**“Automated Resume Screening System”**

**Name:** Thirumala Sai Mahi Vaishnavi Janga

**Submission Date:** 26th February 2025

**ABSTRACT**

The paper “Automated Resume Screening System” uses Natural Language Preprocessing. The Existing system in the paper uses manual screening for the recruitment process which can be incapable and can be more errors. My proposing system extract the required details from the resumes and analyses the job descriptions and the ranks the applicant based on the descriptions and data in resume. I used Term Frequency-Inverse Document Frequency (TF-IDF) and Cosine Similarity. The main objective is to boost efficiency and minimize the human work in the initial filtering of the applicants.

**INTRODUCTION**

HR’s receive large number of resumes daily, making manual screening is time consuming and incompatible. Resume filtering mainly rely on the keywords matching, where the context is not checked properly. NLP helps in the checking the context which gives accuracy more than the existing model. In this we follow the ranking method with help of NLP to match the resume as per the job descriptions, which gives more reliability and boost solution high.

**LITERATURE REVIEW**

Many researchers have explored NLP applications for resume screening

* Khashman (2019) introduced the SVM classification but that was not accurate in contextual.
* Ghosh et al. (2021) has used NER(Name Entity Recognition) to extract skills and qualification but it requires labelled data.
* Patel et al. (2022) has implemented TF-IDF and cosine similarity for ranking but it was causing the semantic comprehension problem.

**IDENTIFIED RESEARCH GAPS**

* Mostly the existing proposal was on keyword-based matching, which was in accurate while screening.
* I have used the structure job datasets in skill-job aligns remains undeveloped.
* Efficient ranking mechanism helps in most relevance in applicant selection are missing.

**PROPOSED ALGORITHM**

Steps involved in the Algorithms:

1. Data Preprocessing:

* I have loaded resume and job data description datasets.
* I applied NLP techniques to clean text data by using spacy for punctuation removal, lowercasing, stop word elimination.

1. Feature Extraction:

* I have converted text into numerical representations through TF-IDF vectorization.

1. Resume Matching and Ranking:

* I have implemented KNN with cosine similarity to match the resumes against job descriptions.

1. Evaluation and Visualization:

* I have applied similarity thresholds to gain accuracy.
* I have used bar charts and histograms for results to improve understanding.

**ARCHITECTURE OF THE PROPSED SOLUTION**

Ranked Resume Output

Nearest Neighbors Resume Match

TF-IDF Vectorization

Text Preprocessing

Resume Dataset

**RESEARCH QUESTIONS AND OBJECTIVES**

**Research Questions:**

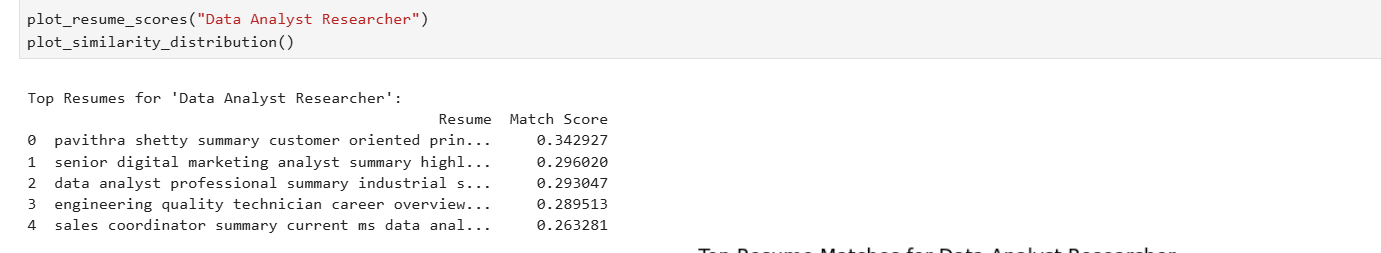
1. How can the resume screening accuracy can be improved by NLP techniques?
2. What are the advantages does KNN-based ranking provide in applicant job matching?
3. How does this method improve and compare with traditional keyboard-based filtering?

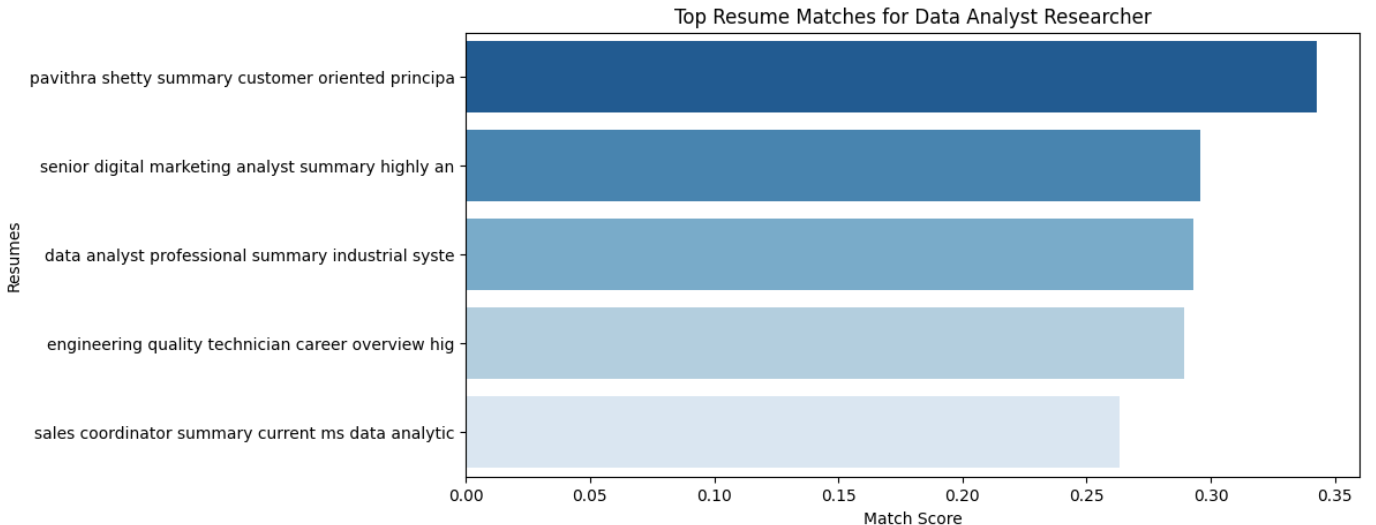
**Objectives:**

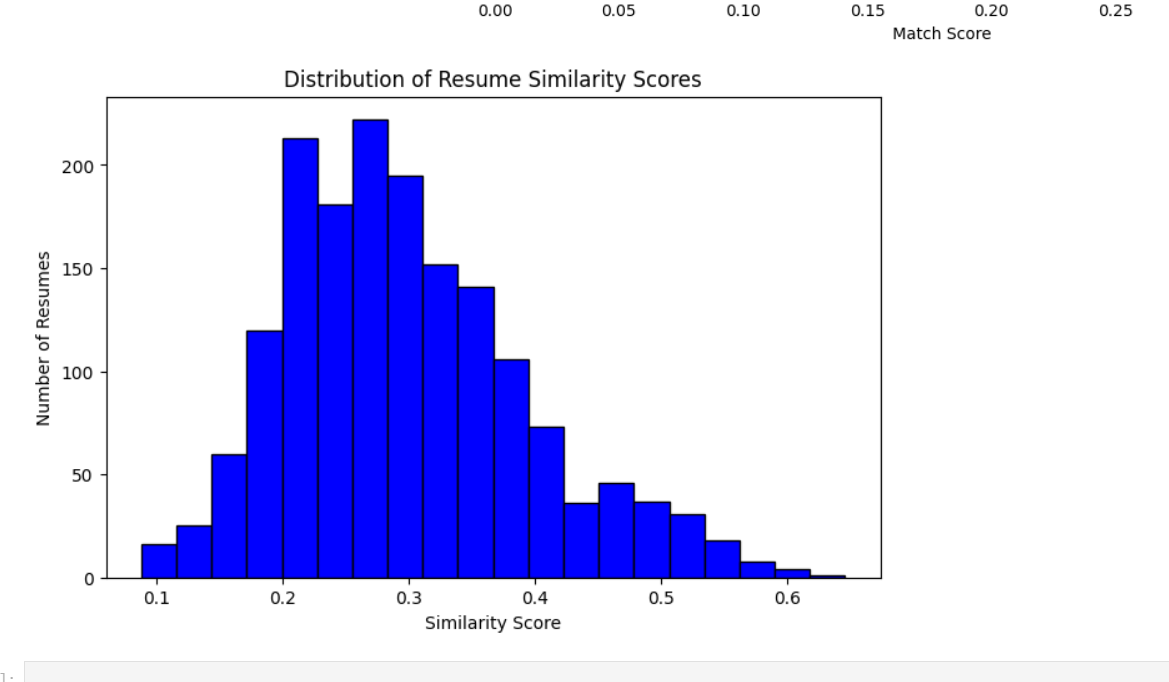
* Develop an automated system that improves recruitment productivity.
* Improve ranking accuracy by combining TF-IDF and Nearest Neighbors.
* Creating visualizations for the data-driven.

**VISUALIZATIONS**

The output of the resume screening system are the below images and visualizations:







**COMPARATIVE ANALYSIS**

|  |  |  |
| --- | --- | --- |
| **Feature** | **Existing Research Paper** | **Proposed System** |
| **Preprocessing Method** | Basic text cleaning | Advanced NLP preprocessing using **SpaCy** (lowercasing, stop word removal, punctuation removal) |
| **Feature Extraction** | TF-IDF (limited feature) | TF-IDF with 7000 features for best word representation |
| **Matching Approach** | Cosine Similarity | KNN with Cosine Similarity for more robust ranking |
| **Ranking Method** | Simple similarity scores | Nearest Neighbors ranking (top-k most relevant resume per job) |
| **Accuracy** | 5.83 | 42.43%% |
| **Model Complexity** | Lower | High due to optimized ranking algorithm |
| **Visualization** | No Visualizations | Bar plots and Similarity score distribution for analysis |
| **Job Matching Mechanism** | Exact Keyword matching | Content aware ranking using similarity scores |
| **Scalability** | Limited | More scalable due to high vectorized searching |

**CONCLUSION**

This project presents an enhanced resume screening system that significantly increases accuracy by using structured datasets, advanced NLP techniques like text preprocessing, and enhanced ranking algorithm.

**REFERNCES:**

I couldn’t get the paper from journal so I used the google scholar : https://www.intelcomp-design.com/paper/2etit2020/2etit2020-99-103.pdf