Nanotechnology Perceptions
ISSN 1660-6795
www.nano-ntp.com

Examining Graduates' Job-Readiness, Employability Skills, and Awareness using machine learning

D. Anuradha[1], Dr. Jnana Raghavendra I[2], P. Pinakapani[3]

[1]*Research Scholar, GIET University, Odisha, India, anuradhaduvvuri1975@gmail.com*
[2]*Supervisor, School of Management Studies, GIET University, Odisha, India*
[3]*Chair Research Cell, GITAM School of Business, GITAM Deemed to be University, Hyderabad*

The importance of higher education institutions (HEIs) understanding the employability of their graduates and the reasons behind it is growing as the number of graduates produced annually from HEIs continues to rise and competition for good jobs heats up. One performance indicator for HEIs is the employability of its graduates. Because it is often used as a measure of success, student employability is vital for educational institutions. On the other hand, globalization, automation, and the latest developments in AI are making the labour market environment more dynamic than ever before. Prior to graduation, students' employability was predicted using machine learning models. Logistic regression, decision trees, random forests, and the K-Means method are all part of this category. Therefore, the purpose of this study is to forecast undergraduates' full-time employability using academic and experience employability attributes, such as CGPA, SIWES, extracurricular activities, gender, and union affiliations prior to graduation.

Keywords: Artificial Intelligence, employability, machine learning, logistic regression; decision tree, random forest Networks.

1. Introduction

Every nation, including India, relies on its higher education institutions (HEIs) to drive economic growth. It is an industry that contributes to the country's gross domestic product (GDP) by providing other sectors with trained workers, who in turn increase the productivity of those other businesses. Higher education institutions (HEIs) used to ignore students' lack of employability and fail to guide them towards more productive course choices. [1].

As early prediction of students' employability is a desirable activity that is continually required for decision making, student employability has recently emerged as a major issue for HEIs. Students need to find ways to stand out in the entry-level job market, both intellectually and experientially, since competition for employment is becoming worse. Data collected from students may also reveal which students have a higher chance of finding employment after graduation and why, which is useful information for colleges that are being held more and more responsible for student outcomes. Although, with more and more schools focusing on student employment, the employment of recent graduates is becoming an important factor in determining the institution's reputation and, therefore, a big worry. The use of data mining, particularly for enrollment prediction, is becoming more common among universities as they compete for students' attention and financial support [2].

Since it is often used as a measure of the institution's success, the employability of its students is a major issue for higher education institutions. Severe social consequences, such as alcoholism and self-destruction, could result from university students' dissatisfaction in getting a job [3].[4]. Graduates may have challenges in the labour market due to factors other than poor academic achievement, such as the development of lifeless prejudices. In order to aid pupils initially with more tailored intervention, it is crucial to understand these biases.

The two main schools of thought in machine learning are supervised and unsupervised. Supervised methods rely on labelled data sets to guide the training process and provide prior knowledge of the target outcome; in contrast, unsupervised methods operate directly on unlabeled data and, in the absence of labels, require the labels to be "discovered" by the algorithm to guide its learning [5] Numerous studies have examined the employability of graduates by demonstrating how their performances are well-suited to HEIs via the use of different machine learning algorithms. Before they can choose the best candidate for the job, HR departments must manually sort or categorise the profiles of the candidates.[6]

There was a significant time commitment involved, however, and it was proportional to the number of applications [7].

2. An overview of machine Learning

A whole new capability was bestowed to computer systems by machine learning. The term "training data" describes the aggregated information gathered from a variety of sources. The accuracy of the model is dependent on the amount and quality of the data used to create it. In order to train machine learning, data must first be prepared, which includes cleaning, randomising, and visualising the data to better comprehend the parameters and their relationships. From the available options, choose a model. In order to build a model that outperforms its baseline, training is essential. Evaluate the model with data that has never been seen before, Improving the model's performance is the goal of parameter adjustment. Last but not least, machine learning's strength is in its ability to replace human judgement with machine learning models to forecast questions' responses.In the past, subjective approaches like interviews, resumes, and self-evaluations were used to evaluate these qualities [8]. Machine learning (ML) and data analytics, however, provide a chance to explore this assessment process in more depth. Machine learning (ML) methods may sift through mountains of data in search of trends, correlations, and prediction models that shed light on college grads' marketability.The purpose of this study is to investigate recent grads' knowledge of, and preparedness for, the workforce by use of machine learning algorithms. We can build thorough profiles of graduates by combining data from several sources, including records of academic achievement, extracurricular activity, internships, and surveys. Not only may technical competences be included in these profiles, but so can soft skills such as communication, problem-solving, and cooperation [9].In addition, ML algorithms may reveal complex interrelationships among many elements impacting graduates' employability. For example, they can determine the optimal mix of academic performance, work experience, and personal qualities for a successful job search[10][11]. Also, you may get real-time insights into how industry trends and skill needs are changing by analysing sentiment on alumni comments or social media debates.

3. Related Work

The employability prediction in [1] was done using several classifiers, including decision tree, K-NN, Naïve Bayes approach, and Random Forest. The naïve Bayes method had the maximum accuracy at 89%, while Random Forest had the lowest accuracy at 69%.

Decision trees had the best accuracy at 85% and KNN the worst at 76% when it came to predicting whether or not an individual will quit their job. The study used well-known and sophisticated machine learning algorithms, including as logistic regression, decision trees, neural networks, and discriminant analysis, to forecast employment before graduation in [2].

Majors and extracurricular activities were also shown to be statistically significant in predicting employment after graduation, according to a sensitivity study. They mapped out the academic and business perspectives on employability and the chasm that exists between them in [3]. For this data, they consulted SPSS, a statistical programme. In the end, they outlined the three factors—"overall capabilities," "behavior qualities," and "precise skills and awareness related abilities"—that are crucial for finding a job. The authors of [4] looked studied how employers rated the employability of undergraduates based on three criteria: personal qualities, core competencies, and subject knowledge. They showed how to use an artificial neural network to HR tasks like assessing a candidate's employability and matching them with potential employers in [5]. In order to forecast which users have received a firm employment offer, an ANN with a two-layer feed forward network using the back propagation method is developed [12]. They created a classification model using algorithms, Nearest Neighbor Naïve Bayes, and Decision Tree to forecast how well Information Science graduates would fit in with the Saudi industry. The data was gathered in relation to Information Science abilities, self-regulated learning (SRL), and academic achievements of students majoring in Information Science. On the IT employment dataset, they contrasted five algorithms using n distinct categorization techniques in [7]. The two methods with the best accuracy rates were logistic regression (78.4% and 76.3%, respectively). the CHAID algorithm was used to make it. Their three suggested predictors—IT Basic, IT Jobwise, and Gender—are likely to have a direct impact on IT employability. The disparity between graduates' employment skills and their academic performance was brought to light in [8] with the use of both quantitative and qualitative approaches. Using benchmarks from the National Student Surveys (NSS) and student evaluation questionnaires (SEQs), as well as data from employers' responses and various technique studies, they analysed the data. As an innovative move towards bettering course creation, they also found that mixed techniques are useful for identifying skill gaps [2].

4. Proposed Methodology

Figure 1 shows the recommended technique. Data collection for this research is detailed, as are the preparation processes used to clean the dataset. These stages include dealing with missing values, removing duplicate or unnecessary variables, and encoding and simplifying the data. Additionally, it details the assessment criteria utilised and how unsupervised clustering machine learning models are created to forecast post-graduation employment.

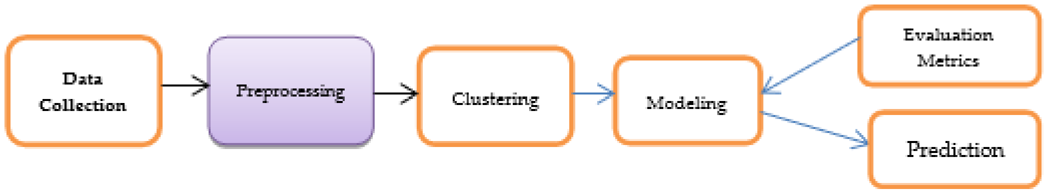

Figure 1: General Methodology of the Proposed System

The suggested model's main focus is on making predictions regarding the student's employability after they graduate. Additionally, this methodology aids the administration in identifying kids who may struggle academically and have limited employment prospects, allowing them to intervene promptly. Jupyter notebook6 was used to implement the two algorithms, which are based on the python programming language and its libraries: pandas for data import from spreadsheets, numpy for numerical analysis, matplotlib for visualisations, and seaborn for rendering.

5. Data acquisition

The quality of the data used to train prediction models is critical to their reliability. Employability prediction relies on the use of resourceful data, which should include a wide range of student information that may shed light on the many facets of the link between students' qualities and their employability abilities. There is a lack of research showing a direct correlation between qualities and talents. In most cases, researchers will use publicly available or readily accessible characteristics to create a collection of features that will feed into their prediction models.

In order to gauge employability, prior research has relied on a wide range of datasets collected from a variety of sources, including online job portals, placement offices, career centre offices, and ministries, among others.

However, it is evident that the majority of researchers took into account students' academic records, which included data retrieved from placement reports, when predicting their employability. In addition to demographics, psychometric characteristics, emotional intelligence, social academic integration, and other personal profile features, several researches took these into account. Lastly, due to the time and effort required, the difficulty of obtaining the data, and the enormous data sets needed to train ML models, relatively few research make use of questionnaire and quiz results. A small number of studies also use job adverts to determine the necessary abilities for different career domains, in addition to student data. The majority of educational institutions today employ a Student Information System (SIS) that makes it easy to access students' academic information, whether they are in high school, college, or university. Student demographics (such as age, gender, and ethnicity) may also be found in SIS, however socio-economic status may not be readily accessible. If that's the case, we might infer it from what we already have or ask students directly via surveys. Data on students' psychological and soft abilities, however, would likely require interviewing them personally. As a result, surveys became the secondary means of gathering information since they permitted the systematic gathering of data. Additionally, a hybrid data gathering strategy was explored, which would have included merging academic records with data collected from surveys to create the dataset.

6. Conclusion

Using machine learning to assess graduates' awareness, employability skills, and job-readiness is a huge step forward in our capacity to comprehend and solve the problems encountered by graduates and those involved in the labour market. This research has introduced a framework for using machine learning to evaluate and improve graduates' employment preparation and shed light on the multi-faceted nature of that readiness.We were able to build thorough profiles of graduates by collecting and preparing expansive datasets that include academic records, extracurricular activities, internship experiences, and industry comments. We have created prediction models that can measure graduates' job-readiness and identify important employability abilities by using feature engineering approaches and choosing relevant machine learning models.

Conflicts of Interest

The authors declare that they have no competing interests.

References

1. Khadilkar, N., & Joshi, D. (2017). Predictive Model on Employability of Applicants and Job Hopping using Machine Learning. International Journal of Computer Applications, 171(1), 37-41.
2. Mishra, T., Kumar, D., & Gupta, S. (2017). Students’ Performance and Employability Prediction through Data Mining: A Survey. Indian Journal of Science and Technology.
3. Hugo, L. S. (2018). Analytics: A Comparison of Machine Learning Models Predicting Student Employment. (Master's thesis). Ohio University.
4. Sui, F., Chang, J. C., Hsiao, H., Chen, S., & Chen, S. (2018). A Study Regarding the Gap Between the Industry and Academia Expectations for College Student's Employability. In IEEE International Conference on Industrial Engineering and Engineering Management (IEEM).
5. Pengnate, W. (2018). Needs of employability skill characteristics based on employers' perception. In 2018 5th International Conference on Business and Industrial Research (ICBIR), 598-601.
6. Thorat, S. G., Bhagat, A. P., &Dongre, K. A. (2018). Neural Network Based Psychometric Analysis for Employability. In 2018 International Conference on Research in Intelligent and Computing in Engineering (RICE).

7. Almutairi, M. M., &Hasanat, M. H. A. (2018). Predicting the suitability of IS students' skills for the recruitment in Saudi Arabian industry. In IEEE International Conference.
8. Hugo, L. S. (2019). Predicting Employment Through Machine Learning. National Association of Colleges and Employers, May 01.
9. Hasan, H. M. R., Islam, M. T., Azad Rabby, A. S., &Hossain, S. A. (2019). Machine Learning Algorithm for Student's Performance Prediction. In 10th ICCCNT 2019.
10. Alghamlas, M., &Alabduljabbar, R. (2019). Predicting the Suitability of IT Students' Skills for the Recruitment in Saudi Labor Market. In 2019 2nd International Conference on Computer Applications & Information Security (ICCAIS), 1-5
11. Panwar, B., Dhuriya, G., Johri, P., Singh Yadav, S., & Gaur, N. (2021). Stock market prediction using linear regression and SVM. In 2021 International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE) (pp. 629-631). Greater Noida, India. doi: 10.1109/ICACITE51222.2021.9404733
12. Pattnaik, Sonali, Hota, SwetaLeena, Kumar, Arya, Hota, Anish Raj &Kiran, Prabha(2024) Awaking deep emotional needs of consumers based on big data and emotional analysis, Journal of Statistics and Management Systems , 27:1, 105–119, DOI: 10.47974/JSMS-1196
13. Pujari, Purvi, Arora, Monika, Kumar, Anuj&Pandey, Anoop(2024) Understanding factors influencing technical inertia in family-run SMEs : A study on technology adoption challenges, Journal of Statistics and Management Systems , 27:1, 155–168, DOI: 10.47974/JSMS-1208