International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878, Volume-8 Issue-5, January 2020

Employability Prediction of Engineering Graduates using Machine Learning Algorithms

Vinutha K, Yogisha H K

Abstract- Number of graduates produced in each year by higher education institutions is increasing. Thus prediction of employability of graduate's plays a vital role for any industry for proper talent acquisition and Utilization and also it helps students in identifying the qualification and skills that they need to improve, before completion of degree to get desired jobs. In this Digital Revolution, informal learning and skill enhancements is happening in unconditional method, relating and converging all this learning's to the employability rate is one of a biggest issue. The main objective is to address this issue by predicting and forecasting the skill acquisition continuously and mapping to industry needs using machine learning Algorithms. The proposed work used different machine learning algorithms like Logistic Regression, Decision tree, k-nearest neighbor, Support Vector Machine and Naïve Bayes for building model where ANN classifier resulted with the highest accuracy of 87.42%. This research would be helpful for all the organization including government, Private and corporations, including students and teachers for employability prediction.

Keywords: Academic result, Employability prediction, Machine learning classifiers, Student Performance Estimate

I. INTRODUCTION

In the world of digitalization, education becoming more and more employment oriented, reputation of any institution is depending on the employment of the students and hence the major concern [9]. The prediction model that suggests which student will and will not be employed will guide them in identifying the individuals in need of support. Instead of normal analytics, the advanced machine learning method, a part of Artificial Intelligence is used to get insights of future outcomes.

Manuscript published on January 30, 2020.
***Correspondence Author**

Vinutha K*, Assistant Professor, Department of Information Science and Engineering, BMS Institute of Technology and Management, Bangalore, Karnataka, India and affiliated to Visvesvaraya Technological University, Belagavi, Karnataka, India . Email: vinuthak_ise2014@bmsit.in

Dr. Yogisha H K, Professor and Head, Dept.of MCA, Ramaiah Institute of Technology, Bangalore, Karnataka, India and affiliated to Visvesvaraya Technological University, Belagavi, Karnataka, India.

Email: yogishhk@gmail.com

An overview of machine Learning

Machine learning granted computer systems entirely a new ability. Figure 1 below shows the overview of machine learning [10]. Data is collected from various sources, often called as training data.

The quality and the quantity of data collected determine how good the model will be. Preparation of data for machine learning training is done, where data cleaning, randomization and visualizations can be made which helps in understanding the relationship between the parameters. Choose a model among various available.

Training helps in developing a model that does better than a baseline, Test the model against previously unseen data, Parameter tuning is to improve the performance of the model. Finally, the power of machine learning is to predict the answers to the questions using machine learning model rather than human judgement.

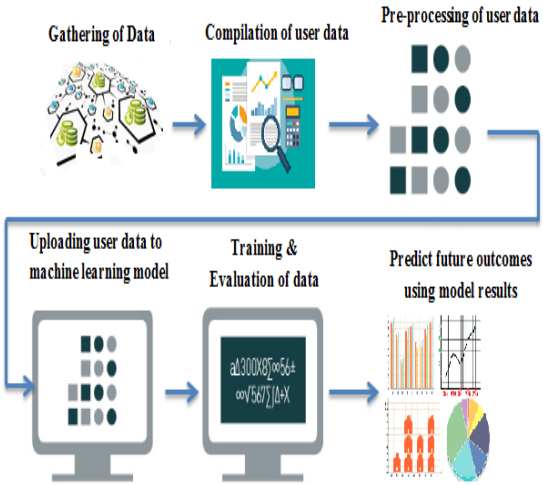

Fig 1: Overview of Machine Learning

II. RELATED WORK

In [1] they used different classifiers such as decision tree, K-NN, Naïve Bayes approach, Random Forest naïve Bayes resulted with a highest accuracy of 89% and Random forest resulted with a lowest accuracy of 69% for employability prediction. They also predicted whether the employee's going to leave the job or not, decision tree provided a highest accuracy of 85% and KNN provided lowest accuracy of 76%. In [2] to predict employment prior to graduation, the research used commonly recognized and advanced machine learning models, including discriminant analysis, decision trees, neural networks and logistic regression.

Results demonstrate that employment prior to graduation can be predicted with73% accuracy with a neural network as the most accurate predictive model. Moreover, a sensitivity analysis identified co-curricular activities and majors as statistically significant variables in predicting employment upon graduation.

In [3] they identified the views of industry and academic world on employability and the hole between the industry and academic world. They Used a statistical software SPSS for data analysis. Finally, they provided the three dimensions which are important for employability like "precise skills and awareness related abilities", "overall capabilities", and "behaviour qualities". In [4] they investigated the perception of employers' on employability abilities of undergraduate scholars in three areas: Individual Qualities; Core Skills; and Topic Knowledge. In [5] they demonstrated how to use Artificial Neural Network to Human Resources for analyzing mental ability of the user with respect to employability, to find the anticipated job. An ANN with two-layer feed forward network that uses Backpropagation algorithm is created for predicting the users who had given with a definite job offer.

In [6] they collected data related to Information Science skills, self-regulated learning (SRL), and academic accomplishment of Information Science students and developed a classification model using Nearest Neighbor Naïve Bayes, Decision Tree and algorithms to predict the aptness of Information Science graduates to the Saudi industry. In [7] they compared five algorithms under n different classification technique on Information Technology employability dataset. Logistic regression provided the highest accuracy of 78.4% and the second highest accuracy of 76.3 %. was reached using CHAID algorithm. They proposed three probable predictors which affects IT employability directly are the IT_Basic, IT_Jobwise and Gender. In [8] highlighted the gap between the graduate results and employability abilities by the use of qualitative and computable methods. To analyze the data, they used employers' response and varied method research and also used National Student Surveys (NSS), and student evaluation questionnaires (SEQs) benchmarks. They also identified that mixed methods are helpful in determining gaps in skills and this is a dynamic step to improve course development.

III. PROPOSED METHODOLOGY

The proposed model is about predicting whether the student is employable after their graduation. This model also helps the administration to identify students at the risk of poor academic performance and low employability and to take timely actions [11].

A. Dataset preparation

New dataset is used for building proposed model. Table I: The Dataset description. For training we have 659 instances of Information science and Engineering and Computer science and engineering students' data with their academic performance from first semester to eighth semester, online courses completed, internship and technical papers presented information.

Table I: The Dataset description

Features	Descriptions	Scores
USN	Unique Student Number	-
Name	Name of the student	-
First to eighth semester marks	Semester wise marks of each student from first semester to eight semester	Out of 100
Online courses completed	Online courses completed by the students	1-completed 0-not completed
Internship	Industry Internships completed by the students	1-completed 0-not completed
Technical papers presented	Technical papers presented by the students in journals/conferences	1-completed 0-not completed

B. Sample Input

Input data for the classifiers are provided from CSV file includes student's marks from first semester to eighth semester, data regarding internship, online courses, and technical papers presented details. Table II: The sample input data. Every input data is in numeric form.

Table II: The sample input data

1st	2nd	3rd	4th	5th	6th	7th	8th	Internship	Online courses	Papers presented	Employment
44	55	44	41	47	52	36	48	0	0	0	0
46	55	58	61	64	74	67	69	1	0	0	0
76	74	67	72	71	70	75	76	0	1	1	1
54	57	54	53	64	74	60	69	1	0	1	1
56	69	63	60	60	63	58	62	1	0	1	1
71	75	73	68	66	69	69	68	1	1	1	0
69	74	76	73	70	69	76	80	1	0	1	1
61	67	59	62	62	69	61	67	1	1	1	1
54	56	61	55	64	68	66	71	0	1	0	0
70	78	76	74	62	74	67	74	1	1	0	1
71	82	78	72	74	81	73	75	0	0	1	1
56	65	66	65	67	68	65	67	0	0	0	0
71	82	82	75	69	70	71	84	1	1	1	1

C. Architecture *of the Model*

The highest objective of study is to predict the employability of engineering graduates. The key elements of the study include:

Identifying employability Parameters: Employers expects particular qualities in students like logical thinking, analysis capabilities, leadership, ethics. Some of these factors are difficult for employers to measure in stipulated time, so they have to rely on few events that infer skills. for example, if presented any technical papers infers he is having good technical skills, communication skills.

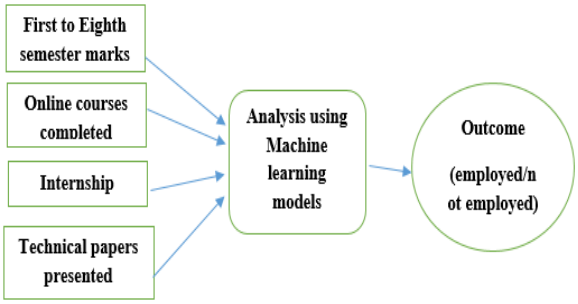

Fig 2: Employability parameters

Classification, Classifier evaluation and Algorithm selection for employability prediction: Classification is a method which categories the data into given number of groups. The groups are often referred to as target, labels, class or categories. The key aim of classification is to find the group or the target to which fresh data will make up.

KNN is the laziest algorithm compared to other in machine learning algorithms. It works by considering the distance from known data points. After gathering the k neighbors we simply take the majority and classify the unknown data into that class. KNN resulted in the accuracy of 82.85% for our dataset.

Decision tree starts with a one node, which divisions into possible results. Each of those results leads to the additional nodes, which branch off into other possibility. This gives a tree like structure. They can be used to understand nonlinearity and provides an algorithm that predicts the best choice statistically. It utilizes the if then rules which are equally exhaustive and mutually exclusive in classification. Decision tree resulted in the accuracy of 81.48% for our dataset.

Naïve Bayes gives the assumption of independence among predictors. The main assumption is existence of a feature in a group is unrelated to the presence of any other feature. Naïve Bayes works on the principle of Bayes' Theorem. Bayes' Theorem finds a probability of an event occurring given the probability of another event that has already occurred. It is called naïve Bayes because the calculation of probabilities for each hypothesis are simplified to make own calculations. Naïve Bayes resulted in the accuracy of 84.21% for our dataset.

Support Vector Machine (SVM) is an interesting algorithm, the minimal Margin classifier is a hypothetical classifier which explains how SVM works practically. In SVM the hyperplane is selected to finest separate the points in the variable space by their class. In two dimensions this can be visualized as a line and in three dimensions it can be a hyperplane. SVM resulted in the accuracy of 80.12% for our dataset.

Logistic Regression uses one or more independent variables to decide a result. The result is measured with the dichotomous variable which means it will have only two possible results. The objective of Logistic Regression is to discover the best fitting association between the dependent variable and set of independent variables. The main disadvantage is it only works when predicted variable is binary. Logistic Regression, this function is also called sigmoid function, input values are joined linearly using weights or coefficient values to predict a result. It is different from linear regression because the output function being modelled is binary, instead of continuous. Logistic Regression resulted in the accuracy of 87.42% for our dataset.

IV. RESULT AND PERFORMANCE OF PROPOSED MODEL

Proposed model got 87.42% accuracy using Artificial Neural Network, 85.2% accuracy using Logistic Regression classifier and 84.21% accuracy using Naïve Bayes classifier on predicting student employability. By studying the outcome, we found proposed model works better for predicting the employability of graduates.

Table III: Accuracy in different Classifiers

Classifier	Classifier Accuracy (%)
ANN	87.42
Logistic Regression	85.20
SVM	80.12
Naïve Bayes	84.21
Decision tree	81.48
KNN	82.85

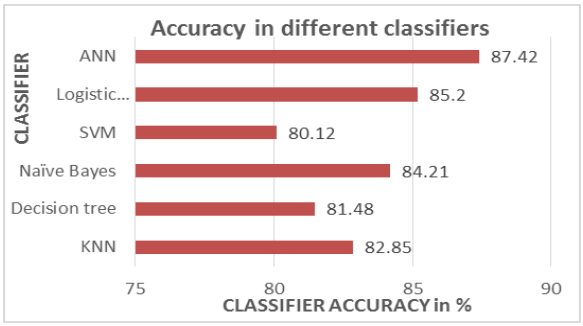

Fig 3: Graph of Accuracy in different Classifiers

The above figure specifies the accuracy rate after implementing the model using different machine learning algorithms. It is observed that the accuracy rate different algorithms like Decision tree, artificial neural network, k nearest neighbor, Naïve Bayes, Support Vector Machine, Logistic Regression is more than 80%.

V. CONCLUSION

Since there is an increase in Engineering graduates produced by the higher education institution in every year, there is a rise in competition for getting employable. This study predicted employability of a graduate based on their academic performance and also employability skills. By applying various machine learning classification algorithms, we were able to create a model for predicting employability of graduates with highest accuracy of 87.42% using Artificial Neural Network, 85.2% accuracy using Logistic Regression classifier and 84.21% accuracy using Naïve Bayes classifier. In future, with the suggestion of knowledgeable, we will improve our datasets with student's cognitive data to increase the accuracy of employment prediction of a graduate.

REFERENCES

1. Neeraj Khadilkar and Deepali Joshi. "Predictive Model on Employability of Applicants and Job Hopping using Machine Learning". International Journal of Computer Applications 171(1):37-41, August 2017.
2. HUGO, LINSEY S., M.S., "Analytics A Comparison of Machine Learning Models Predicting Student Employment", Individual Interdisciplinary Program, Master of Science in Higher Education and Business Ohio University, December 2018
3. Sui, F., J. C., Chang, Hsiao, H., Chen, S. Chen. "A Study Regarding the Gap Between the Industry and Academia Expectations for College Student's Employability", IEEE International Conference on Industrial Engineering and Engineering Management (IEEM), December 2018

4. Pengnate, Wipanee. “Needs of employability skill characteristics based on employers' perception.” 2018 5th International Conference on Business and Industrial Research (ICBIR) (2018): 598-601.
5. S. G. Thorat, A. P. Bhagat and K. A. Dongre, "Neural Network Based Psychometric Analysis for Employability," 2018 International Conference on Research in Intelligent and Computing in Engineering (RICE), San Salvador, 2018, pp. 1-5.
6. Alghamlas, Munirah and Reham Alabduljabbar. “Predicting the Suitability of IT Students' Skills for the Recruitment in Saudi Labor Market.” 2019 2nd International Conference on Computer Applications & Information Security (ICCAIS) (2019): 1-5.
7. Keno C. Piad, Menchita Dumlao, Melvin A. Ballera, Shaneth C. Ambat “Predicting IT employability using data mining techniques” Third International Conference on Digital Information Processing, Data Mining, and Wireless Communications (DIPDMWC) , 2016
8. Mona Masad Almutairi, Mozaherul Hoque Abul Hasanat “Predicting the suitability of IS students’ skills for the recruitment in Saudi Arabian industry” ,IEEE International Conference 2018.
9. Tripti Mishra ,Dharminder Kumar, Sangeeta Gupta, “Students’ Performance and Employability Prediction through Data Mining: A Survey” Indian Journal of Science and Technology, June 2017
10. Linsey S. Hugo, “Predicting Employment Through Machine Learning” National Association of Colleges and Employers, May 01, 2019
11. H.M. Rafi Hasan , Mohammad Touhidul Islam Akm Shahariar Azad Rabby, Syed Akhter Hossain “Machine Learning Algorithm for Student's Performance Prediction”, 10th ICCCNT 2019 July 6-8, 2019 , IIT - Kanpur, Kanpur, India

AUTHORS PROFILE

Vinutha K, is working as Assistant professor in Dept of ISE, BMSIT&M. she has completed her B.E in Channabasaveshwara Institute of Technology Tumkur in 2011 and M.Tech from AMC Engineering College 2013 in CSE. She is having 5 years of teaching experience and pursuing research in the field of prediction models using Machine learning. Her area of interest is Machine learning and Data mining. She has published 7 research papers in various International Conference and Journals and delivered 5 expert talks on various topics of Machine learning, python and data mining. She is a life member of “Indian Science Congress Association”.

Dr. Yogish H K, is working as Professor and Head of the Department of Master of Computer Applications. He completed his B.E in Computer Science and Engineering from PESCE, Mandya, Mysore University, and M.Tech in Computer Engineering from SJCE Mysore, VTU Belagavi and Ph.D. in Computer Science at Bharathiar University, Coimbatore. He has published and presented 57 papers at National/International Conferences/Journals among ELEVEN papers were indexed by SCOPUS, EIGHT IEEE, TWO papers in DBLP. He has authored three text-books and has an extensive experience of 20+ years of Teaching Experience and worked in various capacities from Lecturer to Professor. He is supervising Seven Ph.D Scholar in VTU, Belagavi. A Reviewer of FOUR International Journals and International Conferences, Chaired several International conferences. He is a member of Advisory committee, Staff selection committee, Board of Studies and Board of Exams in various institutions and Universities. He is a life time member of various Professional bodies like FIE, LMISTE, CSI and IAENG. He has also worked as Corporate Trainer for TWO years, trained almost 800 Tech Mahindra employees at CMC Ltd and at Uttara Info Solutions.