International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048
Volume 7, Issue 3, pp 180-190, March-2023
https://doi.org/10.47001/IRJIET/2023.703029

Enhancing Job Matching Accuracy: A Vector Search Approach for Resume-to-Job Description Alignment

[1]Leela Gowtham Yanamaddi, [2]Balaji Kummari

[1]CEO and VP of Engineering, scale.jobs 537 Payne Rd, Woodstock, GA, USA 30188
[2]CTO, scale.jobs 1-84, Beside Venugopala Swamy Temple, Rayanapadu, Vijayawada, AP 521241, India
Authors E-mail: [1]leela.yanamaddi@gmail.com, [2]balajievlo@gmail.com

***Abstract* - The term "resume matching" refers to the practice of comparing an applicant's written work (CV) or resume with specific job qualifications or job descriptions. The purpose of this process is to determine how well a candidate's relevant traits, such as their abilities, qualifications, and experience, match the requirements of the job. Students enrolled in these types of courses often learn how to analyse job postings for essential qualifications and then craft their resumes to emphasise those areas. Human resources (HR) professionals, on the other hand, have the education and experience to sift through stacks of resumes for the best possible fit with their company. It is common practice to use an automated system to compare resume content with job postings, and then to rank or score the results based on how similar the two sets of words are. With a vast pool of candidates and detailed job postings, though, this procedure can quickly grow tedious. Using vector search techniques to align job applicants' resumes with suitable job descriptions, this research proposes a novel strategy for enhancing the job matching process. Employers and job-seekers alike stand to gain from the suggested system's efforts to improve the precision and efficacy of employment referrals. In this article, we offer a dataset that includes software developer resumes culled from an open Telegram channel that is specifically for Israeli high-tech job seekers. In addition, we offer an NLP-based approach to resume matching that makes use of neural representations of phrases, keywords, and named entities to achieve first-rate outcomes.**

We show that our method outperforms the top algorithm for matching resumes with job openings by evaluating it with both human and automated annotations.

Keywords: Vector Search Approach, Resume, Job Matching, Natural language processing (NLP).

I. INTRODUCTION

Despite several improvements to ensure the recruitment process runs well, the shortlisting of candidates is still not completely automated or open-sourced. Combining Natural Language Processing with Machine Learning could be the key to solving these challenges [1].

It is difficult to accept CV and resume data forms in a standardised style because there are no norms for CV and resume authoring, even though the formats used are not completely unstructured.

An Approach to Natural Language Processing: In spite of the many pros and cons, the pursuit of an automated system that allows companies to swiftly choose qualified individuals and allows applicants to demonstrate their creativity by submitting a single application to several organisations has persisted. The capacity to teach computer new tasks and understand and work with unstructured data is prerequisites for document analysis tasks, which includes reviewing resumes [2]. An Approach Based on Machine Learning: Researchers supplemented Natural Language Processing with Machine Learning to enhance the precision and validity of their models. Many different ways to train a model and find solutions to challenges exist due to the abundance of Machine Learning methodologies. For the purpose of determining the goodness or evil of an object or idea, many machine learning-based methods are utilised, including Naive Bayes classifiers, Decision trees, and Logistic regression.

Problems in Manual Screening

Screening a huge volume of resumes by hand requires at least a day.

- A recruiter is unlikely to give further consideration to resumes sent in if they find only four or six that meet their criteria during the initial review. Because of this, your chances of having a well-written resume shortlisted are lower. It takes a lot of time to go over each resume, and it's impossible to manually organise and manage a lot of resumes.
- Prejudice is natural anywhere there has been interaction between humans.

ISSN (online): 2581-3048
Volume 7, Issue 3, pp 180-190, March-2023
https://doi.org/10.47001/IRJIET/2023.703029

Characteristics of Resume Screener

- Time is saved.
- Supports Multiple Formats.
- Better Recruiting.
- Do away with prejudice.
- Installation Made Simple.

How vector search can help you find the perfect candidate

During the process of finding an ideal individual to fill a position. A lot of things need to be thought about. That is, even before the recruiter gets their hands on the CV. There are a lot of checks that it must pass. Typically, an applicant tracking system (ATS) will analyse the resume and apply a keyword filter. Recruiters still don't have enough time to study each résumé, as the old internet adage goes. On average, it takes recruiters 20 seconds to determine if an applicant is a good fit for a position. Due to the fact that work history speaks louder than a résumé, this method is incorrect [3].

In this piece, we'll look at the issues with hiring and how vector databases can help. The same goes for technical recruiters interested in state-of-the-art technology and search enthusiasts working on vector databases and semantic search. If you follow along, we can figure out a critical issue.

Identifying the Problems in the Current Recruiting Condition

Resumes are complex documents that showcase a candidate's professional history, personal endeavours, technical abilities, and extracurricular activities (e.g., charity work, blogging, open-source projects, etc.). A candidate's desire to show that they are valued and interested in being a part of a company's team is also reflected in this.

Vector search and vector databases can alleviate the time-consuming issue of reviewing resumes. It's an ideal resource for identifying top talent. Vector search is important, so let's start by learning how it works.

How Vector Search Works

Vector search is a mathematical technique that converts data (such as text, pictures, or audio) into a numerical representation known as a vector. A database has these vectors. The same procedure is used to convert search queries into vectors. After receiving a vector as a search query, the Vector Database Engine will scan the database for other vectors that are highly similar to it. Finding the distance between the vectors is how we find out how similar they are. The algorithms and vector databases utilised can cause the distances to change. Cosine, Dot Product, Manhattan, Hamming, and the Euclidean are just a few examples of common ones [4]. The most relevant results for the search are the vectors that have the minimum distances. Because of this, vector search will efficiently and accurately retrieve relevant results, regardless of how complicated or nebulous your query is.

Why is Vector Search Important in Recruiting

Bias exists in the human mind. Despite efforts to consciously lessen the prejudices. When making decisions, humans are prone to having blind spots. This carries over into our opinions of candidates in terms of gender, ethnicity, religion, culture, and nation. Those around us can shape our perspective, so we can choose our friends and acquaintances more carefully. And to eliminate this prejudice. Sometimes, the best tool is Vector Search.

Because all of the resumes are numerical or vector-based, the system processes the job description as the search query. Additionally, it finds the best resume vectors to match it with. The findings are quick, objective, and show people who are a good fit for the position without favouring any one set of keywords over another.

Since vector searches treat all data as numbers, they do not discriminate. As a result, you have access to a pool of brilliant individuals from all walks of life who can collaborate to find novel approaches to old challenges.

How Can You Create a Recruiting Solution Using Vector Databases or Vector Search Engines?

In the market, you can find a lot of Vector Databases. Plus, a tonne of them are freely available and open-sourced. Another option is Resume Matcher, a free and open-source program.

First, we may count the number of vector datasets available:

- Milvus
- Qdrant
- Weaviate
- Vespa
- Pinecone

This article by Dmitry Kan goes into much detail regarding them. The databases listed above are not exhaustive, but they are sufficient for our use case.

How to make your resumes work with Vector Databases?

Just like it sounds, this is really easy. Our accessible APIs, documentation, and examples make this a breeze. As far as this article is concerned, my only comment is:

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 7, Issue 3, pp 180-190, March-2023

https://doi.org/10.47001/IRJIET/2023.703029

1. Use a script to parse the set of resumes.
2. Send the resumes to a vector database.
3. Send in your search query.
4. Let the magic happen.
5. You will get the top answers based on what job description you sent in.

Resume Matcher utilises Qdrant to address the identical issue; nevertheless, this deserves its own essay (which will be published shortly). For applicants, it's a win-win: they can enhance their resumes and vice versa. Along with hiring managers in search of more qualified individuals.

II. LITERATURE REVIEW

The challenge of automatically matching resumes or curriculum vitae with open positions or job descriptions is known as the resume-vacancy matching problem [5]. The purpose of this evaluation is to determine whether the applicant possesses the necessary qualifications for the open post. The labour market as a whole can benefit from the time savings, increased efficiency, and streamlined selection process that results from automating a resume-vacancy evaluation.

Due to the complexity of the process and the need to pay close attention to a number of parameters, this activity is far from simple and includes numerous components. Examining the applicant's résumé for pertinent information and aligning it with the job description is what this process is all about. Typically, these components consist of keywords, experience, and skills. The candidate's current skill set should be a good fit for the position, and their previous work history should be evaluated to determine if it meets the requirements of the open position. Keywords are a common way for employers to highlight necessary skills and experience in job descriptions. In order to match resumes with open positions, it is common practice to look for certain keywords in the resume and compare them to those in the job posting. There are many instances, though, where a perfect match just isn't achievable. The candidate's skills and experience may somewhat match the requirements of the position, but not entirely. Using sophisticated methods beyond simple matching is crucial for gaining a better grasp of the candidate's abilities and the job's semantics [6].

The problem of rating resumes according to available positions has been addressed by several automated methods in recent years. Some examples of these techniques are NLP, machine learning, and information retrieval.

The main problem with these systems is that resumes have different forms, which makes it hard to get the right information out of them. Not to mention how fast things change in the job market, thus matching systems need to be able to adapt to new skill requirements and job descriptions.

The rule-based system is simple to create and implement since it uses predefined criteria and keywords to match job openings with applicants' resumes.

More precise matches for their organisation can be achieved when recruiters modify these algorithms according to certain job specifications. However, rule-based algorithms might not be able to adapt quickly enough to deal with unique or rapidly changing job requirements, or with more complex matching scenarios. Using keywords too heavily increases the risk of false positives or mismatches when candidates use different terms or phrases to describe the same skills. One of the earliest systems to employ rule-based ways to match resumes with job postings was Resumix [7]. It used keywords and specified criteria.

The rules for resume-vacancy matching are defined using taxonomies and ontologies in works [8].

Systems that utilise natural language processing (NLP) examine the context and semantics of both the job description and resume in order to enhance the accuracy of the matching process.

They are more effective at matching people because they can find relevant experiences and talents even when no particular keywords are provided. Unfortunately, these algorithms can be swayed by biassed training data, which could lead them to discriminate against certain groups or favour applicants with particular qualities. It can take a lot of effort and processing power to build these systems, and even more time to refine the models. Examples include article [9], which uses word embeddings and the bag-of-words (BOW) model to describe text data, and article [10], which uses the BOW model with TF-IDF weights, which stand for term frequency. Word n-gram vectors [11] are used to represent the text in the paper [12].

To rank resumes or openings, all of these algorithms generally use vector similarity for the specified text format. Various similarity metrics, including cosine, Dice, Jaccard, and Jaccard overlap, are presented in these publications [13]. Some state-of-the-art approaches to text representation use word embedding layers [14] and feature extractors such as recurrent neural networks (RNNs) or convolutional neural networks (CNNs) [15,16].

There is a subset of resume-matching systems to which our suggested approach belongs.

ISSN (online): 2581-3048
Volume 7, Issue 3, pp 180-190, March-2023
https://doi.org/10.47001/IRJIET/2023.703029

General machine learning models are the subsequent category of resume-matching systems. These models can enhance their matching accuracy over time as they process more data and get input from recruiters. These algorithms are flexible enough to respond to the needs of different recruiters and companies by tailoring matches accordingly. However, data security and privacy are issues that arise since these systems rely on large amounts of data.

It is difficult to understand how specific matches are made by some of the algorithms because they operate like black boxes, which can lead to biases. In order to construct end-to-end matching models, systems like as [17] employ neural networks. A lot of research, like [18], treats the resume matching problem like a recommender system that suggests the best job postings based on a person's CV.

For the purpose of the matching, the writers of [19] make use of historical data.

Information retrieval systems employ ranking algorithms like OKAPI BM25 [20] to sort items according to how relevant they are to a certain query. It is one of the most precise computer algorithms using a bag-of-words paradigm that may be used as a foundation for resume evaluation exams.

The goal of the natural language processing task known as extractive summarisation is to create a synopsis of a provided text by pulling out and merging key phrases and sentences [21]. In order to construct the summary, the most informative sentences from the original article are extracted. There have been much advancement in statistical and machine learning methods; a fair overview of the area is provided in [22]. Transformers are essential to current top-level methods for extractive summarisation [23].

Words and phrases that have special significance in relation to a certain field or setting are called keywords. Resume matching and text analysis frequently make use of keywords to denote crucial abilities, credentials, and experience as well as topic-or job-specific terminology. Keywords are crucial for finding and matching important information when analysing text, such as resumes or job vacancies [24].

The goal of research in the area known as "keyword extraction" is to find relevant content and mechanically extract relevant keywords or phrases from it. Applications of natural language processing (NLP) including topic modelling, text categorisation, information retrieval, and document summarisation rely heavily on this process. There is a wide variety of approaches to keyword extraction, including those based on tf-idf, graphs (like TextRank), and co-occurrence (like RAKE and YAKE) algorithms. Last but not least, neural techniques have been the talk of the town recently [25].

Any physical thing or abstract idea with an official name is called a named entity. Anyone, anything, or anything can be referred to by a name or label, including people, places, dates, times, products, and organisations. An essential part of NLP and IR is the ability to identify and extract named things from text. Transfer learning, recurrent neural networks, transformers, conditional random fields, and graph neural networks are among the probabilistic models used by named entity recognition (NER) techniques.

Using summarisation and improving the generated summaries by named entities or keywords detected in texts, we present a strategy for effective resume-vacancy matching in this paper. Every résumé is given a ranking based on the similarity of its upgraded text, which is calculated using statistical and semantic vector representations. This approach does not necessitate training since it is unsupervised. We demonstrate the validity of our strategy through a comprehensive experimental evaluation.

III. ARCHITECTURE OF THE STUDY

Using the capabilities of both SingleStoreDB and OpenAI, this study will demonstrate a real-world application of resume evaluation. We present a new approach to resume matching that goes above and beyond the norm by taking a closer look at the finer points of resume content as well as the unique requirements of each job posting.

We start by reading a PDF résumé, extracting the text, and saving it in a specific database table using SingleStoreDB's inherent data management and querying capabilities. But that's not all! Next, OpenAI's LLM analyses the resume's text and extracts its main points. Our SingleStoreDB table is updated with this condensed version and its vector form.

We convert the job description into vector form so that it may be used to match resumes. To find the embeddings for the resume summaries, we use a dot product operation on the table that stores them. Consequently, we get applications from qualified candidates.

For an extra degree of accuracy, OpenAI's LLM is used to thoroughly examine the matched resumes in conjunction with the job description, providing a thorough evaluation.

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 7, Issue 3, pp 180-190, March-2023

https://doi.org/10.47001/IRJIET/2023.703029

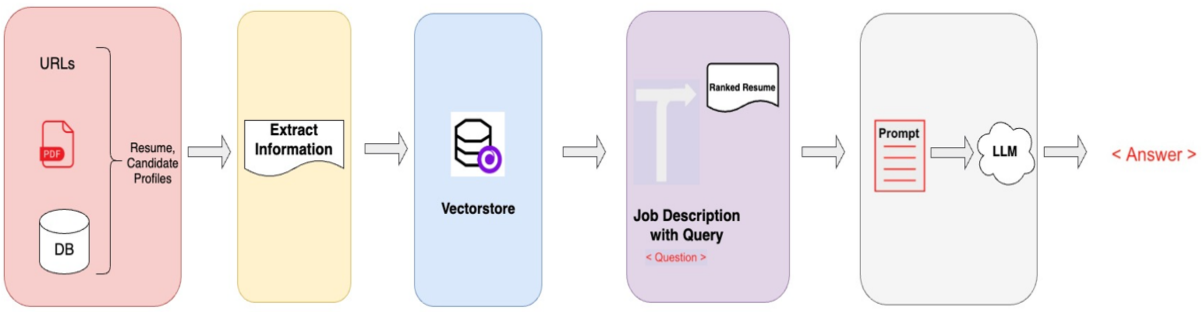

Figure 1: A pipeline for processing resumes and matching them to job descriptions using a combination of information extraction, vector search, and a language model (LLM)

The figure 1 appears to depict a pipeline for processing resumes and matching them to job descriptions using a combination of information extraction, vector search, and a language model (LLM). Here's a breakdown of the process:

1. Input Sources (on the left):

- URLs / PDF / Database (DB): This represents different sources of resumes or candidate profiles, such as URLs, PDFs, or a database.

2. Extract Information:

- The next step involves extracting relevant information from these resumes or profiles. This might involve extracting text, skills, experience, and other relevant fields from the documents.

3. Vectorstore:

- After extracting the information, it is stored in a vectorstore (a database of vectors). This means the resumes or candidate profiles are converted into embeddings (vector representations) that can be used for similarity search.

4. Job Description with Query:

- A **job description** or a **query** (likely a job search query or a job-related question) is input into the system, which is then matched against the vectorized resumes or profiles to get **ranked resumes** that are most relevant to the job description.

5. Prompt to LLM:

- The ranked resumes are then processed, and a **prompt** is generated, which is passed to a **Large Language Model (LLM)**. The LLM likely generates an **answer** or further insight based on the prompt, such as why a particular resume fits a job or suggestions for the next steps.

6. Final Answer (on the right):

- The **LLM** processes the prompt and provides an **answer**, likely used to recommend or rank the most suitable resumes for a given job description.

Cosine Similarity

Overview:

Cosine similarity is a popular metric for comparing two vectors; it's used frequently in ML, text mining, and information retrieval applications. It determines the orientation of two vectors in a three-dimensional space by measuring the cosine of the angle between them.

Definition and Formula:

To find the degree of cosine similarity between two vectors A and B, we use the following formula:

$$A = [a_1, a_2, ..., a_n]$$

Vector A

$$B = [b_1, b_2, ..., b_n]$$

Vector B

$$S(A, B) = \frac{A \cdot B}{||A|| \ ||B||} = \frac{\sum_{i=1}^{n} a_i \cdot b_i}{\sqrt{\sum_{i=1}^{n} a_i^2} \cdot \sqrt{\sum_{i=1}^{n} b_i^2}}$$

Cosine similarity formula

Applications in Similarity Search:

Document Similarity: Used frequently in NLP for text similarity measurement, which helps with document retrieval and clustering.

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 7, Issue 3, pp 180-190, March-2023

https://doi.org/10.47001/IRJIET/2023.703029

Recommendation Systems: Collaborative filtering uses Cosine Similarity to compare item or user profiles and provide suggestions.

Image Comparison: Its primary use is in image feature vector comparisons, a subfield of computer vision.

Advantages:

Angle Measurement: This method is useful for comparing documents of varying lengths since it calculates the cosine of the angle between vectors.

Normalization: The metric is directionally sensitive rather than magnitude sensitive since it normalises vector lengths by design.

Limitations:

Zero Vector Issue: It gets undefined when dealing with zero vectors, therefore it's not really useful.

Not a Metric: Cosine similarity is not a valid metric since it fails to meet the triangle inequality.

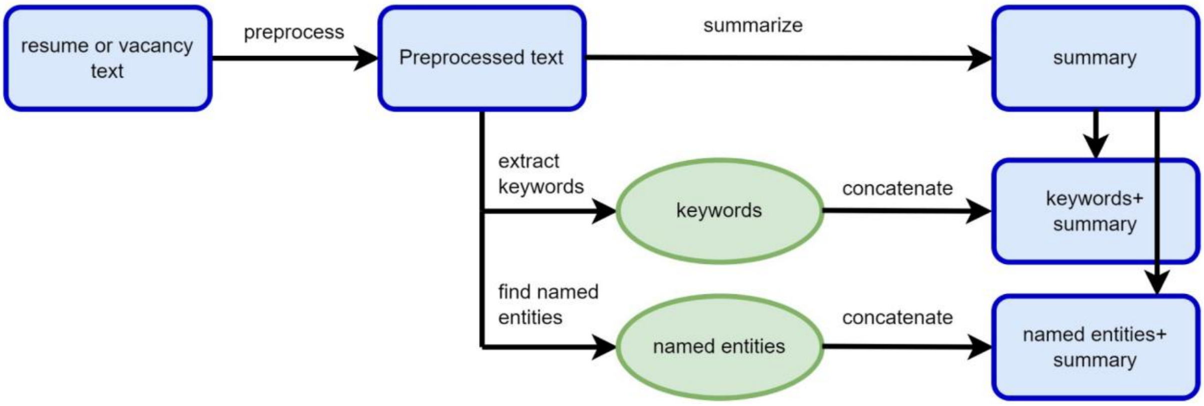

Figure 2: The virtual machine method's text kinds

The figure 2 outlines a flow diagram related to text processing for resumes or job vacancy descriptions using the VM (Vector Matching) method. It describes a pipeline that processes text, extracts relevant information, and combines it for further analysis. Here's a step-by-step description:

1. Input: Resume or Vacancy Text:

- The process starts with a block labeled "resume or vacancy text," representing the initial raw input text from a resume or a job vacancy posting.

2. Preprocessing:

- The text undergoes a preprocessing step, where it is likely cleaned, normalized, or formatted to make it suitable for further analysis.

3. Preprocessed Text:

- After preprocessing, the "Preprocessed text" block is created, which is now in a more structured format.

4. Summarization:

- The preprocessed text is passed through a summarization step, generating a condensed summary of the text that highlights the key points or information from the resume or job description.

5. Keywords Extraction:

- The preprocessed text also undergoes keyword extraction, where important keywords (such as skills, experience, or relevant terms) are identified and extracted. These are stored in the "keywords" block.

6. Named Entity Recognition (NER):

- Another process is used to find named entities, such as names of people, organizations, locations, or specific technical terms. These are stored in the "named entities" block.

7. Concatenation with Summary:

- The extracted keywords are concatenated with the summary to form a combined "keywords + summary" block.
- Similarly, the named entities are concatenated with the summary to create a combined "named entities + summary" block.

IV. RESULTS AND DISCUSSION

Table 1 shows the results of the top ten VM variants when cosine similarity is applied, sorted by Krippendorff's alpha value. A grey backdrop colour indicates the highest scores.

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 7, Issue 3, pp 180-190, March-2023

https://doi.org/10.47001/IRJIET/2023.703029

Table 1: Outcomes for the VM approach using cosine similarity on data annotated by humans

Method	Resume Text Data	Vacancy Text Data	Text Representation	Krippendorff's Alpha	Spearman's Correlation
Used VM cosine Method	Summary of data	Summary of data	Text based on SBERT	0.1718	0.0862
Used VM cosine Method	Full text	Summary of data	Text based on SBERT	0.1891	0.1066
Used VM cosine Method	Summary of data	Full text	Character n-grams	0.3215	0.2515
Used VM cosine Method	Summary of data	Full text	tf-idf + char-ng + word-ng + Text based on SBERT	0.3215	0.2515
Used VM cosine Method	Summary of data	Full text	tf-idf + char-ng + word-ng + Text based on SBERT	0.3740	0.3020
Used VM cosine Method	Summary of data	Full text	Character n-grams	0.3992	0.3360
Used VM cosine Method	Summary of data	Full text	tf-idf + char-ng + word-ng + Text based on SBERT	0.4472	0.3893
Used VM cosine Method	Full text	Full text	tf-idf + char-ng + word-ng + Text based on SBERT	0.5003	0.4507
Used VM cosine Method	Full text	Full text	Character n-grams	0.5183	0.4707
Used VM cosine Method	Summary of data	Summary of data	Text based on SBERT	0.2611	0.1866

With the help of KeyBERT, we were able to extract multi-word keywords from job postings and resumes that had word counts between one and three. See Table 2 for some sample keywords and their corresponding KeyBERT scores for the two data sets.

Table 2: Keyword examples

Keywords	Data Type	Score
Java backend development	Resume	0.6156
Backend java developer	Resume	0.6135
Java backend	Resume	0.6066
Backend java	Resume	0.5917
Now backend java	Resume	0.5800
Software development	Vacancy	0.5359
Using software development	Vacancy	0.5290
Purpose development and	Vacancy	0.5233
Purpose development	Vacancy	0.5101
Maintenance of software	Vacancy	0.5049

Human-Annotated Data Evaluation

In these experiments, for each of the 30 human-annotated resumes, we score five specific job openings. We check the ranks against the mean rating that our two human annotators produced. Table 3 displays the results for our virtual machine technique, OKAPI BM25 and BERT-based rank, our two baselines. The top scores are shown on a grey backdrop.

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 7, Issue 3, pp 180-190, March-2023

https://doi.org/10.47001/IRJIET/2023.703029

Table 3: Baseline and VM technique evaluation findings on human-annotated data

Method	Resume Data	Vacancy Data	Text Representation	Krippendorff's Alpha	Spearman's Correlation
BERT-rank	0.3055	-0.1779	OKAPI-BM25	0.2262	-0.3071
VM	Full text	Summary of data	Character n-grams	0.6287	0.5907
VM	Full text	Summary of data	tf-idf + char-ng + word-ng + Text based on SBERT	0.6179	0.5793
VM	Full text	Keywords + Summary of data	Character n-grams	0.5885	0.5454
VM	Full text	Keywords + Summary of data	tf-idf + char-ng + word-ng + Text based on SBERT	0.5074	0.4554
VM	Keywords + Summary of data	Full text	tf-idf + char-ng + word-ng + Text based on SBERT	0.4917	0.4443
VM	Keywords + Summary of data	Full text	Character n-grams	0.4917	0.4443
VM	Named Entity + Summary of data	Full text	tf-idf + char-ng + word-ng + Text based on SBERT	0.4917	0.4443
VM	Named Entity + Summary of data	Full text	Character n-grams	0.4917	0.4443
VM	Summary of data	Full text	tf-idf + char-ng + word-ng + Text based on SBERT	0.4917	0.4443
VM	Summary of data	Full text	Character n-grams	0.4917	0.4443
VM	Named Entity + Summary of data	Full text	tf-idf	0.4857	0.4375
VM	Full text	Full text	tf-idf	0.4775	0.4272
VM	Full text	Full text	tf-idf + char-ng + word-ng + SBERT	0.4742	0.4211
VM	Keywords + Summary of data	Full text	tf-idf	0.4715	0.4205
VM	Summary of data	Full text	tf-idf	0.4715	0.4206
VM	Full text	Full text	Character n-grams	0.4696	0.4192
VM	Keywords + Summary of data	Summary	SBERT	0.3694	0.3002
VM	Named Entity + Summary of data	Named Entity + Summary	Character n-grams	0.3644	0.2971
VM	Keywords + Summary of data	Named Entity + Summary	tf-idf + char-ng + word-ng + SBERT	0.3622	0.2891
VM	Keywords + Summary of data	Full text	Word n-grams	0.3551	0.2900
VM	Named Entity + Summary	Full text	Word n-grams	0.3551	0.2900
VM	Summary	Full text	Word n-grams	0.3551	0.2900
VM	Full text	Full text	Word n-grams	0.3551	0.2900
VM	Named Entity + Summary	Full text	SBERT	0.3465	0.2772
VM	Summary of data	Named Entity + Summary	Character n-grams	0.3382	0.2624
VM	Summary of data	Named Entity + Summary	tf-idf + char-ng + word-ng + SBERT	0.3311	0.2555
VM	Named Entity + Summary of data	Summary	SBERT	0.3154	0.2402
VM	Keywords + Summary of data	Named Entity + Summary	Character n-grams	0.3153	0.2376
VM	Full text	Full text	SBERT	0.3096	0.2340

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 7, Issue 3, pp 180-190, March-2023

https://doi.org/10.47001/IRJIET/2023.703029

Table 4: Ranking Guidelines

Parameter	Explanation
Full-Text Utilization	Candidates should ensure that their **entire resume** is submitted when applying for job postings, as using the **full text** has been shown to yield more accurate ranking results compared to summarizing only portions.
Summarized Vacancy Descriptions	Focus on job postings that provide **summarized and concise descriptions** of responsibilities, qualifications, and skills. Such descriptions offer higher clarity, which improves matching accuracy.
Character-n-Gram-Based Representation	Candidates should ensure that their resume uses specific terminology and key phrases relevant to the job. **Character-n-grams** (small chunks of text) help improve search and ranking, especially in specialized fields.
Relevance of Content	Ensure that both the **resume content** and **vacancy descriptions** emphasize **important skills, qualifications, and experiences**. Avoid irrelevant information that could lower ranking accuracy.
Precision of Keywords	Use **precise and specific keywords** that match the job requirements. Jobs with well-defined keywords align better with **character-n-gram-based** models, improving ranking accuracy.
Text Representation	Candidates should ensure that their resumes are represented in a way that can be processed efficiently by text-based models. This includes **consistent formatting** and **clear structure** for better ranking performance.
Impact of Data Types	Utilize the best **text data types** for representation, including job requirements, responsibilities, and skills. A well-rounded approach to the text ensures higher relevance and accuracy in matching.
Evaluation of Fit	When evaluating potential job matches, candidates should ensure their **qualifications, skills, and experience** align directly with the **summarized requirements** provided in the job description.

Key Changes:

- Emphasis on **full-text utilization** for better ranking accuracy.
- Focus on **summarized vacancy descriptions** that provide clarity.
- **Character-n-gram-based representation** for better text matching.
- Importance of **precision** in keywords and **relevance** in content to enhance ranking and performance in text-based models.

These new guidelines described in table 4 highlight how different aspects of resumes and job descriptions contribute to ranking accuracy in the VM method.

V. CONCLUSION

Using Vector Databases & Vector Search in recruitment revolutionizes the conventional and sometimes inaccurate process, making it more efficient, unbiased, and simplified. The system transforms resumes and job descriptions into mathematical vectors, enabling it to match candidates to job roles based on multiple factors, not just keywords. This enables better candidate selection, and the whole process can be automated so that whenever you get the best candidate, you can interview them and hire them. Among the many fields that make use of vector similarity search algorithms, Cosine Similarity stands out for its versatility and resilience. In spite of its many benefits, such as being insensitive to size, knowing its limits is essential for making good use of it. It is still useful for data science and AI experts working on recommendation systems, document retrieval, and similarity searches with careful application. Additionally, we compiled a list of the VM approaches that outperform both baselines, ranked by decreasing Krippendorff's alpha. There is a huge disparity between the baselines and the best approach. We also see that the optimal approach represents text using character n-grams, summarised job openings, and full resumes. Our algorithm achieves better ranking accuracy than both baselines, according to the findings of our evaluation. Using character-n-gram-based text representation with full CV and summarised job posting texts yielded the greatest results, according to our findings. This mixture considerably outperformed the baselines in terms of Krippendorff's alpha. We also looked at the effects of several text data kinds and formats on the performance of our VM method.

ISSN (online): 2581-3048
Volume 7, Issue 3, pp 180-190, March-2023
https://doi.org/10.47001/IRJIET/2023.703029

REFERENCES

[1] Do, D.Q.; Tran, H.T.; Ha, P.T. Job Vacancy and Résumé Analysis for Efficient Matching of Job Offers and Applicants. *In Proceedings of the IEEE International Conference on Data Science and Advanced Analytics, Montreal, QC, Canada*, 17–19 October 2016.

[2] Nocker, M.; Sena, V. Big data and human resources management: The rise of talent analytics. *Soc. Sci.* 2019, 8, 273.

[3] Chala, S.A.; Ansari, F.; Fathi, M.; Tijdens, K. Semantic matching of job seeker to vacancy: A bidirectional approach. *Int. J. Manpow.* 2018, 39, 1047–1063.

[4] McKinney, A. Real Resumix & Other Resumes for Federal Government Jobs: Including Samples of Real Resumes Used to Apply for Federal Government Jobs; *Prep Publishing: Fayetteville, NC, USA*, 2003.

[5] Riabchenko, A. Taxonomy-Based Vacancy: CV Matching. Master's *Thesis, LUT University, Lappeenranta, Finland*, 2022.

[6] Senthil Kumaran, V.; Sankar, A. Towards an automated system for intelligent screening of candidates for recruitment using ontology mapping (EXPERT). *Int. J. Metadata Semant. Ontol.* 2013, 8, 56–64.

[7] Pudasaini, S.; Shakya, S.; Lamichhane, S.; Adhikari, S.; Tamang, A.; Adhikari, S. Scoring of Resume and Job Description Using Word2vec and Matching Them Using Gale-Shapley Algorithm. *In Proceedings of the ICOECA 2021, Bangalore, India*, 18–19 February 2021.

[8] Tejaswini, K.; Umadevi, V.; Kadiwal, S.; Revanna, S. Design and Development of Machine Learning based Resume Ranking System. *Glob. Transit. Proc.* 2022, 3, 371–375.

[9] Kawan, S.; Mohanty, M.N. Multiclass Resume Categorization Using Data Mining. *Int. J. Electr. Eng. Technol.* 2020, 11, 267–274.

[10] Sammut, C.; Webb, G.I. Encyclopedia of Machine Learning; *Springer Science & Business Media: Berlin/Heidelberg, Germany*, 2011.

[11] Stefanovič, P.; Kurasova, O.; Štrimaitis, R. The N-Grams Based Text Similarity Detection Approach Using Self-Organizing Maps and Similarity Measures. *Appl. Sci.* 2019, 9, 1870.

[12] Brown, P.F.; Della Pietra, V.J.; Desouza, P.V.; Lai, J.C.; Mercer, R.L. Class-based n-gram models of natural language. *Comput. Linguist.* 1992, 18, 467–480.

[13] Dice, L.R. Measures of the amount of ecologic association between species. *Ecology* 1945, 26, 297–302.

[14] Mikolov, T.; Sutskever, I.; Chen, K.; Corrado, G.S.; Dean, J. Distributed representations of words and phrases and their compositionality. *In Proceedings of the 27th Annual Conference on Neural Information Processing Systems, Lake Tahoe, NV, USA*, 5–10 December 2013; Volume 26.

[15] Qin, C.; Zhu, H.; Xu, T.; Zhu, C.; Jiang, L. Enhancing person-job fit for talent recruitment: An ability-aware neural network approach. *In Proceedings of the 41st International ACM SIGIR Conference on Research & Development in Information Retrieval, Ann Arbor, MI, USA*, 8–12 July 2018; pp. 25–34.

[16] Zhu, C.; Zhu, H.; Xiong, H.; Ma, C.; Xie, F.; Ding, P.; Li, P. Person-job fit: Adapting the right talent for the right job with joint representation learning. *ACM Trans. Manag. Inf. Syst.* 2018, 9, 12.1–12.17.

[17] Yuan, Z.; Lu, S. Application in person-job fit evaluation of BP neural network in knowledge workers. *J. Wuhan Univ. Technol.* 2010, 32, 515–518.

[18] Bing, Y.; Fei, J. Application in person-job fit evaluation of support vector machine. *J. Cent. South Univ. For. Fechnology (Soc. Sci.)* 2011, 5, 92–94.

[19] Alsaif, S.A.; Sassi Hidri, M.; Eleraky, H.A.; Ferjani, I.; Amami, R. Learning-Based Matched Representation System for Job Recommendation. *Computers* 2022, 11, 161.

[20] Koh, M.F.; Chew, Y.C. Intelligent job matching with self-learning recommendation engine. *Procedia Manuf.* 2015, 3, 1959–1965.

[21] R. Ransing, A. Mohan, N. B. Emberi, and K. Mahavarkar, "Screening and Ranking Resumes using Stacked Model," *in 2021 5th International Conference on Electrical, Electronics, Communication, Computer Technologies and Optimization Techniques (ICEECCOT), 2021: IEEE*, pp. 643-648.

[22] Y. Luo, H. Zhang, Y. Wang, Y. Wen, and X. Zhang, "ResumeNet: A learning-based framework for automatic resume quality assessment," *in 2018 IEEE International Conference on Data Mining (ICDM), 2018: IEEE*, pp. 307-316.

[23] Q. Trinh and T.-T. Dang, "Automatic process resume in talent pool by applying natural language processing," *in Proceedings of International Conference on Logistics and Industrial Engineering* 2021, 2021, pp. 234-240.

[24] A.A. Noor, M. Bakhtyar, B. A. Chandio, R. Gull, and J. Baber, "Automatic cv ranking using document vector and word embedding," *Pakistan Journal of Emerging Science and Technologies*, vol. 2, pp. 1-9, 2021.

[25] A.Mohamed, W. Bagawathinathan, U. Iqbal, S. Shamrath, and A. Jayakody, "Smart talents recruiterresume ranking and recommendation system,"

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 7, Issue 3, pp 180-190, March-2023

https://doi.org/10.47001/IRJIET/2023.703029

in 2018 IEEE International Conference on Information and Automation for Sustainability (ICIAfS), 2018: IEEE, pp. 1-5.

Citation of this Article:

Leela Gowtham Yanamaddi, Balaji Kummari, "Enhancing Job Matching Accuracy: A Vector Search Approach for Resume-to-Job Description Alignment" Published in *International Research Journal of Innovations in Engineering and Technology - IRJIET*, Volume 7, Issue 3, pp 180-190, March 2023. Article DOI https://doi.org/10.47001/IRJIET/2023.703029

****