placement criteria [4]. Student employability is crucial for educational institutions as it is often used as a metric for their success. Identifying the significant factors affecting employability, as well as the requirements of the new job market can tremendously help all stakeholders. Knowing their weaknesses and strengths, students might better plan their career. Program managers can anticipate and improve their curriculum to build new competencies, both for educating, training and re-skilling current and future workers. The students' future job placement is a major concern for the institutions offering higher education and a method for early prediction of employability of the students is always desirable to take timely action [5].

Machine learning techniques have been extensively used in various fields of educational data mining. More and more studies are investigating machine learning techniques for the prediction of students' future job placement after graduation

2 Related Literatures

Researchers have conducted studies in relation to students' future job placement after graduation using varied approaches. Amongst were [6] who worked to explore how university students and those who had graduated and been subsequently employed, made career decisions. Their studies employed interviews and focus group discussions with 22 university students and 28 graduates from Australian undergraduate and postgraduate courses in a variety of disciplines. The main findings of their work were that at the enrolment-stage of university and during their studies, most students were pessimistic about their career outcomes and felt largely unsupported in identifying suitable career goals. However, the outcomes after graduation were unexpectedly positive in that, by this point most had identified career goals and were in careers they had desired. The key take away from this research was a set of recommendations for universities regarding how to better support students to make career choices. [7] reported the outcomes of a survey of LIS students undertaken in Slovenia and Australia on their experience of work placements and the benefits it could bring for enhancing their personal portfolios. Students were asked to complete a survey prior to undertaking their placement which sought to determine their expectations as to the usefulness and relevance of the placement in enhancing their portfolios and subsequent career prospects. After undertaking their placement, students completed a second survey as to how well the experience fitted with their expectations and its benefits for their portfolios and professional ambitions. The results of their research confirmed what has generally been reported elsewhere - that placements provided a highly relevant educational experience that is appreciated by students and that generally lives up to their expectations.

[8] conducted a study on computing graduates into the workplace and explored their undergraduate experiences of work placements and subsequent impact on graduate employment. The study involved 14 Scottish universities, the researchers found that graduates had benefited from work experience financially, earning more than those who had not completed placements. They had also found graduate positions more quickly and were more likely to be in work than those who had graduated without completing a placement.

[9] worked on how to assist students with their digital research skills while simultaneously allowing them to research and explore college and career options. The major aim of his systematic literature review was to evaluate the effectiveness of placements for career outcomes and to identify any underpinning core psychological processes and to offer a theoretically grounded framework for future research.

[10] researched to explore the common perception of students about their education for their practical lives, also investigated the relationship between higher education and employment, and the extent higher education predicted employment for students. Number of participants used was 1,210 from public universities in Punjab. The collected data were analyzed through statistical techniques of multiple regression, correlation, t-test, and ANOVA. The results of their study revealed that most of the respondents strongly agreed with the view that higher education was for world of work. The relationship between higher education and employment found significantly positive and it was evident that higher education strongly affected and applied as predictor of employment. On the bases of their findings, it was suggested that Educational ministry must engage its strength for the expansion of higher education and encouragement of proper structural reforms in employment directions. Career guidance and services to search employment inside institutions should be provided to the students for the sake of saving their energies with long time of searching job. Students of postgraduate level should be given more opportunities for placement of work with employment experience. According to [3] the study to determine the extent at which undergraduate and post graduate students were able to gain job opportunities after graduation could predict if a

student secures full-time employment prior to graduation or not. In order to predict employment opportunities prior to graduation, the study used commonly recognized and advanced machine learning models, including logistic regression, discriminant analysis, decision trees, and neural networks. Results demonstrated that employment opportunities prior to graduation could be predicted with 73% 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.

The recent work of [11] on accessing programming skills acquisition of students in computer science and related areas. The study was conducted to look into the correlation between competency level of student in programming skills acquisition and future job placement of student after graduation. As a result of this, qualitative research design was used which targeted graduate students in computer science that are currently working in one firm or the other. 15 participants formed the sample using non-probability sampling technique approach. A structured interview question was used to collect data from the participants via WhatsApp social media platform. The findings and results showed that competency level of students in programming skills could clearly and correctly predict future job placement of student after graduation; similarly, expert skills competency level was noted to be the most powerful required of students that mostly contribute to their future jobs placement over other related skills. It was concluded that computer graduates needed to be adequately sensitized to develop in them the consciousness of acquiring the necessary programming job-related competencies before and after graduation, to enable them secure and sustain good employment in programming firms or related firms.

The paper presented by [12] used a Support Vector Machine to predict a model to determine if prior programming knowledge and completion of in-class and take home formative assessment tasks might be suitable predictors of examination performance. Student data from the academic years 2012 - 2016 for an introductory programming course was captured via ViLLE e-learning tool for analysis. The results revealed that student prior programming knowledge and assessment scores captured in a predictive model, is a good fit of the data. However, while overall success of the model is significant, predictions on identifying at-risk students is neither high nor low and that persuaded them to include two more research questions. However, their preliminary post analysis on these test results showed that on average students who secured less than 70% in formative assessment scores with little or basic prior programming knowledge in programming may fail in the final programming exam and increase the prediction accuracy in identifying at-risk students from 46% to nearly 63%. Hence, these results provide immediate information for programming course instructors and students to enhance teaching and learning process. [1] used students' behavioural data to predict their career choices. Based on the simple premise that the most remarkable characteristics of classes were reflected by the main samples of a category, they proposed a model called the Approach Cluster Centers Based On XGBOOST (ACCBOX) model to predict students' career choices. The experimental results of predicting students' career choices clearly demonstrated the superiority of their method compared to the existing state-of-the-art techniques by evaluating on 13 M behavioral data of over four thousand students.

[5] systematically reviewed the work done in the field of academic performance prediction and employability prediction of students in higher education. Their survey first explained how higher education has become an exciting field of research and why the prediction of academic performance and employability is beneficial for the institutions. They also explained briefly on how many ways higher education was being provided world-wide. The results of the study highlighted and found that prediction of academic performance had progressed a lot but employability prediction is yet to mature. It was concluded that few parameters that has not been considered so far in predicting the performance or employability should be taken seriously and resolved

In the same vein, [4] represented a placement prediction system with the help of Machine learning, in which they used a Support Vector algorithm. The major instrument used was a questionnaire via Google form which contained all students performance like there higher secondary marks diploma score, communication skills, area of interest and most important dream job on the basis they provided them necessary guidance to achieve their goal they then tracked their results each year, also their technical interest and also informed them on how much efforts they needed to achieve their dream job. Even if they missed their dream job due to company criteria the proposed algorithm would develop new data where they could suggest another alternation to students according to their achievements, their skills and also give guidelines at every phase. Hence the future scope and relevance of the system was discussed. The work of [13] used predictive modelling to student likelihood of completing a degree. If students were predicted to be most likely to drop out, interventions could be enacted to increase retention and completion rates. The researcher used University of Nevada, Las Vegas (UNLV), four-year graduation rates of 15% and six-year graduation rates of 39%. To improve these rates, they gathered seven years worth of data on UNLV students who began in the fall 2010 semester or later up to the summer of 2017 which included information from admissions applications, and academic performance. The student group which was reported federally were full-time freshmen beginning in the summer or fall. Their data set included all freshmen and transfer students within the time frame who met the criteria laid down. In the study, they applied data analysis and visualization techniques to understand and interpret the data set of 16,074 students. Predictive modelling such as logistic regression, decision trees, support vector machines, and neural networks were applied to predict whether a student would graduate. In their analysis, decision trees gave the best performance. The work of [14] focused on a comprehensive roadmap, enabling the application of data mining for employability.

In complementing the previous studies, [5] used different classification techniques of data mining, like Bayesian methods, Multilayer Perceptrons and Sequential Minimal Optimization (SMO), Ensemble Methods and Decision Trees, to predict the employability of Master of Computer Applications (MCA) students and found the algorithm which was best suited for this problem. In the course of this, a dataset was developed with the traditional parameters like socioeconomic conditions, academic performance and some additional emotional skill parameters. A comparative analysis concluded that J48 (a pruned C4. 5 decision tree) was most suitable for employability prediction with 70. 19% accuracy, easy interpretation and model building time(0. 02Sec) less than Random Forest, which had slightly better prediction accuracy (71. 30%), higher building time(0. 11) and difficult interpretation. Further, Empathy, Drive and Stress Management abilities are found to be the major emotional parameters that affect employability. In their paper, [15] evaluated and compared the performance of three machine learning classifiers: Support Vector Machines (SVM), Decision Trees (DT) and K-Nearest Neighbor (K-NN) for high resolution satellite image scene classification. In this study, the aim provided insights into the selection of the appropriate classifier and highlighting the importance of the appropriate setting of the classifier parameters. Illustration was made towards the issues of applying scene classification to UC-Merced high resolution satellite image dataset. Image features were obtained through the SURF descriptor and BOVW model.

[16] conducted a comparative study of four well-known supervised machine learning techniques namely; Decision Tree, KNearest-Neighbor, Artificial-Neural-Network and Support Vector Machine respectively. In his paper concentration was channelled towards the key ideas of each technique and its advantages and disadvantages. Practical application was conducted to compare their performance. Some measures were used for evaluating their performance, such as sensitivity and specificity. This study showed that there was no one measure could provide everything about the classifier performance and there was no such classifier that can satisfy all the criteria.

[17] presented paper on the comparative study of six classification algorithms which includes Logistics Regression, Support vector machine, Random forest, K-Nearest Neighbor, Decision Tree and Gaussian Naïve Bayes. Two different smart home datasets were generated and used to train and test the algorithms. The confusion matrix was used to evaluate the outputs of the classifiers. From the confusion matrix, Prediction Accuracy, Precision, Recall and F1-Score of the models were calculated. The Support Vector Machine (SVM) outperformed the other algorithms in terms of accuracy on both datasets with values of 67.89 and 88.56 respectively. The SVM and Logistics Regression also maintained the highest precision of 100.0 as compared to the other algorithms. [18] used various algorithms like KNN, Naïve Bayes, support vector machine (SVM), decision trees and random forest to ascertain their statistical and mathematical aspects of each algorithm, and suitability of the algorithms to certain use cases and the main drawbacks of the corresponding algorithms.

3 Materials and Methods

Students' previous years' historical data in form of gross point average (GPA) was used as dataset for the research work with the use of Dataset Collection Template (DCT). GPA of 20 students was used. Machine learning including: Decision tree algorithm, Support Vector Machine, and K- Nearest Neighbours algorithm were used as techniques to predict the students' future job placement after graduation. It was established that the GPA results used were from validated and reliable sources. Graduating students GPA results was purposefully collected to predict the future job placement of students after graduation

4 Analysis and Results

Table 1: Comparison of classifiers with Time

Time/Classifier	Decision Tree	Support Vector Machine	K-Nearest Neighbour
Time taken to build model	0.08 seconds	0.7 seconds	0 seconds
Time taken to test the model on test split	0.01 seconds	0.09 seconds	0.14 seconds

Table 1 compares the time taken to build model and the time taken to test the model on the test split. K-Nearest Neighbour had the least time taken to build the model; next to this is Decision Tree while Support Vector Machine time taken to build the model is the longest. Similarly, Decision tree had the least time taken to test the model on test split, followed by SVM and finally the KNN in ascending order.

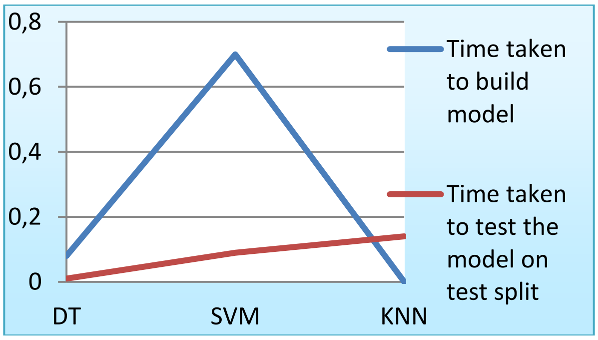

Figure 1: Comparison of time taken to build and to test models

Table 2: Accuracy, Recall, Precision and Statistical errors comparison

	Decision Tree	Support Vector Machine	K-Nearest Neighbour
Correctly Classified Instances	17.9215	9.0559	12.265
Incorrectly Classified Instances	2.3116	11.1772	7.9681
Kappa statistic	0.8256	0.1932	0.3849
Accuracy	88.5753 %	44.7581 %	60.6183 %
Mean absolute error	0.0762	0.3548	0.2747
Root mean squared error	0.276	0.4394	0.4974
Relative absolute error	17.017 %	79.2792 %	61.3695 %
Root relative squared error	58.0579 %	92.4314 %	104.6434 %
Precision	0.896	0.296	0.621
Recall	0.886	0.448	0.606
F-Measure	0.885	0.349	0.606

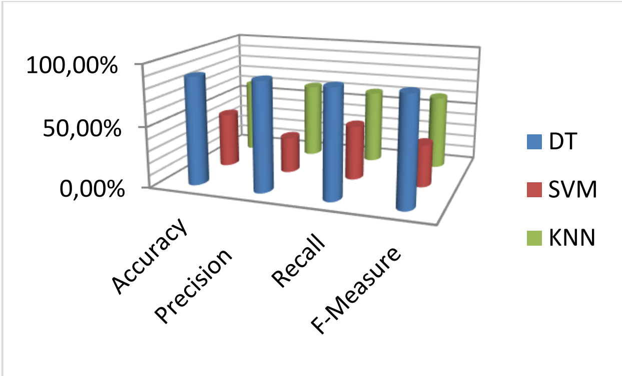

Figure 2: Comparison of Decision Tree, Support Vector Machine and K-Nearest Neighbors with respect to accuracy, precision, recall and F-measure

The accuracy, precision, recall and F-measure across the three classifiers showed that Decision tree had the best accuracy, precision, recall as well as F-measure statistics – 89%, 90%, 89%, and 89% respectively. In comparison with K-Nearest Neighbour which had better accuracy, precision, recall as well as F-measure statistics -61%, 62%, 60%, and 61% respectively. Support Vector Machine had the least accuracy, precision, recall as well as F-measure statistics – 45%, 30%, 45%, and 35% respectively. Likewise, correctly classified instances in Decision tree showed the best classification among others with 18, next was KNN with 12, while SVM had the smallest classification with 9 instances. Conversely, SVM had the highest incorrectly classified instances with 11; next to it was KNN, while Decision tree had the least. Going through the errors from the analysis, it showed that Decision tree had the required least errors that made the classifiers accurate than other classifiers – MAE (0.08), RMSE (0.3), RAE (17%), and RRSE (58%) respectively. The error reported by other two classifiers proved that the two classifiers did not perform to the expectations for instance, SVM classifier had - MAE (0.35), RMSE (0.44), RAE (79%), and RRSE (92%) respectively. Hence, KNN classifier had - MAE (0.27), RMSE (0.50), RAE (61%), and RRSE (105%) respectively

Table 3: Comparison of classifiers on Confusion matrices

Decision Tree		Support Vector Machine			K-Nearest Neighbour			
4.49	0	2.69		0	4.49		0	0.9
0	6.36	0	0	6.36	0	0	2.12	4.24
0	1.41	2.83	7.07	5.66	0	0	2.83	5.66

The comparison of classifiers on table 3 depicted that Decision tree had incorrect classification of (0.9 +1.41+0+0+0+0 = 2.31), Followed by SVM that had incorrect classification of (2.69 +2.83+5.66+0+0+0 = 11.18), while KNN had incorrect classification of (0.9 +4.24+2.83+0+0+0 = 7.97)

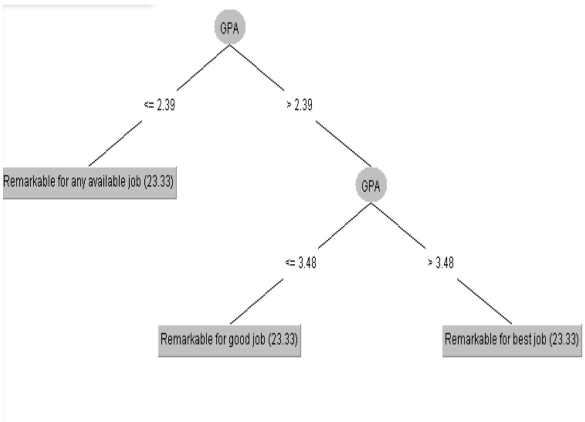

Figure 3: Decision Tree Visualization

The figure 3 complementing accuracy, precision, recall and tree visualization of decision tree with conditions to place students to different three levels of job placements (Remarkable for best job, Remarkable for good job, and Remarkable for any available job) all these were based on the GPA of the students

5 Discussion

It was observed from the analysis that data science techniques used to predict future job placement of students after graduation included Decision Tree, Support Vector Machine, and K-Nearest Neighbour. The findings proved that Decision Tree outperformed other two classifiers, though KNN had the least time taken to build the model.

In comparison, the three classifiers regarding the accuracy, precision, recall, and F-measure; the result showed that Decision Tree had the best classification correctness as well as accuracy measure which included precision, recall, F-measure statistics with – Accuracy (89%), Precision (90%), Recall (89%), and F-measure (89%) respectively. Also, the minimal errors were reported in Decision Tree classifier than other two classifiers.

On confusion matrix, Decision tree had incorrect classification of (0.9 +1.41+0+0+0+0 = 2.31), Followed by SVM that had incorrect classification of (2.69 +2.83+5.66+0+0+0 = 11.18), while KNN had incorrect classification of (0.9 +4.24+2.83+0+0+0 = 7.97). From these statistics, Decision Tree had the minimum of incorrectly classified instances compared to other classifiers

6 Conclusion

The study concluded that students should keep pace with the times, broaden their horizons, learn more knowledge, and adapt themselves with the future employment environment. Gaining job opportunity /employability of students graduating from an institution always determine the value of the institution, researches is required to develop comprehensive models for predicting future job placement with employability tool and develop a system that will be able to predict accurate students' future job placement after graduation. Students will be clear about their career growth and what various options are available in the profession and how far they can improve themselves.

References:

[1] Nie, Min; Xiong, Zhaohui; Zhong, Ruiyang; Deng, Wei & Yang, Guowu, Career Choice Prediction Based on Campus Big Data—Mining the Potential Behavior of College Students, *Applied Science* 2020, 10, 2841; doi:10.3390/app10082841 www.mdpi.com/journal/applsci

[2] Fizer, Darren, Factors Affecting Career Choices of College Students Enrolled in Agriculture, *A Research Paper Presented for the Master of Science in Agriculture and Natural Resources Degree at the University of Tennessee, Martin*, December, 2013

[3] Hugo, Linsey S., A Comparison of Machine Learning Models Predicting Student Employment, *A thesis presented to the faculty of Ohio University, 2018*

[4] Viram, Rathi; Sinha, Swati; Tayde, Bhagyashree & Shinde, Aakshada, Placement prediction system using machine learning, *International Journal of Creative Research Thoughts (IJCRT)*, April 2020, Volume 8, Issue 4, pp. 1507 -1515

[5] Mishra, Tripti; Kumar, Dharminder & Gupta, Sangeeta, Students' Performance and Employability Prediction through Data Mining: A Survey, *Indian Journal of Science and Technology*, June 2017, *Vol 10(24), DOI: 10.17485/ijst/2017/v10i24/110791*

[6] Kinash, S., Crane, L., Capper, J., Young, M., & Stark, A. (2017). When do university students and graduates know what careers they want: A research-derived framework. *Journal of Teaching and Learning for Graduate Employability*, 2017, 8(1), pp. 3– 21

[7] Juznic, Primoz & Pymm, Bob, Students on placement: A comparative study. New Library World, 2011 112(5/6). Doi: 10.1108/03074801111136284

[8] Smith, Sally; Taylor-Smith, Ella; Smith, Colin F & Webster, Gemma, The impact of work placement on graduate employment in computing: Outcomes from a UK-based students, *International Journal of Work-Integrated Learning*, 2018, 19(4), pp. 359-369

[9] Bowers, A. E., Life after Graduation: Exploring the Utility of a College and Career Research Project in Defining Students' Post- High School Goals, (Doctoral dissertation), 2017 Retrieved from https://scholarcommons.sc.edu/etd/4527

[10] Ali, Muhammad Shabbir & Jalal, Hina, Higher Education as a Predictor of Employment: The World of Work Perspective. *Bulletin of Education and Research*, August 2018, Vol. 40, No. 2 pp. 79-90

[11] Muraina, Ismail Olaniyi, Adesanya, Olayemi M., & Agoi, Moses A., Correlation between Students Programming Skills Competency Level and Job Placement after Graduation, *4th International European Conference On Interdisciplinary Scientific Research*, August 8-9, 2021/ Warsaw, Poland. 262 - 270

[12] Kumar, Veerasamy Ashok; D'Souza, Daryl; Lindén, Rolf & Laakso, Mikko-Jussi, Prediction of Student Final Exam Performance in an Introductory Programming Course: Development and Validation of the Use of a Support Vector Machine-Regression Model. *Asian Journal of Education and e-Learning*, 2019, 7(1),

[13] Ploutz, Elliott Collin, Machine Learning Applications in Graduation Prediction at the University of Nevada, Las Vegas. UNLV Theses, Dissertations, Professional Papers, and Capstones, 2018, http://dx.doi.org/10.34917/13568668

[14] Mezhoudi, Nesrine; Alghamdi, Rawan; Aljunaid, Rim; Krichna, Gomathi & Düştegör, Dilek (2021). Employability prediction: a survey of current approaches, research challenges and applications, *Journal of Ambient Intelligence and Humanized Computing*, 2021, https://doi.org/10.1007/s12652-021-03276-9

[15] Bouteldja, Samia & Kourgli, Assia, A comparative analysis of SVM, K-NN, and decision trees for high resolution satellite image scene classification, *Twelfth International Conference on Machine Vision*, 2020, DOI: 10.1117/12.2557563

[16] Mohamed, Amr E., Comparative Study of Four Supervised Machine Learning Techniques for Classification, *International Journal of Applied Science and Technology*, June 2017, 7(2), pp. 5-18

[17] Musa, Martha Ozohu & Oghenekaro, Linda Uchenna, Comparative Analysis of Selected Machine Learning Algorithms Based On Generated Smart Home Dataset. European Journal of Computer Science and Information Technology, 2021, 9(4), pp.42-53, https://doi.org/10.37745/ejcsit.2013

[18] Kumar, K. Uma Pavan; Gandhi, Ongole; Reddy, M. Venkata & Srinivasu, S. V. N, Usage of KNN, Decision Tree and Random Forest Algorithms in Machine Learning and Performance Analysis with a Comparative Measure, *Conference on Machine Intelligence and soft computing*, January, 2021, 473-479

Contribution of individual authors to the creation of a scientific article (ghostwriting policy)

Author Contributions: Please, indicate the role and the contribution of each author:

Example

Ismail Olaniyi Muraina drafted the study, carried out the analysis and implementation of the algorithms and did the final editing and proof-reading of the manuscript.

Moses Adeolu Agoi collected materials, collated and interpreted the data collected. Also, proofread the manuscript

Benjamin Oghomena Omorojor did literature review and referencing.