Smart Employment Navigator - A Job Recommendation System

Aboli Wankhade, Dharma Teja Kolluri, Saranya Gondeli

Department of Applied Data Science, San Jose State University

DATA 240 Sec 12: Data Mining/Analytic

Dr. Shayan Shams

December 7, 2023

Background

In the constantly evolving and competitive landscape of today's job market, individuals often face the formidable challenge of identifying job opportunities that align seamlessly with their unique skill sets and career aspirations. This challenge is further amplified by the overwhelming volume of job listings available online, each with its distinct set of requirements and expectations. Our study embarks on the development of an advanced Job Recommendation System, aimed at bridging this gap. This system will leverage a comprehensive dataset, meticulously compiled through web scraping from various job portals, encompassing a broad spectrum of candidate qualifications and job openings.

Motivation

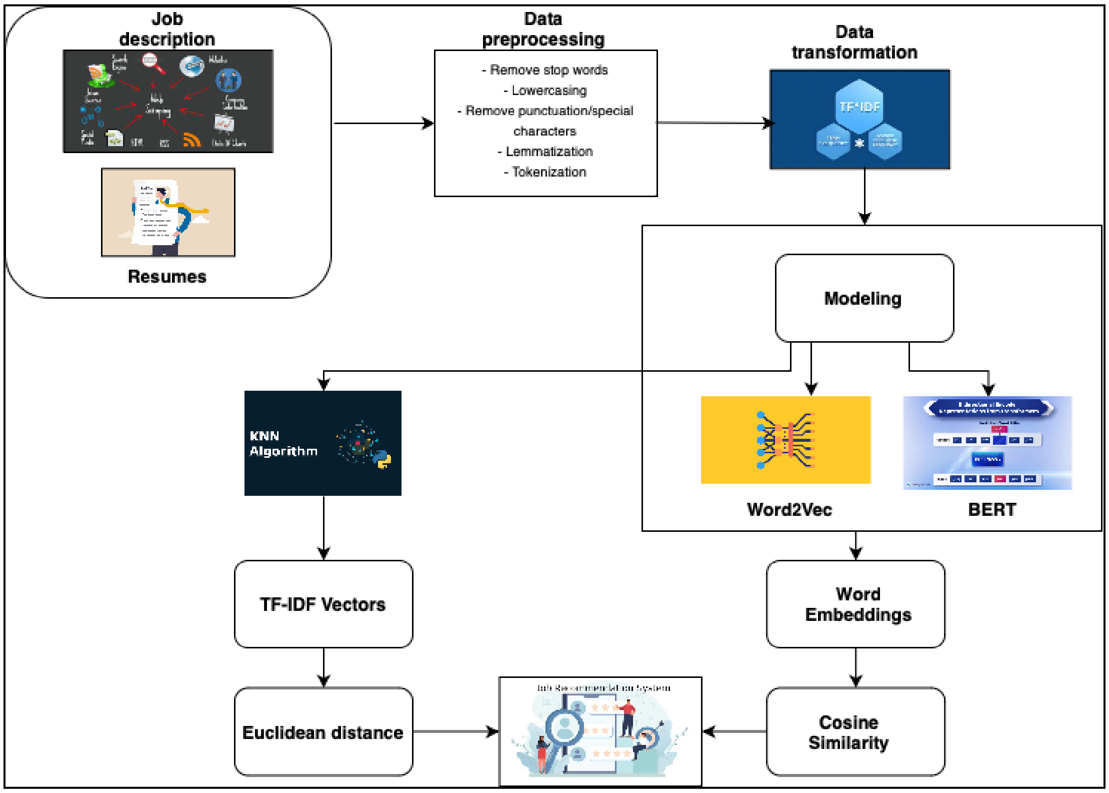
The primary motivation behind our project stems from the urgent need to address the multifaceted challenges confronted by job seekers in the modern digital era. One of the main challenges include navigating through the sheer abundance of online job postings to find relevant opportunities; and counteracting the inefficiencies inherent in current recruitment methodologies, which often lead to protracted hiring processes and missed career opportunities. Our initiative aspires to revolutionize the job-matching process by constructing an effective and intelligent recommendation system. Through meticulous analysis of both candidate profiles and job descriptions, our system is designed to deliver bespoke job suggestions. These suggestions go beyond the conventional criteria of matching skills and instead provide a holistic view by also considering factors such as work experience and technical skills. This comprehensive approach aims to facilitate a more streamlined and personalized job search experience, ultimately leading to more successful and fulfilling career placements.

In summary, our project not only seeks to alleviate the challenges faced by job seekers but also aims to transform the dynamics of the job market by introducing a more refined and user-centric approach to job matching.

Literature review

Chou and Yu (2020) outlined the deployment of a job recommendation system for job seekers and employers at a career fair, utilizing resumes as input and concentrating on 179 available job listings. They applied cosine similarity for similarity assessment and employed content-based, collaborative filtering, and hybrid recommendation techniques for data training. The system generated online reports for employers, presenting the top 20 candidates derived from the term frequency matrix and document vectors. Jain and Kakkar (2019) constructed a job recommendation system that has a textual approach for mapping the skills of the professionals with jobs in the industries, for this they employed the vector space model also known as Tf-Idf. They calculated the Tf, idf, and tf-idf for every skill and the weighted factor by assigning equal probabilities to each skill. They assigned the most suitable job for a candidate with the one that has the highest weighted factor. Chaudary et al. (2020) employed count vectorization and Tf-Idf vectorization for feature extraction and applied the Gradient Boosting algorithm using a one-vs-rest classification approach to extract skills, education, and experience from job descriptions. They achieved an ROC score of 0.80 for count-based and TF-IDF-based classifiers. Wang and Shi (2022) implemented an article recommendation system using Word2Vec and TfIdf methods by extracting the vectors of the articles and then recommending valid articles based on LSH methodology. Word2Vec was implemented beforehand to get the vectorized values and then Tfidf was used to get the most important vectors from the obtained vectors but hasn't performed any validation or evaluation of the results.

Methodology



Data Collection:

In this step, the Job Details dataset was extracted from the Indeed website using Python with libraries such as BeautifulSoup and Selenium to scrape job listings. The key steps in this data extraction process included configuring Selenium WebDriver with Chrome options to navigate and interact with web pages. This setup included handling user-agent strings to prevent detection as a bot. Then each job listing's URL was saved for further detailed scraping, providing a more granular level of data. Using the URLs collected, fetched detailed job descriptions, including job titles, company names, locations, salaries, job ratings, and other relevant details.

Then another job was implemented to retrieve the Skills and Work Experience from the resume which is in PDF format with which the job recommendations will be done

Data Preprocessing:

Data preprocessing involves cleaning and transforming the job description data from the job listing dataset. This cleaning process involves removing the Null value job descriptions, eliminating duplicate job descriptions, lowercasing the text, removing HTML tags, eliminating punctuations, special characters, and stop words. After performing these data cleaning tasks, cleaned data was utilized for tokenization and then lemmatizing techniques were applied to lemmatize the words.

Similarly, the data that was extracted from the resume should be cleaned by lowercasing the text, removing HTML tags, eliminating punctuations, special characters, and stop words. Similar to descriptions, the dataset, tokenization and lemmatization techniques were implemented. These two cleaned datasets will be used for data transformations in the next steps. Since the transformation process is different for each model that we are using, during that modeling of that data, transformations will be performed accordingly.

Modeling and Training:

After preprocessing the data, we employed several models, namely K-Nearest Neighbors (KNN), Word2Vec, and BERT, for our job recommendation system. The specifics of this phase included:

Model Selection: Based on the nature of our dataset, we selected three models for implementation: KNN for its simplicity and effectiveness in recommendation systems, Word2Vec for capturing semantic meanings in job descriptions, and BERT for its advanced capabilities in understanding contextual relationships in text.

KNN Model

After preprocessing the data, we employed TFIDF Vectorizer to transform the job descriptions into numerical vectors based on the TF-IDF scores of words. We used an n-gram approach to capture more context from the text, allowing for a nuanced understanding of job requirements and candidate capabilities.

Building and training the model involved initializing the NearestNeighbors model and we opted for Euclidean distance as the metric for several reasons. Firstly, Euclidean distance is intuitively appealing and easy to understand, as it simply calculates the straight-line distance between two points in the feature space. This characteristic makes it highly suitable for our application, where we need a straightforward yet effective way to measure the similarity between job descriptions and candidate profiles. Secondly, when dealing with high-dimensional data, Euclidean distance provides a reliable measure of closeness, ensuring that our recommendations are both relevant and accurate. The recommendation process involved matching candidate resumes to job descriptions using the trained KNN model. The model identified job descriptions closest to the candidate's resume in the feature space, emphasizing skills and experience compatibility.

Word2Vec Model

After preprocessing, TfidfVectorizer was used to filter out words with lesser significance, retaining those terms in job descriptions that carry more weight and relevance. Subsequently, we train two distinct Word2Vec models on our job descriptions dataset: one employing the Continuous Bag of Words (CBOW) method and the other using the Skip-gram (SG) approach. Each job description is then transformed into a vector by weighted averaging the vectors of the words using TF-IDF scores it comprises, thereby enabling a semantic and context-based representation of the job descriptions. Similar embeddings were obtained for resume data and employed cosine similarity to gauge the similarity between vector representations of resumes and job descriptions. For each resume, the system identifies the top job descriptions with the highest cosine similarity, suggesting the most relevant job opportunities to the candidate. This approach ensures that the recommendations are not only accurate but also highly personalized.

BERT Model

BERT (Bidirectional Encoder Representations from Transformers) was employed for its superior ability to capture context and semantics in text. After retaining the most significant terms using TF-IDF, a pre-trained BERT model was implemented on the Jobs description dataset to get the embeddings that serve as a high-dimensional representation of the text, encapsulating complex semantic relationships and similar to the Word2Vec model, job descriptions is then transformed into a vector by weighted averaging the vectors of the words. Similarly embeddings were extracted from Resume data using the above pretrained model. For matching resumes with job descriptions, we use cosine similarity to measure the degree of similarity between the BERT embeddings of the resumes and the job descriptions. This method quantitatively evaluates the alignment between a candidate's profile and the job requirements. The system then identifies the job descriptions that exhibit the highest cosine similarity with each resume, pinpointing the most fitting job opportunities for the candidates.

Experiments and Evaluation

Experiment Design

Initially, we conducted a test using a Data Engineer's resume as input for our model. The process began with extracting the Skills and Experience sections through a specific function. This extracted text was then pre processed and modified to suit the model's input format. Subsequently, we utilized the KNN algorithm along with TF-IDF to generate vectors. Based on these vectors, similarity metrics were applied to recommend the top 5 job matches. Additionally, we employed Word2Vec and BERT models, training them with job descriptions, to derive embeddings and recommend the top 5 job options. Upon analyzing the results, we observed that the majority of the recommended jobs pertained to Data Engineering, confirming the effectiveness of each model in proposing job matches aligned with the resume's content.

To theoretically assess the effectiveness of our models, we undertook a thorough evaluation of our job recommendation system, which incorporates the three detailed models previously described. This evaluation aimed to determine the precision with which each model recommends pertinent jobs when provided with different resumes.

Data Collection and Preparation: Our dataset comprised a diverse range of job descriptions, covering various job scopes. We meticulously preprocessed this data and categorized all the job descriptions into 4 categories namely Data Scientist, Data Analyst, Data Engineer and Software Engineer and created a new column in the job description dataset. This column will further be used in evaluating the results of the model.

Resume Processing and Recommendation Generation: Each model processed resumes from the different job categories and subsequently generated job recommendations based on the analysis of the resume content. While uploading the resume to the system for getting recommendations, it will first check the folder name from which the resume was uploaded and compare it to the category of the job description that was given in the recommendation. If all of the recommended categories are matching with the resume category we get 100% accuracy, if not we get varied accuracy according to the recommended jobs.

Evaluation and Results

The key metric for our evaluation was accuracy defined as the ratio of relevant job recommendations to the total number of recommendations made for each resume and MAP@K defined as a measure of precision across multiple queries at top K results respectively. A recommendation was deemed relevant if it matched the job scope of the resume.

Models	Accuracy	MAP@K
KNN	76%	69.9%
Word2Vec	82%	78.4%
BERT	80%	76.6%

KNN Model Results:

The KNN model showed varying accuracy across different job categories, reflecting its capability to align resumes with suitable job recommendations based on syntactic similarity.
An overall accuracy of 76% was achieved by averaging the individual accuracies for each resume. MAP@K score of 69.9% was achieved.

Word2Vec Model Results:

The Word2Vec model, focusing on semantic analysis, also displayed differential accuracy across job categories. The overall system accuracy of 82% was achieved through the average accuracy across all processed resumes. MAP@K score of 78.4% was achieved.

BERT Model Results:

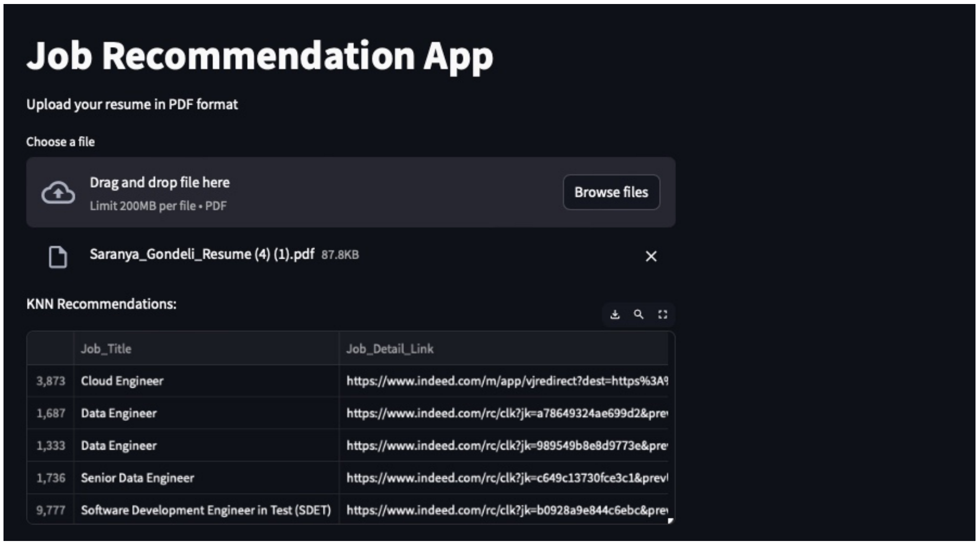
Leveraging its deep contextual understanding, the BERT model exhibited its proficiency in matching resumes to job descriptions, with varying degrees of accuracy. An overall accuracy of 80% was achieved for the BERT model and calculated as an average of individual resume accuracies. MAP@K score of 76.6% was achieved.

User Interface

The job recommendation system is made accessible through a Streamlit-based application, providing a user-friendly interface for candidates to upload their resumes in PDF format. Once a resume is uploaded and processed, the system dynamically displays the top job recommendations. These recommendations are comprehensive, including vital details such as job title, location, and direct links to the job listings.

Results

Below are the screenshots of the UI where we upload the resume in the system and recommendations of all the models are displayed.



Job Recommendation App

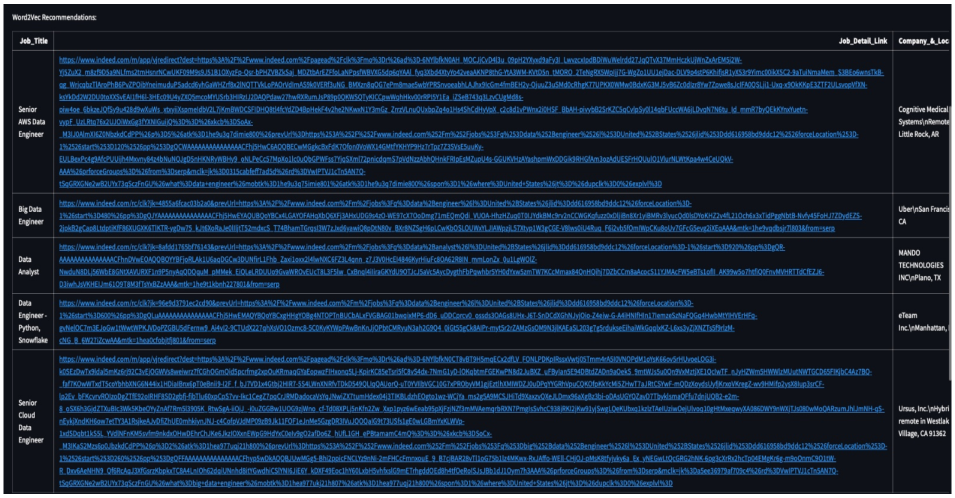
Upload your resume in PDF format

Choose a file

Saranya_Gondeli_Resume (4) (1).pdf 87.8KB

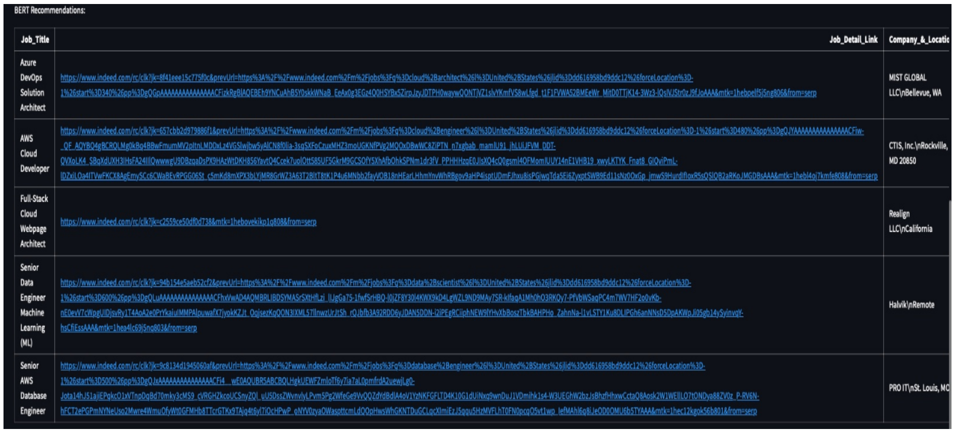
KNN Recommendations:

	Job_Title	Job_Detail_Link
3,873	Cloud Engineer	https://www.indeed.com/m/app/vjredirect?dest=https%3A
1,687	Data Engineer	https://www.indeed.com/rc/clk?jk=a78649324ae699d2&pre
1,333	Data Engineer	https://www.indeed.com/rc/clk?jk=989549b8e8d9773e&pre
1,736	Senior Data Engineer	https://www.indeed.com/rc/clk?jk=c649c13730fce3c1&prev
9,777	Software Development Engineer in Test (SDET)	https://www.indeed.com/rc/clk?jk=b0928a9e844c6ebc&pre



Word2Vec Recommendations:

Job_Title	Job_Detail_Link	Company_&_Loc
Senior AWS Data Engineer	https://www.indeed.com/m/app/vjredirect?dest=https%3A%2F%2Fwww.indeed.com%2Fpagead%2Fclk%3Fmo%3Dr%26ad%3D-6NYlbfkN0AH_MOCJjCvD4l3u_09pH2YXyxd9aFy3l_LwvzcxIpdBDiWuWelrdd27JqQTvX37MmHczkUjWnZxArEM5j2W-Yj5ZuX2_m8zf9D5a9NLfms2tmHsnrNCwUKF09M9s9J51B1OXyzFp-Qsr-bPHZVBZkSaj_MDZtbArEZFfoLaNPpsfWBVXG5dp6qYAAl_fyq3Xbd4XtyYo42veaAKNP8thG-YtA3WM-KVtD5n_tMORO_2TeNgRX5WpIjj7G-WgZo1UU1ejDac-DLV9p4stP6KhifjsR1yX53r9Yimc00ikX5C2-9aTuiNmaMem_S3BEo6wnsTkB-qg_WrjcqbzTlAroPhB6PvZPOibYneimuduP5adcd6yhGaWHZrf8x2lNQTTVkLoPAOrVdlmAS9k0VERf3uNG_BMXzn8qOG7ePm8mae5wbYPRSnvoeabhLAJhx9IcGm4fmBEH2y-QjuuZ3uSMd0cRhgK77UPKX0WMw0BdxKG3MJ5vB6Zc0dIzr8Yw7Zpwe8sJcIFA0QSLji1-Uxq-x9OkKKpE3ZTF2ULsvopVfXN-ksYkDd2W2DU0toXXSvEAI1fH6l-3HEc09U4yZXQ5mcoMYUSrb3HIRzIJ2DAQPdaw27hwRXRumJsP89p0QKWSQTyKICCpwWqhHkv00rRPI5Y1Ea_iZSeB743q3LzvCUgMd8s-piw4oe_6bkzeJQfSy9u428d9wXuWs_xtxyiiXspmeIdbV2L7jKmBWDCSFIDHQ8tl4fcYdZD4BpHekF4v2he2NKwxN1Y3mGz_ZrrzVLnuQUxbpZq4o1Hs4ShCdHyVpX_c2c8d1vPWnx2i0HSF_BbAH-pivybB2SrKZC5qCvlpSy0l14qbFUccWA6jLDvqN7N6tu_Id_mmR7byQEkKYnxYuetn-vypF_UzLRtq76x2UJOiWxGg3fYXNIGuijQ%3D%3D%26xkcb%3DSoAx-_M3IJ0AlmXI6Z0NbzkdCdPP%26p%3D5%26atk%3D1he9u3q7dimie800%26prevUrl%3Dhttps%253A%252F%252Fwww.indeed.com%252Fm%252Fjobs%253Fq%253Ddata%252Bengineer%2526l%253DUnited%252BStates%2526jlid%253Ddd616958bd9ddc12%2526forceLocation%253D-1%2526start%253D120%2526pp%253DgQCWAAAAAAAAAAAAAAACFhj5HwC6AQQBECwMGgkcBxFdK7Ofon0VoWX14GMtfYKHYP9Hz7rTpz7Z3SVsE5uuKy-EULBexPc4g9AfcPUUijh4Mxvny84z4bNuNQJgD5nHKNRyWBHy9_oNLPeCc57MpXo1lc0uQbGPWFss7YjqSXml72pnicdqmS7pVdNzzAbhQHnkFRIpEsMZupU4s-GGUKVHzAYashpmWxDDGik9RHGfAm3ozAdUESFrHQUulO1VlurNLWtKpa4w4CeUQkV-AAA%26prforceGroups%3D%26from%3Dserp&mclk=jk%3D0315cabfeff7ad5d%26rd%3DVwIPTVJ1cTn5AN7Q-tSqGRXGNe2wB2UYx73qSczFnGU%26what%3Ddata+engineer%26mobtk%3D1he9u3q75imie801%26atk%3D1he9u3q7dimie800%26spon%3D1%26where%3DUnited+States%26jt%3D%26dupclk%3D0%26explvl%3D	Cognitive Medical Systems\nRemote Little Rock, AR
Big Data Engineer	https://www.indeed.com/rc/clk?jk=4855a6fcac03b2a0&prevUrl=https%3A%2F%2Fwww.indeed.com%2Fm%2Fjobs%3Fq%3Ddata%2Bengineer%26l%3DUnited%2BStates%26jlid%3Ddd616958bd9ddc12%26forceLocation%3D-1%26start%3D480%26pp%3DgQJYAAAAAAAAAAAAAAACFhj5HwEYAQUBQoYBCx4LGAYOFAHqXbQ6XFj3AHxUDG9s4zO-WE97cX7OoDmg71mEQmQdi_VUOA-HhzHZuq0T0IJYdkBMc9rv2nCCWGKqfuzz0xDljiBn8Xr1yiBMRv3lyucQd0lsDYoKHZ2v4fL21Och6x3xTidPggNbtB-Nvfy45FoHJ7ZDydEZS-2jpkB2gCap8LtdptiKfF86XUGXK6TlKTR-ygDw75_kJt6XoRaJe0lIljtT52mdxcS_T74BhamTGrqsI3W7zJxd6yawiQ8pDtN80v_BXr8NZSgH6pLCwKbOSLOUWxYLJIAWpzjL57Xtyp1W3gCGE-V8lws0iU4Ruq_F6i2vb5fOmIWpCKu8oUv7GFcG5evg2iXEqAAA&mtk=1he9vqdbsjr7l803&from=serp	Uber\nSan Francis CA
Data Analyst	https://www.indeed.com/rc/clk?jk=8afdd1765bf76143&prevUrl=https%3A%2F%2Fwww.indeed.com%2Fm%2Fjobs%3Fq%3Ddata%2Banalyst%26l%3DUnited%2BStates%26jlid%3Ddd616958bd9ddc12%26forceLocation%3D-1%26start%3D920%26pp%3DgQR-AAAAAAAAAAAAAAACFhnDVwEOAQQBOYYBFjoRLAk1U6aqDGCw3DUNfirL1Fhb_Zaxi1oxx2l4lwNXC6FZ3L4qnn_z7J3V0HcEI4846KyrHiuFc8OA62R8IN_mmLonZx_0u1LgWOlZ-NwduN8DLj56WbE8GNtXAVURXF1n9P5nyAqQDQquM_pMMek_EiQLeLRDUUo9GvaWROvEUcT8L3F5lw_CxBnql4iliraGKYdU9OTJcJ5aVc5AycDygthFbPqwhbr5YH0dYxw5zmTW7KCcMmax84QnHQihj7DZbCCm8aAcocS11YJMAcFW5eBTs1ofIl_AK99w5o7htfiQ0FnvMVHRTTdCfEZJ6-D3iwhJsVKHEIJm61O9T8M3fTsYxBZzAAA&mtk=1he9t1kbnh227801&from=serp	MANDO TECHNOLOGIES INC\nPlano, TX
Data Engineer - Python, Snowflake	https://www.indeed.com/rc/clk?jk=96e9d3791ec2cd90&prevUrl=https%3A%2F%2Fwww.indeed.com%2Fm%2Fjobs%3Fq%3Ddata%2Bengineer%26l%3DUnited%2BStates%26jlid%3Ddd616958bd9ddc12%26forceLocation%3D-1%26start%3D600%26pp%3DgQLuAAAAAAAAAAAAAAACFhj5HwEMAQYBQoYBCxgHHgYOBg4NTOPTnBUCbALxFVGBAG01bwqixMP6-dD6_uDDCprcv0_ossds3OAGs8UHx-J6T-SnDCdXGhNJyjOio-Z4eiw-G-A4iHNIfHIn17lemzeSzNaFQGq4HwbMtYIHVErHFq-gvNelOC7m3EJoGw1tWwtWPKJVDoPZGBU5dFernw9_Ai4vI2-9CTUdX227qhXsVO1Ozmc8-5C0KyKYWpPAwBnKnJjOPbtCMRyuN3ah2G9Q4_0iGt55gCk8AIPr-myt5r2rZAMzGsOM9N3jlKAEaSL203g7g5rdukseEihaiWkGqqlxKZ-L6xs3yZjXNZTs5f9rlzM-cNG_B_6W27iZcwAA&mtk=1hea0cfobjtfj801&from=serp	eTeam Inc.\nManhattan,
Senior Cloud Data Engineer	https://www.indeed.com/m/app/vjredirect?dest=https%3A%2F%2Fwww.indeed.com%2Fpagead%2Fclk%3Fmo%3Dr%26ad%3D-6NYlbfkN0CT8vBT9H5mqECx2dfLV_FONLPDKpIRssxVwtj05Tmm4rA5I0VNOPdM1oYsK66ov5rHUvoeLOG3i-k0SEzDwTx9ldal5mKz6rj92C3vEjOGWVs8weiwrz7fCGhOGmQid5pcrfmg2xpOuKRmaqGYaEopwzFIHxonq5Lj-KpirKC85eTsri5fC8vS4dx-7NmG1yD-lOKqbtmFGEKwPN8d2JuBXZ_uFBylan5E94DBtdZADn9aOekS_9mtWJs5u0Qn9VxMztjXE1QcIwTF_nJyHZWm5HWWlzMUutNWTGCD65FIKjbC4Az7BQ-_faf7KOwWTxdT5coYbhbXNG6N44ix1HDiaIBnx6pT0eBnii9-I2F_f_bJ7VD1x4Gtbj2HIR7-5S4LWnXNRfvTDkD549QLIqQAUorQ-uT0YVIlbVGC10G7xPRObyVM1gjEztlhXMIWDZJ0uDPqYYGRhVpuCQKOfpKkYcMi5ZHwT7aJRtCSYwF-mQDzXoydsUyfjKrxoVKregZ-wv9HMifp2ysX8Iup3srCF-lp2Ev_bFKcvrvROIzoDgZTfE92oIRHF8SD2gbfj-fjbTlu60xpCpS7vv-Ikc1CegZ7pqCrJRMDadocaVsYqJNwiZX7tumHdex04j3TIKBLdzhEOgto1wz-WCjYa_ms2g5A9MCSJHiTd9XaxzvQXeJLDmx96aXgBz3bi-oDAsUGYQZavD7TbyklsmaQFfu7dnjUQB2-e2m-8_oSX6h3GjdZTXuBlc3Wk5KbeOYyZnAf7Rm5l3905K_RtwSgA-iiOjJ_-i0uZGGBw1UOG9zjWno_cf-Td08XPLj5nKfn2Zw_Xxp1pyz6wEeab95pXjFzjNZf3mMVAemqrbRXN7PmgIsSvhcC938jRKI2jKw91yjSwgLQeKUbxq1kzlzTAeIUziwOejUlvoq10gHtMxeqwyXA086DWY9nWXjTJs080wMoQARzumJhlJmNH-qS-nEykjXndKH6ow7etTY3A1RsjkeAJvDfiZhUE0mhkiynJNJ-c4CofpVJdMP09zB9Jk11FOF1eJnMe5GzgDR3IVuJQOQaIG9t73USfs1gE0wLGBmYxKLWVp-1xdSDqbt1kS5L_YVdlNFnKM5svfm9nkdxOHwDEhrChJKe6JkzIOXxnEWpG9HdYxC0eIv9gO2afDo6Z_hUfL1GH_ePBtamamC4mQ%3D%3D%26xkcb%3DSoCx-_M3IKaS2Mzs6p0JbzkdCdPP%26p%3D2%26atk%3D1hea977uqj21h800%26prevUrl%3Dhttps%253A%252F%252Fwww.indeed.com%252Fm%252Fjobs%253Fq%253Dbig%252Bdata%252Bengineer%2526l%253DUnited%252BStates%2526jlid%253Ddd616958bd9ddc12%2526forceLocation%253D-1%2526start%253D260%2526pp%253DgQFFAAAAAAAAAAAAAAACFhyp5wDkAQQBJUwMGg5-Bhi2ppicFNCLYz9nNi-2mFHCcFmnxouE_9_B7cjBAR28vTl1oG75b1lz4MKwx-RxJAffo-WEll-CHjOJ-pMsK8tfyjvky6a_Ex_yNEGwLtQcGRG2hNK-6pg3cXrRx2hcTp04EMgKr6g-m9oOnmC9O1tW-R_Dxv6AeNHN9_Qf6RcAqJ3XfGsrzKbpkxTC8A4LnIOh62dqiUNnhd8itYGwdhiCSlYNI6JiE6Y_kDXF49Eoc1hY60LxbH5vhfxslG9mETrhgddOEd8h4tfOeRoISJsJBb1dJ1Oym7h3AAA%26prforceGroups%3D%26from%3Dserp&mclk=jk%3Da5ee36979af709c4%26rd%3DVwIPTVJ1cTn5AN7Q-tSqGRXGNe2wB2UYx73qSczFnGU%26what%3Dbig+data+engineer%26mobtk%3D1hea977ukj21h807%26atk%3D1hea977uqj21h800%26spon%3D1%26where%3DUnited+States%26jt%3D%26dupclk%3D0%26explvl%3D	Ursus, Inc.\nHybri remote in Westlak Village, CA 91362



BERT Recommendations:

Job_Title	Job_Detail_Link	Company_&_Locatio
Azure DevOps Solution Architect	https://www.indeed.com/rc/clk?jk=8f41eee15c775f0c&prevUrl=https%3A%2F%2Fwww.indeed.com%2Fm%2Fjobs%3Fq%3Dcloud%2Barchitect%26l%3DUnited%2BStates%26jlid%3Ddd616958bd9ddc12%26forceLocation%3D-1%26start%3D340%26pp%3DgQGpAAAAAAAAAAAAAAACFizkRgBIAQEBEh9YNCuAhB5Y0skkWNaB_EeAx0g3EGz4Q0HSYBx5ZirpJzyJDTPH0waywQQNTjVZ1slvYKmfVS8wLfgd_t1F1FVWA52BMEeWr_MitD0TTjK14-3Wz3-lQsIVJStr0zJ9fJoAAA&mtk=1hebpelf5j5ng806&from=serp	MIST GLOBAL LLC\nBellevue, WA
AWS Cloud Developer	https://www.indeed.com/rc/clk?jk=657cbb2d979886f1&prevUrl=https%3A%2F%2Fwww.indeed.com%2Fm%2Fjobs%3Fq%3Dcloud%2Bengineer%26l%3DUnited%2BStates%26jlid%3Ddd616958bd9ddc12%26forceLocation%3D-1%26start%3D480%26pp%3DgQJYAAAAAAAAAAAAAAACFiw-_QF_AQYBQ4gBCRQLMg0kBq4BBwFmumMV2pltnLMDDxLz4VGSlwjbw5yAlCN8f0lia-3sqSXFoCzuxMHZ3moUGKNfPVg2MQOxDBwWC8ZiPTN_n7xgbab_mamlU91_jhLUiJFVM_DDT-QVXoLK4_SBqXdUXH3IHsFA24IlIQwwwgU9DBzqaDsPX9HAzWtDKH856YavtQ4Ccek7uolOtt585UF5GkrM9GCSOfY5XhAfbOhkSPNm1dr3fV_PPHHHzqE0JIsXQ4cQ0gsml4OFMomlUUY14nE1VHB19_xwyLKTYK_Fnat8_GlQviPmL-lDZxiLOa4ITVwFKCX8AgEmySCc6CWaBEvRPGG06St_c5mKd8mXPX3bLYjMR8GrWZ3A63T2BltT8tK1P4u6MNbb2fayVOB18nHEarLHhmYnvWhRBgov9aHP4isptUDmFJhxu8isPGjwqTda5Ei6ZyxptSWB9Ed11sNz0OxGp_jmwS9HurdIfloxR5sQSlQB2aRKoJMGDBsAAA&mtk=1hebl4oj7kmfe808&from=serp	CTIS, Inc.\nRockville, MD 20850
Full-Stack Cloud Webpage Architect	https://www.indeed.com/rc/clk?jk=c2559ce50df0d738&mtk=1hebovekikp1q808&from=serp	Realign LLC\nCalifornia
Senior Data Engineer Machine Learning (ML)	https://www.indeed.com/rc/clk?jk=94b154e5aeb52cf2&prevUrl=https%3A%2F%2Fwww.indeed.com%2Fm%2Fjobs%3Fq%3Ddata%2Bscientist%26l%3DUnited%2BStates%26jlid%3Ddd616958bd9ddc12%26forceLocation%3D-1%26start%3D600%26pp%3DgQLuAAAAAAAAAAAAAAACFhxVwAD4AQMBRLIBDSYMASrSXtHfLzi_lIJgGa75-1fwfSrHBQ-l0jZF8Y30l4KWX9kD4LgWZL9ND9MAy7SR-klfaqA1Mh0hO3RKQy7-PfVbWSaqPC4m7WV7HF2o0vKb-nE0evV7cWpgUIDjsvRy1T4AoA2e0PrYkaiuIMMPAlpuwafX7jyokKZJt_OqjsezKqQON3IXML57llnwzUrJtSh_rQJbfb3A92RDD6yJDAN5DDN-i2iPEgRCiiphNEW9lYHvXbBoszTbkBAHPHo_ZahnNa-l1vLSTY1Ku8DLIPGh6anNNsD5DpAKWpJi05gb14ySyinvqY-hsCfiEssAAA&mtk=1hea4lc69j5nq803&from=serp	Halvik\nRemote
Senior AWS Database Engineer	https://www.indeed.com/rc/clk?jk=9c8134d1945060af&prevUrl=https%3A%2F%2Fwww.indeed.com%2Fm%2Fjobs%3Fq%3Ddatabase%2Bengineer%26l%3DUnited%2BStates%26jlid%3Ddd616958bd9ddc12%26forceLocation%3D-1%26start%3D500%26pp%3DgQJxAAAAAAAAAAAAAAACFi4__wE0AQUBR5ABCBQLHgkUEWFZmloTf6y7ia7aL0pmfrdA2uewjLg0-Jota14hJ51ajiEPqkcO1xVTnpDqBd70mky3cMS9_cVRGHZkcoUCSnyZQl_uU5DssZWvnvlyLPvm5Pg2WfeGe9VvQQZdYdBdlA4oV1YzNKFGFLTD4K10G1dUiNxq9wnDuJ1VDmihk1s4-W3UEGhW2bzJsBhzfHhxwCctaQ8Aosk2W1WEllLO7tONDya88ZV0z_P-RV6N-hFCT2ePGPmNYNeUso2Mwre4WmuOfyWt0GFMHb8TTcrGTKx9TAjq4t6yl7iOcHPwP_oNYV0zyaOWaspttcmLdQOpHwsWhGKNTDuGCLqcXImiEzJ5qqu5HzMVFLhT0FN0pcqO5vt1wp_lefMAhl6q8iJeOD0OMU6b5TYAAA&mtk=1hec12kgok56b801&from=serp	PRO IT\nSt. Louis, MC

Discussion

Our experiments highlighted each model's unique strengths: KNN and Word2Vec in capturing syntactic elements and BERT in understanding deeper contextual meanings. The variance in accuracy across different job scopes suggests that the effectiveness of each model may depend on specific job characteristics and data quality. These results illuminate the potential of these models in automated job recommendation while also indicating areas for improvement, especially in handling diverse job descriptions and complex resume data. In summary, the evaluation of our job recommendation system using KNN, Word2Vec, and BERT models has provided valuable insights into their effectiveness. These findings will be instrumental in further refining our system, aiming to enhance accuracy and personalization in our job matchmaking process.

Future Improvement

This recommendation system works better when considering the exact experience of the candidate which is one of the most important and complex characteristics to implement in job recommendations. Integration of Education details for the recommendations could enhance in recommending more user specific jobs. Collaborative filtering can be implemented for the current work which will become a hybrid model to recommend the job using Single Value Decomposition (SVD) modeling techniques.

Bibliography

Jain, H., & Kakkar, M. (2019). Job Recommendation System based on Machine Learning and Data Mining Techniques using RESTful API and Android IDE. *Job Recommendation System Based on Machine Learning and Data Mining Techniques Using RESTful API and Android IDE*. https://doi.org/10.1109/confluence.2019.8776964

Chou, Y., & Yu, H. (2020). Based on the application of AI technology in resume analysis and job recommendation. *Based on the Application of AI Technology in Resume Analysis and Job Recommendation*. https://doi.org/10.1109/iccem47450.2020.9219491

Chaudary, A., Nasar, Z., Mubasher, M. M., & Qounain, S. W. U. (2020). Extraction of Useful Information from Crude Job Descriptions. *Extraction of Useful Information From Crude Job Descriptions*. https://doi.org/10.1109/inmic50486.2020.9318132

R. Wang and Y. Shi, "Research on application of article recommendation algorithm based on Word2Vec and Tfidf," 2022 IEEE International Conference on Electrical Engineering, Big Data and Algorithms (EEBDA), Changchun, China, 2022, pp. 454-457, doi: 10.1109/EEBDA53927.2022.9744824.