A COMPARATIVE STUDY OF STUDENT'S EMPLOYABILITY PREDICTION MODEL USING DATA MINING TECHNIQUE

[1]Tejal M Vasan, [2]KirtiKumar J Sharma, [3]Dr. N C Chauhan
[1]PG Student, [2]Assistant Professor, [3]Professor
[1]Dept. of Computer Engineering,
[1]Birla Vishwakarma Mahavidyalaya, V.V. Nagar, Gujarat, Anand, Gujarat

Abstract: Employability is someone view for improving and maintaining employment. Actually, student's employability rests on BASIC FOLLOWING FACTORS: EDUCATION, UNDERSTANDING POWER, AND PERSONAL DEVELOPMENT. IT IS THE CAPABILITY OF THE STUDENT TO ACHIEVE the basic employment, and then continue it and to find a different one if it is necessary. The aim of this study is to represent associate degree ex-student employability system with numerous data processing techniques to spot a talent that important for employment. Also, many data mining techniques are used for the prepared prediction model. Methods like a decision tree, Neural Network, Support Vector Machine and Bayer's Method.

Index Terms - **Data mining, Student's Graduate employability, Employability Skills and Factors, Educational Data Mining (EDM).**

I. INTRODUCTION

1.1 WHAT IS EMPLOYABILITY?

In the last era, the various numbers of higher studies of universities, organizations, and colleges have cause to grow or increase rapidly manifolds. In terms of bulky numbers of graduates and postgraduates higher study's universities /institutes /colleges, produce students every year. They may follow as much as possible best of the pedagogies: the method and practice of teaching; but still, nowadays all of them face the challenging problems of dropout, detained and irregular students, low successes and jobless students. Student's Employability can be known as "Doing Value Creating Work, Getting Paid For It And Learning At The Same Time, Enhancing The Talent To Get Work In The Future" [1]. For employers, it force to making a working platform that can be responsible for wide range of opportunities for the student's particular and professional progress, within management environs where it's understood that gifted, growing folks mean talented, growing organizations, etc. [17]
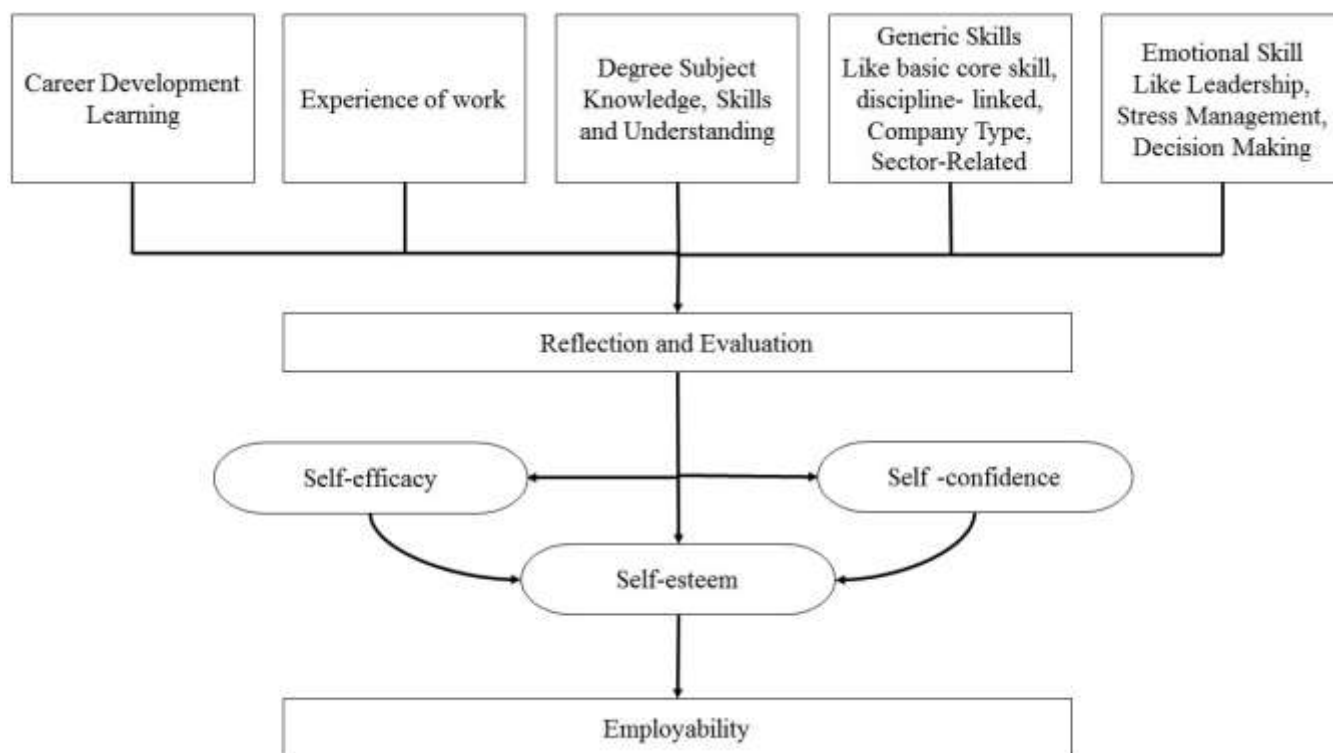
Figure-1 Student Employability Diagram [1]

In data mining Classification method is most significant approach specifically in the research region of the predicting. The efficiency of classification an initial proposal not only manages with a bulk of data to fetch or identify secreted patterns and generating relationships that useful in decision creation, It additionally reduces their physical property the data-generation structure, no matter problem, nice prognosticative and in some scenario, instructive potential. [2]

Table-1 Education Data Mining Level Analysis

	Level Of Analysis	User Benefitted
Education Data Mining	Course Level: Design and development of Intelligent Curriculum, Conceptual Enhancement and development	Students, Faculty
	Departmental: Predictive modeling of Success/ Failure across the departments	Students, Faculty
	Institutional: Academic Performance and overall Institution Growth and Reputation	Administrators, Funders
	Regional: Comparisons between various System	Administrators, Funders
	National and International	Educational Authorities, Governments

Table-1 EDM Users at various levels [16]

1.2 Exiting Problems to Identify Student Employability

- Employment and Employability are not the same (but the two ideas are often combined). Maybe because they both begin with employ!
- If a graduate gains a graduate position, do they have employability - Possibly, but not necessarily?
- If a graduate does not gain a graduate position does that mean they do not have the employability – Possibly, but not necessarily?

1.3 Research Problem Statement

In our country, many students are taking admissions for undergraduate and postgraduate courses. But employment is not available to all after completion of their higher studies. The government also wish knows the employment status of different graduates. Hence, in the present work, it has been planned to prepare an employability model based on students' demographic, academic and skill details. It is also important to identify parameters affecting employability.

II. RELATED WORK AND LITERATURE SURVEY

The Economist Intelligence Unit (EIU) implies that Thai colleges are missing developed graduates that have great dialect aptitudes, specialized and data innovation abilities. In addition, both of business and workers uncover that the holes in nonspecific conduct abilities comprising correspondence, administration, social aptitudes, time overseeing, cooperation, and versatility [3] [10] [16].

Gurule et al [7] perceived that particular understudy attributes are related to their prosperity as indicated by the Grade Point Average (GPA) by utilizing a Microsoft Decision Tree grouping strategy [8]. These examinations have uncovered a few utilizations of characterization in information mining strategy of the instructive area that selections valuable material from gigantic informational collections. Information mining and systematic apparatuses can help administrators' entrance current information for the basic leadership process. [10].

Keno C. Paid, Men Chita Dumlao, Melvin A. Ball era, Shaneah C. Ambit, This paper predicts the employability of IT graduates exploitation 9 criteria. First different classification algorithms in data mining tested making logistic regression with an accuracy of 78.4% it is implemented. Based on logistic regression analysis, three academic criteria directly affect; IT_Core, IT_Professional, and Gender identified as significant predictors of employability. The data were collected based on the 5-year profiles of 515 students randomly selected at the placement office tracer study. [4]

G. Vadivu*, K. Sornalakshmi, in such manner to enhance the understudies' execution, the scholarly execution has been examined and anticipated utilizing the calculations KNN and guileless Bayes. The algorithms are used to the dataset of 250 students with 59 attributes. The accuracy achieved after analysis for KNN is 95.33% and for the naïve Bayes is 97.67%. [5]

Tripti Mishra, Dharminder Kumar, Sangeeta Gupta, here, The paper uses several classification ways of knowledge mining to predict the employability of "Master of Computer Applications" (MCA) students and discovery the algorithmic program that is best suited to this drawback. An information set is a developed with the old-style parameters like socioeconomic conditions, instructional performance, and a few further emotional talent parameters. [6]

Bangsuk Jantawan, Cheng-Fa Tsai, The aim of study presents a graduate employability model that uses Bayesian strategies to search the foremost necessary issue of graduate employability. Also, compare the accuracy of every algorithmic program beneath Bayesian strategies as well as Naïve Bayesian straightforward, Naïve Bayesian, Averaged One-Dependence Estimators, Averaged One-Dependence Estimators with subsumption (minor Promise) resolution, Bayesian networks, and Naïve Bayesian Updateable. [9][12][13]

Shahul Hameed, Nileena G.S., The pilot program was first tested in the college to ensure that the model could indeed create well-rounded students and be means of securing internships for the students. The successes of the model prove that this could be a successfully implemented in other colleges as well thus improving the employability of students and bringing the student quality on equivalence with the tier-1 college students. [7][14]

Tripti Mishra, Dharminder Kumar, and Sangeeta Gupta, in this paper talk about the work implemented in EDM in traditional education. They focused on research areas of traditional education. The core effect of student's emotional skills on academic performance and employability needs to be an explored further. Still, future work suggests a survey of various tools that are available for prediction of student's overall academic performance. [11][15]

III. PROPOSED WORKFLOW AND DATASET
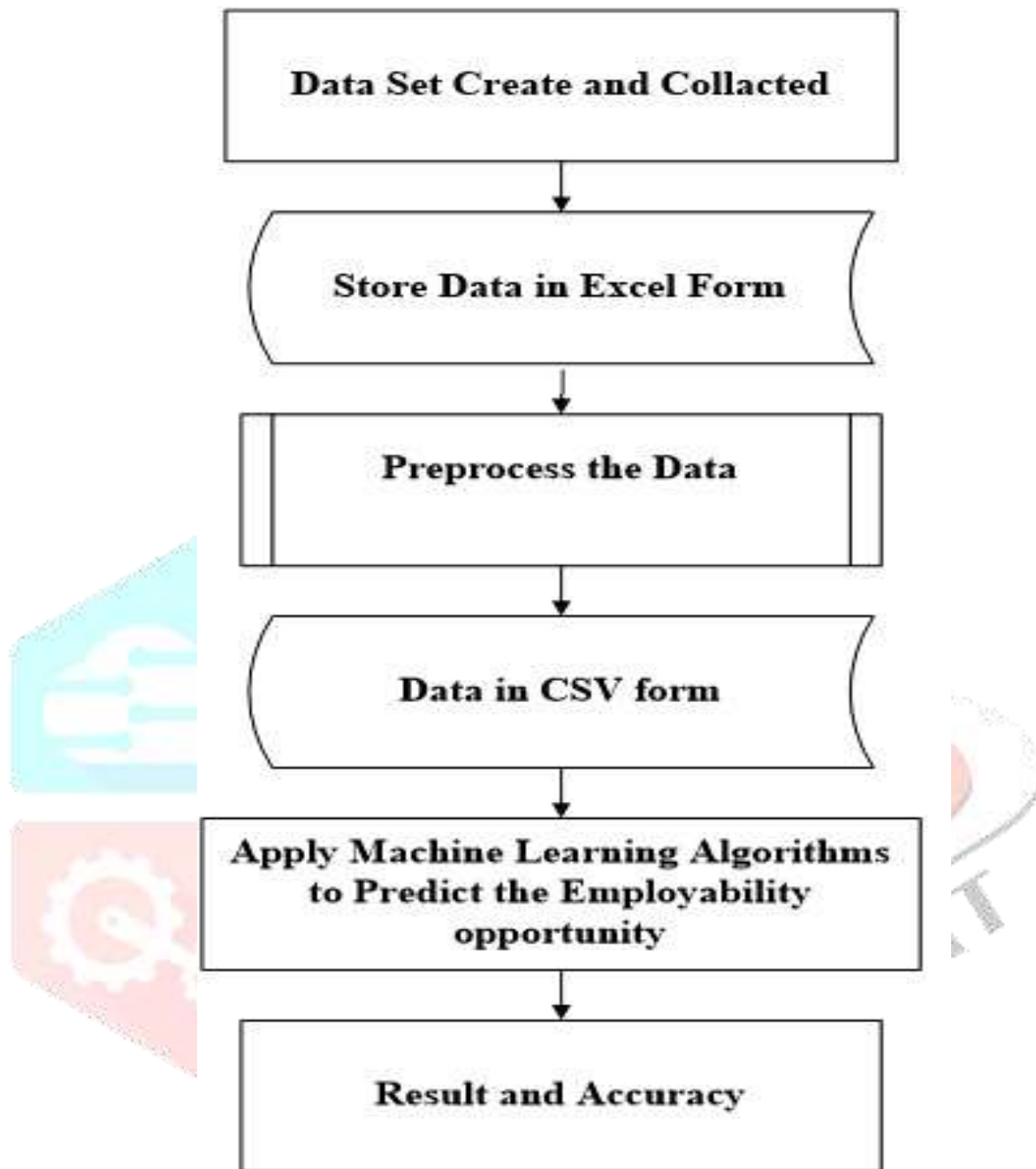
3.1 Proposed Workflow

Figure-3 Workflow

3.2 Sample Dataset
Here, we focused on following factors:

1. **Communication Skill Record**
   - CPD Grade
   - CS Grade
   - 10th Passing Board Medium
   - 12th Passing Board Medium
2. **Technical Record or Academic Record**
   - C
   - DBMS
   - C++ / OOPC
   - JAVA

- Algorithms
- Web Application Development
- Computer Network
- Distribute System
- Advance Technology

3. Aptitude Record
- 10th Passing Board Medium
- 12th Passing Board Medium
- Math's 1,2,3,4 Grades
- Programming Subjects Grades

Table-2 and 3 Companies criteria and selected student number

Sr No	1	2	3	4	5
Company Name	**TCS**	**Einfochips**	**Matrix**	**Collabra**	**L & T InfoTech**
Age Limit	18 to 28 Years	18 to 26 Years	18 to 28 Years	18 to 28 Years	<25 Years
10th %	60%	60%	60%	55%	60%
12th %	60%	60%	60%	55%	60%
Diploma CPI	6.5	6.5	6.5	6	6.5
Diploma Stream	All Branch	All Branch	All Branch	IT/CE/EC	All Branch
Degree CPI	6.5	6.5	6.5	6	7
Degree CGPA	6.5	6.5	6.5	6	7.5
Degree Stream	All Branch	All Branch	All Branch	IT/CE/EC	All Branch
Year Of Gap	<3Years	0	<1 Year	Any	0

Table-2 Companies criteria for Student's Job & Placement

Sr No	Company Name	Computer	IT
1	TCS	20	11
2	Alstom	X	X
3	Searce	1	1
4	L & T ECC	X	X
5	Info stretch	0	0
6	Streebo	0	1
7	GSFC	X	X
8	Einfochips	1	0
9	Zeus	0	0
10	Infibeam	0	0
11	IGATE	4	1
12	Matrix	0	0

Table-3 Student Company Place Record [18]

IV. METHODOLOGY

The study used different classification algorithms to generate data models. These models were used to determine how potential attributes correlated significantly to the target variable of Information Technology employability.

1. Decision Tree

A decision tree is a shape that includes a root node, branches, and leaf nodes. Each inner node denotes a check on a characteristic, each department denotes the outcome of a check, and every leaf node holds a category label. The topmost node in the tree is the foundation node. Each node in the tree is connected to one or greater nodes the usage of branches, the final node within the tree that incorporates no outgoing branches is called leaf node. The leaf node shows to termination or the results cost. The terminology of such classification technique is to preserve asking query until a conclusion is reached. The set of questions and

solutions ought to form a decision tree with set of nodes: first, root node having a 0 or extra outgoing nodes and no incoming nodes, as well as containing the testing circumstance that separate the information; 2d, everyday nodes, those nodes are inner nodes and every has one and handiest one incoming node and or greater outgoing edges. It additionally includes the testing circumstance that separate statistics and thirdly, Leaf nodes, those nodes keep the class labels, don't have any outgoing edges, and most effective one incoming side.

In the choice analysis, a selection tree and the carefully related have an impact on the diagram are used as a visible and analytical decision aid device, where the anticipated values (or predicted utility) of competing options are calculated. A decision tree consists of three types of nodes:

1. Decision nodes – Represented by squares
2. Chance nodes – Represented by circles
3. End nodes – Represented by triangles.

2. Naïve Bayer

Naïve Bayes classifier is real classifiers, which can predict class interest probabilities with the ultimate objective that the probability of a given tuple falls into a particular class. Simple Bayes classifier relies upon Bayes' speculation. Naïve Bayes is a fundamental procedure for creating classifiers: models that dole outmaneuver names to issue illustrations addressed as vectors of feature regards, where the class marks are drawn from some constrained set. It is not alone figuring for planning such classifiers, yet a gathering of counts in perspective of an ordinary standard: all honest Bayes classifiers acknowledge that the approximation of a particular segment is free of the estimation of some other component, given the class variable. For example, a characteristic item may be believed to be an apple if it is red, round, and around 10 cm in separate over. An honest Bayes classifier considers each of these features to contribute uninhibitedly to the probability that this regular item is an apple, paying little regard to any possible connections to be tween's the shading, roundness, and separation crosswise over features.

3. Bayesian Method

The classification tasks in these methods involve arranging a class variable in light of an arrangement of property factors. This is a kind of measurable examination in which the earlier dissemination is assessed from the information before any new information is watched; hence, every parameter is allotted an earlier likelihood appropriation [10]. The Naïve Bayesian calculation acts as takes after: Let D be a preparation set of tuples and their related class marks. Obviously, each tuple is spoken to by an n-dimensional property vector, $X = (x1, x2,…, Xn)$, delineating n estimations made on the tuple from n qualities, separately, A1, A2, … , An. Assume that there are m classes, C1, C2, … , Cm. Given a tuple, X, the classifier predicts that X has a place with the class with the most noteworthy back likelihood, molded on X; that is, the Naïve Bayesian calculation predicts that tuple X has a place with the class Ci if and just if $P(Ci|X) > P(Cj|X)$ for $1 \leq j \leq m; j \neq I$. Subsequently, we amplify P (Ci|X). The class Ci for which P (Ci|X) a boosted is known as the greatest posterior theory. The Naïve Bayesian Updateable algorithmic program uses a default preciseness of zero.10 for numeric attributes once build classifier is termed with zero coaching instances. The Naïve theorem easy sculptured numeric attributes by a traditional distribution.

V. CONCLUSION

In traditional education, performance prediction is in the matured state with contribution from many researchers. As both performance and employability of students graduating from an institution decide the market value of the institution, research is required to develop comprehensive models for performance and employability tool and develop a system that will be able to predict both performance and employability. From the literature review, it is clear that most commonly used predictors are a socio-economic /demographic profile and past academic record of the students.

Apart from this, a number of hours dedicated to studies, the individual researchers in their studies have considered a distance of the institution from home, loan, internet facility etc. Thus, in general researchers in the field of EDM have focused on the academic and social integration of students for performance and employability prediction. The effect of emotional skills on academic performance and employability needs to be an explored further.

REFERENCES

[1] Evaluation of the lasting impacts on employability of co-operative serious game-playing by first-year Computing students: An exploratory analysis, Jyoti Bhardwaj, 978-1-4799-3922-0/14/$31.00©2014 IEEE
[2] Economist Intelligence Unit, "Skilled Labor Shortfalls in Indonesia, the Philippines, Thailand, and Vietnam: A custom research report for the British Council", Retrieved November 29, 2013, from http://www.eiu.com

[3] Predicting IT Employability Using Data Mining Techniques, Keno C. Paid, Menchita Dumlao, Melvin A. Ballera, Shaneth C. Ambat, IEEE Systems Journal

[4] Applying Machine Learning Algorithms for Student Employability Prediction Using R, G. Vadivu, K. Sornalakshmi, Int. J. Pharm. Sci. Rev. Res., 43(1), March - April 2017; Article No. 11, Pages: 38-41

[5] Students' Employability Prediction Model through Data Mining, Tripti Mishra, Dharminder Kumar, Sangeeta Gupta, International Journal of Applied Engineering Research ISSN 0973-4562 Volume 11, Number 4 (2016) pp 2275-2282

[6] A Classification Model on Graduate Employability Using Bayesian Approaches: A Comparison, Bangsuk Jantawan, Cheng-Fa Tsai, International Journal of Innovative Research in Computer and Communication Engineering, Vol. 2, Issue 6, June 2014

[7] Gurule, H., Istanbullu, A., and Karahasan, M., "A new student performance analyzing the system using knowledge discovery in higher educational databases", Computers & Education, Vol. 55, pp. 247-254, 2010.

[8] Kumar, V. and Chadha, A., "An Empirical Study of the Applications of Data Mining Techniques in Higher Education", International Journal of Advanced Computer Science and Applications, Vol. 2, pp. 80-84, 2011.

[9] IEEE Student Quality Improvement Program: To improve the employability rate of students, Shahul Hameed, Nileena G.S., 978-1-4799-6876-3/14/$31.00c 2014 IEEE.

[10] Bangsuk Jantawan, Cheng-Fa Tsai, A Classification Model on Graduate Employability Using Bayesian Approaches: A Comparison, IIRC, An ISO 3297: 2007 Certified Organization, Vol. 2, Issue 6, June 2014

[11] Tripti Mishra, Dharminder Kumar, and Sangeeta Gupta, Indian Journal of Science and Technology, Vol 10(24), DOI: 10.17485/ijst/2017/v10i24/110791, June 2017

[12] Minaei-Bidgoli, B., Kashy, D. A., Kortemeyer, G., and Punch, W. F., "Predicting student performance: An application of data mining methods with an educational Web-based system", 33rd Frontiers in Education Conference, pp. 13-18, 2003.

[13] Gokuladas Vk, Technical and Non-Technical Education and the Employability of Engineering Graduates: An Indian Case Study, International Journal of Training and Development. May 2010

[14] Nikita Gorad, Ishani Za life, Aishwarya Nandi, Deepali Nayak, Career Counseling using Data Mining, IJESC, 2017

[15] Romero C, Ventura S. Educational data mining: a review of the state of the art. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews). 2010 Nov; 40(6):601–18. Crossref

[16] Leena Khanna, Dr. Shailendra Narayan Singh, Dr. Mansaf Alam, Educational Data Mining and its Role in Determining Factors Affecting Students Academic Performance: A Systematic Review, IEEE 2016

[17] https://en.wikipedia.org/wiki/Employability

[18] Training and placement cell, BVM Engineering College.