iJETRM
International Journal of Engineering Technology Research & Management
Published By:
https://www.ijetrm.com/

A MACHINE LEARNING BASED JOB MATCHING AND SKILL GAP DETECTION SYSTEM

Beulah J Karthikeyan
Assistant Professor, Department of Artificial Intelligence and Machine Learning,
J.B. Institute of Engineering and Technology, Hyderabad, Telangana, India

Yeerapaneni Anurag
G. Deekshith
P. Abhishek
B. Sarika
B. Jatin Raj
UG Student Department of Artificial Intelligence and Machine Learning
J.B. Institute of Engineering and Technology, Hyderabad, Telangana, India

ABSTRACT
This research proposes an innovative approach to assist individuals in aligning their skill sets with the current labor market demands by utilizing machine learning techniques for job forecasting and skill gap evaluation. The system leverages classification algorithms such as Random Forest, Naive Bayes, and Support Vector Machine (SVM) to predict suitable job roles based on the user's existing skill set. Additionally, specialized models analyze skill gaps and suggest targeted recommendations for additional skills required to secure high-demand job opportunities. The system evaluates model performance using key metrics such as accuracy, precision, recall, and F1 score, ensuring reliable predictions and robust analysis. Through interactive visualizations, users can understand their skill gaps and make informed decisions to enhance their career trajectories. The platform not only predicts potential job matches but also provides a comprehensive skill development pathway, empowering users to reduce the gap between their current capabilities and market expectations. This system is designed to increase employability by equipping users with relevant insights, enabling continuous learning and career advancement. The initiative ultimately serves as a bridge between individual aspirations and industry needs, fostering a more adaptive and skilled workforce

Keywords:
Job matching, skill gap analysis, machine learning, random forest, naive Bayes, support vector machine (SVM)

1. INTRODUCTION

In today's rapidly evolving job market, staying relevant and competitive requires continuous skill enhancement and adaptation to emerging trends. With the rise of automation, artificial intelligence, and digital transformation, traditional job roles are being redefined, and new skill sets are in high demand. However, individuals often face challenges in identifying relevant job opportunities that align with their current capabilities or understanding the skill gaps that prevent them from securing desirable positions. To address this, advanced technological solutions such as machine learning can play a pivotal role in guiding individuals toward suitable career paths and equipping them with the necessary skills to thrive in dynamic work environments. This research introduces a machine learning-based system that predicts appropriate job roles based on a user's existing skill set and identifies skill gaps that may hinder them from securing high-demand positions. The system leverages classification algorithms, including Random Forest, Naive Bayes, and Support Vector Machine (SVM), to analyse a user's skill profile and match it with suitable job roles. By utilizing historical data from job postings and industry trends, the system generates accurate predictions and recommendations tailored to individual career aspirations. A significant feature of this system is its ability to evaluate skill gaps and provide actionable insights to bridge them. Once a user's skill set is analysed, the system highlights areas where additional expertise

iJETRM
International Journal of Engineering Technology Research & Management
Published By:
https://www.ijetrm.com/

is required and suggests relevant skills that align with industry expectations. These suggestions are based on the analysis of current labor market trends and the requirements of in-demand job roles, allowing users to enhance their competencies and stay relevant in their respective fields. To ensure the effectiveness of the predictive models, the system employs rigorous evaluation metrics, including accuracy, precision, recall, and F1 score. Comprehensive visualizations allow users to interpret the performance of these models and gain a deeper understanding of how their skill profiles align with available job roles. By offering data-driven recommendations and personalized learning paths, this system empowers individuals to take proactive steps toward career growth and skill enhancement. Moreover, the platform not only benefits job seekers but also assists educational institutions and training providers in designing tailored programs that address skill gaps and industry requirements. As a result, the system contributes to building a more resilient and future-ready workforce that can effectively navigate the complexities of the modern job market. Through this initiative, individuals are equipped with the knowledge and resources needed to align their aspirations with industry expectations, fostering a culture of continuous learning and professional growth.

2. OBJECTIVES

The objective is to develop a web-based application that enables users to:

- Assess their current skills and receive job recommendations accordingly.
- Identify skill gaps relative to desired job roles.
- Explore job opportunities that match their skill profiles.
- Access career roadmaps outlining necessary skills for progression

3. SYSTEM ANALYSIS

3.1 Existing system:

Current platforms offer job listings and generalized career advice but often lack personalized analysis based on an individual's unique skill set. They may provide extensive data but fail to deliver actionable insights tailored to the user's career development needs..

3.2 Proposed System:

The proposed system utilizes machine learning techniques to predict job roles and evaluate skill gaps based on a user's current skill set. It incorporates classification algorithms such as Random Forest, Naive Bayes, and SVM to identify suitable employment opportunities. Upon analyzing the user's skill profile, the system highlights skill deficiencies and recommends additional skills necessary to meet industry demands. Visual representations of model performance metrics such as accuracy, precision, recall, and F1 score enable users to assess the reliability of the predictions. The system empowers users by providing personalized skill development paths, thereby enhancing their employability and ensuring a better alignment with the labor market...

4. METHODOLOGY

This section presents the steps and procedures followed to implement and evaluate job matching models using both machine learning (ML) and naïve bayes. techniques. The methodology is divided into three main stages: dataset description, pre-processing, feature extraction, model selection, and evaluation. Each stage is explained in detail below.

4.1 Feature Extraction

Feature extraction played a vital role in transforming raw textual data into meaningful numerical representations that could be effectively processed by machine learning models. Since job descriptions and user skills were primarily in textual form, advanced natural language processing (NLP) techniques were employed to extract key features that captured the underlying relationships between skills, job roles, and required qualifications.

4.2 Pre-processing

To ensure accurate and efficient job role predictions, data pre-processing played a crucial role in refining the textual data before model training. Raw job descriptions and user-provided skills often contain inconsistencies, redundant information, and noise, which can negatively impact model performance. Various cleaning techniques were applied to standardize and optimize the dataset, ensuring that only meaningful and relevant features were retained for analysis. This involved removing stop words, punctuation, and special characters, as well as normalizing text through lemmatization and stemming. Additionally, missing values were handled using appropriate imputation techniques to prevent data sparsity issues.

iJETRM
International Journal of Engineering Technology Research & Management
Published By:
https://www.ijetrm.com/

4.3 Model selection

This work tested several machine learning and deep learning models. Text classification performance and capacity to handle high-dimensional, sparse data like textual content led to the selection of these models. Traditional ML models exist. Machine learning models include Text categorization commonly uses support vector machines (SVM) due to their ability to handle high-dimensional feature spaces well. Naive Bayes (NB), a basic probabilistic classifier that assumes feature independence, tends to classify text effectively. This study employed logistic regression, a linear model for binary classification, to determine if a review is honest. The ensemble technique Random Forest averages numerous decision trees to decrease variation and avoid over fitting. We tried a variety of machine learning and deep learning models. We selected these models for text categorization and their ability to handle high-dimensional, sparse data. Traditional ML and DL models exist. ML models include Text categorization uses SVM because it excels at handling high-dimensional feature spaces. Naive Bayes (NB),

4.4 Construction of Use case diagrams:

A use case diagram in the Unified Modeling Language (UML) is a type of behavioral diagram defined by and created from a Use-case analysis. Its purpose is to present a graphical overview of the functionality provided by a system in terms of actors, their goals (represented as use cases), and any dependencies between those use cases. The main purpose of a use case diagram is to show what system functions are performed for which actor. Roles of the actors in the system

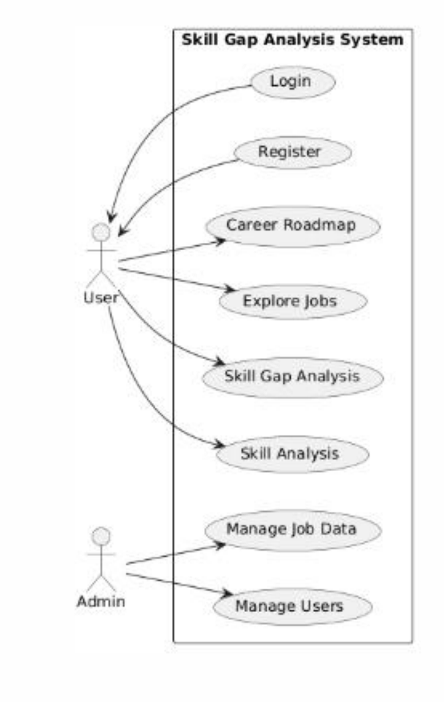

Fig.4.1.1 Use Case Diagram

5. RESULTS AND PERFOMANCE EVALUATION

Table 5.1 shows the performance comparison of Random Forest, Naïve Bayes, and SVM using accuracy, precision, recall, and F1-score. Random Forest achieved the highest accuracy, followed by SVM, while Naïve Bayes performed the lowest.

Model	Accuracy	Precision	Recall	F1 Score
RandomForest	0.924658	0.911301	0.924658	0.912769
NaiveBayes	0.787671	0.689314	0.787671	0.722026
SVM	0.910959	0.877854	0.910959	0.891389

International Journal of Engineering Technology Research & Management
Published By:
https://www.ijetrm.com/

Table 5.1: Model Performance Comparison

Figure 6.2 illustrates the accuracy of three machine learning models—Random Forest, Naïve Bayes, and Support Vector Machine (SVM). Random Forest and SVM demonstrate higher accuracy, both exceeding 90%, while Naïve Bayes lags behind with an accuracy below 80%. This comparison highlights the superior classification performance of Random Forest and SVM, making them more reliable choices for the given dataset.

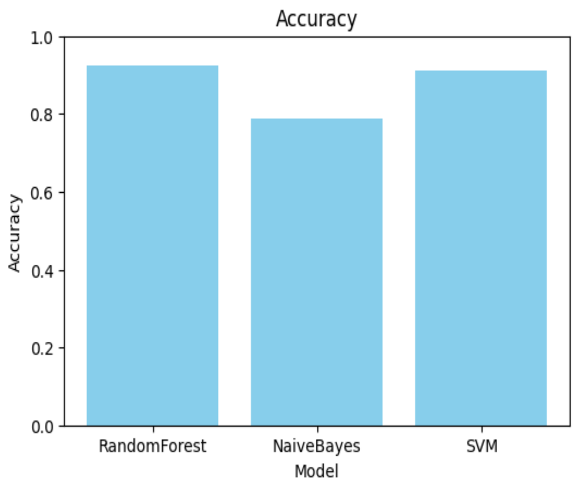

Fig. 5.2 Accuracy graph

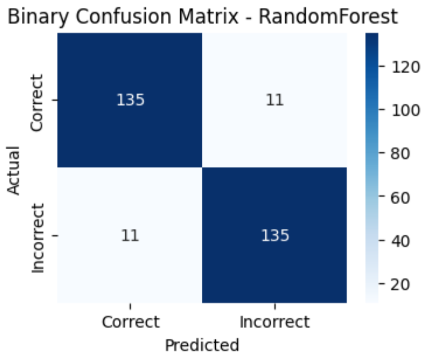

Fig.5.3 Random forest Binary Confusion Matrix

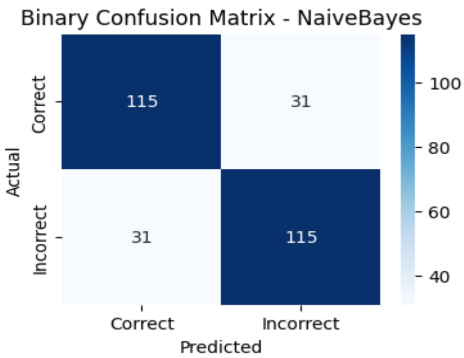

Fig.5.4 Naïve Bayes Binary Confusion Matrix

iJETRM
International Journal of Engineering Technology Research & Management
Published By:
https://www.ijetrm.com/

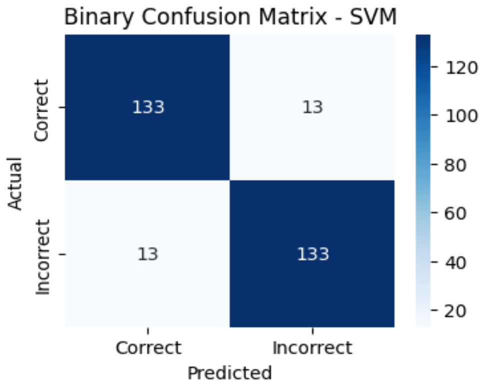

Fig.5.5 SVM Binary Confusion Matrix

The confusion matrices in Figures 5.3, 5.4, and 5.5 provide insight into the classification performance of the Random Forest, Naïve Bayes, and Support Vector Machine (SVM) models, respectively.

- **Random Forest (Figure 5.3):** demonstrates strong predictive performance, correctly classifying 135 instances in both positive and negative classes, with only 11 misclassifications in each category. The balanced distribution of correct predictions indicates a well-generalized model with minimal errors.
- **Naïve Bayes (Figure 5.4):** exhibits the highest misclassification rate, with 31 false positives and 31 false negatives. This suggests that the model struggles with class separation, leading to reduced accuracy and reliability compared to the other models.
- **SVM (Figure 5.5):** performs comparably to Random Forest, with 133 correct classifications in both categories and only 13 misclassifications each. The lower error rate highlights its robustness and suitability for the given dataset.

6. CONCLUSION

This research introduces a machine learning-based system that effectively bridges the gap between users' current skill sets and the demands of the labor market. By utilizing classification algorithms like Random Forest, Naive Bayes, and SVM, the system predicts suitable job roles and provides targeted recommendations for addressing identified skill gaps. Comprehensive performance evaluation ensures the reliability of these predictions, empowering users to make informed career decisions. The system not only enhances employability but also contributes to the development of a skilled workforce capable of adapting to evolving industry trends. Ultimately, this initiative promotes lifelong learning and career growth by aligning individual aspirations with market expectations. Addressing the skill gap in the IT industry requires innovative solutions that empower individuals to take charge of their career trajectories. By providing a platform that combines skill assessment, gap analysis, job exploration, and career mapping, this project offers a comprehensive tool for users to align their competencies with market needs. Such alignment not only enhances employability but also contributes to a more efficient and responsive workforce.

7. REFERENCES

[1] Agrawal, S.R., Mittal, D., 2017. Need of an online e-waste market in India. Int. J. Environ. Waste Manag. 19, 21–41. https://doi.org/10.1504/IJEWM.2017.083560.

[2] Asekol, 2019. Collection points < ASEKOL | Asekol [WWW Document]. URL https://www.asekol.cz/en/asekol/collection-points/ (accessed 8.5.19).

[3] Aziz, F., Arof, H., Mokhtar, N., Mubin, M., Abu Talip, M.S., 2015. Rotation invariant bin detection and solid waste level classification. Measurement 65, 19–28. https://doi.org/10.1016/j.measurement.2014.12.027.

[4] Bin Aziz, F., Arof, H., Mokhtar, N., M. Shah, N., Khairuddin, A., Hanafi, E., Sofian Abu Talip, M., 2018. Waste level detection and HMM based collection scheduling of multiple bins. PLOS ONE 13, e0202092. https://doi.org/10.1371/journal.pone.0202092.

iJETRM
International Journal of Engineering Technology Research & Management
Published By:
https://www.ijetrm.com/

[5] Cao, J., Xu, J., Wang, H., Zhang, X., Chen, X., Zhao, Y., Yang, X., Zhou, G., Schnoor, J., 2018. Innovating collection modes for waste electrical and electronic equipment in China. Sustainability 10, 1446. https://doi.org/10.3390/su10051446.

[6] Chi, X., Wang, M.Y.L., Reuter, M.A., 2014. E-waste collection channels and household recycling behaviors in Taizhou of China. J. Clean. Prod. 80, 87–95. https://doi.org/10.1016/j.jclepro.2014.05.056.

[7] Chopdar, P. Kr., Korfiatis, N., Sivakumar, V.J., Lytras, M.D., 2018. Mobile shopping apps adoption and perceived risks: A cross-country perspective utilizing the unified theory of acceptance and use of technology. Comput. Hum. Behav. 86, 109–128. https://doi.org/10.1016/j.chb.2018.04.017.

[8] Kim, M., Kim, J., Choi, J., Trivedi, M., 2017. Mobile shopping through applications: understanding application possession and mobile purchase. J. Interact. Mark. 39, 55–68. https://doi.org/10.1016/j.intmar.2017.02.001.

[9] Cucchiella, F., D'Adamo, I., Lenny Koh, S.C., Rosa, P., 2015. Recycling of WEEEs: An economic assessment of present and future e-waste streams. Renew. Sustain. Energy Rev. 51, 263–272. https://doi.org/10.1016/j.rser.2015.06.010.

[10] Dat, L.Q., Linh, D.T.T., Chou, S.-Y., Vincent, F.Y., 2012. Optimizing reverse logistic costs for recycling end-of-life electrical and electronic products. Expert Syst. Appl. 39, 6380–6387.

[11] Esmaeilian, B., Wang, B., Lewis, K., Duarte, F., Ratti, C., Behdad, S., 2018. The future of waste management in smart and sustainable cities: A review and concept paper. Waste Manag. 81, 177–195. https://doi.org/10.1016/j.wasman.2018.09.047.

[12] European Commission, 2012. Directive 2012/19/EU of the European Parliament and of the Council of 4 July 2012 on waste electrical and electronic equipment (WEEE)Text with EEA relevance - LexUriServ.do.

[13] Girshick, R., Donahue, J., Darrell, T., Malik, J., 2014. Rich feature hierarchies for accurate object detection and semantic segmentation. In: 2014 IEEE Conference on Computer Vision and Pattern Recognition. Presented at the 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Columbus, OH, USA, pp. 580–587. https://doi.org/10.1109/CVPR.2014.81.

[14] Gu, F., Zhang, W., Guo, J., Hall, P., 2019. Exploring ''Internet+Recycling'': Mass balance and life cycle assessment of a waste management system associated with a mobile application. Sci. Total Environ. 649, 172–185. https://doi.org/10.1016/j.scitotenv.2018.08.298.

[15] Gundupalli, S.P., Hait, S., Thakur, A., 2017. A review on automated sorting of source-separated municipal solid waste for recycling. Waste Manag., Special Thematic Issue: Urban Mining and Circular Economy 60, 56–74. https://doi.org/10.1016/j.wasman.2016.09.015.

[16] Hannan, M.A., Abdulla Al Mamun, Md., Hussain, A., Basri, H., Begum, R.A., 2015. A review on technologies and their usage in solid waste monitoring and management systems: Issues and challenges. Waste Manag. 43, 509–523. https://doi.org/10.1016/j.wasman.2015.05.033.

[17] Hannan, M.A., Arebey, M., Begum, R.A., Basri, H., Al Mamun, Md.A., 2016. Content-based image retrieval system for solid waste bin level detection and performance evaluation. Waste Manag. 50, 10–19. https://doi.org/10.1016/j.wasman.2016.01.046.

[18] Howard, A.G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., Adam, H., 2017. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications.

[19] Krizhevsky, A., Sutskever, I., Hinton, G.E., 2017. ImageNet classification with deep convolutional neural networks. Commun. ACM 60, 84–90. https://doi.org/10.1145/3065386