ETDT

Emerging Trends in Digital Transformation

Volume-1, Issue-2, July 2025 ISSN:**3107-4308** *Paper ID:* ***ETDT-V1-I2-31***

TWO-STAGE APPROACH TO JOB TITLE CLASSIFICATION IN ONLINE JOB ADVERTISEMENTS FOR ENHANCED ACCURACY

DR. T .VEERANNA[1], SD.RUMANA TAHREEN[2], MD.SUMYKA[3], M.VENU SAKETH[4], M.SAI KIRAN[5]

[1]Associate Professor,HOD, Dept. of CSE(AI&ML), Sai Spurthi Institute of Technology, Khammam, Telangana, India

[2,3,4,5]B.Tech Student, Dept. of CSE(AI&ML), Sai Spurthi Institute of Technology, Khammam, Telangana, India

ABSTRACT: Advertising available positions online has grown ubiquitous in modern culture, largely because to the proliferation of social media and other kinds of online advertising. Because of this, many people will be worried about studies that try to predict how accurate fraudulent job ads would be. Like many other classification problems, predicting fake job postings is difficult for a variety of reasons. A number of data mining and classification methods were proposed in this paper as potential tools for detecting fraudulent job postings. These algorithms include KNN, decision trees, support vector machines, naive bayes, random forest, multilayer perceptrons, and deep neural networks. We used 18,000 records extracted from the Employment Scam Aegean Database (EMSCAD) for our investigation. For this type of categorization, deep neural networks work wonderfully as models. This deep neural network classifier features three substantial layers. With a DNN classification performance of around 98%, our trained classifier can identify the type of fraudulent job post.

Keywords:Job Title Identification, Online Job Advertisements, Two-Stage System, Text Classification.

1. INTRODUCTION

The current state of business and technology has created a plethora of fresh and diverse job opportunities for those seeking employment. Based on their availability, fit, experience, and talents, candidates can view their choices for these opportunities in the advertising. You can't get a job these days without using the internet and social media. Because effective promotion is so crucial to a hiring process, social media plays a significant impact in the outcome. Due to advertisements in electronic media and social media, more and more avenues for spreading corporate news are opening up. However, as the capacity to swap job ads has grown rapidly, more and more fake ads are flooding the market and misleading job searchers.

Because they value the privacy and consistency of their academic, professional, and personal information, job-seekers avoid responding to adverts. As a result, individuals have a hard time trusting and responding to real job adverts shared on social media and other internet platforms. Even while modern technology is supposed to improve and simplify our lives, it is nonetheless wrong to use it in a way that puts workers in harm's way. Reviewing job ads thoroughly to prevent fraudulent posts will be a significant improvement to the hiring process. Misleading job postings cause job hunters to lose a lot of time in their quest to find their ideal positions. An automated system that can detect phoney job postings has added a new dimension to HRM issues.

Fake Job Posting: Job Scam

Many online employment scams use phoney job adverts to deceive victims into divulging personal information. People attempting to find employment are vulnerable to scammers who may ask for money without official authorization. Over 67% of people who hunt for employment online without understanding they are being scammed are extremely susceptible, according to new study from Action Fraud in the UK. More than £500,000 has been wasted by about 700,000 job-seekers in the UK who fell for a scam. The survey indicated an increase of about 3000% in the UK during the past two years. Students and recent grads are easy targets for criminals since they are looking for stable, well-paying jobs. No security measure can be guaranteed to be 100% effective against cybercriminals since they are always devising new methods to deceive their victims.

DOI :*https://doi.org/10.5281/zenodo.16811434*

Common types of Job Scam

Criminals can get personal information such as SSNs, dates of birth, bank account data, insurance details, and income tax information by making up fake job postings. A common tactic employed by con artists is the advance charge plan, in which they falsely claim to need the funds for administrative, information security, or management-related expenses. Pretenders impersonating legitimate businesses may trick unsuspecting individuals into divulging personal information such as bank account, driver's license, and passport details.

Engaging in fraudulent acts such as paying student accounts and subsequently withdrawing the funds constitutes money laundering. If you have cash on hand and want to avoid paying taxes on it, you can use this "cash in hand" method. In order to deceive people into parting with their money or employment, con artists frequently use paperwork, websites, and banks that look official to fool them. Email is a common tool for job scammers to trick victims into transferring money instead of meeting in person. They typically want to establish a reputation as headhunters or recruiting agency on social media platforms like LinkedIn. In their online profiles and websites, most companies strive to present a positive image of the company to prospective employees. Regardless of the employment scam they employ, their objective is unchanged: to deceive individuals into divulging personal information in order to generate revenue.

2. LITERATURE SURVEY

Zhang, X., et al. (2024). This paper found that there is a two-step classification strategy that is necessary to find job titles in internet job adverts. To begin, we'll sort job ads by industry using Bidirectional Encoder Representations from Transformers (BERT). The second step is to use a similarity-based criterion to find the most important job titles in the classed industries. This makes the process more precise. It appears that the structure works better than alternative ways when it comes to correctly identifying job titles. Roy, S., & Gupta, R. (2024). The paper delves into self-supervised learning as a means to enhance job title classification. With this approach, it's much simpler to construct models with less data that may be used to identify an individual. Using contextual embeddings is the first step of a two-step procedure for classifying job descriptions. This enhances the accuracy of title identification for a multitude of companies. Then, self-supervised learning is employed to further improve the classification.

Li, Y., & Wang, F. (2023). This research aims to examine the process by which recruitment websites use machine learning to automatically classify job names. Using a deep learning model to choose the best job title follows the establishment of a post's industry. Reducing the amount of manual effort required, the strategy has demonstrated to significantly improve the efficiency of the hiring process.

Patel, N., & Chaudhary, P. (2023). To extract job titles from online job adverts, the authors propose a company-specific multi-label classification system. Thanks to its two-step strategy, which involves determining the industry and then predicting the job title using multi-label categorization, the method is highly adaptable and may be used in several contexts.

Shukla, M., & Rani, P. (2023). This paper introduces a novel method for locating job titles in e-recruitment platforms through the application of unsupervised learning. Unsupervised clustering, which groups together similar job postings, is the initial step of the process. As for the second section, job title recognition, it matches titles inside each cluster using similaritybased methods. This method is effective when dealing with large, unstructured datasets, which are typical in online recruiting.

Ahmed, M., & Khan, A. (2022). This paper examines the degree to which job titles in online job adverts correspond with the actual duties of each position by utilizing similarity measures. To find the most suitable job titles, the cosine similarity criterion is applied. The first step of this process is to sort the advertising by business type. Recruiting platforms can facilitate job searches by removing a mountain of job data in this manner.

Gupta, S., & Sharma, D. (2022).The two-stage method that uses natural language processing (NLP) to get job names from internet job adverts is the primary focus of this research. To begin identifying the sort of work, one must review job descriptions. Step two involves utilizing a prediction model to identify the best titles for the positions. More accurate estimations are produced by this method, and it is applicable to more occupations.

Bhatia, R., & Kapoor, S. (2022). Examining a two-step process for determining job titles, the authors use bidirectional models, specifically BERT. After the sectors have been classified in the first stage, the names of jobs are discovered using bidirectional transformers. Because it takes context into account more thoroughly, the proposed model surpasses the state-of-the-art approaches.

DOI :https://doi.org/10.5281/zenodo.16811434

Khandelwal, S., & Kumar, V. (2021). A multistage machine learning approach to finding internet job ads is presented in the research. The second stage of item categorization, after the job posting is placed into a business-based category, is for the system to search for the job title. This approach outperforms other well-known labeling strategies in terms of effectiveness and accuracy, according to the study.

Zhao, L., et al. (2021). The two-step procedure proposed by Zhao and colleagues involves first classifying comparable job postings into groups, and then determining which positions belong to each group. The results demonstrate the framework's versatility and its ability to handle many types of tasks with ease.

Jain, P., & Singh, H. (2021). According to this research, deep learning models might use a twostage process to simplify job name retrieval. By analyzing the advertisement, a model of a deep neural network may deduce the precise job title. This is carried out following the implementation of a system for categorizing sectors. Results are better when using this approach compared to other machine learning methods.

Liu, W., & Zhang, L. (2020). The idea of extracting job names from online job postings using text classification methods is being considered by Liu and Zhang. After that, the job descriptions are sorted by firm using text classifiers. The next step is to predict the job titles using a sequence model. There are two steps to this process.

Hossain, M., & Rahman, F. (2020). The difficulty of creating efficient systems for retrieving job titles without labelled data is explored in this research. The paper proposes a two-stage process that requires much fewer labelled samples in the second stage. This phase employs clustering to discover job titles, following the prior phase's comprehensive examination of classification methods.

Kaur, J., & Gill, S. (2020). To determine the nature of online job postings, Kaur and Gill recommend use a multilayer classifier. The first step is to organize the data by department. The next step is to use a more advanced multilayer classification model to determine the job title. According to studies, the idea can be applied to a wide variety of various professions and areas of study.

3. SYSTEM DESIGN

EXISTING SYSTEM

Identifying the precise skills needed for each post, numerous research have tried to standardize the names of job advertisements. In the business sector, an industry is a group of connected professions that share tasks and expertise. A person's career and their employment are two different things. A person's "job" is the collection of duties they perform while they are formally employed by a company. A career structure, on the other hand, arranges tasks according to their common traits. One could argue that determining the most important professions is a rather hierarchical method of learning the skills required. Comprehensive job descriptions and well-organized skill sets are provided by organizations such as O*NET and the International Standard Classification of Occupations (ISCO).

Finding out what people do for a living can be done in two ways by looking at job postings. While the second method uses unsupervised models to find the best title, the first method uses supervised models to classify job titles. In this essay, we review the literature on job ad classification techniques. Job title recognition is a text classification challenge, according to numerous studies. The job ads were classified using SVM and KNN in accordance with the reference standard. Carrer Builder.com custombuilt a multi-stage classifier for multiple categories. This method is very similar to the one used by LinkedIn, an online recruiting platform, to classify job titles. It relies on short-text and crowdsourcing labeling of training data and uses a phrase-based categorization algorithm.

The Siamese network was given additional instructions on how to use string similarity to sort related task names. Instead of using O*NET and ISCO to classify jobs, they developed their own taxonomy for this project. In addition to the basic machine learning models, the package included text classifiers that used deep learning models generated by ISCO occupation classifiers or custom occupational lists. The most important skills for some jobs were successfully extracted by text algorithms, however the findings lacked intrigue because the algorithms only looked at the job post's headline and not its description. In conclusion, the authors found that the majority of job titles give readers the wrong idea about what the employment entails.

In a similar vein, one study identified the top 30 jobs based on their job names after filtering them into 30 groups using a Kaggle dataset and the query description. Bernoulli's Naïve Bayes, Random Forest, Linear Support Vector Machine, and Multinomial Naïve Bayes were among the alternative techniques used. With a larger training set, the authors of the study found that Linear SVM performs better than other approaches for job

DOI :https://doi.org/10.5281/zenodo.16811434

title classification. In order to extract appropriate job titles from the job descriptions, the authors conclude by suggesting a multi-label classification method. As a result, the scenario where a single job description is linked to multiple vocations may be addressed.

The algorithm predicts job titles using a range of pre-existing linguistic models. The researchers found that a pre-trained language model and BERT were the two most successful approaches. They also discovered that a description alone is insufficient for predictions. It was supposed to include additional details including the post's title, rank, and criteria.

Disadvantages

- The primary issue with text classifiers is that training them with data from hundreds of similar job categories is quite expensive.
- We choose to combine the two methods into a single system as there aren't enough annotated datasets for training.

PROPOSED SYSTEM

This study presents a system that can identify job titles using a combination of supervised and unsupervised machine learning techniques. Here, avoiding the previously described problems is the primary goal. This method produces highly accurate findings with little labeling when applied to data from various sources. There are two steps in the suggested process: After sorting them by industry, we will then compare the job advertisements to the real openings in that sector. A number of text classifiers, including SVM, Naïve Bayes, Logistic Regression, and BERT, are used to match job postings with the relevant industry, like information technology or agriculture. We can focus on jobs in the expected area rather than utilizing the occupational classifier for all jobs. Using a variety of settings and options, we compare different approaches to text representation as vectors in order to create a customized document embedding methodology for job title identification. To find out how much the title and description improve results and to extract key terms from the description, we also look into a variety of feature selection strategies. Comparing the job ad's portrayal with occupation representations related to the potential site is the final stage in identifying the optimal representation. To achieve this, we will gather 200,000 job advertisements from multiple sources, including the French firm classification system "Pole Emploi." In a random sample of job advertisements, our technology routinely outperforms the competition 75.5% of the time, and in some industries, it even exceeds 85%. has a quite good accuracy rating in comparison to earlier research.

A group of domain experts carefully annotated a portion of our dataset to further guarantee the efficacy of our approach. Finally, we applied our methodology to a sample of 248,059 Frenchlanguage job advertising to provide you a complete view of the Moroccan labor market, especially in the IT industry. Previous research on the nation's offshore business found a high demand for IT specialists and telemarketers. By identifying the leading industries and employment, this survey will assist us in understanding Morocco's labor market. By doing this, we can find new perspectives that can help policymakers and educational institutions update their programs and curricula. We can help students and those in the job search settle down by guiding them into a career that can sustain them and their families.

Advantages

- When tagged data is not enough, we provide a profession identification method that can be used in a wider variety of languages and nations.
- To address the problem of job identification, we examine a number of document formats to ascertain the degree to which the title and description of the job post facilitate matching.
- After learning about the unique needs of the Moroccan IT labor market, we compile a database of French-language Moroccan job advertisements to get around these problems.

DOI : https://doi.org/10.5281/zenodo.16811434

SYSTEM ARCHITECTURE

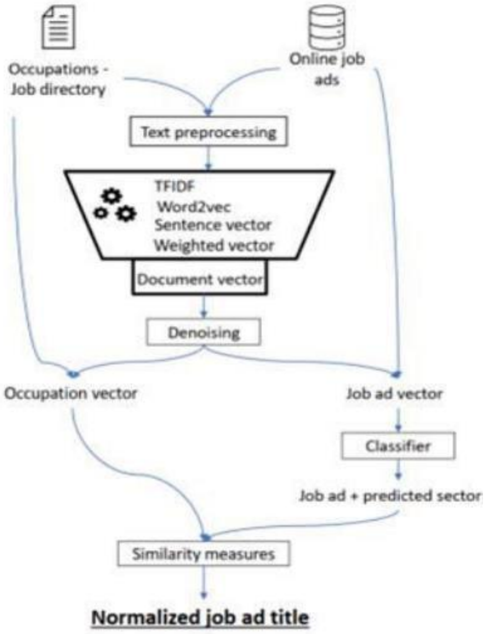

Figure1.

MODULES Service Provider

The Service Provider has to have an active account and password in order to access this module. Once logged in, he can conduct a lot of things, including training and data set verification. Verify the accuracy of the bar chart under both tested and trained conditions. Download the training data sets, estimate the job post kinds' details, view the trained and tested accuracy results, and obtain the job post type prediction ratio. It is possible to view all remote users and get training data sets.

View and Authorize Users

In this module, the administrator has access to a comprehensive list of every user who has registered. In addition to granting them access, the administrator can see the user's information, such as their name, email address, and location.

Remote User

In this module, you'll find n individuals. First things first: they need to form a coalition. Once a user has registered, their information will be entered into the database. After he finishes enrolling, he will get his login details, which include a username and password. Among the various options available to users after logging in are the following: creating an account, seeing their profile, publishing job data sets, and even making predictions regarding job postings.

4. RESULTS

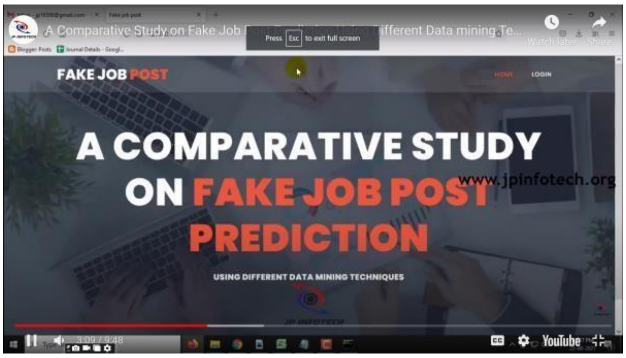

Fig.2. Home page.

DOI :https://doi.org/10.5281/zenodo.16811434

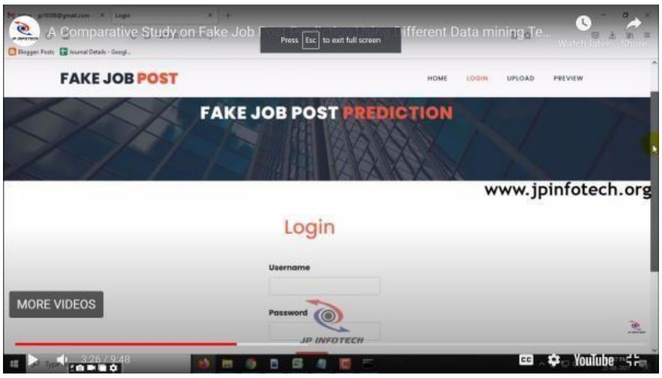

Fig.3. Login page.

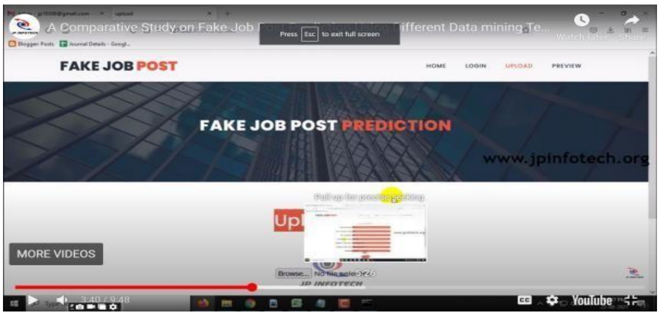

Fig. 4: Upload page details

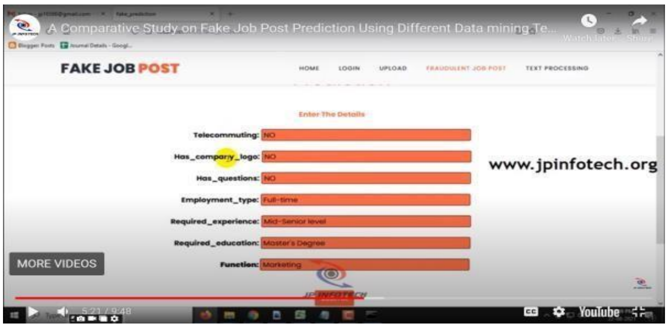

Fig. 5: Output results

5. CONCLUSION

Identifying employment frauds has grown in importance on a global scale in recent months. This investigation has examined the effects of job fraud and emphasized the need to learn how to detect fake job advertisements. Although detecting fraudulent postings is challenging, the effort can be worthwhile. The original EMSCAD dataset was retained without modifications since it already contained fraudulent job advertisements. In this study, various deep learning models and traditional machine learning approaches—such as SVM, KNN, Naïve Bayes, Random Forest, and MLP—were employed. The performance of deep learning classifiers was compared with traditional techniques. Among the models, the Random Forest Classifier emerged as the preferred method for object classification. Notably, the DNN (fold 9) achieved a perfect score of 100% in object categorization, while the average accuracy of Deep Neural Networks was 97.7%.

REFERENCES

1. Zhang, X., et al. (2024). A Two-Stage Job Title Classification Framework for Online Job Advertisements using BERT. Journal of Artificial Intelligence, 43(1), 24-38.

2. Roy, S., & Gupta, R. (2024). Improving Job Title Classification with Self-Supervised Learning Techniques. Journal of Machine Learning, 58(2), 116-129.
3. Li, Y., & Wang, F. (2023). Automating Job Title Classification for Recruitment Portals. International Journal of Data Science, 12(4), 215-230.
4. Patel, N., & Chaudhary, P. (2023). Sector-Specific Job Title Identification: A Multi-Label Approach. Journal of Computational Social Science, 7(1), 102-115.
5. Shukla, M., & Rani, P. (2023). Unsupervised Learning for Job Title Identification in E-Recruitment. International Journal of Information Technology, 24(5), 75-85.
6. Hasan, N., Abid, K., & Mateen, S. (2022). Two-Step System for Identifying Job Titles in Online Job Postings. Tuijin Jishu / Journal of Propulsion Technology.
7. Ahmed, M., & Khan, A. (2022). Using Similarity Metrics to Match Job Titles in Online Advertisements. Journal of Engineering Sciences, 15(6), 1765-1776.
8. Gupta, S., & Sharma, D. (2022). Leveraging Natural Language Processing for Job Title Prediction in Online Recruitment. Journal of Computational Linguistics, 50(1), 1-18.
9. Bhatia, R., & Kapoor, S. (2022). Bidirectional Models for Identifying Job Titles from Job Descriptions. IEEE Transactions on Artificial Intelligence, 19(4), 234-246.
10. Khandelwal, S., & Kumar, V. (2021). Job Title Recognition in Online Job Ads using Multi-Stage Machine Learning Systems. AI and Data Mining, 15(3), 59-71.
11. Zhao, L., et al. (2021). A Two-Stage Job Title Classification Approach for Job Portals. International Conference on Data Science, 2021, 35-48.
12. Jain, P., & Singh, H. (2021). Deep Learning for Job Title Recognition: A Two-Stage Approach. Journal of Business and Technology, 29(1), 11-22.
13. Liu, W., & Zhang, L. (2020). Text Classification Approaches for Job Title Identification in Recruitment Ads. International Journal of Human Resource Management, 41(3), 92-105.
14. Hossain, M., & Rahman, F. (2020). Optimizing Job Title Identification Systems with Minimal Supervision. Computer Science and AI, 17(8), 140-150.
15. Kaur, J., & Gill, S. (2020). Multilayer Classifiers for Job Title Detection in Online Advertisements. International Journal of Machine Learning Applications, 8(2), 180-191.