Designing a Bidirectional Job Matching Model Using Artificial Intelligence: the Case of Ministry of Labour and Skills (MoLSs)

Tewodros Desta and Million Meshesha

October 25, 2024

Designing a Bidirectional Job Matching Model Using Artificial Intelligence: The Case of Ministry of Labour and Skills (MoLSs)

Addis Ababa University

The School of Information Science: Past, Present, and Future

Tewodros Desta
teddyexe2017@gmail.com
School of Information Science
Addis Ababa University

Million Meshesha (Ph.D)
million.meshesha.aau.et
School of Information Science
Addis Ababa University

ABSTRACT

The rapid advancement of AI and big data presents both challenges and opportunities for global labor markets. Urban youth are highly plagued by frictional unemployment. This incredibly high level of frictional unemployment has spatial and informational components, driven by high costs. However, currently existing but undiscovered matches between job seekers and open positions account a smaller portion of urban unemployment. This study explores leveraging digital tools to enhance the effectiveness of Ethiopia's public employment services, which face limitations in job matching, labor market information, and policy implementation. The research aimed to design an AI based job matching model to improve the public employment service support provided by Ethiopia's Ministry of Labor and Skills. The study followed a hybrid knowledge discovery process (KDP) data mining model, which includes understanding the problem domain, collecting historical data, preparing datasets, and four logically designed experimentations for selecting best fit job matching model for the public employment service. The designed job matching model match and suggest the best fit jobs to the jobseeker or vis versa by calculating a similarity index using vector representations job seeker's and the available job postings features in the dataset to describe features (documents and corpus of texts) using n-dimensional vectors which each dimension representing the frequency of a certain term in a document in the available skill sets and then ranking them according to their Doc2vec algorithm and cosine similarity measure. This designed model has been implemented in python. The selected Doc2Vec model with selected text embedding's from job seeker and vacancy datasets achieved a cosine similarity of 0.99 to 1, revealing the desired result for the best-fit job matching model. The high cosine similarity result suggests the model is performing well in matching jobs to the jobseeker.

KEYWORDS: AI, hybrid KDP, frictional unemployment, job matching, Doc2vec, cosine similarity

1. INTRODUCTION
1.1 Background

This current era where social, economic and political activities are dependent on information and communication technologies is also known as the information age and the digital era. It refers to the time period in which personal computers and other subsequent technologies are introduced to provide users the ability to easily and rapidly transfer and access information in human history to the next phase, characterized by the shift to the fourth Industrial Revolution [1], that moves business and society to the next phase of efficient productivity and improved quality of life.

Over the past few years, the development of Artificial Intelligence tools and technologies coupled with the age of big data has tremendous importance and have been providing improved accuracy and efficiency in research and innovations that has become very important to the public at large than ever before in the era of non-Artificial Intelligence[2]. However, these advanced digital technologies pose significant challenges and opportunities for the labor markets around the world, and Public employment services have the potential to play a central role in mediating these dynamics due to the fact that they are involved [3].

Job matching is the process of assigning the job seeker that fulfills the requirements of a job. According to Greenberg, job matching is the process of matching the right person to the right job based upon the individual's inherent motivational strengths that requires a thorough understanding of the vacancy and the job seeker. It is important to study job matching in order to address adverse phenomena that occur in the job market such as unemployment, underemployment, long-term unemployment and unfilled job vacancies, each of which cause further undesired economic and social consequences[4].

Public employment service provision in Ethiopia [5] is a function of the Ministry of Labour and Skills (MoLSs), which was part of the Ministry of Labour and Social Affairs. According to [3] the public employment service should play a key intermediary role in the delivery of labour market services and policies. The services include job brokerage by disseminating job vacancies to be filled and facilitating matches between labour market supply and demand; provision of labour market information by collecting data on job vacancies and potential applicants; making necessary market adjustments by implementing labour market policies aimed at regulating labour demand and supply; and managing labour migration by coordinating the geographic mobility across borders. However, currently the effectiveness of the service remains very limited due to an ineffective and inefficient labour market information system, limited capacity and job search skills.

1.2 Problem Statement

Youth unemployment is the key problem for both developed and developing countries in the world that currently turns to be policy issue. To alleviate this global problem, developed countries are promoting national employment program based on macroeconomic and sectoral policies in the use of digital technology to support the labour market with an emphasis on inclusive job creation through their active labour market policies. However, the overall macroeconomic environment, in many developing countries, has become less supportive and fiscal space is now constrained which will make it harder to promote an inclusive job creation in the world poorest countries. Nevertheless, due to the passive labour market policies, weak Public employment service provision which lack job matching technologies designed and implemented in structural transformation, most countries in Africa are unable to provide up-to-date job information and quality matching for both job seekers and employers in the labour market. Hence, due to the absence of correct job match, most African countries are unable to provide up-to-date job information and quality job match in public employment service provisions in promoting job creation and employment resulting to discrepancies in the measurement and monitoring of most of the Sustainable Development Goals (SDG) indicators in the continent [6].

Ethiopia is one of the many African countries that do not have a working digital job matching platform to support public employment service providers responsible for registering jobseekers and providing and overseeing job matching and placement services that use a common national information system to meets the demand and supply in the labour market at the national and international standards[7]. There is no central, compulsory and permanent digital job matching platform for public employment service provision in Ethiopia[3]; Hence, the absence of this employment information system to alleviate youth unemployment though efficient job information indicate the mismatches problem in the demand and supply of skills. It also has efficient-reducing effect on knowledge and skill, productivity and socioeconomic deterioration in the country [3][7].

The registration of job seekers and vacancies is part of a public employment service provision system for the benefit of getting centrally coordinated employment information that also include a supportive role for other services like guidance and labour market information provision. However, the system vary from simply taking notes on registration cards manually to uncoordinated systems of registration at different national and regional government offices which shows inconsistency and non-uniformity in the way they do so. Hence, their contribution for relevant, comparable, and regularly updated administrative job seekers data and employers skills need to provide information about job information, labour supply and demand, and to describe developments of the labour market is not significant[7].

Job seekers, searching for work in the formal sector, need to have easy access for job information to get employment opportunities, understand labor market conditions, and possess an accurate understanding of their own skill level and how firms and markets might value those skills. However, the labor market is analogue, not digital, with more than 50% of all unemployed urban youth are relying on centrally located physical job boards and newspapers as the most common method of job search [8]. Nevertheless, urban youth unemployment is the result of frictional issues. This frictional unemployment is a result of the time it takes for potentially matched employers and job seekers to find one another. Job searches cost money and time, increasingly so across larger distances which makes it expensive and unaffordable, this in turn breaks job search intensity in short. The relatively large amounts of money and time spent on job searches are less efficiently used when job seekers and employers don't have enough current information. The flow of information is also slowed by non-travel costs such as newspapers cost, broker's charge, application fees and credentials copy[9].

Information slowdowns also create misperceptions of both the labor supply and demand. These misperceptions are additional causes of information-based frictional unemployment. These interplay between distance and information drives job match and employment to an inefficient level. Hence the absence of digital job match makes the youth job seekers and firms hard to find easily accessible job information and real time job matching connected to urban labour market[3][10].

Hirbaye Mokona et al [11]found out that there is a strong evidence that shows the adverse effects of frictional unemployment causes to youth, namely psychological distress, poor health condition, consequent lower wages, underemployment and slow career progress. These mismatches problem in the demand and supply of skills also causes interrelated impact on socio-economic, political and moral consequences at both individual and societal level in the country. This corroborates the urge to applying efforts to alleviate skill mismatch.

Recently, there has been an increase in public sector initiatives that use AI-based job matching in their public employment services. Belgium partnered with a private company to develop a job matching algorithm, known as [12], a deep learning model that predicts the probability of a match from associations between keywords in the jobseekers profiles and vacancy descriptions. Bolet etal [12] created a model based on data from the general population. The purpose of the model was to display alternative relevant occupations and associated job opportunities to job seekers. The developers found that their recommendations helped broaden the job search for users and increased the number of job interviews they were able to secure. However there is a growing interest in developing countries to adopt a similar approaches. More recently, the Ministry of Labor in Kenya is working with research institutions to the advanced digitalization of public employment services provider responsible for registering jobseekers and providing and overseeing job matching and placement services (both domestic and foreign) in most areas[3]. As to the researcher knowledge, there is no attempt in Ethiopia to come up with AI based digital Public employment services and apply technology for jobseekers registration, job matching and placement which are considered too complex to apply in urban area[7]:

Therefore, to fill this gap in Ethiopia's employment service system, this study aims to design a job matching model using Artificial Intelligence and Data Mining techniques to generate real-time job matching by connecting registered job seekers' profiles with online available job postings.

1.2.1 General Objective

The general objective of this study is to design a job matching model using Artificial Intelligent tools and technology in collection, integration and representation of vacancy and job seeker data as an integral part of the public employment service support for MoLSs.

1.2.1.1 Specific Objectives

For the attainment of the general objective of the study, list of specific objectives are the following:

- To review literature so as to understand and select techniques, algorithms and approaches well-suitable for the collection, integration and representation of vacancy and job seeker data
- To understand the problem and collect historical jobseekers data and online vacancy data from the right source
- To generate quality data by applying data preparation tasks for structured and unstructured data.
- To design best fit model that can match job seekers profile and vacancy for given features.
- To evaluate the performance of the preferred model using effectiveness measures.

1.3 Literature Review

In online job board, job matching is a crucial aspect of the labour market, benefiting both job seekers and employers. The search and matching model, developed by economists Dale Mortensen and Kenneth Burdett, explains the process of finding a job as an information gathering and matching process. Mortensen's Nobel Prize winning work highlighted that job-seeker and employer information is often imperfect and costly, leading to a time-consuming matching process. This information asymmetry can be particularly challenging for young people who lack social capital and networks, potentially excluding them from unadvertised job opportunities [16].Several recent studies have proposed different approaches to enhance job matching and help job seekers find suitable employment opportunities.

Choudhary et al. [22] developed a data mining system that uses collaborative filtering methods to recommend jobs to job seeker based on their skills and abilities, depending upon jobseekers profile and by calculating conditional probability for similarity index using Euclidian distance of two skill sets occurring together and then ranking them accordingly using naïve Bayes algorithm. The recommendation system was implemented in python. The designed system was able to successfully recommend jobs based on a user's current skill set by combining it with the similar skills in the global data set acquired. However, this model used an asymmetric similarity function. The limitations of using an asymmetric similarity function is that each item will tend to have high conditional probabilities with items that are being purchased frequently for solution inspired from the inverse-document scaling performed in information retrieval systems.

Mamadou [19] developed content-based recommender system which suggests jobs to LinkedIn users. Experiments recommend that to forecast the users interests for jobs, using basic resemblance measures together with their relations data collected by Work4 can be upgraded upon. The second portion of this study presents a method to estimate the significance of each field of users and jobs in the task of job recommendation. Lastly, the third part is devoted to the use of a machine learning algorithm in order to increase the results gained with resemblance measures: the authors trained a linear SVM (Support Vector Machines). The results indicate that using this supervised learning procedure increases the performance of the content based recommender system. The proposed recommender systems only use users' interactions data and jobs descriptions to predict users' interests for jobs.

Other studies have proposed on developing new models and algorithms to enhance the job matching process: Morgan [17] defined job matching as the process of assessing a job seeker's characteristics to determine if they meet the requirements of a job. He studied on the development and evaluation of a web-based instrument to assess degree of match among employment preferences . Le et al. [23] developed an adapting network that determines job preferences and career priorities based on historical job request. Joseph et al [24] developed an algorithm that adds new weighted arcs to the digraph of endorsements based on the relation of endorsements.

Furthermore, several studies have employed machine learning algorithms to improve job matching. Lee, Kim and Na [25], developed a method for career matching amidst university students and companies using Artificial Intelligence based Design platform (AID). They analyzed the results from the model with statistical methods like least squares, Pearson correlation, and

Manhattan distance. In their experimentation they found that their model gave them zero miss-matching between student's skills and company's need on the other hand statistical method gave 30% miss-matching. Similarly, Leeprechanon et al.[26], in "Job Matching Using Artificial Intelligence" developed a model that match candidates with the right skills and workers with the right company. They used the Stochastic Gradient Descent, a supervised learning algorithm in their research. Furthermore, they developed their algorithm, Hiring History, which will review a person's previous hiring details.

Theoretically the use of AI algorithms for matching is appealing, practical research in PES job matching are rare worldwide. This could be the result of several factors. One is that the field of job matching has seen the launch of numerous private websites to bring jobs and jobseekers together [12].Moreover, a related local research is very limited until this study is carried out and hence the only one available local research on labour market is the work of selam[18].

Selam Abebe[18] developed a web based context aware personalize job recommender model. She integrated the job seeker's preferences and context in addition to their resume and the job description. The author implemented different types of recommenders; collaborative filtering, content-based and context/preference-based to improve job recommendations for a jobseeker. Her Evaluation showed that personalized job recommendations can be enhanced by integrating contextual information into the user profile. Overall, her thesis proposed a novel context-aware personalized job recommendation system. However, the proposed system and algorithms are not described in detail, so further work is needed to fully implement and evaluate the approach. In addition, integrating additional contextual factors beyond just location preferences could be explored to further improve recommendation accuracy.

The related works mentioned earlier used various classification algorithms but job seekers data sources were not from the same origin; some from resume, some social media etc., and they failed to test various preprocessing strategies such as test cleaning, feature engineering, categorical feature encoding etc., for the job seekers and vacancy data acquired and for the suggested classifiers resulting in poor classifier accuracy.

2. Methodology

This study aims to give solution regarding the problem of frictional unemployment in public employment service. Such issues include, what is the profile standard of jobseekers and classification standard of vacancy what is the profiling of jobseekers and vacancy? What are the determinant components and features of job matching that makes the Public employment service inaccurate and inefficient?

Developing a solution to a given problem requires adherence to scientific principles and guidelines, from the initial stages to the final outcome. The study systematically addresses the research question in order to achieve the objective of the investigation. To this end, there are various methodologies available to be followed, depending on the characteristics of the research or the discipline. The methodology outlines the step-by-step procedure that the research must follow, defining the concepts or phenomena that require investigation in a manner that provides a clear and understandable picture of the complexity of the problem at hand[13].

2.1 Research design

This study follows experimental research. As stated in [13], experimentation is a method used to discover facts and to test ideas. The science that proceeds with experimentation will produce proves of hypothesis by experiment.

For conducting systematically the experiment, the study uses the hybrid DM process model [14].This model is selected because of the following reasons; emphasize on understanding of the problem domain and data mining, it is research-oriented structure and the model has several feedback, and the extracted knowledge extended for another application domain. The hybrid process model has six-steps; understanding of the problem, understanding of the data, preparation of the data, data mining, evaluation, and use of the Knowledge.

This study is fully implemented in Python language under the Jupyter Notebook framework. Data transformations are carried out using Pandas library whereas Data mining algorithms and Statistics methods applied come from Scikit-learn, Scipy and Numpy libraries. Doc2vec is applied following the implementation on Gensim library.

2.2 Understanding of the problem domain

The LMIS is an AI-based application for labour market management, developed by the Ministry of Labour and Social Services (MoLSs). It was launched to provide PES (Public Employment Services) including job placement, monitoring, and administration.The LMIS application runs on a Linux operating system with Python, JavaScript, and Hasura at the backend, and a hybrid database management system. It provides real-time monitoring and administration of PES.
The three main components are:

- Job seekers: Registered through a simple process, providing 65 attributes. This data is used for job matching and generating labor market intelligence.
- Employer companies: Can post job vacancies on the National Job Portal, providing 23 attributes.
- MoLSs: Utilizes the automation and technological capabilities of LMIS to improve the quality, efficiency, accessibility, and availability of PES.

The LMIS application has three main portals:

- Registration Portal: For job seekers and employers
- National Job Portal: For posting and managing job vacancies
- National Recruitment Portal: For government job providers to manage the recruitment process

The research proposes developing an AI-based job matching model for the PES to better match job seekers to vacancies based on skills and competencies. The expected benefits include improved job access, better job-skill matching, and increased trust in PES for both job seekers and employers.

The proposed model involves a "bidirectional" matching process, using AI tools to match key attributes like education, work experience, skills, and location between job seekers and vacancies. The central LMIS platform will serve as the hub for job seekers to build profiles and employers to post jobs and screen candidates.

2.3 Understanding of Data
2.3.1 Understanding of jobseeker Data

The research considers job seekers from urban Ethiopia, using historical data gathered from the Ministry of Labour and Social Affairs' (MoLSs) database containing job seekers' profiles registered online for the Public Employment Service (PES) on the Labour Market Information System (LMIS) [47]. The data includes information about job seekers' profiles registered online across 12 regions of urban Ethiopia, with a total of 2,137,932 job seekers registered on the LMIS this year [27].

The MoLSs LMIS stores job seekers' profiles in a database for PES data monitoring, using various tables for targeting, job matching and placement, as well as LMI reporting [27]. The original data attributes available in the LMIS database are presented in Figure 3.2 [27], and the researcher, in collaboration with domain experts, selected a subset of 208,013 job seekers' profiles containing the following key attributes: resident address (region, zone, woreda, kebele, and house number), education background (highest education grade, TVET level, study or department, country, institution name, graduated year, licenses and certifications, and skills), work experience (employment status, weekly working hours, unemployment period, organization name, title/position, employment type, salary, start and end dates, and termination reason), licenses and certifications (name, skills, issuing institute, start and end dates), and employment preferences (job search, enterprise preference, employment type, sector, and city) [27].

The job seeker profile data extracted from the raw source has a structured format that forms feature categories.These categories include personal information (resident address), education, experience, preferences, and skills, along with their respective details, except for data that may affect privacy.

2.3.2 Understanding of vacancy Data

The data used for this study on vacancy analysis was collected from online job postings on the Ethiopian Reporter Jobs website (www.ethiopianreporterjobs.com) [15]. The job advertisements were scraped using a Selenium-based web scraper with a Chrome driver to extract the necessary data elements [48]. This "real-time labor market information" provides up-to-date details on job postings as soon as employers list them online [15].

The collected data was structured into various attribute categories, including job ID, company, education requirements, job title, experience level, job description, skills, location, and industry. The web scraping process was designed to be ethical and considerate of the website's terms and conditions, with measures taken to avoid overwhelming the remote server [28].

The focus of this study is on the IT (Information Technology) sector, as it offers a broad range of educational qualifications and dynamic job requirements. The data was collected from February 2024 to May 2024, and the extraction of useful data from the raw vacancy metadata was performed for each attribute.

The 12 regions of Ethiopia and their respective major cities, obtained from the ethiopia-city-region.csv and region.csv files, are used as a secondary dataset input for the job matching process in this study, with the source for the regions and cities being Wikipedia [29].

2.3.3 Data Preparation

Data Preparation for Structured Data

As noted by Cios et al. [14]this step concerns deciding which data could be used as input for DM methods in the subsequent step. Major tasks in data preparation are Data Cleaning, Data Integration, Data Reduction, and Data Transformation. It also includes sampling derivation of new attributes (discretization), and summarization of data (data granularization).The key tasks applied in this study were data cleaning, transformation and derivation of new attributes.

Handling null values: Removed columns with significant null values, such as jobid and uniq_id.
Derivation of new attributes:
Location: Extracted region from job location address, imputed with "Ethiopia" if blank [14].
Experience ranges: Derived features like years of experience, experience level (entry-level, mid-level, senior).
Education level: Created attributes like bachelor's, master's, doctorate, unknown, and other specializations.
Transformation: Discretized experience into ranges (0-1 years, 1-5 years, 5-10 years, 10+ years).

Text Pre-processing on Unstructured Data

As noted by Cios et al. [14] this step concerns deciding which data could be used as input for DM methods in the subsequent step. In the context of data mining, the tasks of preprocessing text data typically involve the following steps: Tokenization, Stop Word Removal, Stemming or Lemmatization, Normalization, Feature Extraction, Feature Selection, Handling Missing Data, and Dimensionality Reduction.
Data cleaning: Removed special characters, HTML tags, and tokenized, lemmatized, and removed stop words [14].

Vector representation (Bag of Words): Converted text data into a numeric representation to analyze word counts and common themes [41].
Dimensionality reduction: Explored techniques like count vectorizer and TF-IDF, but found them to be suboptimal for the dataset.

In this study, the researcher performed a fuzzy string matching on the "title" column of the jobseeker dataset to help standardize and consolidate the different variations of job titles [21]. Standardizing job titles is an important step in improving the accuracy of job matching algorithms, which try to connect job seekers with relevant job openings.

The fuzzy string matching algorithm, using the "fuzzywuzzy" library's "process.extract" function, finds the most similar job titles for each unique title based on the token sort ratio in Python, which implements the Levenshtein distance. This allows grouping together job titles that are semantically similar, even with slight variations in spelling or phrasing.

Similar to preprocessing vacancy text data, the study aims to create 20 vectors to train the model for job-seeker matching. The text data used for training the jobs model includes job title, job description, and skills. To train the jobseeker model, the study utilizes similar text data, such as jobseeker title, description, skills, and location, from the jobseeker dataset.

2.3.4 Building the model

In this study, the data miner used various data mining methods to derive knowledge from the preprocessed data [14]. Predictive analytic algorithms were employed to uncover patterns for designing a bidirectional job matching model for MoLSs. The tool used for the data mining task was Python, which provides a variety of machine learning algorithms.

The selected predictive model algorithms were Doc2vec and Cosine similarity. The Doc2Vec model architecture was chosen to create a matrix of vector representations for training both the jobseeker and vacancy datasets separately. Doc2Vec is a powerful technique for document embedding, which represents documents such as jobseekers and vacancies in a vector space. This allows the model to capture the semantic meaning and contextual relationships within the text, which is valuable for the goal of job matching.

The trained models were used to build job matching models on an experimental basis, starting from a basic content-based filtering approach and moving towards a more sophisticated model involving text embedding techniques to better capture the semantic similarity between job seekers and vacancies. The results of these models were compared using the Cosine similarity score to find the closest match and recommend the top (n) matches, allowing the selection of the best-fit model to address the research goal.

The experiments were divided into four groups, using different combinations of features, such as professional (main) skills, locations, text vectors, and varying the number of epochs for jobseeker and vacancy vector training. The cosine similarity score was applied to each technique to compare the degree of match at each level, with the similarity column being used to rank or filter job postings based on their relevance to a jobseeker's skill-based profile and preferences.

According to MoLSs [27], professional skills (hard skills) are job-specific abilities obtained through education, training, or work experience, and these skills are the main skills used for skill-based profiling, job matching, and placements in MoLSs.

3. Experimentation

Summarizing the four experiments and their respective results are;
Experiment 1: Using Professional (Main) Skills Features (Control)
In this experiment, the job matching model was built using only the professional (main) skills features of the jobseekers and vacancies. This served as the control group for the subsequent experiments. According to the result shown above, the similarity scores range from 0.607 to 0.707, indicating a reasonable level of match, but a less ideal fit.

Experiment 2: Using Professional Skills and Locations
This experiment added the location feature to the professional skills, aiming to improve the job matching accuracy by considering both the skills and the geographic preferences of the jobseekers. The summary of result shows that, the similarity scores range from 0.881 to 0.996, indicating a very high degree of match between the job requirements and jobseeker profiles. The top 7 listings have similarity scores above 0.95, suggesting an exceptionally close fit.

Experiment 3: Adding Text Vectors
The third experiment incorporated text embedding techniques, using the Doc2Vec model to capture the semantic similarity between jobseekers and vacancies. This approach was expected to provide better job matching results compared to the previous experiments. The similarity scores for all 10 listings are an impressive 1.0, indicating a perfect match between the job requirements and jobseeker profile. This is a significant improvement compared to the previous set of matches on experiment 4.5, where the scores ranged from 0.881 to 0.996. The addition of the text vector comparison has further strengthened the evaluation of these job matches.

Experiment 4: Adding Text with Lesser Job Features and More Epochs
In this final experiment, the researchers reduced the number of job features used, but increased the number of training epochs for the jobseeker and vacancy vector representations. The goal was to explore the impact of prioritizing the text-based similarity over the specific job features. Evaluating the updated set of 10 best job matches using fewer job features and 200 epoch. The similarity scores range from 0.999179 to 0.999227, which is exceptionally high and indicates a very strong match between the job requirements and jobseeker profile.

4. Result and Discussion

The results of these four experiments were analyzed and compared using the cosine similarity score to identify the best-fit job matching model. The experiments demonstrated the progressive improvement in job matching accuracy as the researchers incorporated more advanced techniques, such as text embedding and increased training epochs. The final model, which utilized the text-based similarity with fewer job features and more training epochs, was found to be the most effective in addressing the research goal of designing a bidirectional job matching model for MoLSs.

Evaluation includes understanding the results, checking whether the discovered knowledge is novel and interesting, interpretation of the results, and checking the impact of the discovered knowledge [26]. This research job matching model using AI were developed, to match registered jobseekers and online vacancy posted by employercompany, as part of PES in MoLSs. Doc2vec based on text embedding techniques with lesser job features exhibited best fit model performance.

Considering the goal of this study, the utilization of Doc2Vec techniques for finding a best fit job matching model reveals distinct performance variations across different techniques conducted. Text embedding technique revealed outstanding performances compared to job marching on selected job features.

The professional skill based features and location are also important features common for both jobseekers and employer company demands for best job match. Furthermore, the result confirmed that both the skill and location preference of job seeker can be satisfied in a job match as compared to other researches revealed in the review literatures.

With respect to the hyper parameters, the models achieved better accuracy on increasing epoch up to 200, compared to their non-hyperparameter tuned counterparts. This indicates that the fine-tuning of hyperparameters allowed the models to better capture the underlying patterns and relationships in the data, resulting in improved performance metrics.The following figure 4.1 is the model of this research.

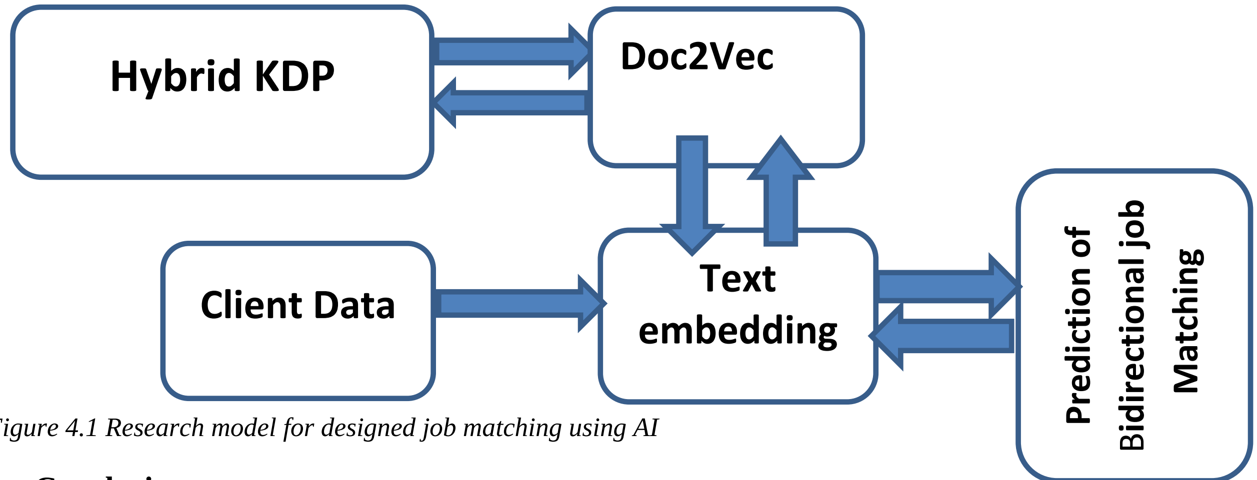

Figure 4.1 Research model for designed job matching using AI

5. Conclusion

In conclusion, this study aimed to address the persistent issue of frictional unemployment among urban youth populations by developing an AI-based job matching model. By leveraging digital tools and advanced job matching algorithm, the model has the potential to significantly reduce the time and friction associated with the job search process. The researchers utilized a hybrid knowledge discovery process, focusing on data mining techniques in the use of Doc2Vec text embedding approach, to design an effective bidirectional job matching model.

The key findings of the study demonstrate the suitability of the Doc2Vec technique for building an intelligent bidirectional job matching model. The distributed, continuous vector representations in text embedding enabled the model to capture the semantic meaning and contextual relationships between vacancy and jobseeker with lesser job features and more epochs, leading to nuanced and accurate matches beyond simple keyword-based comparisons.

However, the study was limited by the availability of data, focusing solely on the IT sector and relying on self-reported jobseeker profiles, which may introduce biases. Additionally, the lack of longitudinal data restricted the ability to examine trends over time. These methodological weaknesses and data limitations represent challenges in drawing robust and generalizable conclusions.

Despite these limitations, the insights gained from this research contribute to the ongoing efforts to address frictional unemployment among urban youth. The successful implementation of the proposed AI-based job matching model could help public employment services improve labor market efficiency and create better outcomes for this key demographic.

Reference

[1] Iqbal H. Sarker., “AI-Based Modeling: Techniques, Applications and Research Issues Towards Automation, Intelligent and Smart Systems,” *A SPRINGER Nat. Journal,* 2022.

[2] A. H. Gandomi, F. Chen, and L. Abualigah, “Big Data Analytics Using Artificial Intelligence,” *MDPI, Basel, Electron.*, vol. 12, no. 4, Feb. 2023, doi: 10.3390/electronics12040957.

[3] SINCE Programme, “Public Employment Services Case Study,” 2020.

[4] Herbert M. Greenburg, “Job Matching to Hire Motivated Employees,” National Precast Concrete Association, 2010.

[5] The Federal Democratic Republic of Ethiopia, Ethiopian Statistics Service, "The 2022, 1st Round urban employment unemployment survey (A Comparative Analysis with 2016, 2018 UEUS and 2021 LMS Survey Results)," 2022.

[6] ILO, *Global Employment Policy Review 2023 Global Employment Policy Review 2023*. 2023.

[7] International Labour Organization, “Public Employment Services Provision and Labour Market Information Collection and Utilization ETHIOPIA,” 2018.

[8] Ketema Mezigebu, “The Socio-economic Challenges and Job Searching Practices of Unemployed University Graduates: The Case of Adea Berga District, West Shoa Zone, Oromia Region.” M.A. thesis, Addis Ababa University, Department: Department of Sociology, 2020

[9] World Bank Group, “Why so idle? Wages and Employment in a Crowded Labor Market.” p. 79, 2016.

[10] G. Abebe *et al.*, “Job Fairs: Matching Firms and Workers in a Field Experiment in Ethiopia,” *Job Fairs Matching Firms Work. a F. Exp. Ethiop.*, no. January 2024, 2017, doi: 10.1596/1813-9450-8092.

[11] H. Mokona, K. Yohannes, and G. Ayano, “Youth unemployment and mental health: Prevalence and associated factors of depression among unemployed young adults in Gedeo zone, Southern Ethiopia,” *Int. J. Ment. Health Syst.*, vol. 14, no. 1, pp. 1–11, 2020, doi: 10.1186/s13033-020-00395-2.

[12] J. Körtner and G. Bonoli, “Predictive Algorithms in the Delivery of Public Employment Services,” 2021.

[13] P. Kanellis and T. Papadopoulos, “Conducting Research in Information Systems,” *Inf. Syst. Res. Methods, Epistemol. Appl.*, no. August 2001, 2011, doi: 10.4018/9781605660400.ch001.

[14] K. J. Cios, W. Pedrycz, R. W. Swiniarski, and L. A. Kurgan, “Data mining: A knowledge discovery approach,” *Data Mining: A Knowledge Discovery Approach*. pp. 1–606, 2007. doi: 10.1007/978-0-387-36795-8.

[15] H. E. N. Tilahun, “Federal Democratic Republic of Ethiopia Labour Market Intelligence (LMI) in Ethiopia,” 2022.

[16] D. T. Mortensen, “Markets with search friction and the DMP model,” *Nobel Lect. Econ. Sci. 2006 - 2010*, pp. 307–331, 2014, doi: 10.1142/9789814635585_0005.

[17] R. L. Morgan, "Job matching: Development and evaluation of a web-based instrument to assess degree of match among employment preferences," Journal of Vocational Rehabilitation, vol. 29, pp. 29–38, 2008.

[18] S. A. Abera, "Context-Aware Personalized Job Recommendation," Master's thesis, Dept. Comput. Sci., Addis Ababa Univ., Addis Ababa, Ethiopia, 2020.

[19] M. Diaby, J. Recommandation, and N. Social, “Methods for Job Recommandation on Social Networks Mamadou Diaby To cite this version : HAL Id : tel-01493679 MÉTHODES POUR LA RECOMMANDATION D ’ OFFRES D ’ EMPLOI,” 2017.

[20] L. M. Kureková, M. Beblavý, and A. Thum-Thysen, “Using online vacancies and web surveys to analyse the labour market: a methodological inquiry,” *IZA J. Labor Econ.*, vol. 4, no. 1, pp. 0–20, 2015, doi: 10.1186/s40172-015-0034-4.

[21] S. A. Alwidian, H. A. Bani-Salameh, and A. N. Alslaity, “Text data mining: A proposed framework and future perspectives,” *Int. J. Bus. Inf. Syst.*, vol. 18, no. 2, pp. 127–140, 2015, doi: 10.1504/IJBIS.2015.067261.

[22] S. Choudhary, S. Koul, S. Mishra, A. Thakur, and R. Jain, “Collaborative job prediction based on Naïve Bayes Classifier using python platform,” *2016 Int. Conf. Comput. Syst. Inf. Technol. Sustain. Solut. CSITSS 2016*, no. November, pp. 302–306, 2016, doi: 10.1109/CSITSS.2016.7779375.

[23] R. Le, T. Zhang, W. Hu, D. Zhao, Y. Song, and R. Yan, “Towards effective and interpretable person-job fitting,” *Int. Conf. Inf. Knowl. Manag. Proc.*, no. November, pp. 1883–1892, 2019, doi: 10.1145/3357384.3357949.

[24] H. Pérez-Rosés, F. Sebé, and J. Maria Ribó, “Endorsement deduction and ranking in social networks,” *Comput. Commun.*, vol. 73, no. July, pp. 200–210, 2016, doi: 10.1016/j.comcom.2015.08.018.

[25] D. Lee, M. Kim, and I. Na, “Artificial intelligence based career matching,” *J. Intell. Fuzzy Syst.*, vol. 35, no. 6, pp. 6061–6070, 2018, doi: 10.3233/JIFS-169846.

[26] S. Kitichalermkiat, V. Singh, S. Chaowiang, S. Tangprasert, P. Chiraprawattrakun, and N. Leeprechanon, “Job Matching using Artificial Intelligence,” *Int. J. Soft Comput. Artif. Intell. Appl.*, vol. 11, no. 3, pp. 1–12, Aug. 2022, doi: 10.5121/ijscai.2021.11301.

[27] The Federal Democratic Republic of Ethiopia, Ministry of Labour and Skills, "http://www.mols.gov.et/project," 2024.

[28] R. N. Landers, R. C. Brusso, K. J. Cavanaugh, and A. B. Collmus, “A primer on theory-driven web scraping: Automatic extraction of big data from the internet for use in psychological research,” *Psychol. Methods*, vol. 21, no. 4, pp. 475–492, 2016, doi: 10.1037/met0000081.

[29] “Regions of Ethiopia - Wikipedia.” [Online]. Available: https://en.wikipedia.org/wiki/Regions_of_Ethiopia