To Design A Software Application For Assessing The Abilities Of A Student With Respect To Employability

Prof. G.S.Thakare
Professor, Department of CSE
Sipna college of engineering and Technology
Amravti, India

Harshada Dhumone
Undergraduate Student of CSE
Sipna College of Engineering and Technology
Amravati, India

Prachi Sarap
Undergraduate Student of CSE
Sipna College of Engineering and Technology
Amravati, India

Shivani Rotey
Undergraduate Student of CSE
Sipna College of Engineering and Technology
Amravati, India

Ashwini Chavhan
Undergraduate Student of CSE
Sipna College of Engineering and Technology
Amravati, India

Sunita Metkar
Undergraduate Student of CSE
Sipna College of Engineering and Technology
Amravati, India

Abstract— In this project we proposed an automatic decision support system for TNP administrator and students. The proposed decision support system will find out student's job abilities on the basis of their IQs, so that the TNP admin will be able to decide which companies to be called for students' placement. We proposed a machine learning model by using decision tree algorithm. On the basis of student's IQ our machine learning model will find out suitable job types for every student and the recommendation model will recommend the skills that are missing in student's profile which may improve student's profile to get suitable job easily.

Keywords— Classification, Data Mining, Employability Prediction, Emotional Skills Modeling.

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 . 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. In machine learning, 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. In this project, we proposed IQ based student employability finding model using machine learning algorithms. To find out IQ based matching job types and missing courses we proposed decision tree machine learning algorithm. In addition with matching job types finding and missing courses recommendation we also proposed to develop a decision support system based on k means clustering algorithm for TNP administrator.

2.RELATEDWORK:

[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%.

[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.

[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".

[4] they investigated the perception of employers' on employability abilities of undergraduate scholars in three areas: Individual Qualities; Core Skills; and Topic Knowledge.

[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.

[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.

[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.

[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.

PROPOSED SYSTEM

In this project we proposed an intelligent system for TNP admin and students. We proposed a decision support system for TNP administrator. Our model will suggest TNP admin about how many students are their suitable for particular job type. We will use K means clustering algorithm to generate decision support system. Following is a working diagram of proposed system. In this system there are four types of users(administrator, TNP admin, students and company admin) Company admin and students will do registration by their own and administrator will approve their account so that no fake user will be able to do registration in our system. Administrator will create TNP admin login from his login. The TNP admin will be able to register job types and register IQ questions for IQ test. Students will solve IQ test and as per the result, decision tree model will suggest suitable job types and missing courses. Company admin will upload dataset according to the requirements. The dataset will be a csv file which will be stored on python server. When any user solve IQ test, control will be transferred to python server to find out suitable job types and missing courses. For decision support reports we will use k means clustering algorithm.

II. METHODOLOGY

A. Decision Tree Algorithm

Classification is a two-step process, learning step and prediction step, in machine learning. In the learning step, the model is developed based on given training data. In the prediction step, the model is used to predict the response for given data. Decision Tree is one of the easiest and popular classification algorithms to understand and interpret. In Decision Trees, for predicting a class label for a record we

start from the **root** of the tree. We compare the values of the root attribute with the record's attribute.

Decision Tree algorithm belongs to the family of supervised learning algorithms. Unlike other supervised learning algorithms, the decision tree algorithm can be used for solving regression and classification problems too.

The goal of using a Decision Tree is to create a training model that can use to predict the class or value of the target variable by learning simple decision rules inferred from prior data(training data).

In Decision Trees, for predicting a class label for a record we start from the root of the tree. We compare the values of the root attribute with the record's attribute. On the basis of comparison, we follow the branch corresponding to that value and jump to the next node.

Mathematics behind Decision tree algorithm: Before going to the Information Gain first we have to understand entropy.
Entropy: Entrop is the measures of impurity, disorder or uncertainty in a bunch examples.

Purpose of Entropy:
Entropy controls how a Decision Tree decides to split the data. It affects how a Decision Tree draws its boundaries.
"Entropy values range from 0 to 1",
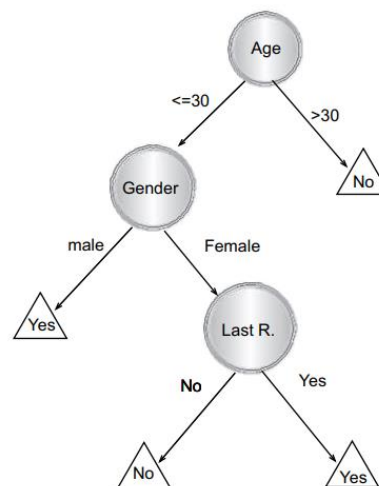Less the value of entropy more it is trusting able.

Figure:-Decision tree Algorithm

B. K- mean Algorithm

The K-means clustering algorithm computes centroids and repeats until the optimal centroid is found. It is presumptively known how many clusters there are. It is also known as the flat clustering algorithm. The number of clusters found from data by the method is denoted by the letter 'K' in K-means. In this method, data points are assigned to clusters in such a way that the sum of the squared distances between the data points and the centroid is as small as possible. It is essential to note that reduced diversity within clusters leads to more identical data points within the same cluster. K-Means Clustering is an unsupervised learning algorithm that is used to solve the clustering problems in machine learning or data science. In this topic, we will learn what is K-means clustering algorithm, how the algorithm works, along with the Python implementation of k-means clustering.

Few things to note here:

- Since clustering algorithms including k-means use distance-based measurements to determine the similarity between data points, it's recommended to standardize the data to have a mean of zero and a standard deviation of one since almost always the features in any dataset would have different units of measurements such as age vs income.
- Given k-means iterative nature and the random initialization of centroids at the start of the algorithm, different initializations may lead to different clusters since k-means algorithm may *stuck in a local optimum and may not converge to global optimum*. Therefore, it's recommended to run the algorithm using different initializations of centroids and pick the results of the run that that yielded the lower sum of squared distance.

Working of K-Means Algorithm -   The working of the K-Means algorithm is explained in the below steps:

Step-1: Select the number K to decide the number of clusters.

Step-2: Select random K points or centroids. (It can be other from the input dataset).

Step-3: Assign each data point to their closest centroid, which will form the predefined K clusters.

Step-4: Calculate the variance and place a new centroid of each cluster.

Step-5: Repeat the third steps, which means reassign each datapoint to the new closest centroid of each cluster.

Step-6: If any reassignment occurs, then go to step-4 else go to FINISH.

Step-7: The model is ready.

Applications

k-mean algorithm is very popular and used in a variety of applications such as market segmentation, document clustering, image segmentation and image compression, etc. The goal usually when we undergo a cluster analysis is either:

1. Get a meaningful intuition of the structure of the data we're dealing with.
2. Cluster-then-predict where different models will be built for different subgroups if we believe there is a wide variation in the behaviours of different subgroups. An example of that is clustering patients into different subgroups and build a model for each subgroup to predict the probability of the risk of having heart attack.

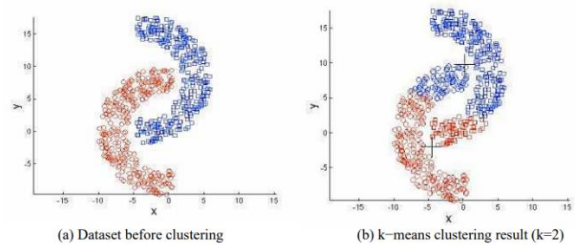(a) Dataset before clustering   (b) k−means clustering result (k=2)

Figure : k-means result for non-convex shaped clusters

IV.   Model Building

Analyzing the related works presented in this study, we notice two approaches utilized for employability prediction approaches, namely data mining and machine learning. While data mining is a subset of business analytics, it refers to exploring an existing large dataset to discover previously unknown patterns, relationships and anomalies that are present in the data. Thus, it is used to find new insights on the data. Machine learning, on the other hand, is a subset of artificial intelligence (AI), where computers analyze large datasets to 'learn' patterns that will making predictions for new data. However, the frontier between both approaches is blurry, as machine learning may use some data mining techniques to build models and find patterns, so that it can make better predictions. Data mining also can sometimes use machine learning techniques to produce more accurate analysis. Thus, in this study we look at both approaches inter changeably. Machine learning uses two types of techniques: unsupervised learning, which finds hidden patterns in input data without labeled responses, and supervised learning, which trains a model on known input and output data so that it can predict future output. Unsupervised learning has the potential to reveal interesting unknown relationship or patterns. However, in this study, since we are interested in employability prediction, that is to say, we have a well-defined target variable, supervised learning was considered. Commonly identified studies in the literature deployed supervised learning techniques. Supervised learning uses classification and regression techniques to develop predictive models. While regression techniques predict continuous responses, classification techniques predict discrete responses. Since regression would need the ability to assign a continuous score of employability, all studies define the employability prediction problem as a classification problem. Classification models classify input data into categories. Most studies define two classes, namely *employable*, and *unemployable*; some studies also define a third class to represent students that are in-between, as *unknown*.

III. RESULT ANALYSIS
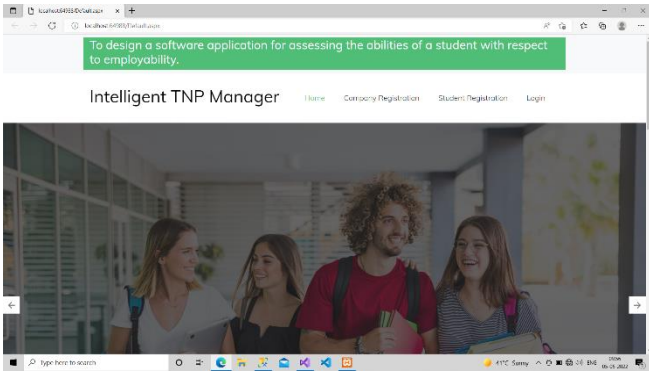
Here, is the result which shown portal of project.

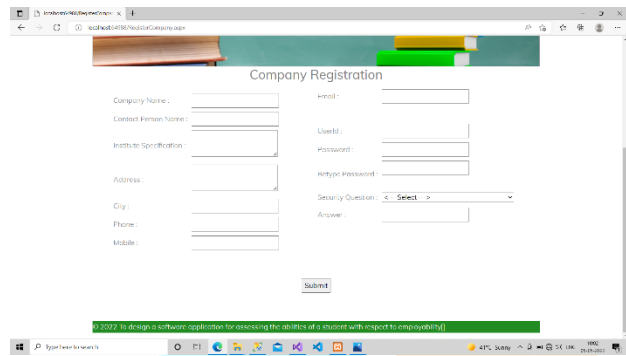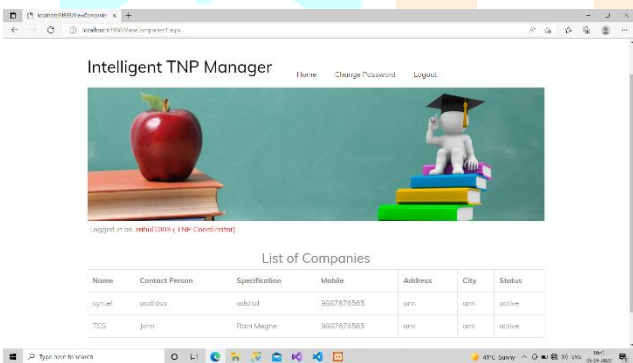

Figure: Company Registration

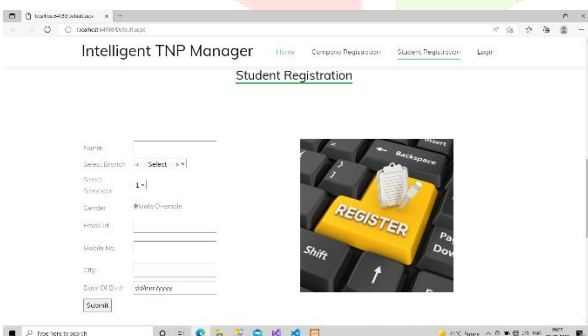

Figure: List of Companies

Figure : Student Registration
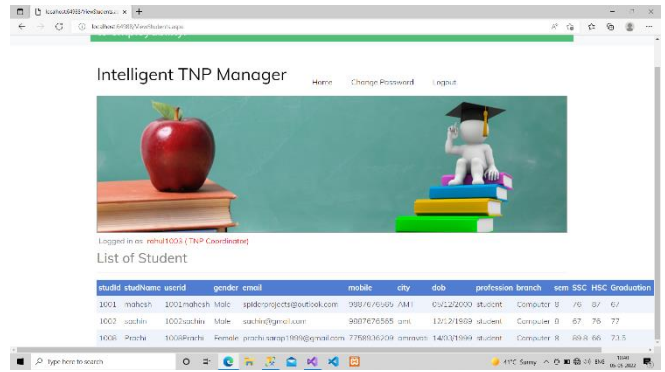

Figure: Student Registred

Figure: Student List

IV. CONCLUSION

The purpose of this research was to identify effective strategies for student abilities with respect to employability. In this project we proposed an automatic decision support system for TNP administrator and students. The proposed decision support system will find out student's job abilities on the basis of their IQ's, so that the TNP admin will be able to decide which companies to be called for students' placement. TNP admin will design IQ tests for all job types on the basis of skills required for the particular job. Students will solve the scheduled tests and as per the test result proposed model will find out student's ability with respect to employability. We proposed a machine learning model by using decision tree algorithm. Along with the decision support system we proposed a recommendation module for students. On the basis of student's IQ our machine learning model will find out suitable job types for every student and the recommendation model will recommend the skills that are missing in student's profile which may improve student's profile to get suitable job easily.

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 major concern. 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. In machine learning, 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.

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. Paris IH, Affendey LS, Mustapha N. Improving academic performance prediction using voting technique in data mining. World Academy of Science, Engineering and Technology. 2010 Feb; 62:820–3.

12. Kovacic Z. Early prediction of student success: Mining students' enrolment data. 2010; 647–65.

13. Ramaswami M, Bhaskaran R. A CHAID based performance prediction model in educational data mining. ArXivpreprint arXiv. 2010 Feb; 7(1):1–9.

14. Guruler H, Istanbullu A, Karahasan M. A new student performance analysing system using knowledge discovery in higher educational databases. Computers & Education. 2010 Aug; 55(1):247–54. Crossref

15. Delen D. A comparative analysis of machine learning techniques for student retention management. Decision Support Systems. 2010 Nov; 49(4):498–506.

16. Affendey LS, Paris IH, Mustapha N, Suleiman MN, MudaZ. Ranking of influencing factors in predicting students academic performance. Information Technology Journal. 2010; 9(4):832–7.

17. Baradwaj BK, Pal S. Mining educational data to analyze students' performance. ArXiv preprint arXiv: 1201.3417. 2012 Jan; 2(6):63–9.