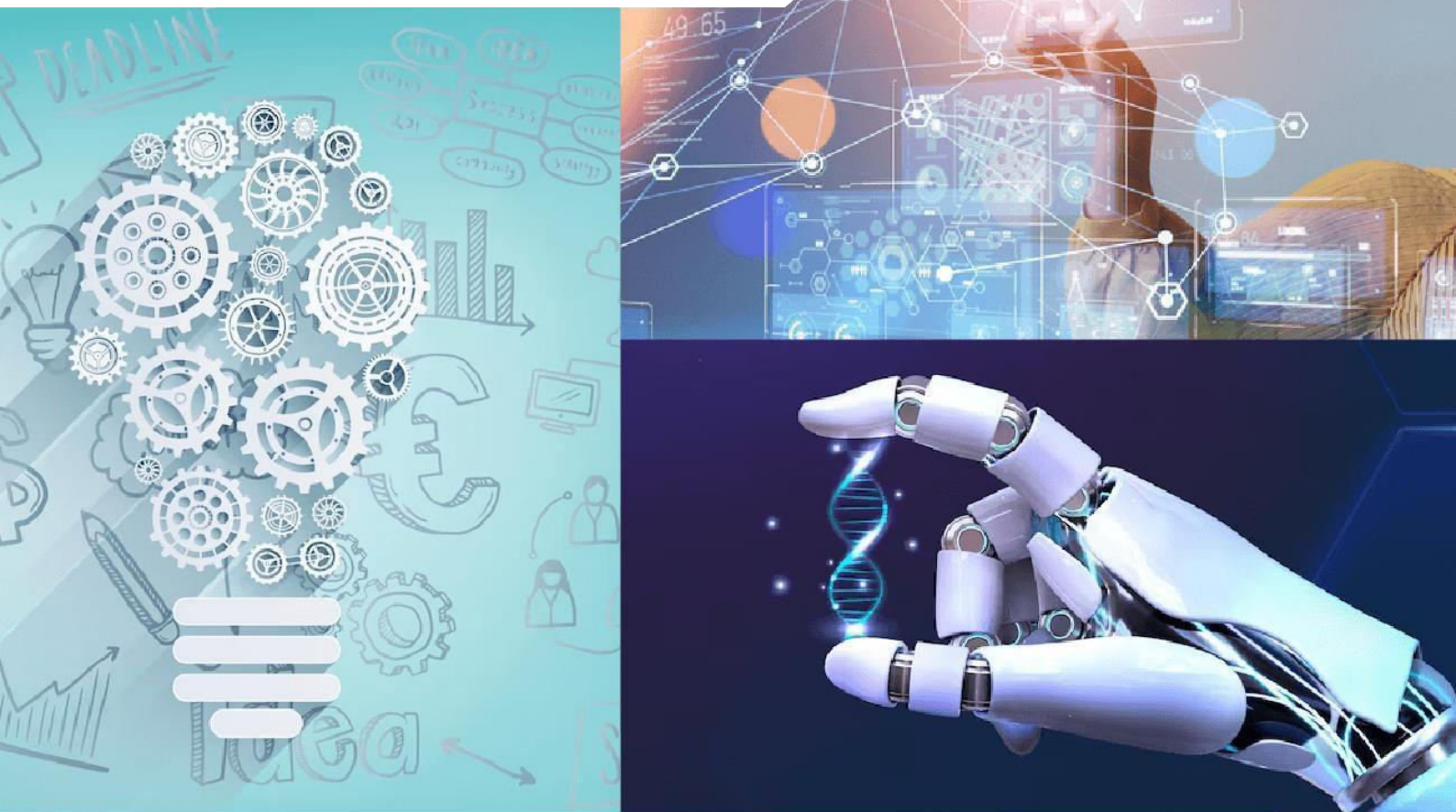




International Journal of Advanced Research in Education and Technology (IJARETY)

Volume 13, Issue 2, March-April 2026

Impact Factor: 8.152



An Efficient Bi-Encoder-Based Skill Classification and Employer Notification Model for Smart Job Market Analysis

Mr. Shaikh Mohamad¹, Mr. G. Balasubhramanyam², Mr. K. Narsimha³, Ms.

MD. Sameena Thahasin⁴

Assistant Professor, Department of CSE, Guru Nanak Institute of Technology, Hyderabad, Telangana, India¹

Student, Department of CSE, Guru Nanak Institute of Technology, Hyderabad, Telangana, India^{2,3,4}

ABSTRACT: The modern labour market increasingly depends on extracting meaningful insights from unstructured job advertisements to map job roles to required skill sets. Existing solutions, such as Context-match, leverage contrastive learning with token-level attention to tackle the extreme multi-label classification task of skill extraction. While these methods deliver strong performance and state-of-the-art accuracy, they often rely heavily on large language models (LLMs), making them computationally expensive and less practical for real-time applications. To address these challenges and enhance the practical applicability of labour market analysis, we propose a comprehensive AI-enabled Job Portal System designed with a lightweight, scalable architecture and integrated employer-employee interaction modules. Our system builds on the conceptual foundation of ConTeXT-match and Skill-XL by offering a smart recruitment platform where employees register with detailed educational and skill data, which is then analysed and processed to match relevant job opportunities. A unique enhancement in our approach is the integration of Gmail-based notification services, where applicants receive immediate email updates indicating whether their profile is under process or rejected. This ensures real-time feedback and keeps applicants informed throughout the recruitment cycle without requiring constant portal access.

I. INTRODUCTION

Labour market analysis plays a critical role in addressing global workforce challenges, including talent shortages, skill gaps, and rapidly evolving job requirements driven by technological advancement. Job advertisements constitute a valuable and continuously updated source of labour market intelligence, as they provide detailed information regarding job titles and corresponding skill requirements across industries and regions. However, the heterogeneous terminology used in job advertisements presents significant challenges for large-scale analysis, particularly in the identification and normalization of occupations and skills. To address this issue, the European Commission introduced ESCO(European Skills, Competences, Qualifications and Occupations), a multilingual classification framework that standardizes nearly 14,000 skills and over 3,000 occupations. Within this context, two key natural language processing tasks are essential: job title normalization and skill extraction. These tasks enable the transformation of unstructured job advertisement data into structured and comparable representations, thereby supporting robust and scalable labour market analysis. This work proposes an efficient and scalable framework for labour market analysis through three major contributions. First, ConTeXT-match, a contrastive learning approach with token-level attention, is introduced for sentence-level skill extraction in an extreme multi-label classification setting, achieving state-of-the-art performance while maintaining computational efficiency. Second, the Skill-XL benchmark is developed to support exhaustive sentence-level annotation of job advertisements, explicitly accounting for redundancy among skill labels and improving the robustness of model evaluation. Third, JobBERT V2 is presented as an enhanced and simplified model for job title normalization, delivering performance comparable to more complex state-of-the-art approaches. Collectively, these contributions establish a practical and effective methodology for extracting and standardizing labour market information from job advertisements, thereby enabling cost-effective analysis at scale for researchers, organizations, and policymakers.

II. LITERATURE SURVEY

Title: Learning job title rep presentation from job description aggregation network.

Year: 2024

Authors: Napat Laosaengpha, Thanit Tativannarat, Chawan Piansaddhayanon, At tapol Rutherford, and Ekapol Chuangsuwanich.

Description: Learning job title representation is a vital process for developing automatic human resource tools. To do so, existing methods primarily rely on learning the title representation through skills extracted from the job description, neglecting the rich and diverse content within. Thus, we propose an alternative framework for learning job titles through their respective job description (JD) and utilize a Job Description Aggregator component to handle the lengthy description and bidirectional contrastive loss to account for the bidirectional relationship between the job title and its description. We evaluated the performance of our method on both in-domain and out-of-domain settings, achieving a superior performance over the skill-based approach.

Title: Data science for job market analysis: A survey on applications and techniques.

Year: 2024

Author: Ibrahim Rahhal, Ismail Kassou, and Mounir Ghogho.

Description: The job market is evolving continuously due to changes in economic landscapes, technological improvements, and skill requirements. In the era of digitalization, a wealth of data is becoming available, opening up new opportunities for labor market analysis. Many stakeholders can make informed decisions if they benefit from accurate and timely insights about the job market. However, traditional data sources and methods used for labor market analysis often fall short of capturing the diversity and trends of the evolving job market. Recently, researchers started exploring various data sources by leveraging data science techniques, which makes information extraction achievable. This survey reviews recent research published between 2015 and 2022 on labor market analytics through data science techniques and discusses future research directions. 101 primary studies were classified and evaluated to identify the data sources utilized for job market analysis; the skill extraction methods and their type; the occupation and sector identification methods; and the application of the study conducted. Finally, we explore potential avenues for future research in this area. The growth of digital technologies has resulted in an explosion of job market data, providing unprecedented opportunities to gain insights into workforce needs. This data includes information on job openings, required skills, job titles, salaries, and more. Additionally, job seekers' online profiles and resumes provide valuable insights into their skills, education, and experience. However, this high volume of data also presents challenges in the processing and analysis steps. To keep up with the rapidly changing job market, traditional methods of data analysis may not be sufficient. Innovative data science techniques such as natural language processing (NLP) and machine learning algorithms can help in analyzing the vast amount of data and extracting meaningful insights.

Title: In-context learning for extreme multi-label classification.

Author: Karel D'Oosterlinck, Omar Khattab, François Remy, Thomas Demeester, Chris Develder, and Christopher Potts.

Year: 2024

Description: Multi-label classification problems with thousands of classes are hard to solve with in-context learning alone, as language models (LMs) might lack prior knowledge about the precise classes or how to assign them, and it is generally infeasible to demonstrate every class in a prompt. We propose a general program, Infer--Retrieve--Rank, that defines multi-step interactions between LMs and retrievers to efficiently tackle such problems. We implement this program using the DSPy programming model, which specifies in-context systems in a declarative manner, and use DSPy optimizers to tune it towards specific datasets by bootstrapping only tens of few-shot examples. Our primary extreme classification program, optimized separately for each task, attains state-of-the-art results across three benchmarks (HOUSE, TECH, TECHWOLF). We apply the same program to a benchmark with vastly different characteristics and attain competitive performance as well (BioDEX). Unlike prior work, our proposed solution requires no finetuning, is easily applicable to new tasks, alleviates prompt engineering, and requires only tens of labeled examples.

III. METHODOLOGY

EXISTING SYSTEM

The existing labor market analysis systems predominantly rely on large language models (LLMs) to extract insights from job advertisements, particularly for mapping job titles to required skills. These systems utilize advanced natural language processing (NLP) techniques and pre-trained models to identify and classify skills at sentence or token level. While methods like ConTeXT-match have improved classification efficiency through contrastive learning and attention mechanisms, they are still integrated into platforms that offer traditional web-based recruitment solutions. In such systems, employers register on a web portal, manually post job listings, and await responses from job seekers. Candidates are required to log in regularly to check application status, often without receiving timely updates. Though

skill matching is effective, these platforms generally lack real-time interactivity, dynamic notification mechanisms, and lightweight alternatives for large-scale deployment.

EXISTING SYSTEM DISADVANTAGES

- High Computational Cost: Models based on LLMs require substantial computational resources, making them unsuitable for lightweight, real-time systems.
- Lack of Real-Time Feedback: Job applicants do not receive immediate notifications about the status of their applications, leading to uncertainty and poor user experience.
- Web-Based Notification Limitations: Status updates are often only available through the portal, requiring manual login and reducing user engagement.
- Minimal User Interaction: Existing platforms lack interactive features like AI-guided assistance or automated helpdesks.

PROPOSED SYSTEM

The proposed system introduces an AI-enhanced job analysis and notification platform that replaces traditional web-based notification mechanisms with intelligent Gmail-based communication for real-time interaction. Leveraging ConTeXT-match, a contrastive learning-based lightweight bi-encoder model with token-level attention, the system efficiently extracts skills and classifies job titles with high accuracy while significantly reducing computational costs. To enhance user experience, the system immediately informs job applicants of their application status—either processed or not processed—via automated Gmail notifications. Additionally, the platform is equipped with a professional admin dashboard UI/UX and supports smooth animations and a chat-based AI interface for enhanced user engagement. This system minimizes user waiting time, maximizes real-time communication, and supports scalable labour market analysis with minimal infrastructure overhead.

PROPOSED SYSTEM ADVANTAGES

- Gmail-Based Notifications: Immediate email alerts enhance communication efficiency and eliminate the need for manual portal checks.
- Lightweight and Fast: Uses an optimized bi-encoder (ConTeXT-match), ensuring reduced computational cost and faster processing.
- Improved Skill Extraction: High-performance token-level classification ensures more accurate and relevant skill-job mapping.
- Real-Time Interaction: Users receive instant updates on application status, improving transparency and user trust.
- Scalability: Suitable for large-scale deployment due to lightweight architecture and asynchronous processing.

SYSTEM ARCHITECTURE

System Architecture



IV. IMPLEMENTATION

ConTeXT-match (Contrastive Token-level Attention Matching) :

ConTeXT-match is a skill extraction approach designed to handle the extreme multi-label classification problem in labor market analysis. Traditional methods often rely on large, computationally expensive models (LLMs), but ConTeXT-match provides a lightweight and efficient alternative.

Core Idea: It uses contrastive learning to train models so that job advertisement sentences and their corresponding skills are embedded closely together in a shared vector space, while unrelated pairs are pushed apart.

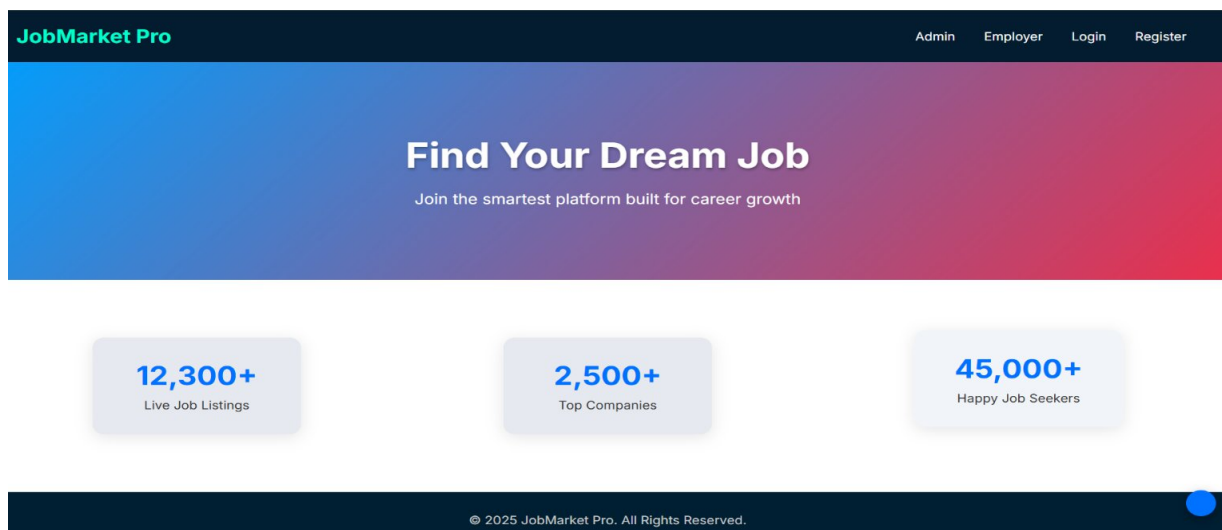
Token-level Attention: Instead of treating the whole sentence equally, it assigns attention weights to each token (word), identifying which parts of the text are most important for predicting a specific skill. This improves both accuracy and explainability.

Efficiency: Unlike large LLMs, ConTeXT-match is built with a bi-encoder architecture (one encoder for sentences, one for skills), making it lightweight, faster, and suitable for real-time job market applications.

Performance: It achieves state-of-the-art results by not only being accurate but also scalable for large datasets like job ads, where there can be thousands of possible skills.

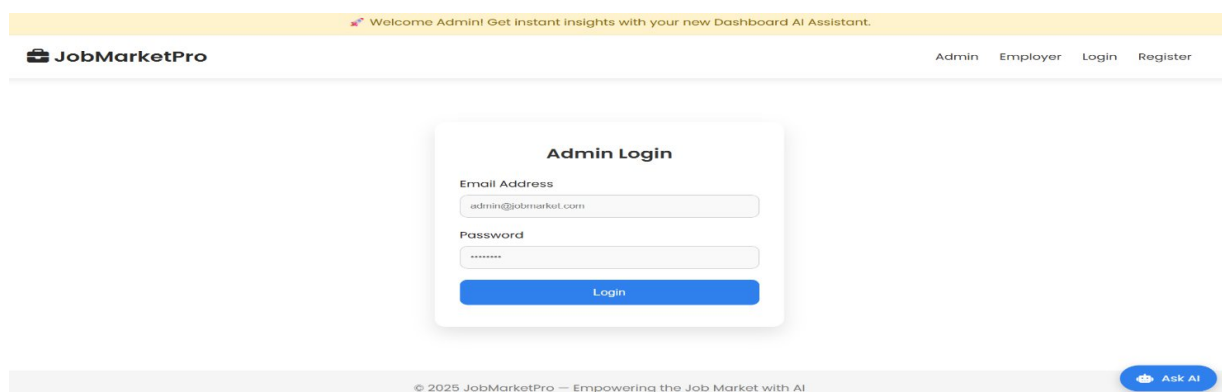
V. EXPERIMENTAL RESULTS

HOME PAGE :



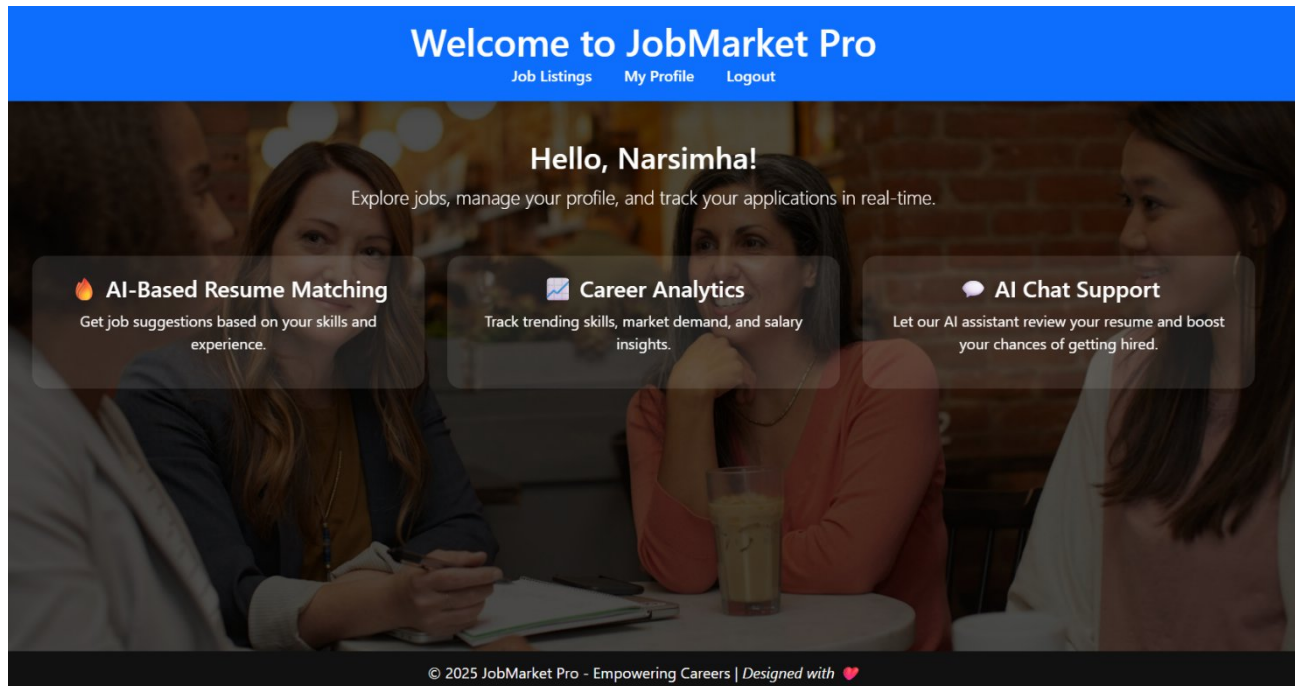
The figure shows the home page of the JobMarket Pro web application, designed to provide an easy and user-friendly interface for job seekers, employers, and administrators.

ADMIN LOGIN PAGE:



The figure shows the Admin Login interface of the JobMarket Pro application, enabling secure access for administrators through credential verification.

EMPLOYEE LOGIN PAGE:



The figure shows the Job Seeker Dashboard of the JobMarket Pro system, providing personalized access to job listings, profile management, and application tracking.

RESULT PAGE:



VI. CONCLUSION

In conclusion, this work presents a scalable and efficient approach to labor market analysis by addressing two critical tasks: skill extraction and job title normalization. The proposed ConTeXT-match framework enhances both accuracy and explainability in skill extraction, outperforming LLM-based systems while significantly reducing computational costs. Complementing this, JobBERT V2 achieves state-of-the-art performance in job title normalization by leveraging high-quality skill data generated through the extraction model. To support robust evaluation, the introduction of Skill-XL, a comprehensive benchmark with sentence-level annotations, establishes a new framework for measuring both accuracy and practical usefulness of skill extraction methods. Together, these contributions demonstrate the potential for large-scale, real-time labor market analysis, enabling more practical applications in recruitment systems. Future research can expand this work to non-English languages and broader domains, while the open-sourcing of models and benchmarks encourages continued innovation and collaboration in the field.

FUTURE ENHANCEMENT

Several feature enhancements can further improve the usability, accuracy, and scalability of the AI-enabled Job Portal. One potential enhancement is the integration of personalized job recommendations, where the system not only matches applicants to jobs based on extracted skills but also ranks opportunities according to past applications, profile preferences, and skill proficiency. Another improvement could be advanced analytics dashboards for employers and admins, providing insights on skill demand trends, applicant success rates, and job posting effectiveness. Incorporating multi-language support would broaden the system's applicability across diverse labour markets. Additionally, the notification system can be extended to include SMS or mobile push notifications alongside Gmail updates, ensuring faster communication with applicants. Finally, leveraging incremental learning in the skill extraction engine could allow the model to continuously improve by learning from new job advertisements and feedback, keeping the system up-to-date with emerging skills and industry trends. Collectively, these enhancements can significantly increase engagement, efficiency, and trust in the recruitment process while making the platform more robust for large-scale labour market analysis.

REFERENCES

- [1] Ibrahim Rahhal, Ismail Kassou, and Mounir Ghogho. Data science for job market analysis: A survey on applications and techniques. *Expert Systems with Applications*, 251:124101, 2024.
- [2] Wendy V Cunningham and Paula Villaseñor. Employer voices, employer demands, and implications for public skills development policy connecting the labor and education sectors. *The World Bank Research Observer*, 31(1):102–134, 2016.
- [3] Imane Khaouja, Ismail Kassou, and Mounir Ghogho. A survey on skill identification from online job ads. *IEEE Access*, 9:118134–118153, 2021.
- [4] Benjamin Clavié and Guillaume Soulié. Large language models as batteries-included zero-shot ESCO skills matchers. In Mesut Kaya, Toine Bogers, David Graus, Chris Johnson, and Jens-Joris Decorte, editors, *Proceedings of the 3rd Workshop on Recommender Systems for Human Resources (RecSys in HR 2023)* co-located with the 17th ACM Conference on Recommender Systems (RecSys2023), Singapore, Singapore, 18th–22nd September 2023, volume 3490 of CEUR Workshop Proceedings. CEUR WS.org, 2023.
- [5] Karel D'Oosterlinck, Omar Khattab, François Remy, Thomas Demeester, Chris Develder, and Christopher Potts. In-context learning for extreme multi-label classification, 2024.
- [6] Parisa Safikhani, Hayastan Avetisyan, Dennis Föste-Eggers, and David Broneske. Automated occupation coding with hierarchical features: a data-centric approach to classification with pre-trained language models. *Discover Artificial Intelligence*, 3:6, 02 2023.
- [7] Decorte, Jens-Joris and Van Haute, Jeroen and Demeester, Thomas and Develder, Chris. JobBERT : understanding job titles through skills. In *FEAST, ECML-PKDD 2021 Workshop, Proceedings*, page 9, 2021. [8] Meng Zhao, Faizan Javed, Ferosh Jacob, and Matt McNair. Skill: A system for skill identification and normalization. *Proceedings of the AAAI Conference on Artificial Intelligence*, 29(2):4012–4017, Jan. 2015. [9] Shanshan Jia, Xiaolan Liu, Ping Zhao, Chang Liu, Lianying Sun, and Tao Peng. Representation of job-skill in artificial intelligence with knowledge graph analysis. In *2018 IEEE Symposium on Product Compliance Engineering- Asia (ISPCE-CN)*, pages 1–6, 2018.
- [10] Luiza Sayfullina, Eric Malmi, and Juho Kannala. Learning representations for soft skill matching. In WilM.P.vanderAalst,VladimirBatagelj,Goran Glavaš, Dmitry I. Ignatov, Michael Khachay, Sergei O. Kuznetsov, Olessia Koltsova, Irina A. Lomazova, Natalia Loukachevitch, Amedeo Napoli, Alexander Panchenko, Panos M. Pardalos, Marcello Pelillo, and Andrey V. Savchenko, editors, *Analysis of Images, Social Networks and Texts*, pages 141–152, Cham, 2018. Springer International Publishing.

International Journal of Advanced Research in Education and Technology

ISSN: 2394-2975

Impact Factor: 8.152