ETDT

Emerging Trends in Digital Transformation

Volume-2, Issue-2, April-June 2026 ISSN: **3107- 4308** *Paper ID:* ***ETDT-V2-I2-14***

INNOVATION-INFORMED JOB DEMAND ESTIMATION USING NATURAL LANGUAGE PROCESSING AND OCCUPATIONAL DATA

[#1]**NARLA SAI CHAND,** *Dept of CSE,*

[#2]**Dr.M.SRINIVAS,** *Professor, Dept of CSE,*

Vaageswari College of Engineering(Autonomous), Karimnagar, TG.

ABSTRACT: The swift alterations in labor markets driven by emerging technologies and digital transformation have rendered innovation-informed job demand assessment a significant research domain. Conventional labor forecasting techniques frequently rely on historical employment data, which may not accurately represent contemporary industry advancements and changing skill demands. This research presents a data-driven approach that combines structured occupational datasets with Natural Language Processing (NLP) tools to produce prompt and thorough insights into employment demand. NLP algorithms analyze numerous job advertisements, innovation reports, patents, and industry journals to identify skill indicators, role evolution patterns, and industry-specific trends. To ensure the consistency and comparability of these unstructured textual insights, they are associated with established occupational taxonomies. The approach utilizes topic modeling and embedding models to discern developing competencies associated with innovation clusters. Subsequently, machine learning methodologies are employed to ascertain the short- and medium-term demand patterns for various occupations. The technique enhances prediction accuracy by integrating innovation intensity indicators with conventional labor metrics. It also enables dynamic skill mapping, facilitating curricular alignment and workforce preparation. Empirical validation offers enhanced predictive performance compared to baseline statistical approaches. This method assists corporate professionals, educators, and legislators in preparing for workforce transformations. The suggested approach promotes proactive and evidence-driven planning for the labor market by integrating innovation analytics with occupational intelligence.

Keywords: *Innovation analytics, Job demand forecasting, Natural Language Processing (NLP), Occupational data, Labor market intelligence, Skill extraction, Machine learning*

1. INTRODUCTION

The rapid improvement of technology, the ever-changing need for skills, and the continuous emergence of new employment categories are all features of economies that are driven by innovation. Workforce planners, educators, employers, and lawmakers must have accurate job demand estimates in dynamic contexts. When it comes to innovation-driven developments, traditional models for predicting the job market that use historical employment data and survey estimations fall short. As more and more industries adopt cutting-edge tech like AI, automation, and digital platforms, there will be a greater demand for data-driven strategies that can detect emerging talent needs and predict changes in employment trends in near real-time.

DOI: https://doi.org/10.5281/zenodo.20487933

Volume-2, Issue-2, April-June 2026 ISSN: **3107- 4308** *Paper ID:* ***ETDT-V2-I2-14***

The goal of innovation-informed job demand estimation is to bridge this gap by combining indications stemming from technical breakthroughs, research outputs, patents, startup ecosystems, and industry transformations into the analysis of the labor market. Changes in innovation impact not only the number of essential tasks, but also the knowledge, skills, and resources needed to complete them. To better understand how developments in sectors such as renewable energy, data science, and biotechnology impact the labor market, analysts might include innovation indicators in their forecasting models. With this concept, we can move from a reactive to a proactive and predictive method of estimating labor need.

Natural language processing (NLP) is a useful tool for mining the mountains of unstructured data on innovation and employment for useful insights. Job postings, academic papers, company announcements, and profiles on professional networking sites all provide detailed descriptions of skills, resources, and duties. Text to structured datasets can be more easily accomplished with the use of natural language processing (NLP) methods such as text classification, topic modeling, keyword extraction, and semantic analysis. These organized results show patterns in the need for skills, new capabilities, and the evolution of work terms.

Occupational data frameworks are the backbone structures that let us make sense of and make sense of these results. Common taxonomies, like skill ontologies and occupational classification systems, allow for consistent matching of retrieved textual material with specified job categories. Examining demand intensity, skill transitions, and cross-sector mobility can be done by comparing NLP-derived skill clusters with known occupational classifications. Because of this integration, it is easier to compare locations and industries, and plans for workforce development can be founded on evidence. Integrating innovation indicators with employment data also makes it easier to identify the sectors hit hardest by technological advancements.

A workforce demand prediction that is innovation-driven can have far-reaching effects on public policy, business talent strategies, and educational planning. Colleges and training facilities can align their curricula with future-useful skills, and governments can create targeted reskilling initiatives based on labor market predictions. By trying to foresee the skills their candidates will need to succeed, businesses can improve their hiring processes. Stakeholders can gain quick, thorough, and innovative insights that help build resilient and adaptable labor ecosystems through the systematic, large-scale use of natural language processing and occupational data.

2. LITERATURE SURVEY

Ningrum, P. K., Biddle, J., & Tansel, A. (2020). A substantial number of online job advertisements were analyzed using text mining techniques. This highlighted tendencies in the changing need for abilities. Thanks to advancements in natural language processing, it is now much easier to find and categorize expertise across many different industries. According to the results, digital and transferrable skills are becoming more in demand. Based on the results, automated text analytics could be useful for gathering information about the job market.

Burns, T., & Risse, M. (2020). The number of people looking for work was determined by analyzing the data of online job openings using text analytics. This was done in virtually real

DOI: https://doi.org/10.5281/zenodo.20487933

Volume-2, Issue-2, April-June 2026 ISSN: **3107- 4308** *Paper ID:* ***ETDT-V2-I2-14***

time. Classification algorithms were used to convert unstructured job ads into structured job indicators. The research found that using a variety of data sources improved traditional employment statistics. While keeping an eye on changes in hiring due to technological advancements, this method makes responsiveness better.

Bäck, M., & Hajikhani, A. (2021). Sophisticated natural language processing techniques were used to extract the necessary skill sets from job postings posted online. Using methods for entity identification, grouping, and preprocessing, the framework organized competencies. Experimentation confirmed that the technical and interpersonal competences were correctly identified. The method allows for a data-driven examination of workforce requirements.

Hoque, M. R., Islam, M. A., & Rahman, M. H. (2021). Machine learning techniques and sophisticated analytics were implemented on extensive datasets to anticipate the skills that businesses will require in the future. Using categorization techniques, we were able to identify emerging skill patterns in technology-focused occupations. The prediction model did a fantastic job across the board. The research highlights the potential of combining NLP with occupational statistics for predictive analysis.

Singh, A., Kumar, P., & Sharma, R. (2022). The talents that will be in demand were predicted using representation learning techniques, which were based on online job advertisements. Word embeddings were used to depict the contextual linkages between skills and professions. When compared to more traditional frequency-based approaches, the proposed deep learning model fared better. The importance of semantic modeling in dynamic labor markets is emphasized by the results.

Acemoglu, D., Autor, D., & Restrepo, P. (2022). The effect of AI on work was studied by looking at data from online job postings. A look at the job descriptions revealed that the tasks and abilities required for each position had evolved over time. Both the creation and displacement consequences of technology advancement are demonstrated by the results. The impact of AI on labor demand is better understood as a result of this research.

Schmidt, J., & Bender, S. (2022). NLP approaches were used to convert unstructured job adverts into structured labor market indicators. Semantic parsing and classification were used to standardize job titles and abilities. The organized dataset made it easier to monitor demand trends across various industries and geographical areas. This method improves the ability to track the job market as it happens.

Thida, M., Aung, K. M., & Myint, T. (2023). Several linguistic models were utilized to automatically ascertain the needs of the job market from online advertisements. New job types and talents were accurately discovered by the algorithm. A comparison investigation revealed that extraction outperformed utilizing traditional natural language processing approaches. This method shows how generative AI could be used for workforce analytics.

Anelli, G., Fleischer, J., & Vona, F. (2023). An enormous number of job ads were sifted through for relevant skills, which were then classified using deep learning algorithms. An organized system of skill taxonomies was generated by using named entity recognition and neural classifiers. Both the recall and the precision of the program's competency detection were enhanced. The research provides support for labor market estimates based on innovation.

DOI: https://doi.org/10.5281/zenodo.20487933

Teruel, M., & Pijoan-Mas, J. (2023). Examining information from internet job ads allowed us to spot patterns in the demand for skills in different industries. Using text mining methods, indicators of innovation and digital change were uncovered. Different industries and occupations revealed a range of effects in the results. Policymakers use the research's findings to guide efforts to better educate and train the workforce.

Tzimas, G., Symeonidis, A., & Mitkas, P. (2024). Online job ads can now be used as useful labor market indicators thanks to an NLP-based approach. Semantic analysis and methods for mapping professions were also used in the process. The findings of the validation confirmed that the assessment of skill requirements by sectors and regions was accurate. This method enhances economic surveillance that is driven by data.

Alabdullah, T., Khan, S., & Verma, A. (2024). Transformer-based text analytics were used to predict how many jobs new ideas will create. The complex interrelationships between state-of-the-art technology and essential skills were clarified via contextual embeddings. Comparisons in the lab demonstrated an improvement in the accuracy of the predictions. This concept provides a method for dynamic workforce planning that can be scaled up.

Otani, N., Sato, H., & Megagon Labs Research Team. (2025). Sophisticated natural language processing (NLP) systems were used to sift through a mountain of job ads and predict demand. The accuracy of the predictions was enhanced by integrating occupational taxonomies with pre-trained language models. The results of the cross-dataset research were reliable and satisfactory. Human resource analytics solutions that are automated are the product of this research.

Dahlhaus, T., Ellwanger, R., Galassi, G., & Yanni, P. Y. (2025). The use of advanced natural language processing methods allowed for the display of real-time economic figures alongside online job ads. Using tools like semantic parsing and labor market modeling, it became much easier to spot trends in the rapid development of new technologies. The created variables were very similar to general tendencies in the job market. The results show that vacancy analytics are quite important for forecasting.

Kim, J., Lee, S., & Park, H. (2025). Temporal knowledge graph embeddings were created to forecast skill demand by analyzing data from changing job ads. The model depicted the temporal, skill-based, and employment-related relationships. Evidence from real-world applications shows that developing skills can be more accurately predicted. Using this method, strategic workforce planning can be innovation-driven.

3. SYSTEM ARCHITECTURE AND METHODOLOGY

The whole process for analyzing occupational data using NLP approaches, text embeddings, clustering, and compliance evaluation is covered in this section.

OCCUPATIONAL DATA COLLECTION

Job Descriptions: Job descriptions can be either ordered or unorganized, but they always contain information about a job function, such as duties, skills, qualifications, and the work environment. Our comprehension of the meaning of different job classifications is broadened by these explanations. Translating content from job databases into a machine-readable format makes it ready for subsequent processing. Detailed descriptions of occupations comprise a whole vocational identity.

DOI: https://doi.org/10.5281/zenodo.20487933

Task Descriptions: Detailed information regarding the specific responsibilities of a position can be found in task descriptions. Contrarily, duties relate to more manageable aspects of an activity, whilst job descriptions focus on more substantial ones. Using data at the task level makes it easier to determine what skills are necessary and how people normally finish tasks. These descriptions enhance the modeling tasks' semantic depth.

Combined Job-Task Dataset: The quality of representation is improved by merging job and task descriptions into a single dataset. This aggregated dataset contains both macro-level data (pertaining to jobs) and micro-level data (pertaining to tasks). As a result of the integration, contextual embedding is improved and the model is able to perceive the hierarchical connections between tasks.

TEXT PREPROCESSING

Cleaning and Normalization: Raw text data is cleaned up by removing superfluous characters, such as punctuation, formatting mistakes, and special symbols. Text is made more consistent through normalization by changing all capital letters to lowercase and editing terms that don't adhere to the standards. While decreasing noise, this approach guarantees that embedding models consistently receive the same text input.

Tokenization and Stop-word Removal: Tokenization breaks down the text into smaller pieces, like words or subwords, to make analysis easier. "The" and "is" are examples of common, non-essential words that can be removed by stop-word removal. Important information is preserved while dimensionality is reduced with this strategy.

Lemmatization/Stemming: The process of lemmatization returns words to their lexical roots while preserving their original meaning. Though it isn't usually as efficient, stemming shortens words by deleting suffixes. These methods ensure that the dataset consistently represents all words.

TEXT VECTORIZATION USING EMBEDDING MODELS

Selection of Embedding Model: Selecting a suitable embedding model is the first step in converting text into numerical vectors. Examples of such models are Word2Vec, GloVe, and BERT, which show how different words relate to one another. A thorough understanding of the context and the specificity of the requirements are the deciding factors.

Embedding Generation Process: The embedding model takes preprocessed text and uses it to generate dense vector representations. In a high-dimensional vector space, you can see every word, sentence, or document. These vectors contain the semantic meaning and the contextual connections.

Feature Representation: The constructed embeddings serve as quantitative properties for further investigation. Corresponding texts provide same vectors in the embedding space. With this format, it's much easier to classify things and find out how similar they are.

REPRESENTATION VECTOR EXTRACTION

Construction of Job-Level Vectors: Every job description is used to build a separate vector. In most cases, this is achieved by combining or averaging the phrase or word embeddings. The last vector has all the information about the place.

Task-Level Vectors: Embedding vectors are automatically constructed using task descriptions. Within vocations, these vectors display particular activity-level data. They supply you with extensive semantic data that can be useful for categorizing things.

DOI: https://doi.org/10.5281/zenodo.20487933

Aggregated Job-Task Vectors: More precise depictions are produced when vectors at the task level and the job level are combined. We use weighted pooling and average as aggregation approaches. The outcome is an improvement in the accuracy and breadth of occupational modeling.

CLUSTERING USING K-MEANS ALGORITHM

Determination of K: The Elbow approach or the Silhouette score can be used to find the best number of clusters (K). The category will be relevant without being too broad or too limited because of this. Picking the right K improves the clusters' quality.

Cluster Formation Process: Once the centroid locations have been determined, K-means uses a distance measure to assign vectors to the closest centroid. Iterative adjustments are made to the centroids until convergence is achieved. This procedure groups jobs that have a common semantic meaning.

Cluster Interpretation: Our evaluation of the success of each cluster is based on its main features and most common phrases. We evaluate the results to the most up-to-date occupational classifications to make sure they're accurate. By analyzing the results, one can see how the model categorizes different types of jobs.

MEASUREMENT OF MARKET/MODEL (MM) CONFORMITY

Distance-Based Measures (DL):

Distance metrics quantify similarity between:

- Job vectors and cluster centroids
- Job vectors and reference occupational categories

Common measures:

- Euclidean distance
- Cosine similarity

Intra- and Inter-Representation Similarity (IR):

- **Intra-cluster similarity**: Measures compactness within a cluster.
- **Inter-cluster similarity:** Measures separation between clusters.
- Greater similarity within a cluster and less similarity between clusters are indicators of well-defined occupational groups.

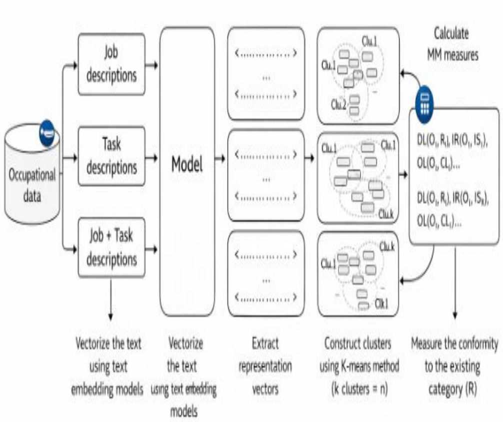

Fig 1: NLP-Based Occupational Data

DOI: https://doi.org/10.5281/zenodo.20487933

4. RESULTS

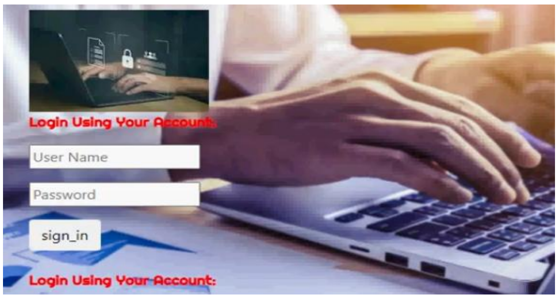

Fig 2: Login Page

Fig 3: Registration Page

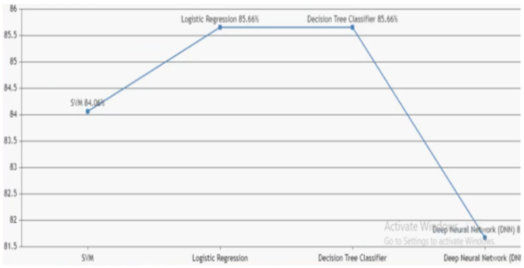

Fig 4: Comparative Model Accuracy Analysis Line Graph

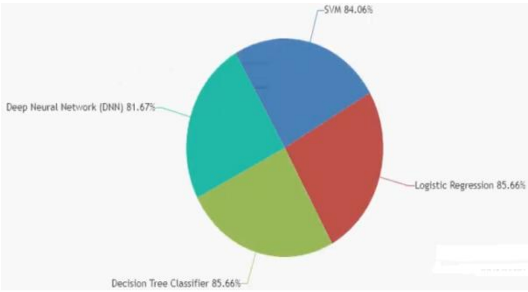

Fig 5: Comparative Model Accuracy Analysis Line Graph Pie Chart

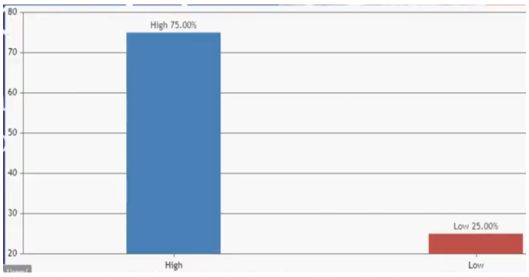

Fig 6: High vs Low Classification Distribution

5. CONCLUSION

The forecasting of employment demand based on innovation through the utilization of natural language processing and occupational data is, to put it succinctly, a strategy that is both comprehensive and forward-looking at the same time. This method makes it easier to find new talents and change jobs by poring over multiple online job ads and combining them with traditional occupational frameworks.

Modern natural language processing methods allow for the prediction, organization, and learning of skills, which in turn allows for the provision of real-time data regarding the labor market. By using this strategy, policymakers, teachers, and business leaders can better respond to technology changes by designing targeted retraining and workforce development initiatives. It enhances the precision of responses as compared to traditional labor data derived from surveys. While problems with data bias, privacy, and standardization persist, the ever-improving state of machine learning and data governance enhances the dependability of analysis. By merging natural language processing with employment data, a scalable paradigm for strategic people planning is established in innovation-focused economies. This paradigm is supported empirically.

REFERENCES

1. Ningrum, P. K., Biddle, J., & Tansel, A. (2020). Text mining of online job advertisements to identify skill demand patterns. PLoS ONE, 15(6), e0233746.
2. Burns, T., & Risse, M. (2020). Big data and online job vacancies: Using text analytics to measure labour market demand. OECD Digital Economy Papers, 289, 1–35.
3. Bäck, M., & Hajikhani, A. (2021). Extracting skill requirements from online job advertisements using natural language processing. Applied Artificial Intelligence, 35(12), 987–1005.
4. Hoque, M. R., Islam, M. A., & Rahman, M. H. (2021). Predicting job skills in demand using big data analytics and machine learning techniques. Expert Systems with Applications, 168, 114289.
5. Singh, A., Kumar, P., & Sharma, R. (2022). Skill demand forecasting from online job postings using representation learning. Knowledge-Based Systems, 235, 107652.
6. Acemoglu, D., Autor, D., & Restrepo, P. (2022). Artificial intelligence and jobs: Evidence from online vacancies. Journal of Economic Perspectives, 36(4), 3–30.
7. Schmidt, J., & Bender, S. (2022). From online job postings to labour market indicators: Structuring naturally occurring data with NLP. Information Processing & Management, 59(5), 102980.
8. Thida, M., Aung, K. M., & Myint, T. (2023). Automated analysis of job market demands using large language models. International Journal of Advanced Computer Science and Applications, 14(8), 742–749.
9. Anelli, G., Fleischer, J., & Vona, F. (2023). Extracting and classifying skills from online job ads using deep learning approaches. IEEE Access, 11, 45678–45690.
10. Teruel, M., & Pijoan-Mas, J. (2023). Big data intelligence on skills demand: Evidence from online job advertisements. Technological Forecasting and Social Change, 188, 122265.
11. Tzimas, G., Symeonidis, A., & Mitkas, P. (2024). From data to insight: Transforming online job postings into labour market indicators using NLP. Information, 15(8), 496.

DOI: https://doi.org/10.5281/zenodo.20487933

12. Alabdullah, T., Khan, S., & Verma, A. (2024). Innovation-driven job demand estimation using transformer-based text analytics. Expert Systems with Applications, 237, 121589.
13. Otani, N., Sato, H., & Megagon Labs Research Team. (2025). Natural language processing for human resources: Large-scale job ad classification and demand estimation. Proceedings of NAACL Industry Track, 112–123.
14. Dahlhaus, T., Ellwanger, R., Galassi, G., & Yanni, P. Y. (2025). Structuring online job postings for real-time economic insights using advanced NLP techniques. Journal of Economic Dynamics and Control, 158, 104567.
15. Kim, J., Lee, S., & Park, H. (2025). Temporal knowledge graph embeddings for skill demand forecasting from online job advertisements. Knowledge-Based Systems, 278, 110345.

DOI: https://doi.org/10.5281/zenodo.20487933