STUDENT PERFORMANCE PREDICTION USING ADVANCED MACHINE LEARNING TECHNIQUES

Ch.Akhil, Student, Department of Electronics and Computer Engineering, Vignan's Institute of Information Technology, India.
V.Divya, Student, Department of Electronics and Computer Engineering, Vignan's Institute of Information Technology, India.
V.Madesh, Student, Department of Electronics and Computer Engineering, Vignan's Institute of Information Technology, India.
D.Charvika, Student, Department of Electronics and Computer Engineering, Vignan's Institute of Information Technology, India.
S.Siva Prathyusha, Assistant Professor, Department of Electronics and Computer Engineering, Vignan's Institute of Information Technology, India.

Abstract:
Selecting an relevant career is one among the foremost major decisions for the scholars and with the rise within the number of career opportunities and paths, making this decision became most difficult for the scholars . this may help them in repose on their performance and motivating their interests in order that they're going to be directed towards their chosen career and obtain settled therein .Also these quite career recommender systems help the recruiters while recruiting the candidates during which job role they ought to be kept in supported his/her performance and other analysis . These paper mainly consider career area prediction of scholars .In this we use Advanced machine learning techniques and algorithms.

Keywords— *Advanced machine learning, career prediction, learning algorithm, Decision tree, SVM, One-hot encoder, XG boost.*

INTRODUCTION:

There are many learning fields during this technology era. during this era multiple ways of applying machine learning methodologies for varied educational purposes. One of the most focuses of those fact finding is to spot high risk students also on identify features which influence the performance of scholars .As we all know there are several career options in every single course structure .This creates hesitation to the scholars who are studying in X or XII standards choose one career option. the most cause for his or her confusion is that they are unaware of there self-talent ,self-personality traits and another career options. This kind of confusing results in choose them a wrong career option .Finally the resultant consequences of the erroneous decision might be work poor performance ,work dissatisfaction etc.so it's essential to summation the scholar performance from budding stage of their learning path and identify their attentiveness and assess how closely they're to their desire ,whether they are within the fair path or not, and this assess directs towards their target.

Not just for students this career prediction also can be useful for recruiter while in process of recruiting the candidate ,they can easily evaluate the candidate in several parameters as per the recruitment of the work role .There are many sort of roles like software architect ,software developer, product manager etc .All the roles required some imperative knowledge in them to rent for a specific job role therefore the recruiter can look over the candidate skills ,talents accordingly .These quite prediction system make hiring process easy.

Fig 1: Overview of various Advanced Machine Learning Algorithms

LITERARTURE SURVEY:

[1]. Roshani Ade & P. R. Deshmukh (2014). In this paper for classification of students using psychometric tests. They used incremental naive bayes algorithm. And the results were TPRate_0.896, FP Rate_0.01, Precision_0.903, Recall_0.896, F-Measure_0.893 and ROCArea_0.99. In future naïve bayes algorithm can be used as a weak classifier in the ensemble concept for incremental learning.[2]. Ahmad F. Subahi (2018). He proposes a data collection strategy to build the required career path prediction dataset for a promising data driven system. A new artificial neural network (ANN) approach for career path prediction was used.[3]. Beth Dietz-Uhler & Janet E. Hurn (2013). So, they have used a learning analytics to predict student success through a perspective of faculty. In this paper, they defined about learning analytics, how educational institutions has been used it, what learning analytics tools are available and how faculty can make use of data in their courser to improve the performance of students.[4]. Min Nie, ET AL (2020). In past, professional career appraisers used questionnaires to suggest the best career path for a student, instead of that they have created a career choice prediction based on campus big data mining the potential behaviour of college students. Algorithm used is XGBOOST (ACCBOX). Accuracy of ACCBOX was 0.638.[5]. Amer Al-Badarenah & Jamal Alsakran (2016). As we know that there are recommendation systems for the recommendation purpose while online shopping, movies, songs, etc. In that way they have created an automated recommender system for course selection which will be easy for students to choose the right subject for them.[6]. Nikita Gorad, ET AL (2017). Keeping in mind that selecting the right career is one of the important decisions. Some students end up selecting wrong decision. For that purpose, they have created a career counselling model using data mining. They used adaptive boosting algorithm which gave around 94% of accuracy.[7]. Lakshmi Prasanna & DR.D.Haritha (2019). Keeping recommender system in mind, they have created a smart career guidance and recommendation system. This paper proposes feasible predictions for student's field selection based on their marks and choice of interest. Ten to eleven machine learning algorithms were used for the predictions. In which logistic regression gave 82% accuracy. In future we can use clustering methods for better understanding.[8]. K. Sripath Roy, ET AL (2018). They have created a student career prediction model using advanced machine learning techniques. Algorithms used are support vector machine (SVM), XG boost and decision tree. SVM gave more accuracy with 90.3 percent and then the XG Boost with 88.33 percent accuracy.[9]. Mubarak Albarka Umar (2019). A case study of student academic performance prediction using artificial neural networks was presented. This study presents a neural network model capable of predicting student's GPA using students' personal information, academic information, and place of residence. Thus, the model correctly predicts 73.68% of student performance and specifically, 66.67% of students that are likely to dropout or experience delay before graduating.[10]. Ezenkw.C.P, ET AL (2017). In this paper, an Automated Career Guidance Expert System (ACGES) has been developed using case-based reasoning (CBR) technique. AC-GES is to assist high school students in choosing career paths that best

suit their abilities based on their previous performances in some selected subjects, using Nigerian students as a case study.[11]. Sudheep Elayidom, ET AL (2009). They have applied data mining on dataset using statistical techniques for career selection. This will help the students in a great way in deciding the right path for them for a bright future. The software developed is simple to use besides being reasonably accurate. Moreover, the user-friendly interface used in this project turns out to be easy to handle and avoid complications.[12]. Maha Nawaz, ET AL (2014). In this paper they have created an automated career counseling system for students using case-based reasoning (CBR) and J48. This model presents an automated system that copies a one-to-one meeting with a professional career counselor. Out of the two algorithms tested, CBR gave the highest accuracy and Decision tree J-48 gave the lowest accuracy. The results indicate that the system is capable of correctly proposing majors with approximately 80% accuracy when presented with sufficient data and features.[13]. Leaf Abu Amirah, ET AL (2016). They have used data mining technique in educational data to predict student's academic performance using ensemble methods. They have used bagging, boosting and random forest (RF) and set of classifiers such as artificial neural network, naïve Bayesian and decision tree. The obtained results reveal that there is a strong relationship between learner's behaviours and their academic achievement.[14]. Vivek Kumar Mourya, ET AL (2020). They have created a career guide application using machine learning. Through this application students can easily choose a best career path for them. The machine learning algorithm used for predicting is a clustering algorithm named as K-means algorithm.[15]. Ye Liu, ET AL (2016). They have created a career path prediction model for career path instead of going to the fortune tellers. They have collected the information from various social networks. And the future work is to extend the model to consider the source descriptiveness and learn the source confidence adaptively.

METHODOLOGY:

Data Collection :

Data collection is that the major part to create machine learning model. Data collection is that the process of collecting and measuring data on selected variables in a long time system. So, the algorithms efficiency and reliability depends upon the correctness and quality of knowledge gathered. for college kids future performance prediction many parameters are needed like students academic grades in various subjects, specializations, programming and logical capabilities, other details like interests, sports, competitions, hackathons, workshops, certifications and lots of more. Data is collected in some ways . Some data is collected from employees working in several corporations, some amount of knowledge is gathered through various social media platforms, some amount of knowledge is randomly generated and other from college academic database. As of these factors play crucial role choose student improvement towards a career, of these are taken into consideration.

Data Pre-processing:

Data pre processing may be a important and necessary stage whose main aim is to accumulate final data sets which will be considered correct and useful for further algorithms. Data pre processing describes any sort of processing executed on data to organize it for an additional processing method. Collecting the info is one task and making that data applicable is another crucial task. Data gathered from various means are going to be in an unorganized format and there could also be lot of null values, invalid data values and unwanted data. Even data collected may contain entirely garbage values. it's going to not be in exact format or way that's propose to be. All such cases must be confirm and replaced with another values to form data purposeful and useful for further processing. Cleaning of these data and replacing them with relevant or estimated data and removing null and missing data and restoring them with some fixed alternate values are the essential steps in pre processing of knowledge . data pre processing modifies the info into a format which will be more easily and effectively processed for the aim of the user. Data must be kept during a categorize format.

One-hot Encoding:

One hot encoding may be a method by which unreserved variables are become a form that would be provided to Machine learning algorithms to try to to a far better add prediction. One Hot Encoding may be a technique by which categorical values present within the data gathered are converted into numerical or other ordinal format in order that they will be provided to ML algorithms and obtain better output of prediction. This method works fine with most machine learning algorithms. Few algorithms like random forest handle categorical values alright . In such cases One Hot encoding isn't required. Process of 1 Hot encoding could seem difficult but latest day machine learning algorithms lookout of that. One Hot encoding is required for proper representation of definite elements of a variable. It makes representation of categorical variables to be more meaningful.

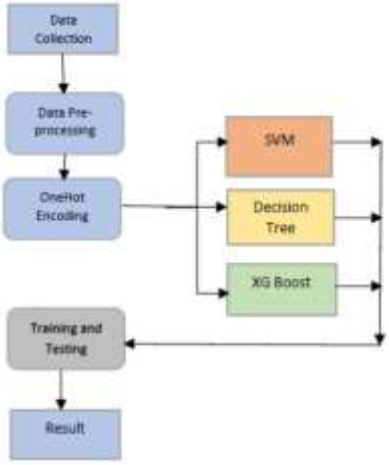


Fig 2: Process Flow Diagram of proposed system

MACHINE LEARNING ALGORITHMS:

SVM:

SVM may be a branch of machine learning and SVM is additionally called as support vector machine or support techniques that's related to the training algorithms which analyze the dataset used for classification and multivariate analysis . It is one of the acceptable methods of linear and nonlinear data type. Each student has various identifiers, and every of them is represented as a multidimensional items. Hyperplane separates the info from one class to a different class. SVM find the hyperplane and using vectors and edges. Many data analysts says that SVM takes longer for training; but the accuracy level is high compared with other techniques. SVM approach is best method when the training samples are classified by using great deal of arguments. Using nonlinear approach compare selected argument of 1 student with others to forecast the performance of the scholars . The sample data falls on HP(hyperplane) are called as support vectors. Normally HP value is closer to MMH (maximum marginal hyperplane).

Implement the ensemble support vector machine (ESVM) classification technique, which was built by combining multiple diversity structures of SVM classifiers and thus has high generalization efficiency and classification precision. The proposed SVM ensemble learning model is formed from two different SVM classifier structures and five different kernel functions. The diversity of the ensemble members, especially , is predicated totally on different kernel function options and therefore the structure of the SVM classifiers.

Every educational institute now a days requires an efficient student academic performance prediction model. However, resolving data quality problems in student success prediction models is usually the foremost difficult task. This study proposed a model for predicting student performance hooked in to the supervised learning technique ensemble SVM.

XGBOSST:

XGBOOST was introduced by TIANQI CHEN, a great machine learning library. It extends Grade Boosting Machines, which has been conceded as one of the most effective supervised literacy

algorithms. XGBoost stands for “Extreme Gradient Boosting," according to the abstract. Grade Boosting Machines are one of the stylish performing algorithms in supervised literacy, and XGBoost is an extension of them. XGBoost stands for “Extreme Gradient Boosting," according to the abstract. The model and parameters are the abecedarian aspects of XGBoost, as it's a supervised literacy strategy. The model is a fine model that's used to prognosticate issues grounded on input values, and the parameters must be learned from the data set. We are trying to produce an objective function to quantify the performance of a particular model with specific parameters since we need to determine the optimal parameters for a specific training data set. obj (Θ) = L (θ) Ω (Θ). The loss of the training and the regularization element are two crucial factors of the objective function. The training loss is a metric that measures how well the model fits the training data. The model's complexity is measured via regularization.

Decision Tree:

A Decision Tree may be a supervised learning algorithm. Decision Tree algorithm can be wont to solve both regression and classification problems in Machine Learning. that's why it's also referred to as CART or Classification and Regression Trees. because the name suggests, in Decision Tree, we form a tree-like model of decisions and their possible consequences. Decision trees laid basic foundation for several advanced algorithms like bagging, gradient boosting and random forest. The XG Boost algorithm discussed above is that the advanced version of this general decision tree. The commonly used decision trees are CART,C4.5,C5 and ID3.

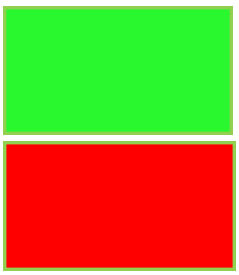
The typical scenario that a choice tree follows is first selecting a root node. Calculate information gain or entropy for every of the nodes before the split. Select the node that has more information gain or less entropy. Further split the node and reiterate the method . the method is iterated until there's no possibility to separate or the entropy is minimum.

TESTING OF DATA:

Later Performing training on the data our ultimate goal is for testing it . Here the performance of the algorithm , required output also as the info quality appears as output . From the given data set 80 percent of the info is employed at testing purposes and therefore the remainder of the 20 percent of the info is employed at Conclusion of the tested data . Training is defined because the capability of creating further predictions on which the training it taken within the process of creating machine to learning .Coming with testing it's already initially defined with a predefined data set with output also of previous labelled data and this model is tested whether it's perfectly working or isn't working and checks whether it gives the accurate prediction or not. When most predictions are correct then the model is about to possess an excellent accuracy of percentage and said to be reliable to vary with other models too. If there's any addition of latest data to the prevailing model data set then the model set are going to be more accurate and powerful for further predictions.

RESULT AND DISCUSSIONS:

The data is performed in all the algorithms and is perfectly trained and tested with best accuracy . Among these more accuracy is found in SVM of 91.2 percent and XG Boost of 89.33 percent. So the algorithm with high accuracy will be preferred for the further data predictions.



Correctly Predicted

Incorrectly Predicted

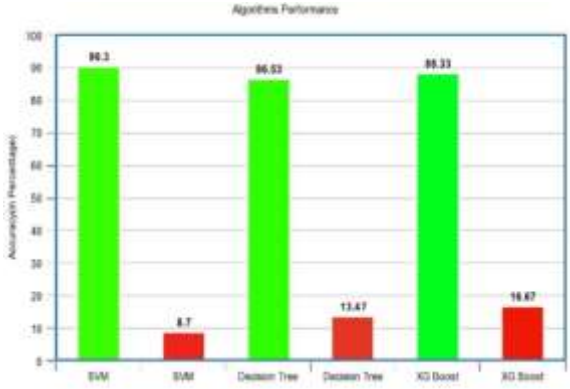


Fig 3: Final Output Graphs

CONCLUSION AND FUTURE WORKS:

By using the machine learning algorithm is a confront to predict student's performance by assessing their abilities from initial standards of learning platform .In addition ,the excessive of various datasets that are stepped in various types and also provides opportunities at the same period of time for the expansion of the use of machine learning algorithm for prediction of student's performance .The conclusion from this literature are expected to put up to understanding for Educational institutions to be able to use the most relevant machine learning algorithm build on their needs and datasets to make certain achievement of the utmost learning goal, specially the learning success of candidate.

They are often used for filtering the candidates for the work roles which provides more impact for many of the company's . Not only within the Company's but also within the education institutes it's easy to prefer students with a more accurate percentage for a high job required notifications .

***REFERENCES*:**

1. Dr . N. Mangathayaru, Dr . M. Sreenivasa Rao, V. Krishna Rao, Prediction and Evolution of Radical Multi Dimensional E-Learning System with Clustering data Processing Techniques, IOP Conf. Series: Materials Science and Engineering 1055 (2021) 012122.
2. Ali Daud, Naif Radi Aljohani, "Predicting Student Perfor-mance using Advanced Learning Analytics", 2017 International World Wide Web Conference Committee (IW3C2).
3. Marium-E-Jannat,SaymaSultana,Munira Akther, "A Probabilis-tic Machine Learning Approach for EligibleCandidate Se-lection", International Journal of Computer Applications (0975 – 8887)Volume 144 –No.10, June 2016.
4. Sudheep Elayidom, Dr. Sumam Mary Idikkula, "Applying Data mining using Statistical Techniques for Career Selec-tion", International Journal of Recent Trends in Engineer-ing, Vol. 1, No. 1, May 2009.
5. Dr. Mahendra Tiwari ,Manmohan Mishra, "Accuracy Esti-mation of Classification Algorithms with DEMP Model", International Journal of Advanced Research in Computer and Communication Engineering Vol. 2, Issue 11, November 2013.
6. RutvijaPandya Jayati Pandya , "C5.0Algorithm to Im-proved Decision Tree with Feature Selection and Reduced Er-ror Pruning", International Journal of Computer Applications (0975 –8887) Volume 117 –No. 16, May 2015.
7. AhmedSharafElDen ,Malaka A. Moustafa, Hany M. Harb and AbdelH.Emara (2013), "Adaboost Ensemble With Simple Genetic Algorithm For Student Prediction Model", International Journal of Computer Science & Information Technology (IJCSIT) , Vol. 5, No 2, pp 73-85.
8. S.A. Oloruntoba1 and J.L.Akinode (2017), "Student Academic Performance Prediction Using Support Vector Machine", International Journal Of Engineering Sciences & Research Technology, pp. 588-598.

9. AnkitaKadambande, Snehal Thakur, AkshataMoholand A.M.Ingole (2017), "Predict Student Performance by Utilizing Data Mining Technique and Support Vector Machine", International Research Journal of Engineering and Technology (IRJET), Vol. 04, No. 05, pp 2818-2821.
10. K. B. Eashwar, R. Venkatesan and D. Ganesh (2017), "Student Performance Prediction Using SVM", International Journal of Mechanical Engineering and Technology (IJMET) Vol. 8, No. 11, pp. 649–662.
11. Min Nie , ZhaohuiXiong , RuiyangZhong , Wei Deng and Guowu Yang (2020), " Career Choice Prediction Based on Campus Big Data—Mining the Potential Behavior of College Students" , Journal of Appl. Sci. pp. 1-14.
12. Bo Guo , Rui Zhang, "Predicting Students Performance in Educa-tional Data Mining",International Symposium on Educational Technology Vol.143, No.8 Augest 2016.
13. Anuj Karpatne, Gowtham Atluri, "Theory- Guided Data Science: A New Paradigm for Scientific Discovery from Data", IEEE trans-actions on knowledge and data engineering, vol.29, no. 10,october 2017.
14. Subahi A., F. (2018). Data Collection for Career Path Prediction Based on Analyzing Body of Knowledge of Computer Science Degrees. Journal of Software. Volume 13.
15. Ezenkwu C.P., Johnson E.H., Jerome O.B. (2017).Automated Career Guidance Expert System Using Case-Based Reasoning Technique. Cisdijournal. Volume 8, No. 1.
16. Umar M. A. (2019). Student Academic Performance Prediction using Artificial Neural Networks: A Case Study. International Journal of Computer Applications (0975 – 8887) Volume 178..
17. Mourya V., Phatale S., Thakur S., Mane P. (2020). Career Guide Application using ML.International Research Journal of Engineering and Technology (IRJET). Volume: 07 Issue: 09.
18. Uhler B. D., Hurn J. E. (2013). Using Learning Analytics to Predict (and Improve) Student Success: A Faculty Perspective. Journal of Interactive Online Learning. Volume 12.