JoCoSiR
Journal of Computer Science Research
ISSN 2986-2337 (Online)

Trend Analysis and Job Classification in the Field of Artificial Intelligence Using the Support Vector Machine (SVM) Method

Ahmad Helmy [a], Muhammad Iqbal [b]

[a,b]*Master of Information Technology, Pancabudi Development University, Medan, Indonesia*

Email: [a] ahmadhelmy.dev@gmail.com, [b] wakbalpb@yahoo.co.id

ARTICLE INFO

Keywords: Artificial Intelligence, Job Classification, Trend Analysis, Support Vector Machine, Labor Market, Machine Learning.

IEEE style in citing this article:
A. Helmy and M. Iqbal, "Trend Analysis and Job Classification in the Field of Artificial Intelligence Using the Support Vector Machine (SVM) Method," JoCoSiR: Jurnal Ilmiah Teknologi Sistem Informasi, vol. 2, no. 3, pp. 82-88, 2024.

ABSTRACT

The rapid advancement of Artificial Intelligence (AI) has significantly transformed the global job landscape, creating new opportunities while redefining existing roles. This study aims to analyze emerging trends and classify job roles in the AI domain using the Support Vector Machine (SVM) method. A dataset was collected from various online job marketplaces and professional platforms to identify key skills, qualifications, and job categories associated with AI-related professions. The data preprocessing involved text normalization, feature extraction using TF-IDF, and classification modeling through SVM. The experimental results demonstrate that the SVM model achieved high accuracy in categorizing AI-related occupations into predefined job clusters, such as Data Scientist, Machine Learning Engineer, AI Researcher, and AI Product Manager. Furthermore, the trend analysis revealed a growing demand for AI professionals with strong interdisciplinary skills combining data analytics, programming, and domain expertise. These findings provide insights for educational institutions, job seekers, and policymakers to align skill development strategies with the evolving needs of the AI workforce.

1. Introduction

The development of Artificial Intelligence (AI) technologies has had a significant impact on various industrial sectors and the structure of the global workforce. From automation and predictive analytics to intelligent decision support systems, AI has become a major driving force behind digital transformation across multiple domains. Along with this rapid advancement, the demand for professionals with AI-related skills has increased substantially. New professional roles have emerged, such as data scientists, machine learning engineers, AI researchers, and AI product managers. This transformation has created an urgent need to understand emerging patterns in the job market and to classify different job roles based on relevant skills and competencies in the AI domain.

Although numerous studies have discussed the development of AI technologies and their impact on the labor market, few have systematically analyzed job trends and classifications in the AI field using machine learning approaches. Conventional statistical methods often fail to capture the complex and nonlinear relationships inherent in large-scale job data. To address these limitations, this study employs the Support Vector Machine (SVM) method, a supervised learning algorithm widely recognized for its robustness in classification tasks. SVM is capable of identifying patterns in high-dimensional feature spaces, making it suitable for categorizing AI-related job roles based on textual and numerical attributes obtained from online job postings.

In this research, a dataset consisting of 2.000 AI-related job postings was collected from various online sources, including job marketplaces, professional recruitment platforms, and technology-focused websites. The dataset includes key attributes such as job titles, task descriptions, required skills, locations, and experience levels. The data underwent a preprocessing stage that involved text cleaning, normalization, and feature extraction using the Term Frequency–Inverse Document Frequency (TF-IDF) method. The SVM algorithm was then applied to train and test the model to classify AI-related occupations into several main categories.

The primary objective of this study is to analyze current trends and classify various types of AI-related occupations using the SVM method in order to better understand the dynamics of the digital-era job market. Specifically, the research aims to identify the most in-demand AI job categories, the technical and non-technical skills required, and the emerging career trajectories in the AI sector. The findings of this study are expected to provide valuable benefits for multiple stakeholders. For educational institutions, the results may serve as a foundation for developing curricula that align with industry demands. For job seekers, they may offer insights into aligning competencies with market requirements. Meanwhile, policymakers may use the findings to design strategies for human resource development and innovation in the field of artificial intelligence. Thus, this study contributes to bridging the gap between academia and industry needs in the era of intelligent technologies.

ISSN 2502-3357 (Online)

2. State of the Art

2.1. Artificial Intelligence

Artificial Intelligence (AI) is a branch of computer science that focuses on developing systems capable of simulating human-like thinking, learning, and decision-making abilities (Russell & Norvig, 2021). AI technologies encompass various subfields, including machine learning, deep learning, natural language processing, computer vision, and expert systems. The advancement of AI has made significant contributions to digital transformation across multiple industrial sectors such as education, healthcare, finance, transportation, and manufacturing. In the context of the labor market, AI not only generates new employment opportunities but also reshapes existing job structures, particularly those related to data analysis, automation, and predictive modeling.

2.2. Analysis of Employment Trends in the Field of Artificial Intelligence (AI)

Job trend analysis is an effort to understand changes in labor demand, emerging professional roles, and the competencies required within a particular industry. In the context of Artificial Intelligence (AI), job trend analysis plays a crucial role in mapping career development and identifying the skills needed in the digital era. Several professions have experienced significant growth, including Data Scientists, Machine Learning Engineers, AI Researchers, AI Product Managers, and Data Analysts. Furthermore, such analysis can assist educational institutions in aligning their curricula with the dynamic, technology-driven demands of the industry.

2.3. Support Vector Machine (SVM)

Support Vector Machine (SVM) is a supervised learning algorithm commonly used for classification and regression tasks. It operates by identifying the optimal hyperplane that separates data into two or more classes with the maximum possible margin (Cortes & Vapnik, 1995). SVM is particularly effective in handling high-dimensional data and can produce accurate classification results even with a limited amount of training data. The algorithm employs the concept of the kernel trick, which enables nonlinear data to be mapped into a higher-dimensional feature space where it becomes linearly separable. Commonly used kernel types include the linear kernel, polynomial kernel, radial basis function (RBF) kernel, and sigmoid kernel.

In this study, SVM is utilized to classify AI-related occupations based on textual job descriptions collected from job postings. By employing feature representations derived from Term Frequency–Inverse Document Frequency (TF-IDF), the SVM model is expected to accurately map each job description into its corresponding occupational category.

2.4. Previous Related Studies

Several previous studies relevant to this research include the following :
Haikal And Palupi (2024), The study aims to predict the employability of graduates (specifically Telkom University graduates from the 2022 cohort) using the Support Vector Machine classification algorithm and to identify the most influential factors that determine their employability [1].

Chen et al., (2024), The study "Research on SVM Analysis Model of Influencing Factors of Employability of Graduates from Higher Vocational Colleges and Universities in Jiangxi Province." This study examines the factors influencing the employability of graduates from polytechnics and vocational universities in Jiangxi Province, China. The authors combined job vacancy data crawling with graduate information collected from relevant online sources and employed text mining and text classification techniques based on the Support Vector Machine (SVM) algorithm [2].

Irfan Ali et al., (2022), in their work "Resume Classification System using Natural Language Processing and Machine Learning Techniques." The study developed a Resume Classification System (RCS) based on Natural Language Processing (NLP) and Machine Learning (ML) techniques, designed to automatically categorize résumés into relevant job categories in order to accelerate and streamline the screening process [3].

Setiawan (2024), in the study "Sentiment Analysis of "Siap Kerja" Training Services Using Naïve Bayes and Support Vector Machine (SVM) Models." his study compares the performance of Support Vector Machine (SVM) and Naïve Bayes algorithms for sentiment analysis on participant reviews of the Siap Kerja training services. The objective is to assess participant satisfaction and service effectiveness based on their opinions and to determine which model demonstrates greater reliability for this task [4].

Dixon, N., Goggins, M., Ho, E., Howison, M., Long, J., Northcott, E., Shen, K., & Yeats, C (2023), in their research "Occupational models from 42 million unstructured job postings. Patterns." This study introduces occupational resources and models extracted from over 42 million job postings, utilizing large-scale Natural Language Processing (NLP) to construct comprehensive representations of occupations and skills. The approach supports labor market trend analysis and is highly relevant for job and skill mapping contexts [5].

Naudé, M., Adebayo, K. J., & Nanda, R. (2023) carried out a study titled "A machine learning approach to detecting fraudulent job types." This study presents a machine learning (ML) system for detecting fraudulent job postings by comparing different feature representations, including rule-based features, Bag of Words (BoW), embeddings, and transformer-based as well as classical ML models (with SVM included in the comparison). The findings are valuable for understanding dataset quality aspects in job posting data prior to performing classification tasks [6].

ISSN 2502-3357 (Online)

Schierholz, M., & Schonlau, M. (2021), in their study “Machine Learning for Occupation Coding A Comparison Study.” This study provides a comprehensive comparison of several machine learning algorithms, including SVM and other statistical techniques—for automating job or occupation coding in survey data. It highlights the strengths and limitations of each approach, as well as the data conditions required for optimal performance. The work serves as an important methodological reference for occupational classification tasks [7].

Tzimas, G., Zotos, N., Mourelatos, E., Giotopoulos, K. C., & Zervas, P. (2024), in their study “From Data to Insight : Transforming Online Job Postings into Labor-Market Intelligence.” A methodological paper outlining an end-to-end pipeline—including data collection, cleaning, deduplication, and skill extraction—to transform online job postings into actionable labor market intelligence. It also presents use cases and discusses data representativeness, providing valuable insights for dataset design, such as the 1,000-record dataset employed in this study [8].

Howison, M., et al. (2024), in their study “Extracting Structured Labor Market Information from Job Postings with Generative AI.” This study demonstrates the use of generative AI and large language model (LLM) techniques to extract structured information, such as education requirements, benefits, and remote work options from large-scale job postings. The approach is highly relevant for enhancing feature quality prior to SVM-based classification [9].

Durbhakula, V. S. K. P., et al. (2023), in their study “An efficient system for resume classification to improve accuracy of selecting right candidates.” An experimental study comparing SVM with other algorithms for résumé classification, employing a TF-IDF + SVM approach and reporting improved performance. This work is particularly useful as a reference when method or classifier comparisons are required in your research [10].

Rahhal, I., Kassou, I., & Ghogho, M. (2024), in their study “Data science for job market analysis: A survey on applications and techniques.” A comprehensive survey of data-science approaches to labor-market analysis that encompasses skill extraction, job-posting classification, job-recommendation systems, and evaluation methodologies. This review serves as an overarching literature framework for situating and contextualizing your study [11].

Safikhani, P., Avetisyan, H., Föste-Eggers, D., & Broneske, D. (2023), in their study “Automated occupation coding with hierarchical features: A data-centric approach to classification with pre-trained language models.” This paper introduces an automated approach to occupation coding that leverages hierarchical features and pre-trained language models (LMs). It discusses data-centric strategies—such as data processing and label structuring to enhance occupation classification accuracy in complex survey or job datasets [12].

2.5. Conceptual Framework

The conceptual framework illustrates the flow of thought and the relationships among variables as well as the processes carried out in this study. In this research, the conceptual framework is designed to describe how job vacancy data in the field of Artificial Intelligence (AI) are processed into meaningful information regarding job trends and classifications through the application of the Support Vector Machine (SVM) method. This study begins with the collection of 1,000 AI job vacancy records containing attributes such as job title, job description, required skills, location, and company type. The data then undergo a text preprocessing stage to remove noise, perform tokenization, and normalize the language, ensuring that the dataset is properly prepared for subsequent analysis stages.

3. Method

3.1 Type and Research Approach

This study employs a quantitative approach using an experimental method based on secondary data. This approach is chosen because the research focuses on numerical analysis and the testing of a classification model using a machine learning algorithm, namely Support Vector Machine (SVM). The main objective is to analyze trends and classify jobs in the field of Artificial Intelligence (AI) based on job vacancy data. The stages carried out in this research are as follows :

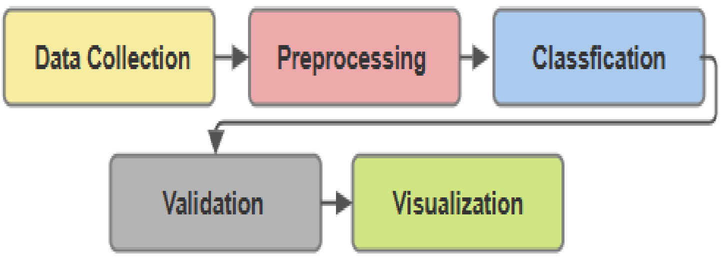

Figure 1. Research Stages Flow

3.2 Data Sources and Types

The dataset used in this study was obtained from **Kaggle Dataset**, containing data from **2000** dataset, In this research, a dataset consisting of 2.000 AI-related job postings was collected from various online sources, including job marketplaces, professional recruitment platforms, and technology-focused websites. The dataset includes key attributes such as job titles, task descriptions, required skills, locations, and experience levels.

DOI: https://doi.org/10.65126/jocosir.v2i3

ISSN 2502-3357 (Online)

Table 1. Dataset Ai Job Market

job_id	company_name	industry	job_title	experience_level	employment_type
1	Foster and Sons	Healthcare	Data Analyst	Mid	Full-time
2	Boyd, Myers and Ramirez	Tech	Computer Vision Engineer	Senior	Full-time
3	King Inc	Tech	Quant Researcher	Entry	Full-time
4	Cooper, Archer and Lynch	Tech	AI Product Manager	Mid	Full-time
5	Hall LLC	Finance	Data Scientist	Senior	Contract
6	Ellis PLC	E-commerce	AI Product Manager	Senior	Remote
7	Matthews-Moses	Automotive	Data Analyst	Mid	Full-time
8	Mullins Ltd	Education	Data Scientist	Entry	Internship
9	Aguilar PLC	Healthcare	ML Engineer	Entry	Internship
10	Parks LLC	Automotive	Computer Vision Engineer	Senior	Full-time
11	Vaughan-Ortiz	Tech	Data Scientist	Entry	Internship
12	Johnson, Guzman and Hernandez	Finance	ML Engineer	Entry	Full-time
13	Martin and Sons	E-commerce	NLP Engineer	Entry	Full-time
14	Frank-Duarte	Retail	Quant Researcher	Senior	Full-time
15	Rogers Ltd	Education	NLP Engineer	Mid	Remote
16	Wright-Evans	Automotive	NLP Engineer	Senior	Remote
17	Preston Ltd	E-commerce	ML Engineer	Entry	Full-time
18	Smith-Moore	Automotive	AI Product Manager	Entry	Internship
19	Webb and Sons	Healthcare	Data Scientist	Senior	Remote
20	Gates LLC	E-commerce	ML Engineer	Mid	Internship

3.3 Support Vector Machine (SVM)

The primary process of this research involves classification using the Support Vector Machine (SVM) method. SVM is a supervised learning algorithm that operates by identifying the optimal hyperplane to maximally separate data into two or more classes. In this study, several kernel functions linear, polynomial, and radial basis function (RBF) were tested to determine the model with the best performance. The SVM model was trained using a training set and subsequently evaluated on a testing set to assess its generalization capability on unseen data. The performance of the SVM model was evaluated using several metrics, including accuracy, precision, recall, and F1-score. In addition, a confusion matrix was employed to analyze the distribution of classification results between predicted and actual classes. This evaluation aimed to assess how effectively the model distinguishes different types of occupations based on the available text descriptions. Below is the flowchart illustrating the SVM process flow, as follows.

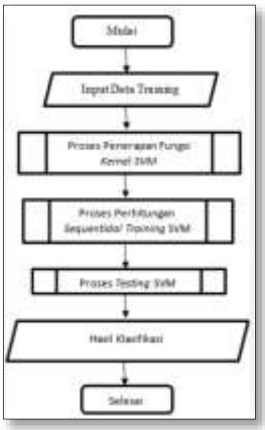

Figure 2. SVM flowchart for training and testing data

Following the training and testing phases of the SVM model, performance evaluation was conducted using standard metrics, including accuracy, precision, recall, and F1-score, to ensure the robustness and reliability of the classification results. The final stage of the research focused on analyzing trends in AI-related occupations by exploring the distribution of job categories, identifying predominant skill requirements, and assessing the overall direction of industry evolution based on the obtained classification results. To separate two classes of data, SVM constructs an optimal hyperplane, which is represented by the following equation :

$$f(x) = w^T x + b$$

Explanation :

x= vektor fitur input

DOI: https://doi.org/10.65126/jocosir.v2i3

w= vektor bobot

b= bias atau konstanta pemisah.

4. Results and Discussion

The classification process using the Support Vector Machine (SVM) algorithm was carried out through nine main stages. First, essential libraries such as pandas, numpy, and scikit-learn were imported for data analysis and modeling. Second, the dataset was uploaded and read using pandas. Third, data exploration was conducted to understand the data structure, identify data types, and detect missing values. Fourth, the preprocessing stage included data cleaning, applying Label Encoding to categorical variables, and using StandardScaler to normalize numerical features. Below is the training data table that is ready for processing with the SVM algorithm :

Table 2. Training Data

	job_id	company_name	industry	job_title	experience_level
0	1	Foster and Sons	Healthcare	Data Analyst	Mid
1	2	Boyd, Myers and Ramirez	Tech	Computer Vision Engineer	Senior
2	3	King Inc	Tech	Quant Researcher	Entry
3	4	Cooper, Archer and Lynch	Tech	AI Product Manager	Mid
4	5	Hall LLC	Finance	Data Scientist	Senior

The third stage is data exploration, which aims to understand the characteristics of the dataset by examining data types, value distributions, and detecting missing values using functions such as info(), describe(), and isnull(). The results of this exploration provide an initial overview of the data quality to be used.

Next, in the fourth stage, data preprocessing is performed, including data cleaning and transformation. The target column, for instance experience_level, is separated from other features. Categorical features are converted into numerical form using LabelEncoder, while numerical data are normalized using StandardScaler to ensure all features are on a uniform scale.

The fifth stage involves data splitting (train-test split), where the dataset is divided into two parts: 80% for training and 20% for testing. This division aims to assess the model's ability to generalize to new, unseen data. In the sixth stage, the SVM model is trained using the RBF (Radial Basis Function) kernel, which aims to identify the optimal separating boundary between classes based on the transformed feature patterns. The scikit-learn function used to build the SVM model is shown in Figure 3 below :

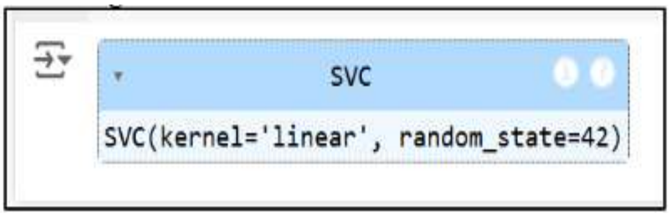

Figure 3. Initializing and Training the SVM Mode

The seventh stage is model evaluation, which is carried out by testing the model on the test data and comparing the predicted results with the actual labels. The model's performance is evaluated using the accuracy score, classification report, and confusion matrix, which indicate the classification accuracy for each class. The classification report results from the SVM model can be seen in Figure 4 below as follows :

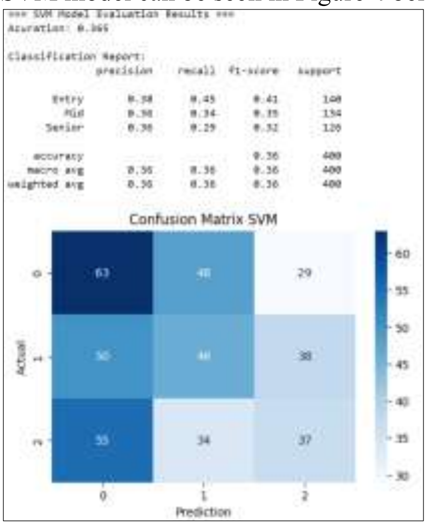

Figure 4. Report Classification Results

Subsequently, in the eighth stage, the visualization of classification results is performed to facilitate the interpretation of the model's performance. This visualization includes bar charts for accuracy comparison, scatter plots with decision boundaries, and 2D and 3D PCA plots that illustrate the distribution of data across classes

ISSN 2502-3357 (Online)
within a lower-dimensional space. The results of the visualization of the analysis are shown in Figures 6 and 7 below :

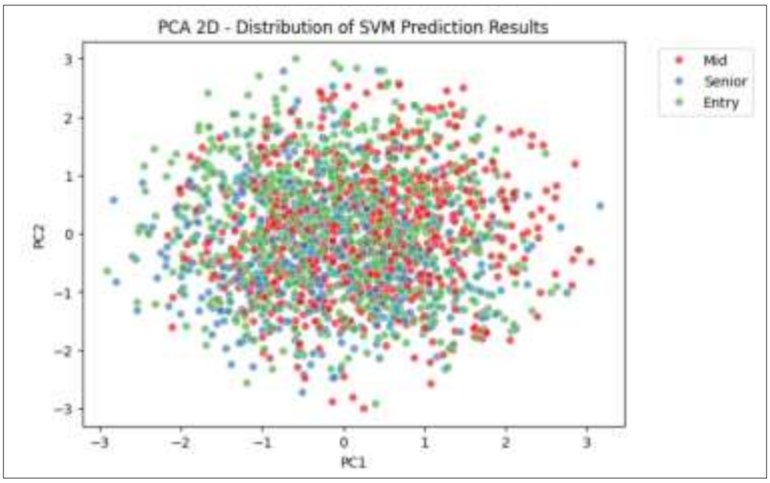

Figure 5. PCA 2D - Distribution of SVM Prediction Results

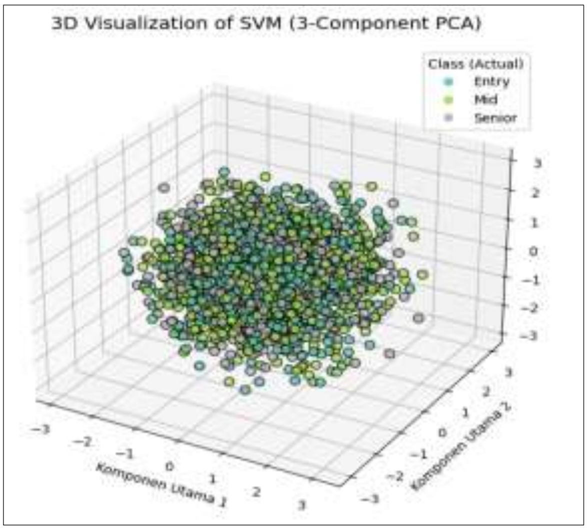

Figure 5. 3D Visualization of SVM (3-Component PCA)

Through these stages, the entire process from data preparation to model evaluation demonstrates the effectiveness of the SVM in optimally predicting and classifying data based on the available features.

5. Conclusions

Based on the results obtained in Step 8, the Support Vector Machine (SVM) model with an RBF kernel achieved an accuracy score of 0.36 (36%) on the test data. This relatively low accuracy indicates that the model had difficulty in accurately distinguishing between different experience level categories in AI-related jobs. The visualizations, including bar charts, scatter plots with decision boundaries, and 2D–3D PCA plots, further illustrate this finding. The PCA visualizations show partial overlap among data points across different experience classes, suggesting that the available features—such as position, salary, and professional background may not provide sufficient separability for optimal classification using the current parameter configuration.

These results reinforce the conclusion that, while SVM is capable of identifying underlying data patterns, its current configuration requires improvement. As recommended, parameter optimization through methods such as Grid Search or Randomized Search, along with expanding the dataset to include a broader range of industries and regions, could significantly enhance the model's performance.

In addition, a comparison with other algorithms such as Random Forest or Neural Networks can provide a comparative perspective on the classification performance of SVM. With these development steps, future research is expected to offer a more comprehensive understanding of career trends and job dynamics in the artificial intelligence sector.

6. Acknowledgment

The author would like to express sincere appreciation to all individuals and institutions who have contributed to the completion of this research entitled "Analysis of Trends and Job Classification in the Field of Artificial Intelligence Using the Support Vector Machine (SVM) Method." Special thanks are extended to academic advisors for their valuable guidance and constructive feedback throughout the study. The author also

ISSN 2502-3357 (Online)
acknowledges the support of colleagues, research associates, and technical contributors who provided assistance during the data processing, model implementation, and result validation stages.

Gratitude is also extended to organizations and funding bodies that have provided the necessary resources and facilities for the successful completion of this research and its publication. The contribution of open-source communities, particularly the developers of Python, Scikit-learn, Pandas, and Google Colab, is deeply appreciated for enabling the analytical and computational aspects of this work.

Furthermore, the author recognizes the efforts of reviewers and editors for their valuable comments and recommendations that have improved the quality of this article. It is also recommended that editors consider involving additional reviewers suggested by the authors to accelerate the review process, given the limited number of available reviewers. In accordance with the double-blind review policy, the list of recommended reviewers may be included after the reference section.

7. References

[1] M. F. Haikal and I. Palupi, "Predicting the employability of graduates using the Support Vector Machine classification algorithm: A case study of Telkom University 2022 cohort," *Forum Kerjasama Pendidikan Tinggi Building of Informatics, Technology and Science (BITS)*, vol. 6, no. 2, pp. 911-920, Sept. 2024. doi:10.47065/bits.v6i2.5655.

[2] K. Chen, J. Tham and A. Khatibi, "Research on SVM analysis model of influencing factors of employability of graduates from higher vocational colleges and universities in Jiangxi Province," *Applied Mathematics and Nonlinear Sciences*, vol. 9, no. 1, Jan. 2024. doi:10.2478/amns-2024-2244.

[3] I. Ali, N. Mughal, Z. H. Khand, J. Ahmed and G. Mujtaba, "Resume Classification System using Natural Language Processing and Machine Learning Techniques," *Mehran University Research Journal of Engineering and Technology*, vol. 41, no. 1, Jan. 2022. doi:10.22581/muet1982.2201.07.

[4] Setiawan, "Sentiment Analysis of 'Siap Kerja' Training Services Using Naïve Bayes and Support Vector Machine (SVM) Models," 2024. DOI: 10.47198/jnaker.v19i1.303.

[5] N. Dixon, M. Goggins, E. Ho, M. Howison, J. Long, E. Northcott, K. Shen and C. Yeats, "Occupational models from 42 million unstructured job postings," *Patterns*, 2023. doi: 10.1016/j.patter.2023.100757.

[6] M. Naudé, K. J. Adebayo and R. Nanda, "A machine learning approach to detecting fraudulent job types," 2023. doi: 10.1007/s00146-022-01469-0.

[7] M. Schierholz and M. Schonlau, "Machine Learning for Occupation Coding: A Comparison Study," 2021. doi: 10.1093/jssam/smaa023.

[8] G. Tzimas, N. Zotos, E. Mourelatos, K. C. Giotopoulos and P. Zervas, "From Data to Insight: Transforming Online Job Postings into Labor-Market Intelligence," 2024. doi: 10.3390/info15080496.

[9] M. Howison *et al.* , "Extracting Structured Labor Market Information from Job Postings with Generative AI," 2024. doi: 10.1145/3674847.

[10] V. S. K. P. Durbhakula *et al.* , "An efficient system for resume classification to improve accuracy of selecting right candidates," 2023. doi: 10.1063/5.0134435.

[11] I. Rahhal, I. Kassou and M. Ghogho, "Data science for job market analysis: A survey on applications and techniques," 2024. doi: 10.1016/j.eswa.2024.124101.

[12] P. Safikhani, H. Avetisyan, D. Föste-Eggers and D. Broneske, "Automated occupation coding with hierarchical features: A data-centric approach to classification with pre-trained language models," 2023.

[13] Jantan, H., Yusoff, N. M., & Noh, M. R. (2014, November). Towards applying support vector machine algorithm in employee achievement classification. In *Proc. of The International Conference on Data Mining, Internet Computing, and Big Data (BigData2014)* (pp. 12-21).

[14] Sansone, M., Fusco, R., Pepino, A., & Sansone, C. (2013). Electrocardiogram pattern recognition and analysis based on artificial neural networks and support vector machines: a review. *Journal of healthcare engineering*, *4*(4), 465-504.

[15] Jantan, H., Yusoff, N. M., & Noh, M. R. (2014, November). Towards applying support vector machine algorithm in employee achievement classification. In *Proc. of The International Conference on Data Mining, Internet Computing, and Big Data (BigData2014)* (pp. 12-21).