

RESUME RANKING SYSTEM

Using Natural Language Processing

Prepared by – Lakshitha Kumudu Ranga

Index – 200523J

Department of Computer Science and Engineering

University of Moratuwa

Contents

1. Introduction 3

 1.1 Problem Description 3

 1.2 Objective 3

2. Literature Review 3

3. Methodology..... 5

4. Tools and Technologies 5

5. Expected Outcomes 5

6. Improvements..... 5

7. Conclusions 6

8. References..... 6

Figures

Figure 1- Project Process..... 3

Figure 2- Old process 4

Figure 3- New Process..... 4

1. Introduction

1.1 Problem Description

Many companies use a large amount resume screening process during the recruitment process in order to find qualified applicants. This manual method is costly, time-consuming, and biased by human error. Recruiters often read through resumes for extended periods of time, which can cause weakness and result in inconsistent evaluations. This inefficiency may cause the hiring process to take longer and result in the loss of qualified candidates.

1.2 Objective

The goal of this project is to create a system for ranking resumes based on a job description that will evaluate and rank resumes automatically. The system aims to,

- Reduce the time required for traditional resume screening
- Improve the accuracy of resume evaluations
- Build a scalable system which can manipulate large number of resumes

Process,

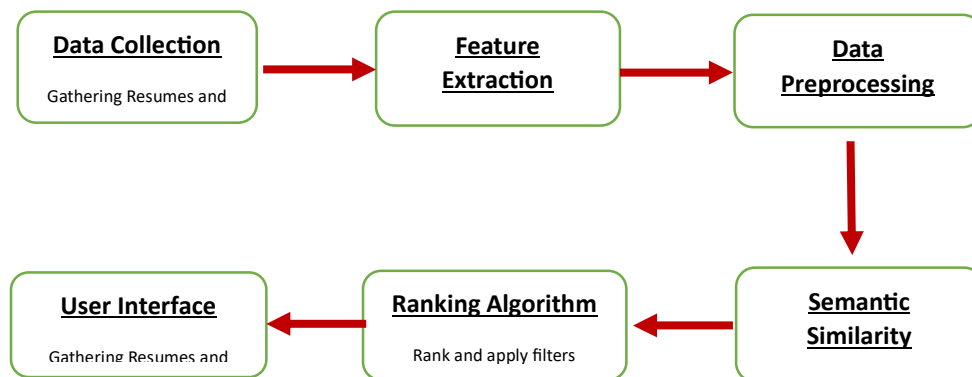


Figure 1- Project Process

2. Literature Review

Researchers have been actively studying how Natural Language Processing (NLP) can be used in recruiting. In the past, people have looked at different ways to get information from resumes and match it to job listings. Older methods typically focus on finding specific keywords and using

predefined rules. However, these approaches might miss the bigger picture and the actual meaning behind the words.

Example:

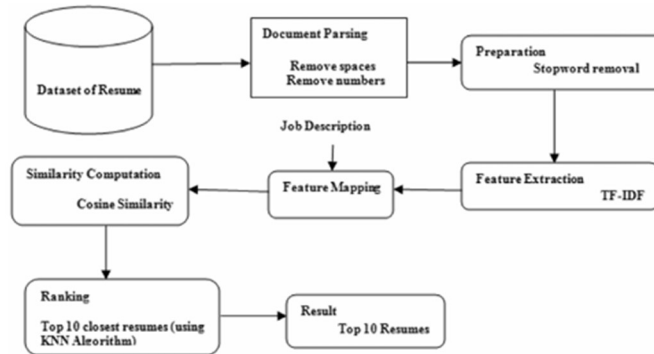


Figure 2- Old process

New developments in NLP, especially with large language models like BERT and GPT, are proving very successful in capturing semantic meaning of a texts. These models can create a kind of code that captures the essence of sentences, allowing for a more accurate comparison between resumes and job descriptions.

Example:

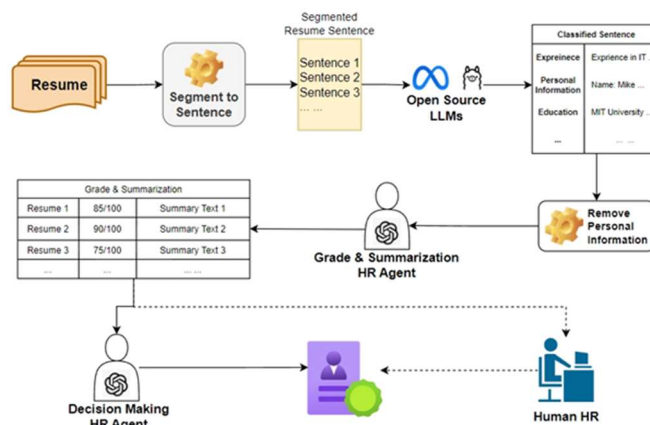


Figure 3- New Process

3. Methodology

The project is divided into several stages:

- **Data Collection:** Gather resumes and the job description.
- **Data Preprocessing:** Convert resumes to a uniform text format and clean the text data.
- **Feature Extraction:** Extract relevant features such as skills, experience, and education using NLP techniques.
- **Semantic Similarity Measurement:** Measure the similarity between the job description and each resume.
- **Ranking Algorithm:** Rank resumes based on similarity scores and additional qualifications.
- **User Interface:** Provide a user-friendly interface for inputting job descriptions and uploading resumes, and display the ranked list of resumes.

4. Tools and Technologies

- **Programming Languages:** Python
- **Libraries and Frameworks:** (Hugging Face), torch, TensorFlow
- **Data Processing:** PyPDF2, python-docx, re
- **UI Development:** Streamlet, HTML/CSS
- **Pretrained Models:** BERT, GPT
- **Others:** PyPDF2 for PDF processing, python-docx for DOCX processing

5. Expected Outcomes

- **Automated Resume Screening:** An automated system to rank resumes based on their relevance to the job description.
- **Improved Efficiency:** Significant reduction in the time required to screen and rank resumes.
- **Accuracy:** Enhanced matching accuracy through semantic understanding using LLMs.
- **User Interface:** An easy-to-use web interface for recruiters to upload job descriptions and resumes and receive ranked results.

6. Improvements

To further enhance the accuracy and comprehensiveness of the resume ranking system, we can integrate additional data sources such as GitHub and LinkedIn profiles. Many candidates include links to their GitHub repositories and LinkedIn profiles in their resumes. By extracting and

analyzing these links, we can gain deeper insights into the candidates' skills and professional experiences.

7. Conclusions

The suggested resume ranking system automates and improves the resume screening process by utilizing LLMs and cutting-edge NLP techniques. The system may comprehend the context and semantics of resumes and job descriptions by utilizing the power of models like BERT, which leads to more precise and effective candidate matching. This project shows how artificial intelligence and machine learning can revolutionize the recruitment industry, benefiting employers and candidates alike.

8. References

- 1) <https://www.sciencedirect.com/science/article/pii/S2666285X21001011>
- 2) <https://arxiv.org/abs/2401.08315>
- 3) <https://core.ac.uk/download/pdf/55305289.pdf>