN represents a conventional machine learning approach, while BERT, a transformer-based design, is optimized for NLP tasks. Though BERT effectively captures context, it comes with a higher number of parameters and increased computational demands. Both models were trained to classify resumes into 15 distinct job categories (classes) based on the job category (class) they represent.

3.4.1. FFNN Model

The FFNN was designed specifically for multiclass classification. The selection of its architecture and other hyperparameters was determined based on literature insights and our trial-and-error experiments, particularly focusing on the model's architecture, optimizer choice, learning rate and rho values. Its architecture consists of an input layer with 512 units, aligned with the feature vectors generated by the USE4 model. USE4, a pretrained embedding model, converts textual data into fixed-size feature vectors, encapsulating semantic nuances, enabling the FFNN to discern data intricacies without directly interfacing with the raw text. Following the input layer, the FFNN incorporates a hidden layer of

256 units using ReLU activation. The output layer, with its 15 units, corresponds to the distinct job categories and utilizes softmax activation. For configuring the model, we employed the tf.keras.Model function, selecting Adadelta as the optimizer with a learning rate of 0.15 and a rho of 0.95. The model compilation leveraged 'sparse_categorical_crossentropy' for the loss and 'accuracy' as the primary metric. The training spanned 40 epochs, using a dataset partitioned into training and validation subsets. The test subset size was set at 20% of the training dataset.

3.4.2. BERT Model

Our BERT model, sourced from Hugging Face's library [31], was tailored for text classification. BERT, a transformer-based model, generates embeddings that reflect contextual relationships within text, making it a powerful tool for capturing semantic nuances in large textual datasets. These embeddings are derived from multiple transformer layers, allowing the model to consider both the current word and its surrounding context in the text. For data preparation, we curated datasets for training and evaluation. The text underwent preprocessing, converting to lowercase and removing numbers and special characters. Using Hugging Face's BERT tokenizer, the text was tokenized with a defined maximum length of 512 tokens. Our encode_data function then converted this tokenized text into 'input_ids' and 'attention_mask', essential for feeding data into the BERT model. Our model was built upon the 'bert-base-uncased' configuration from Hugging Face, selected for resume classification due to its advanced pretrained contextual embeddings and the advantages of transfer learning, which, together, guaranteed a precise resume analysis. We established two input layers to channel the tokenized and encoded data into BERT. From the output of BERT, we exploited only the first element, representing the embeddings from the terminal hidden layer, which contained rich contextual information. Building on these embeddings, we added a 1D convolutional layer with 256 filters and ReLU activation, a dropout layer at a rate of 0.5, a global max-pooling layer and a dense layer with 15 units using softmax activation. These units represented the probability distribution among job categories (classes).

For training, we employed a learning rate with an exponential decay, initiating at 1×10^{-5}. The Adam optimizer was chosen for the model compilation. After a training duration of 2 epochs with a batch size of 8, we assessed the model's efficacy on the evaluation dataset, extracting labels via the argmax function and computing the accuracy metric.

3.5. Training Datasets and Experimental Setup

Our research objective encompassed a main use case: resume classification enhanced with augmented data. At the core of our study, we aimed to explore and assess the feasibility and advantages of enriching real-world job resume data, sourced from Indeed with synthetic resumes created with ChatGPT. Our focus then shifted to performing the multiclass classification task on this real-world data, particularly in categories characterized by sparse samples.

To perform the multiclass classification of resumes, five datasets were created:

- One synthetic dataset from ChatGPT (ChatGPT, Section 3.3).
- One dataset (Indeed, Section 3.2) with real-world data from Indeed.com.
- A training dataset (Indeed_60) derived from the Indeed dataset (approximately 60%).
- An evaluation dataset (Indeed_40) derived from the Indeed dataset (approximately 40%). The selection of instances for both Indeed_60 and Indeed_40 was performed randomly, ensuring against a biased distribution or the exclusion of difficult-to-learn or classify instances.
- An augmented dataset (Indeed augmented) was constructed by combining the "Indeed_60" dataset, which contained 919 resumes, with an additional 1331 resumes generated with ChatGPT. This resulted in a total of 2250 resumes in the augmented dataset. A primary objective behind this augmentation was to ensure a balanced representation across resume categories within the dataset.

This balance was realized through a thorough curation process, where each job class received an approximately equal number of entries. The primary motivation behind this augmentation was twofold: firstly, to enhance the model's training, promoting consistent prediction accuracy and, secondly, to address potential inconsistencies arising from overlaps in similar resumes that might describe different job categories. This reinforced the model's ability to discern subtle differences and distinctions among professions.

In our experiments, we used occupation abbreviations for clarity and simplicity in presenting the procedure and results. Table 1 provides an overview of the datasets we constructed. The term "Job category/class" refers to the class represented in the multiclass classification task.

Table 1. Overview of all created datasets and job categories.

Job Category/Class	Occupation Abbr.	O*NET Alt. Titles	Indeed Dataset	ChatGPT Dataset	Indeed Augmented	Indeed_60	Indeed_40
Heating, Air Conditioning and Refrigeration Mechanics and Installers	HVAC	9	99	296	150	59	40
Human Resources Specialists	HRS	10	139	342	150	83	56
Driver/Sales Workers	Driver	10	133	284	150	80	53
Landscaping and Groundskeeping Workers	Landscaper	10	86	424	150	52	34
Marketing Managers	Mrkt.Mgr	10	101	373	150	60	41
Tellers	Tellers	10	85	302	150	51	34
Automotive Service Technicians and Mechanics	Auto.Mech.	10	53	296	150	32	21
Dentists, General	Dentist	7	101	324	150	60	41
Customer Service Representatives	CSR	10	95	278	150	57	38
Chefs and Head Cooks	Chef	10	87	308	150	52	35
Electrical Engineers	Elec.Eng.	10	175	272	150	105	70
Software Developer	Soft.Dev.	16	86	326	150	52	34
Civil Engineers	Civ.Eng.	5	96	362	150	58	38
Database Administrators	DBA	5	101	278	150	60	41
Lawyers	Lawyer	8	95	326	150	57	38
Total	15		1533	4791	2250	919	614

4. Results

4.1. Previous Work Results

In our previous research [32], we investigated the use of synthetic data generated using ChatGPT for training within the context of resume multiclass classification. Evaluating the model's performance using the Indeed dataset, we found that synthetic datasets, indeed, had potential value in the domain of e-recruitment.

In Table 2, we showcase the results (experiments one and two) pertaining to the FFNN employing USE4 embeddings when tasked with resume classification according to job categories. To ensure a comprehensive examination of this task, we conducted an analogous experiment leveraging the BERT model.

4.2. Main Use Case and per Class Results

In this study, we assessed the performance of both a BERT and an FFNN model on a resume multiclass classification task. The optimal results from our extended experiments are presented in Table 2.

The experiments for both the BERT and FFNN models were conducted as follows:

- Training with the Indeed_60 and evaluating with the Indeed_40 dataset (experiments three and four).

- Training with the Indeed augmented and evaluating with the Indeed_40 datasets (experiments five and six).

Table 2. Prediction accuracy and model performance for all experiments per class.

Exp.	Model	Training Dataset	Evaluation Dataset	Accuracy	Precision	Recall	F1-Score
1	FFNN	ChatGPT	Indeed	0.85	0.85	0.85	0.85
2	BERT	ChatGPT	Indeed	0.85	0.85	0.85	0.85
3	FNN	Indeed_60	Indeed_40	0.84	0.86	0.84	0.84
4	BERT	Indeed_60	Indeed_40	0.85	0.88	0.86	0.86
5	FNN	Indeed augmented	Indeed_40	0.85	0.86	0.85	0.85
6	BERT	Indeed augmented	Indeed_40	0.92	0.92	0.92	0.92

The experimental results showcased in Table 2 demonstrated the impact of the different models and training datasets on the accuracy of resume multiclass classification and presented the average performance metrics, offering a comparison between these methods in a per-class manner.

In the third experiment, utilizing the FFNN model with the Indeed_60 training dataset, we achieved an accuracy of 84%. The fourth experiment introduced the BERT model to the same training dataset, resulting in an improvement of 85% accuracy. Experiment five maintained the FNN model but employed an augmented version of the training data. Indeed, the augmented dataset improved the accuracy to 85%.

The most improved accuracy was observed in the sixth experiment, where the BERT model was combined with the Indeed augmented dataset, resulting in a 92% accuracy. These results underscored the significance of both model selection and the quality of the training data, highlighting the potential of advanced models like BERT when coupled with augmented datasets for enhancing the accuracy of resume classification tasks.

In our evaluations, the models trained on ChatGPT data were assessed using the full Indeed dataset, as discussed in experiments from the paper [32], while other models were tested on the Indeed_40 dataset. Figure 2 presents the F1 scores of the FFNN when trained on three different datasets: ChatGPT data, Indeed_60 data and the Indeed augmented dataset. The model seemed to perform best when trained on the augmented dataset for the classes Auto.Mech. and Elec.Eng., with F1 scores reaching as high as 0.95. However, for the classes CSR and Landscaper, the model's performance was inconsistent across the datasets. Notably, for the CSR class, the model had its lowest performance when trained on Indeed_60 data. For the Landscaper class, the F1 score dipped when trained on Indeed_60, but improved again with the augmented dataset. It is also worth noting that some classes like Dentist, Civ. Eng. Elec. Eng. and Soft. Dev. maintained consistently high F1 scores across all datasets, indicating the model's robust performance for these categories, regardless of the training data source. The tellers class had a significant improvement when trained on the Indeed_60 dataset compared to the ChatGPT dataset. Overall, the results of the FFNN suggested that the Indeed augmented dataset generally enhanced the model's performance, but the effectiveness could vary depending on the specific profession.

Figure 3 presents the results of the BERT model trained on the same three datasets: the ChatGPT dataset, Indeed_60 dataset and the Indeed augmented dataset. The results showed that some classes, Dentist, Elec. Eng. and DBA, consistently achieved high scores across all three datasets. This suggested that the BERT model was proficient at correctly classifying these classes. These categories might have had distinctive language patterns or keywords that the model captured effectively, leading to a robust performance.

Dataset influence: The variation in F1 scores across the three datasets highlighted the impact of data quality and diversity. The augmentation technique was effective in improving the model's performance. This suggested that the Indeed augmented dataset was well-constructed and contributed positively to the model's training. In contrast, the F1 scores for the ChatGPT dataset showed slight variability, potentially reflecting the unique language patterns and context of the generated resumes.

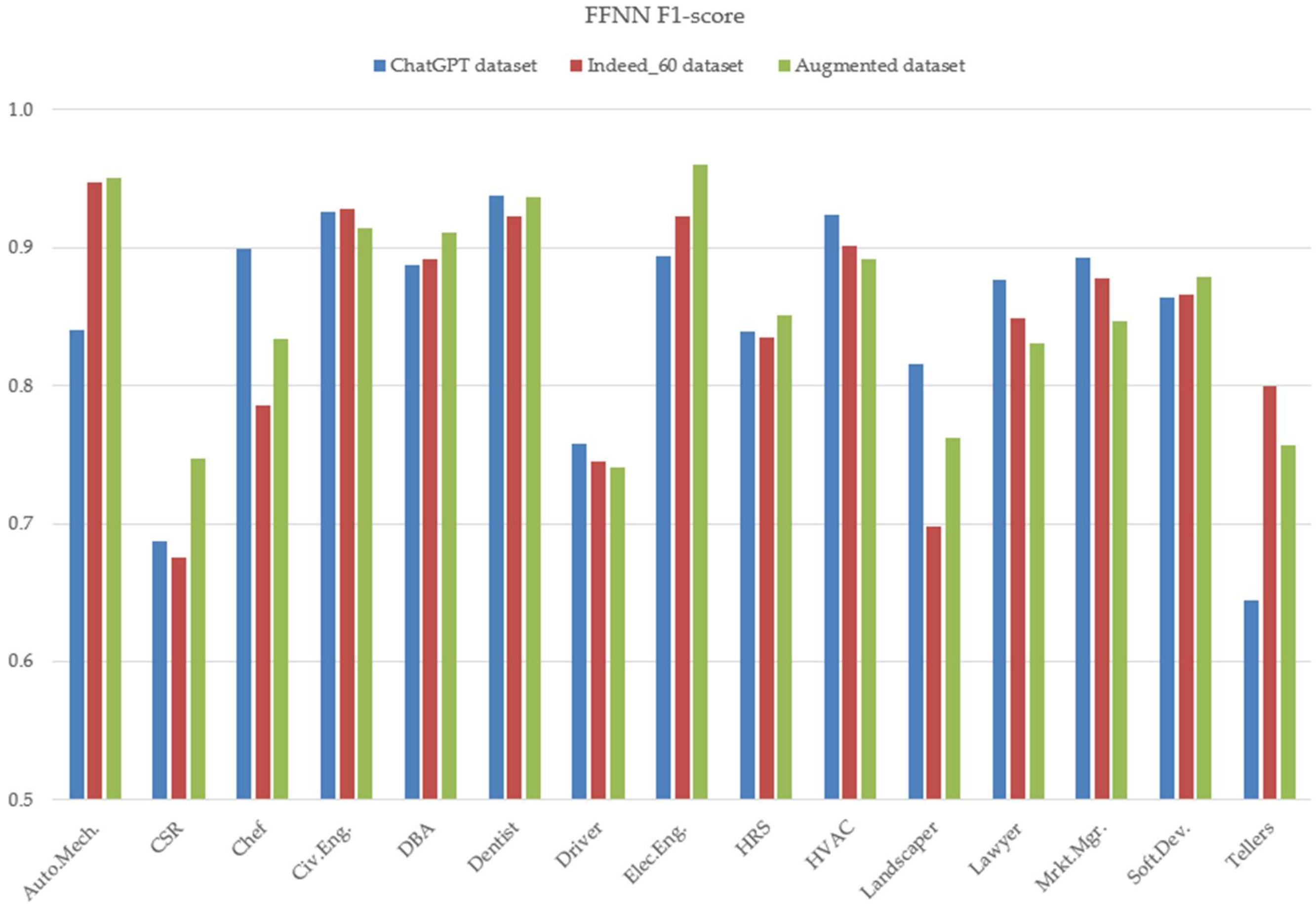

Figure 2. FFNN overall F1 score results for main use case.

As anticipated, the FFNN exhibited shortcomings in accurately categorizing specific classes due to the presence of overlapping descriptions within the resumes. Conversely, the BERT model effectively addressed these instances of overlap because it benefited from pretrained contextual embeddings and its enhanced ability to capture language nuances, which gave it a significant advantage in text classification tasks.

In both the FFNN and BERT models, the presence of classes with fewer samples (sparse classes) had a noticeable impact on the accuracy results. These classes, such as landscaper and tellers, exhibited lower F1 scores compared to classes with more training data. The introduction of the Indeed augmented dataset significantly improved the F1 scores for these sparse categories in both models. The Indeed augmented dataset, which added more examples and diversity to the training data, enabled the models to better capture the nuances and features of these underrepresented classes. This enhancement was particularly evident in the BERT model's results, where the F1 scores for landscaper and tellers showed notable improvements.

Finally, the sparse classes with limited training samples initially had a negative impact on accuracy in both the FFNN and BERT models. However, the use of the Indeed augmented dataset mitigated this effect by providing additional training examples, resulting in improved F1 scores for these sparse classes. The Indeed augmented dataset played a crucial role in helping the models to better capture and classify the under-represented classes.

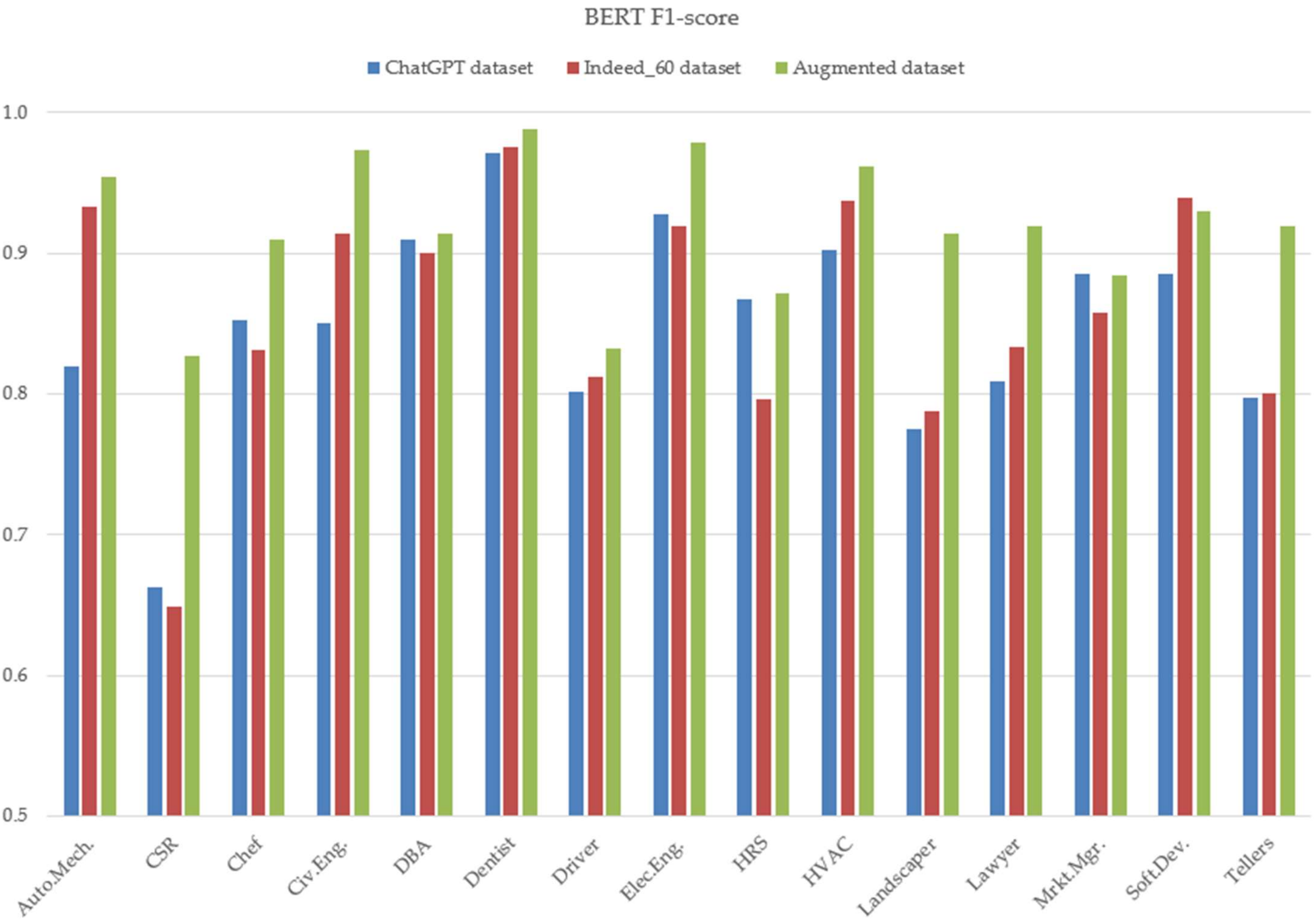

Figure 3. BERT overall F1 score results for main use case.

5. Discussion

In conclusion, the Indeed augmented dataset was proven to add more value to training models for the multiclass classification task. While the BERT model demonstrated exceptional accuracy, the FFNN's performance was commendable, albeit with some limitations, especially in certain job category classes.

Future iterations of this methodology offer potential enhancements to further refine the multiclass classification of resumes. Addressing misclassifications might require the addition of another classification layer to discern more specific characteristics within resumes. In upcoming research, we plan to explore variants of BERT trained on data more aligned with our specific challenge, such as JobBERT [33], or even train the model using our own dataset to evaluate its performance. Moreover, as an alternative direction, by expanding the dataset to include multilingual resumes and employing multilingual variants of BERT, we can enhance the classification's comprehensiveness and global applicability.

An exciting prospect is the fusion of real-world and synthetic datasets, especially in metadata extraction tasks such as entity recognition. Combining synthetic data generated using models like ChatGPT or other open-source large language models (LLMs) with real-world data could herald significant advancements in e-recruitment automation, with potential applicability across various domains.

Additionally, integrating LLMs, known for their expertise in natural language processing (NLP) tasks, offers the potential to enhance their ability to capture complex patterns in unstructured data, like business descriptions or job postings. The goal would be to streamline the matching of resumes with job postings. This direction holds promise for significantly improving e-recruitment processes and may find utility in broader contexts.

Author Contributions: G.T., P.S. and P.Z. conceptualized the research. The methodology was framed by P.Z. and P.S. Data curation and formal analysis were performed by P.S. and P.Z., with software development conducted by P.S. and P.Z.; P.S. drafted the original manuscript, with subsequent reviews performed by P.Z. and G.T. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data and the notebooks presented in this study are available on request to the corresponding author.

Acknowledgments: This paper incorporated information from O*NET Web Services [27,28] by the U.S. Department of Labor, Employment and Training Administration (USDOL/ETA). O*NET® is a trademark of USDOL/ETA.

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

References

1. Ye, J.; Chen, X.; Xu, N.; Zu, C.; Shao, Z.; Liu, S.; Cui, Y.; Zhou, Z.; Gong, C.; Shen, Y.; et al. A Comprehensive Capability Analysis of GPT-3 and GPT-3.5 Series Models. *arXiv* **2023**, arXiv:2303.10420. [CrossRef]
2. Kuchnik, M.; Smith, V.; Amvrosiadis, G. Validating Large Language Models with ReLM. ArXiv [Cs.LG]. *arXiv* **2022**, arXiv:2211.15458. [CrossRef]
3. OpenAI API. Available online: https://bit.ly/3UOELSX (accessed on 29 September 2023).
4. White, J.; Fu, Q.; Hays, S.; Sandborn, M.; Olea, C.; Gilbert, H.; Elnashar, A.; Spencer-Smith, J.; Schmidt, D. A Prompt Pattern Catalog to Enhance Prompt Engineering with ChatGPT. *arXiv* **2023**, arXiv:2302.11382. [CrossRef]
5. Strobelt, H.; Webson, A.; Sanh, V.; Hoover, B.; Beyer, J.; Pfister, H.; Rush, A.M. Interactive and Visual Prompt Engineering for Ad-hoc Task Adaptation with Large Language Models. *IEEE Trans. Vis. Comput. Graph.* **2023**, *29*, 1146–1156. [CrossRef]
6. Liu, Y.; Deng, G.; Xu, Z.; Li, Y.; Zheng, Y.; Zhang, Y.; Zhao, L.; Zhang, T.; Liu, Y. Jailbreaking ChatGPT via Prompt Engineering: An Empirical Study. *arXiv* **2023**, arXiv:2305.13860. [CrossRef]
7. Gao, A. Prompt Engineering for Large Language Models. *Soc. Sci. Res. Netw.* **2023**, *in press*. [CrossRef]
8. Liu, V.; Chilton, L.B. Design Guidelines for Prompt Engineering Text-to-Image Generative Models. In Proceedings of the CHI '22: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems, New Orleans, LA, USA, 29 April–5 May 2022; pp. 1–23. [CrossRef]
9. Sabit, E. Prompt Engineering for ChatGPT: A Quick Guide to Techniques, Tips, And Best Practices. *TechRxiv* **2023**. [CrossRef]
10. Josifoski, M.; Sakota, M.; Peyrard, M.; West, R. Exploiting Asymmetry for Synthetic Training Data Generation: SynthIE and the Case of Information Extraction. *arXiv* **2023**, arXiv:2303.04132. [CrossRef]
11. Xu, B.; Wang, Q.; Lyu, Y.; Dai, D.; Zhang, Y.; Mao, Z. S2ynRE: Two-stage Self-training with Synthetic data for Low-resource Relation Extraction. In Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Toronto, ON, Canada, 9–14 July 2023; Association for Computational Linguistics: Toronto, ON, Canada, 2023; pp. 8186–8207. [CrossRef]
12. Whitehouse, C.; Choudhury, M.; Aji, A.F. LLM-powered Data Augmentation for Enhanced Crosslingual Performance. *arXiv* **2023**, arXiv:2305.14288. [CrossRef]
13. Jeronymo, V.; Bonifacio, L.; Abonizio, H.; Fadaee, M.; Lotufo, R.; Zavrel, J.; Nogueira, R. InPars-v2: Large Language Models as Efficient Dataset Generators for Information Retrieval. *arXiv* **2023**, arXiv:2301.01820. [CrossRef]
14. Veselovsky, V.; Ribeiro, M.H.; Arora, A.; Josifoski, M.; Anderson, A.; West, R. Generating Faithful Synthetic Data with Large Language Models: A Case Study in Computational Social Science. *arXiv* **2023**, arXiv:2305.15041. [CrossRef]
15. Abonizio, H.; Bonifacio, L.; Jeronymo, V.; Lotufo, R.; Zavrel, J.; Nogueira, R. InPars Toolkit: A Unified and Reproducible Synthetic Data Generation Pipeline for Neural Information Retrieval. *arXiv* **2023**, arXiv:2307.04601. [CrossRef]
16. Bayer, M.; Kaufhold, M.A.; Reuter, C. A Survey on Data Augmentation for Text Classification. *ACM Comput. Surv.* **2022**, *55*, 7. [CrossRef]
17. Shi, Z.; Lipani, A. Rethink the Effectiveness of Text Data Augmentation: An Empirical Analysis. *arXiv* **2023**, arXiv:2306.07664. [CrossRef]
18. Kumar, V.; Choudhary, A.; Cho, E. Data Augmentation using Pre-trained Transformer Models. *arXiv* **2021**, arXiv:2003.02245. [CrossRef]
19. Li, Y.; Wang, X.; Miao, Z.; Tan, W.C. Data augmentation for ML-driven data preparation and integration. *ACM Proc. VLDB Endow.* **2021**, *14*, 3182–3185. [CrossRef]
20. Devlin, J.; Chang, M.W.; Lee, K.; Toutanova, K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *arXiv* **2019**, arXiv:1810.04805. [CrossRef]
21. Malinowski, J.; Keim, T.; Wendt, O.; Weitzel, T. Matching people and jobs: A bilateral recommendation approach. In Proceedings of the 39th Annual Hawaii International Conference on System Sciences (HICSS'06), Kauai, HI, USA, 4–7 January 2006; p. 137c. [CrossRef]

22. Yi, X.; Allan, J.; Croft, W.B. 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, Amsterdam, The Netherlands, 23–27 July 2007. [CrossRef]
23. Tallapragada, V.V.S.; Raj, V.S.; Deepak, U.; Sai, P.D.; Mallikarjuna, T. Improved Resume Parsing based on Contextual Meaning Extraction using BERT. In Proceedings of the 2023 7th International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, 17–19 May 2023; pp. 1702–1708. [CrossRef]
24. Jiechieu, K.F.; Tsopze, N. Skills prediction based on multi-label resume classification using CNN with model predictions explanation. *Neural Comput. Appl.* **2020**, *33*, 5069–5087. [CrossRef]
25. Li, X.; Shu, H.; Zhai, Y.; Lin, Z. A Method for Resume Information Extraction Using BERT-BiLSTM-CRF. In Proceedings of the 2021 IEEE 21st International Conference on Communication Technology (ICCT), Tianjin, China, 13–16 October 2021; pp. 1437–1442. [CrossRef]
26. Vukadin, D.; Kurdija, A.S.; Delač, G.; Šilić, M. Information Extraction from Free-Form CV Documents in Multiple Languages. *IEEE Access* **2021**, *9*, 84559–84575. [CrossRef]
27. O*NET Code Connector. Available online: https://www.onetcodeconnector.org/ (accessed on 29 September 2023).
28. "Welcome to the O*Net Web Services Site!" O*NET Web Services. Available online: https://services.onetcenter.org/ (accessed on 29 September 2023).
29. Anand, Y.; Nussbaum, Z.; Duderstadt, B.; Schmidt, B.; Mulyar, A. GPT4All: Training an Assistant-Style Chatbot with Large Scale Data Distillation from GPT-3.5-Turbo. 2023. Available online: https://github.com/nomic-ai/gpt4all (accessed on 29 September 2023).
30. Miller, G.A. WordNet: A lexical database for English. *Commun. ACM* **1995**, *38*, 39–41. [CrossRef]
31. Hugging Face Libraries. Available online: https://huggingface.co/docs/hub/models-libraries (accessed on 29 September 2023).
32. Skondras, P.; Psaroudakis, G.; Zervas, P.; Tzimas, G. Efficient Resume Classification through Rapid Dataset Creation Using ChatGPT. In Proceedings of the Fourteenth International Conference on Information, Intelligence, Systems and Applications (IISA 2023), Volos, Greece, 10–12 July 2023.
33. Decorte, J.-J.; Van Hautte, J.; Demeester, T.; Develder, C. JobBERT: Understanding Job Titles through Skills. *arXiv* **2021**, arXiv:2109.09605. [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.