INTEGRATED JOB POST VERIFICATION AND PERSONALIZED JOB RECOMMENDATION SYSTEM

Team Members

Himanshi Raturi

Sumeet Suvarna

Shubhangi Dabral

Luddy School of Informatics, Computing, and Engineering
Indiana University Bloomington

CSCI-B 565 DATA MINING - Prof. Yuzhen Ye

Table of Contents

ABSTRACT:

In today's fast-changing job market, it's crucial to have accurate and precise job postings. This project introduces a system that combines two important aspects: checking if job listings are genuine and offering personalized job suggestions. Leveraging advanced data mining techniques and a large dataset with detailed job information, company profiles, required qualifications, and incentives, this system helps job seekers trust the job listings and find the right job for them. With the rise of online job listings, it's challenging to know if they're real. The Job Posting Verification part of our system would carefully check job details, company information, and qualifications to ensure job listings are reliable. It will utilize data mining techniques to extract valuable insights from the data.

In addition, our Personalized Job Recommendation System will use advanced data mining algorithms to suggest jobs that match a person's skills and preferences. It will look at job titles, locations, specific details, and job descriptions to find the best matches, again making use of data mining for enhanced accuracy.

In summary, this project aims to solve the problem of false job information and help job seekers find the right jobs. By combining Job Posting Verification and Personalized Job Recommendations with data mining techniques, we want to create a trustworthy job market where both job seekers and employers can connect easily and honestly. This will contribute to a strong and dependable job market, enriched by data-driven insights.

KEYWORDS:

Job Posting Verification, Personalized Job Recommendations, Data Quality, Job Market, Authentication, Recommendation Algorithms, Skills Matching, Data-Driven Insights, Features, data frame, data preprocessing, Data Mining, Machine Learning

INTRODUCTION:

Recognizing the need for a new and creative solution, this project introduces a special approach that combines the benefits of Job Posting Verification and a Personalized Job Recommendation System. The Job Posting Verification part acts as a protector for job seekers, carefully checking if job listings are real and trustworthy. It's essential for confirming important details like salary ranges, making sure company profiles are complete, and ensuring that listed requirements and benefits are relevant. This verification process is crucial for building trust and honesty in the job market, protecting job seekers from false information and misleading postings. Additionally, the Personalized Job Recommendation System focuses on honesty and integrity. It uses advanced data-driven methods to connect job seekers with opportunities that match their specific skills,

preferences, and geographic interests. By thoroughly analyzing job titles, locations, department details, and detailed job descriptions, this system is excellent at finding job listings that are not only relevant but also genuinely interesting to individual users.

LITERATURE REVIEW:

Data cleaning and processing is the key aspect in the data mining process. The valuable insight about it is provided in research paper [1] which addresses the issue such as missing values, noise, incompleteness, inconsistency, and outliers. Various techniques like cleaning, integration, transformation, and reduction are provided in the paper to enhance data efficiency and facilitate more efficient knowledge discovery.

So to get the better understanding about recommendation system more information is provided in paper [2] and [3] where it focus on methods such sampling, dimensionality reduction, and distance functions in data preprocessing overview of commonly used data mining methods in Recommender Systems, namely classification, clustering, and association rule discovery. The research employs data cleaning and various Machine Learning techniques, with the Random Forest Classifier demonstrating the highest prediction accuracy.
The paper [4] emphasizes the need for accurate job recommendations based on user profiles and preferences, and provides insights into how to tailor your job recommendation system to individual job seekers. The approach of mining rules and content-based matching can be integrated into your recommendation engine to enhance personalization.

The paper [5] on detecting fraudulent job advertisements is essential for the job post verification aspect of your integrated system. It discusses the use of machine learning and classification techniques, including Gradient Boosting, to identify fraudulent job postings. You can incorporate these techniques to filter out fake job posts and ensure the integrity of the job listings.

The paper [6] on predicting fake job postings can be utilized to enhance the job post verification process. By leveraging data mining techniques and deep neural networks, you can develop a more robust mechanism for identifying and flagging potentially fraudulent job advertisements. By combining the knowledge from these papers, we will create an integrated system that verifies job postings for authenticity, detects fraudulent ones, and offers personalized job recommendations to job seekers based on their profiles and preferences. This system will not only help job seekers find genuine job opportunities but also protect them from potentially harmful or misleading listings.

METHODS:

In our project, we employ advanced data mining techniques to tackle job market challenges. We begin with meticulous data collection, creating a comprehensive dataset of job postings and job seeker profiles. Rigorous preprocessing ensures data integrity, handling text cleaning, missing values, and feature extraction. Our system focuses on Job Posting Verification, scrutinizing company details, salary ranges, and benefits to eliminate fraudulent postings. Simultaneously, our Personalized Job Recommendation System analyzes skills and preferences against job attributes, providing tailored suggestions. Rigorous evaluation would enhance accuracy and user experience, fostering a trustworthy, data-driven job market connection
The methods that we will follow for Integrated Job Post Verification and Personalized Job Recommendation System project is as follows

1. **Data Collection:**
   We will gather a comprehensive dataset of Job Postings from Kaggle source, ensuring it contains detailed information such as job titles, location, department, salary_range, company profiles, benefits, etc.

2. **Data Preprocessing:**
   Data Cleaning: We will clean the dataset, handle missing values, outliers, and inconsistencies in job postings data

   Text Processing: Textual data will be preprocessed using techniques like tokenization, stemming, and removing stop words to prepare it for analysis.

   Feature Extraction: Extract relevant features from job postings such as description, company profile, requirements, salary, etc

3. **Job Posting Verification:**
   Data Verification: Utilizing advanced data mining techniques, we will scrutinize job postings to validate their authenticity. By employing methods such as web scraping, named entity recognition, and text pattern matching, we will verify critical details like company information, salary ranges, and benefits against authoritative sources and official databases.

   Data Quality Assessment: In our analysis aimed at distinguishing between genuine and fraudulent job postings, we will assess the data quality by evaluating the completeness, accuracy, and reliability of the information provided. Suspicious or incomplete postings will be flagged or removed after a thorough examination, ensuring the dataset's integrity and enhancing the accuracy of our classification models.

4. **Personalized Job Recommendation System:**

Search Query Analysis: We will Utilize NLP techniques for understanding user intent from search queries. And will extract key elements using tokenization, named entity recognition, and sentiment analysis.

Content-Based Filtering: Based on relevant features like skills and qualifications, we will match user queries with job postings and then apply similarity measures (e.g., cosine similarity) to recommend closely aligned jobs.

Collaborative Filtering: Recommend jobs based on similar users' search patterns and preferences. Utilize user-item interaction data and techniques like user-based or item-based collaborative filtering.

5. **Machine Learning Models:**
   We will train classification models to identify fraudulent job postings based on verified and trustworthy data points. Also, we will use classification models to understand user intent from search queries. And then we would classify queries into categories (e.g., job types, industries) for precise recommendations.

6. **Evaluation:**
   Job Posting Verification: Develop metrics to measure the accuracy of job posting verification. Evaluate false positive and false negative rates to ensure the effectiveness of the verification process.
   Recommendation System: Use evaluation metrics like precision, recall, and F1-score to assess the performance of the personalized job recommendation system.

7. **Documentation and Reporting:**
   We will document and prepare detailed reports all the methods, algorithms, data preprocessing steps and evaluation metrics

By following these methods, we will create a robust Integrated Job Post Verification and Personalized Job Recommendation System, ensuring the authenticity of job postings while providing tailored job recommendations to users based on their personalized search query.

REFERENCES:

Research Papers

- https://www.researchgate.net/publication/319990923_Review_of_Data_Preprocessing_Techniques_in_Data_Mining [1]
- https://www.researchgate.net/publication/225924875_Data_Mining_Methods_for_Recommender_Systems [2]
- https://www.sciencedirect.com/science/article/pii/S2666412721000489 [3]
- https://gdeepak.com/thesisme/Applying%20Data%20Mining%20For%20Job%20Recommendations.pdf [4]
- https://link.springer.com/article/10.1007/s00146-022-01469-0 [5]
- https://www.researchgate.net/publication/349884280_A_Comparative_Study_on_Fake_Job_Post_Prediction_Using_Different_Data_mining_Techniques [6]