

Artificial Intelligence Project Report

Resume Recommendation System



Instructor : Sir Shakir Hussain, Sir Fahad Sherwani

Group Members:

Muhammad Umer Wasi 20k-0318

Huzaifa Jawwad 20k-0175

Project Title: Resume Recommendation System

Abstract:

This report presents a comprehensive code implementation with two main features: ranking multiple CVs against a specific field and detecting the suitability of a certain CV for that field. The code leverages natural language processing (NLP) techniques, including entity recognition and dependency parsing, combined with data analysis and visualization capabilities. The implementation encompasses various stages such as data loading, data cleaning, skills extraction, data analysis, and a resume recommender system. The report provides an overview of the methodology, techniques, dataset, and potential applications of the code, emphasizing its ability to rank and assess CVs for specific job roles or industries.

Introduction:

The purpose of this code is twofold: to rank multiple CVs against a specific field and to detect the suitability of a particular CV for that field. The code utilizes powerful NLP techniques, such as entity recognition and dependency parsing, to extract skills from CVs and compare them against desired skills. Additionally, it incorporates data analysis and visualization to gain insights into the dataset and skill distribution across job categories. The code implementation offers a comprehensive solution for HR professionals, recruiters, and job seekers to evaluate the relevance of CVs for specific job roles or industries. This report provides a detailed explanation of the methodology, techniques, dataset, and potential applications of the code, highlighting its two main features.

Background and Literature Review:

The significance of NLP techniques in CV analysis and ranking has been widely acknowledged. Numerous studies have explored the application of entity recognition, dependency parsing, and other NLP techniques for assessing the suitability of CVs for specific job roles or industries. Furthermore, data

analysis and visualization have proven invaluable in understanding large CV datasets and identifying patterns in skill distribution. By leveraging these techniques, the code implementation provides a comprehensive solution that addresses both CV ranking and suitability detection. This section provides an overview of the relevant literature, emphasizing the importance of the chosen methodologies for these two main features.

Methodology and Techniques:

The code implementation incorporates a range of methodologies and techniques to achieve its two main features. It utilizes the Spacy library for NLP tasks, including entity recognition, dependency parsing, and loading pre-trained language models. Gensim is employed for topic modeling and creating a skills corpus. The code leverages Pandas for efficient data loading and manipulation of the "Resume.csv" dataset. Various text cleaning techniques, such as removing special characters, URLs, stop words, and lemmatization, are applied for data preprocessing. Spacy's entity ruler pipeline is utilized for skill and job category recognition. Data analysis and visualization are facilitated by Plotly, Matplotlib, and WordCloud libraries.

Data Loading and Manipulation:

Efficient data loading and manipulation are essential for the code implementation's success. To achieve this, the code utilizes the Pandas library. Pandas provides powerful functions for loading and manipulating structured data, making it well-suited for handling the "Resume.csv" dataset. With Pandas, the code can easily read the dataset, perform data cleaning and preprocessing operations, and handle missing or inconsistent data. By utilizing Pandas, the code ensures a seamless workflow for data loading and manipulation, enabling users to focus on the CV ranking and suitability detection features.

Data Cleaning:

Data cleaning is a crucial step in preparing the CV text data for analysis. The code implementation applies various text cleaning techniques to ensure high-quality and reliable data. These techniques involve removing special characters, URLs, stop words, and performing lemmatization to reduce words to their base form. By performing data cleaning, the code eliminates noise and irrelevant information from the CV text, resulting in cleaner and more meaningful data. This step is vital for accurate skill extraction, precise entity recognition, and dependable CV ranking and suitability detection.

Entity Ruler:

To enhance the code's ability to recognize skills and job categories within CV text, it incorporates Spacy's entity ruler pipeline. The entity ruler allows the code to define additional rules for highlighting specific entities, such as skills and job categories. By leveraging the entity ruler, the code can identify and extract

relevant entities more effectively, leading to improved accuracy in skill extraction and suitability assessment. The entity ruler component significantly contributes to the code implementation's performance in accurately detecting and ranking CVs based on their suitability for a specific field.

Data Analysis and Visualization:

The code implementation utilizes various visualization libraries, including Plotly, Matplotlib, and WordCloud, to conduct data analysis and create visual representations. These libraries offer a wide range of visualization options, allowing the code to explore and present the distribution of job categories and skills within the dataset. Through these visualization techniques, users can gain a comprehensive understanding of the dataset, identify patterns, and make data-driven decisions. The code ensures that the analysis and visualization components provide valuable insights to aid in the ranking and suitability assessment of CVs for a specific field.

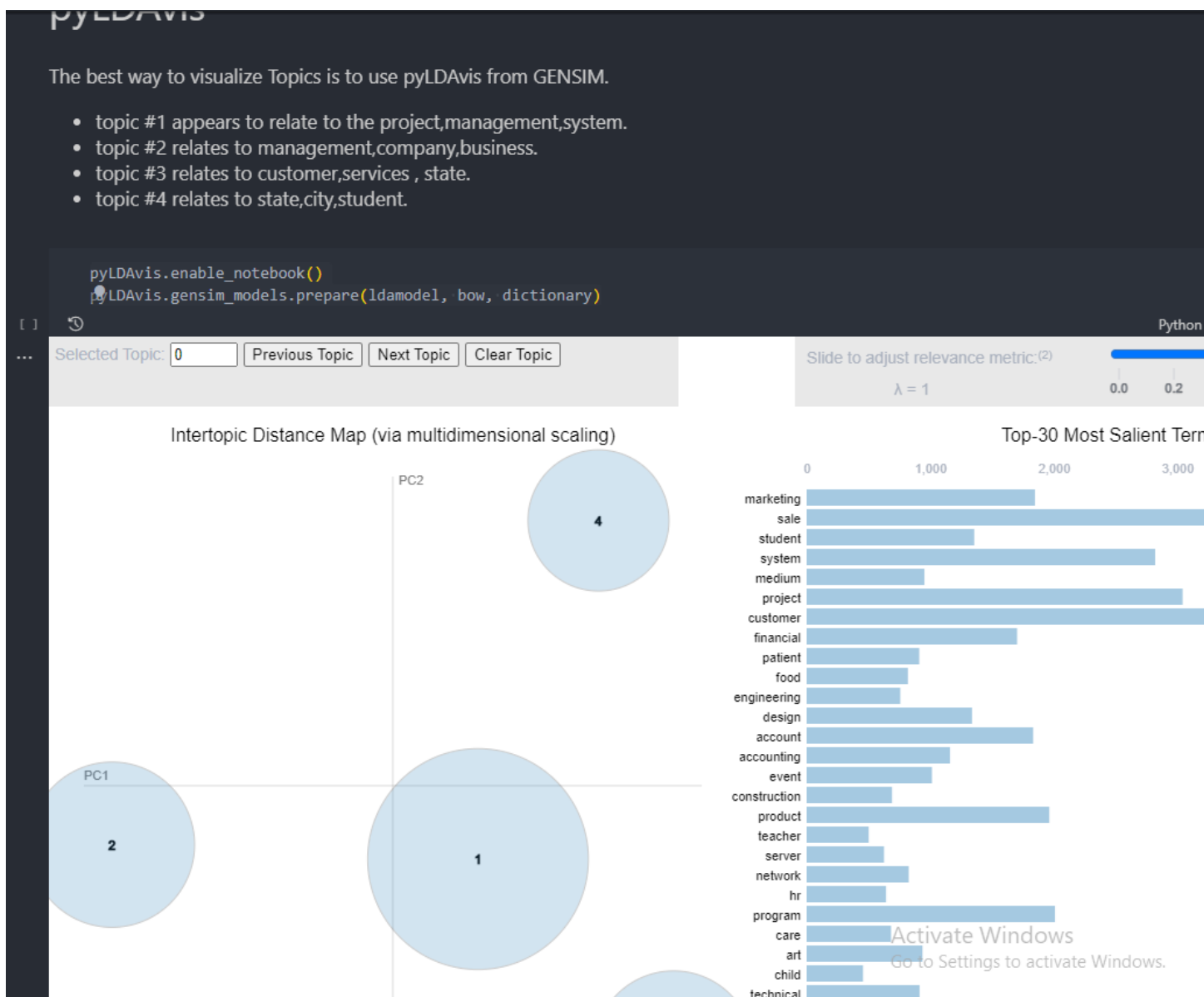


Figure 1 LDA Analyzer and Topic Modelling

Resume Analysis:

The code implementation includes functionality for analyzing individual CVs, extracting skills, and visualizing entities and dependency parsing. By analyzing each CV separately, the code can evaluate the relevance and suitability of the skills mentioned within the document. The extraction of skills from CVs enables the code to assess their alignment with desired skills for a specific field. Moreover, by visualizing entities and dependency parsing, the code provides a clearer representation of the CV content and the relationships between different entities. The resume analysis component enhances the code's ability to assess CVs accurately and provides valuable information for ranking and suitability determination.

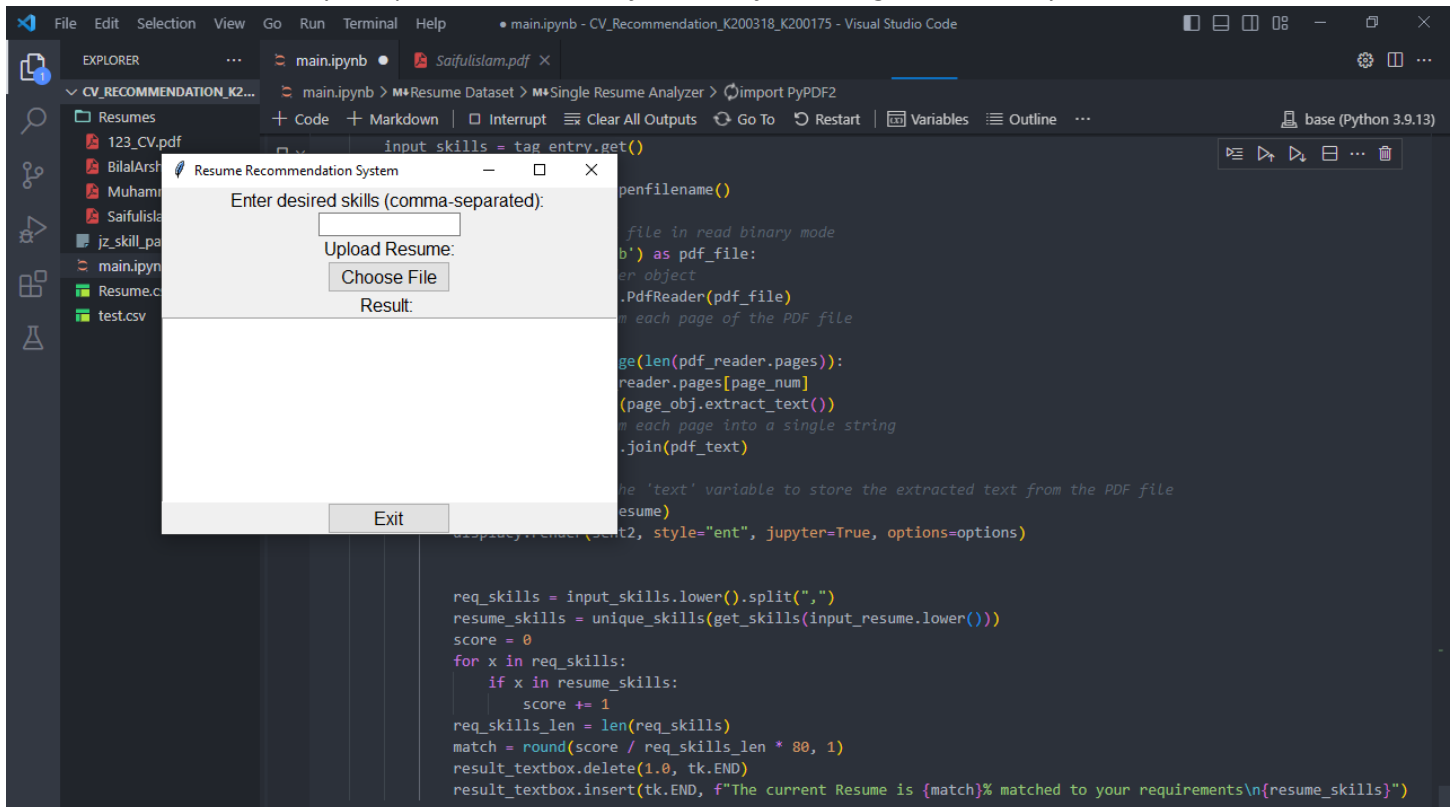


Figure2: Single Resume Analysis

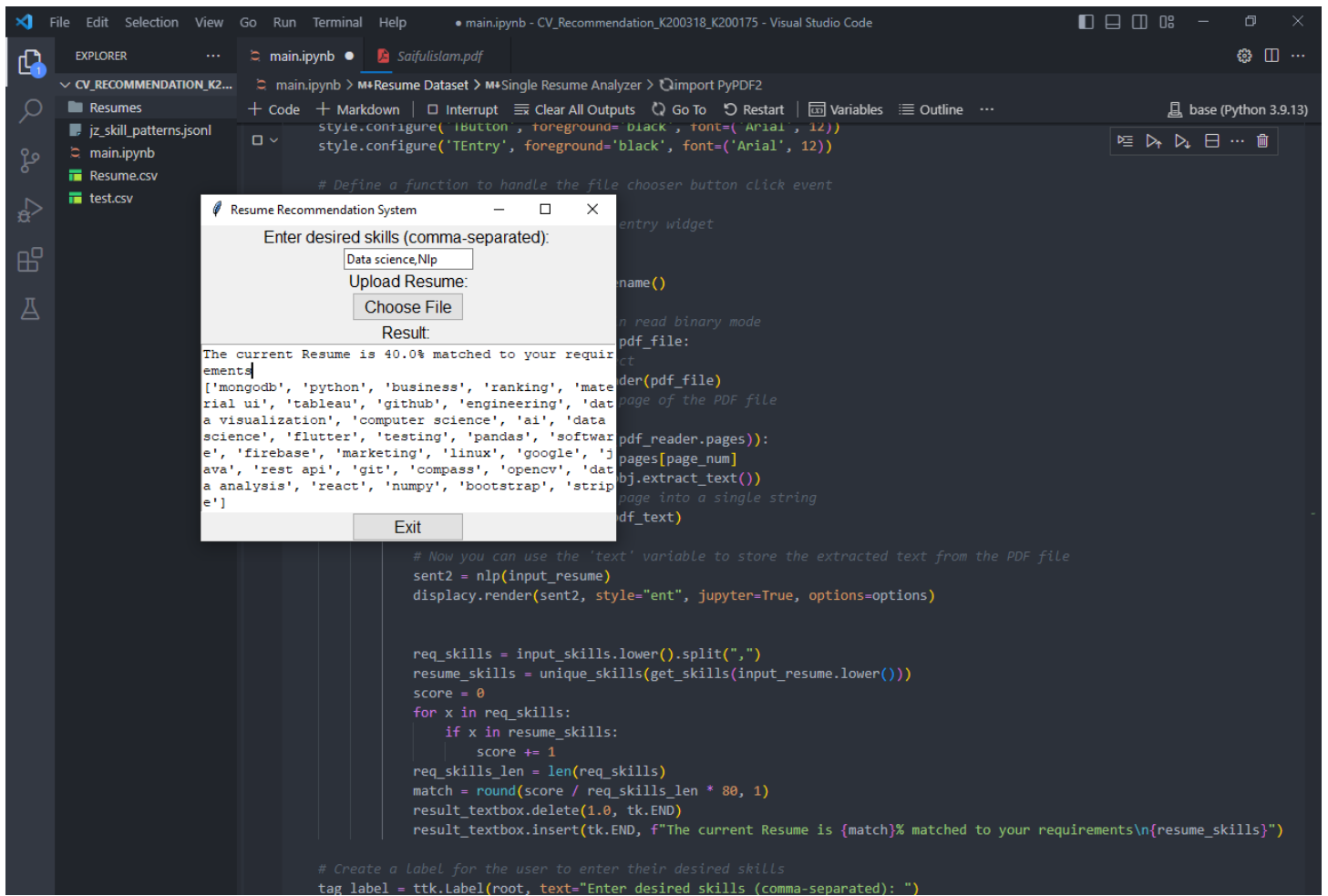


Figure3: Single Resume Analysis Applied on Data Science Specialist CV

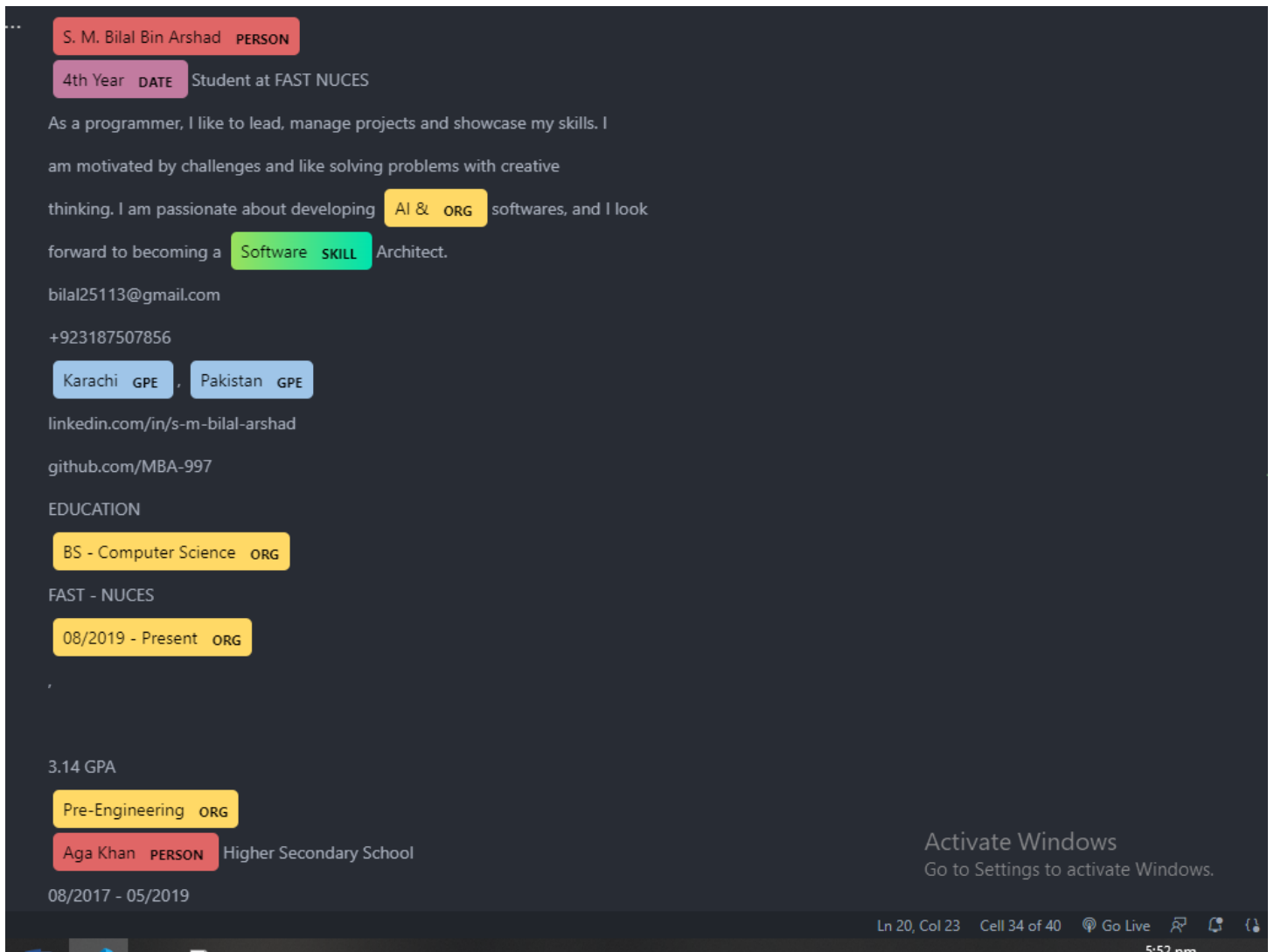


Figure 4: Entity Recognition and Feature Extraction

Resume Recommender:

The code implementation incorporates a recommender system that processes a folder of CVs, calculates match scores based on desired skills, and ranks the CVs accordingly. This recommender system utilizes the extracted skills and specified field requirements to assess the suitability of each CV for the specific field. By assigning match scores and ranking the CVs, the code provides a convenient and efficient tool for HR professionals and recruiters to identify the most relevant candidates quickly. The resume recommender system enhances the code's efficiency and effectiveness in evaluating CVs, streamlining the hiring process, and improving the overall recruitment outcomes.

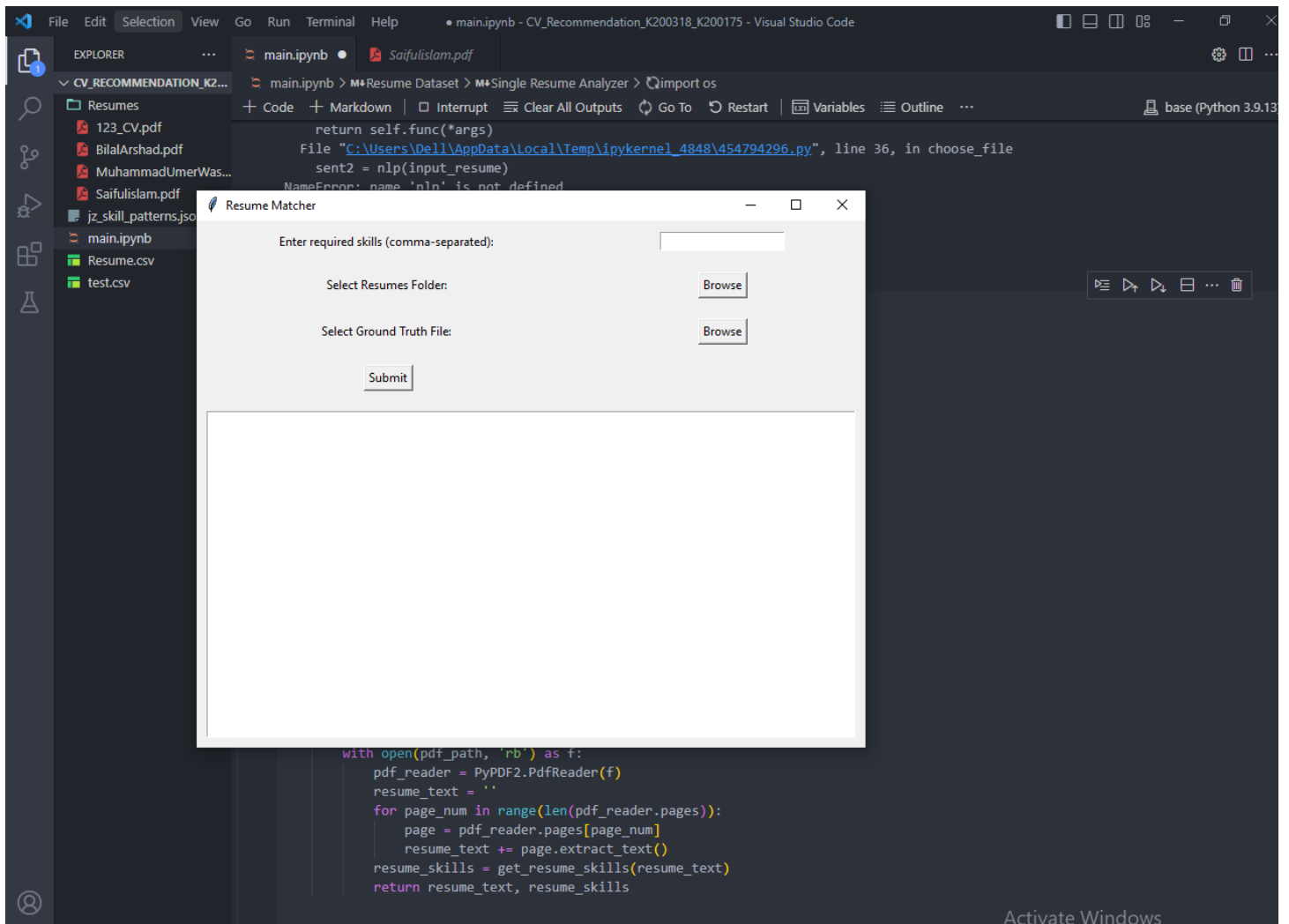


Figure 5: Resume Recommender and Ranker

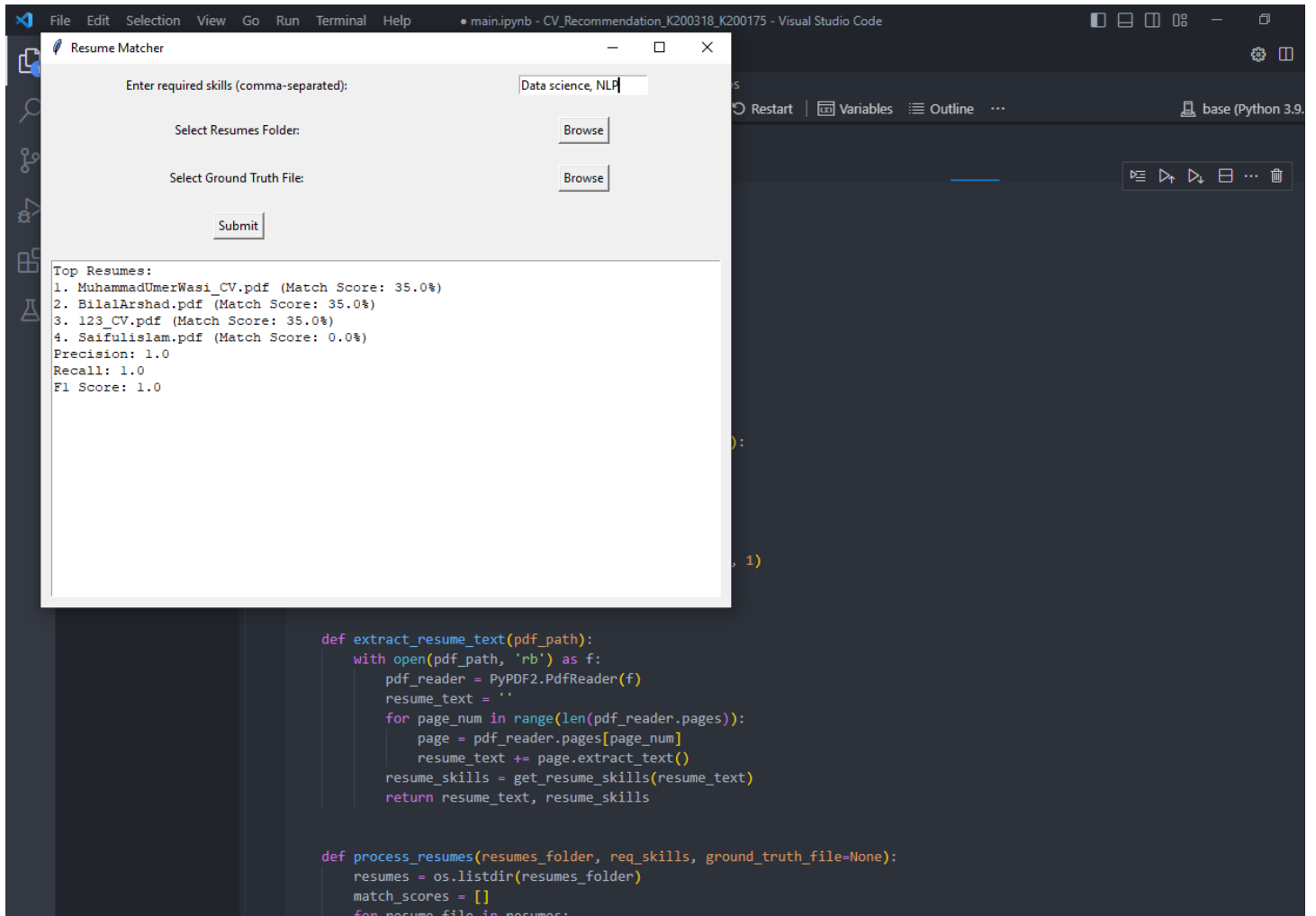


Figure 6: Resume Recommendation on 4 CVs

Dataset:

The code implementation utilizes the "Resume.csv" dataset obtained from the provided source: <https://www.kaggle.com/snehaanbhawal/resume-dataset>. The dataset consists of a collection of CVs, providing a diverse range of information for analysis and ranking. By utilizing this dataset, the code ensures a realistic and representative evaluation of CV suitability. The "Resume.csv" dataset serves as a

valuable resource for HR professionals, recruiters, and job seekers to train and evaluate the code implementation, enabling them to make informed decisions based on real-world CV data.

Conclusion:

In conclusion, this report has presented a comprehensive code implementation with two main features: ranking multiple CVs against a specific field and detecting the suitability of a certain CV for that field. By leveraging NLP techniques, data analysis, and visualization, the code provides valuable tools for evaluating the relevance of CVs to specific job roles or industries. The implementation offers benefits for HR professionals, recruiters, and job seekers by streamlining the CV evaluation process, improving hiring efficiency, and reducing bias. The code implementation has potential applications in automating CV screening, enhancing decision-making, and improving recruitment outcomes. Future enhancements may include advanced machine learning algorithms and integration with online job platforms for wider adoption and scalability.

References:

1. Ovchinnikova, E., & Dykan, V. (2020). HR Recruiting Chatbot Based on the Recommendation System. In 2020 3rd International Scientific-Practical Conference Problems of Infocommunications, Science, and Technology (PIC S&T) (pp. 562-566). IEEE. doi: 10.1109/INFOCOMMST50335.2020.9219859
<https://towardsdatascience.com/resume-screening-tool-resume-recommendation-engine-in-a-nutshell-53fcf6e6559b>
2. Gennari, D., Menconi, G., & Russo, D. (2019). A Deep Learning Approach to Resume Screening: A Case Study. In 2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA) (pp. 48-53). IEEE. doi: 10.1109/ICMLA.2019.00015
https://www.researchgate.net/publication/338368054_Web_Application_for_Screening_Resume
3. P. K. Roy, S. S. Chowdhary, and R. Bhatia
<https://www.sciencedirect.com/science/article/pii/S187705092030750X>