

# Study And Design Behavior Predictors Using Advanced Machine Learning (Knn) For School Education

Yougal Kishore Sharma and Dr. Arpana Bharani

Department of Computer Science, Dr. A. P. J. Abdul Kalam University, Indore

Corresponding Author Email: [yougalscs@gmail.com](mailto:yougalscs@gmail.com)

---

**Abstract** - Education can be used to face and overcome a variety of life challenges. Education-based knowledge can enhance one's employment development prospects. Educational data mining is frequently used to extract useful information from obtained knowledge. Educational data mining is the process of analysing and visualizing an organization's (school's) data using various data mining tools and techniques. It can be used to identify a distinct pattern in students' behaviour and academic performance. The current study uses data mining techniques to enhance students' academic performance in secondary school. The patterns and types of analyses presented in this study should be used to enhance the educational process. We forecast student information and produced a report for the teachers and parents. We provide solutions based on performance to help them perform better and behave better. Due to the vast amount of information contained in educational databases, predicting student behaviours has grown more difficult. The emphasis of this paper is also on how the information that is forecasted can be used to identify the student data's most crucial features. Using educational data, we could genuinely increase student success and achievement in an effective manner.

## 1. INTRODUCTION

The quality of a country's educational system is crucial to its future success economically. The statistics on illiteracy reveal the rates at which students are failing a course or dropping out of school altogether. For instance, in 2010, 60% of pupils in Portugal passed the secondary exam, whereas 40% did not. The global literacy rate has been rising over the previous few decades. Failing a core subject like arithmetic can have a significant impact on a student's ability to learn and do well in other disciplines. In order to keep track of all of their pupils, schools need to keep extensive databases. Questions like, "How many students will assign equal importance to all subjects?" and "What kinds of courses can be used to attract students?" are examples of the kinds of questions that can be answered with this data. Can we anticipate how well pupils will do? We need to know what elements contribute to pupils' success or failure. etc., can be gleaned from the amassed records in storage [1]. With the advent of IT came a corresponding interest in business intelligence and data mining techniques for extracting useful insights from this mountain of data. The database is a treasure trove of useful information, including trends and patterns that may be used to refine strategy and boost odds of success. Knowledge can be gained through data mining

by extracting useful information from vast databases. Data mining can be used for analysis, categorization, and prediction, all with the goal of bettering academic outcomes for students. A form of supervised learning, classification sort's data into categories according to labels applied beforehand. Classification models can be constructed with a variety of data mining methods (Support Vector Machines, Naive Bayes, Decision Tree, etc.). The alternative is to employ computerized methods to examine the data in question in order to draw conclusions for the decision-maker [2].

### **1.1.Student Performance**

Student performance in the educational process can be characterized literally as anything that results from changes in students' behaviour as a result of their experiences; additionally, learning outcomes are a fulfilment of students' potential or ability [5-6]. Students' learning outcomes can be visible in their behaviour, both in the form of understanding knowledge, thinking abilities, or motor skills, and as a result of the process of modifying student behaviour after attending sessions. Students' tangible performance can be observed in their grasp of the material being studied, their expertise in processing information and forming judgments based on certain concepts or motor abilities. Based on those understandings, student performance in the areas of knowledge, attitudes, and abilities can be evaluated and measured after a series of lessons. Students' performance is affected by the teaching and learning processes they go through, hence learning outcomes can be used to improve the quality of the learning process [7]. Learning achievement, on the other hand, is a similar term that is a measure of student achievement after participating in learning activities in the form of an assessment scale (either letters or numbers) (either letters, numbers). Giving a weight or rating to a student's learning achievement necessitates the development of appropriate evaluation indicators, as well as the assurance of validity and reliability. This value can then be used to describe a student's performance throughout a specific time period. When students do different learning procedures, their performance is evaluated after passing numerous measurements (in various sorts of evaluation).

### **1.2.Analytics in Education**

Data is driving and demonstrating the current situation, thanks to the widespread use of the term "analytics" in today's society. Teachers can increase their output while also improving student performance by using educational analytics. To improve instruction and learning, a value-added evaluation of educational data can be used. Data evaluation and analysis can benefit students, faculty, staff, and departments alike. Both students and consultants can use these results as a guide for future academic success. Academic analytics specialists' key tasks include predicting students' academic performance and mining data to support teaching and learning, tutoring, and learning environment improvements. Data and models are utilized to measure students' growth, development, and performance, as well as their capacity to apply that knowledge. He now has a greater knowledge of how students learn as a result of his research, and he can assist other instructors use that understanding in their own classrooms. He now has a greater grasp of how students learn as a consequence of his research, and he can aid other instructors in using that information in their own classroom

## 2. LITERATURE REVIEW

**Jawad, H. M., Tout, S. et al, (2018)**the purpose of this research was to create an evaluation instrument for examining predictors of students' interest in a CS degree, and the methodology used was the Theory of Planned Behaviour. Students' motivation to pursue a CS major was found to be positively correlated with their levels of prior exposure to computer science (PBE), social network participation (SN), and personal computer competence (PCC). For whatever reason, male students were more enthusiastic about pursuing a CS major than their female counterparts. Two models were found through regression analysis to predict students' desire to major in computer science. One model used PBE as the only predictor, whereas the other included PBE and PCC. [1]

**Athani, S. S., Kodli, S. A. et al,(2017)** Many of life's challenges and issues can be overcome with the correct tools. Through greater knowledge, education boosts one's chances for career advancement. Educational data mining is routinely used to extract actionable insights from years of classroom experience. Educational data mining is the discipline of analysing and visualizing a school's data using a variety of data mining tools and methodologies in order to find trends in student accomplishment and behaviour. Our goal is to use data mining techniques to help secondary school students perform better in their classrooms. [2]

**Zeng, T. (2017)**Students are classified based on their level of motivation during their 5S-EDMF experiences. Following that, those certificates will be exposed to additional information prediction. Prediction accuracy improves because the classification model's parameters can be fine-tuned to match more pupils. If we want to forecast specific information other than a student's grade, we must resign the classifier to meet more targets classification application, such as classifying students into numerous levels. However, when we consider "data exhaust" from computer-mediated learning environments as unstructured data, where signals must be identified from the surrounding noise, things get even more difficult.[3].

**Webb, M. E., Fluck, A. et al,(2021)**Machine learning systems are becoming increasingly used in the workplace and the classroom. This review article draws on a synthesis and analysis of existing literature to investigate how current advances in machine learning might affect the way people learn. To begin, this article will draw parallels between the deep learning processes of computers and humans. In the field of artificial intelligence, deep learning is considered a subfield of machine learning. As deep learning relies heavily on backpropagation in weighted neural networks, it is inherently nondeterministic; as the system gains experience or is trained, it learns to adapt and improve. [4]

**Webb, M. E., Fluck, A. et al, [2021]** Machine learning systems are infiltrating our lives and are beginning to become important in our education systems.

**Tedre, M., Toivonen, T. et al, [2021]** Machine learning is increasingly included in computing curricula in higher education, and a quickly growing number of initiatives are expanding it in K–12 computing education, too.

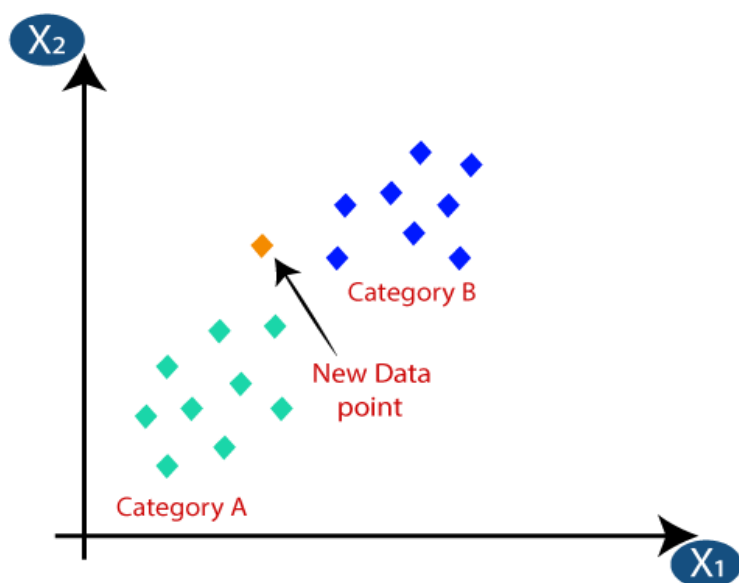
**Feng, G., Fan, M. et al,[2022]** This paper comprehensively uses the relevant theories of clustering, discrimination and convolution neural network to analyze and predict students' academic performance.

## 3. METHODOLOGY

**KNN** K-Nearest Neighbor (KNN) Algorithm for Machine Learning K-Nearest Neighbor is one of the simplest Machine Learning algorithms based on Supervised Learning technique. K-NN algorithm assumes the similarity between the new case/data and available cases and put the new case into the category that is most similar to the available categories. K-NN algorithm stores all the available data and classifies a new data point based on the similarity. K-NN algorithm mostly used for the Classification problems. K-NN is a non-parametric algorithm, which means it does not make any assumption on underlying data. KNN algorithm at the training phase just stores the dataset and when it gets new data, and then it classifies that data into a category that is much similar to the new data.

The K-NN working can be explained on the basis of the below algorithm:

- **Step-1:** Select the number K of the neighbors.
- **Step-2:** Calculate the Euclidean distance of K number of neighbors
- **Step-3:** Take the K nearest neighbors as per the calculated Euclidean distance.
- **Step-4:** Among these k neighbors, count the number of the data points in each category.
- **Step-5:** Assign the new data points to that category for which the number of the neighbor is maximum.
- **Step-6:** Our model is ready.



- Firstly, we will choose the number of neighbors, so we will choose the  $k=5$ .
- Next, we will calculate the Euclidean distance between the data points. The Euclidean distance is the distance between two points, which we have already studied in geometry. It can be calculated as:

KNN is a Supervised Learner for both Classification and Regression.

Supervised machine learning algorithms can be split into two groups based on the type of target variable that they can predict:

1. Classification is a prediction task with a categorical target variable. Classification models learn how to classify any new observation. This assigned class can be either right or wrong, not in between. A classic example of classification is the iris dataset, in which you use physical measurements of plants to predict their species. A famous algorithm that can be used for classification is logistic regression.

2. **Regression** is a prediction task in which the target variable is numeric. A famous example of regression is the Housing Prices Challenge on Kaggle. In this machine learning contest, participants try to predict the sales prices of houses based on numerous independent variables.

### Advantages of KNN Algorithm:

- It is simple to implement.
- It is robust to the noisy training data
- It can be more effective if the training data is large.

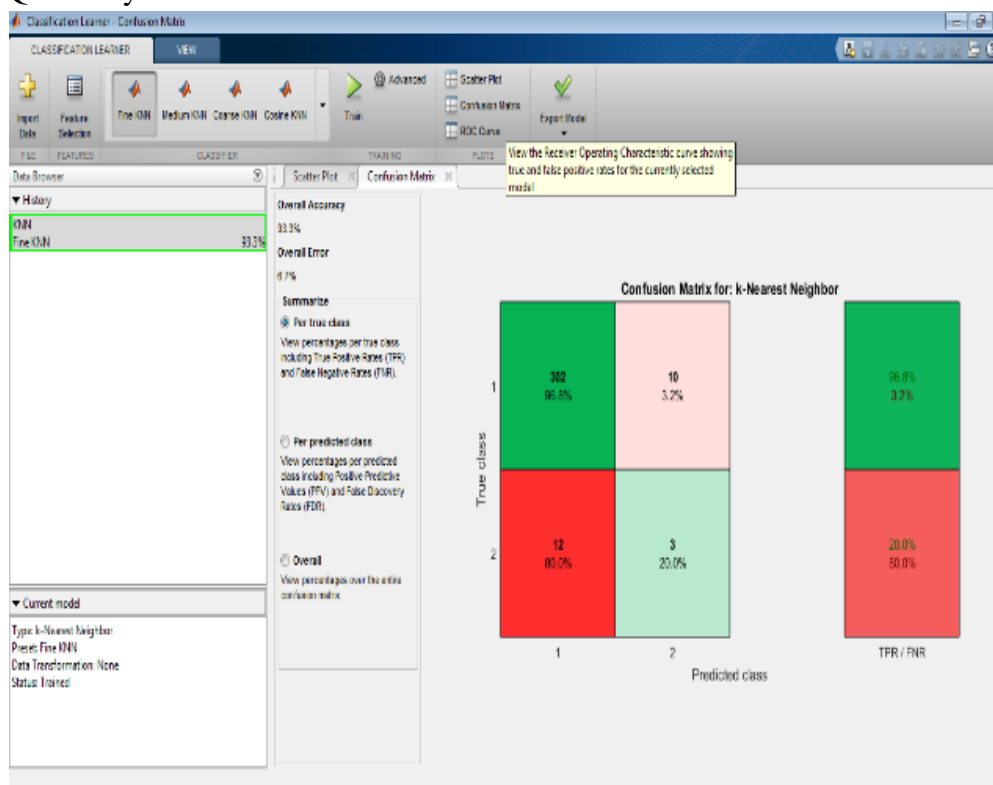
## 4. RESULTS AND DISCUSSION

Steps to implement the K-NN algorithm:

- Data Pre-processing step
- Fitting the K-NN algorithm to the Training set
- Predicting the test result
- Test accuracy of the result(Creation of Confusion matrix)
- Visualizing the test set result.

Now Classification should be conduct each survey question class firstly select training and other survey questions class select a test data.

Q. 1. Do you like school work?



**Figure 1: Dependency of Do you like school work parameters classification accuracy is 93.3%.**

In this figure Shows accuracy in terms confusion matrix using KNN method when survey question class select training and other survey questions class select a test data.

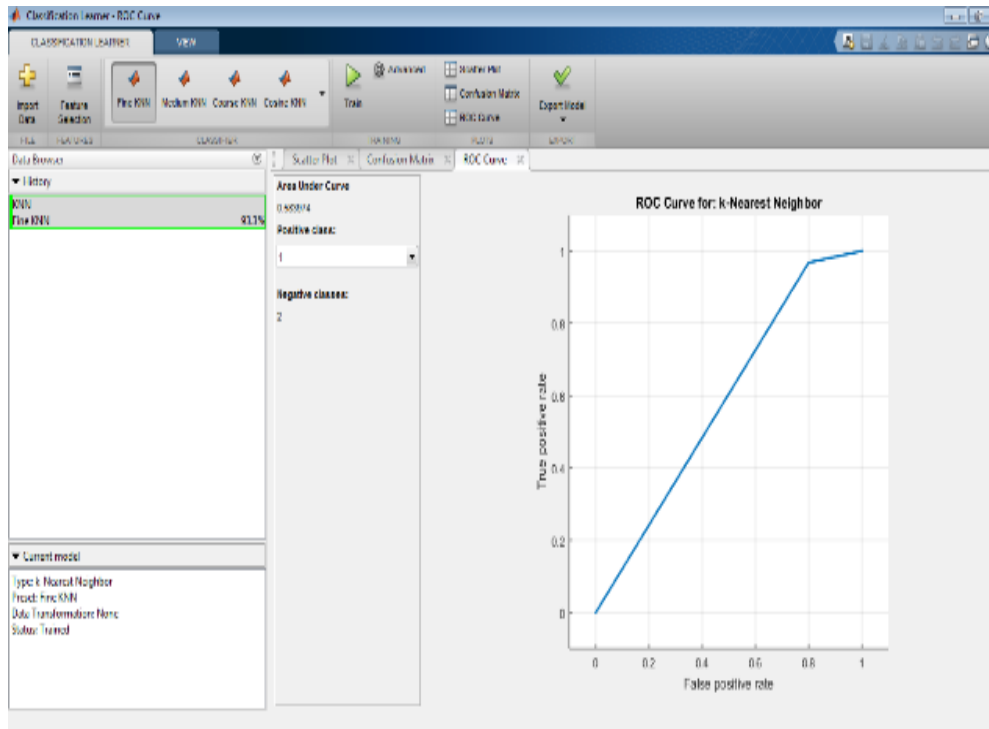


Figure 2: Shows true positive rate in terms ROC curve using KNN method.

Q. 2. Do you have problem with any subject

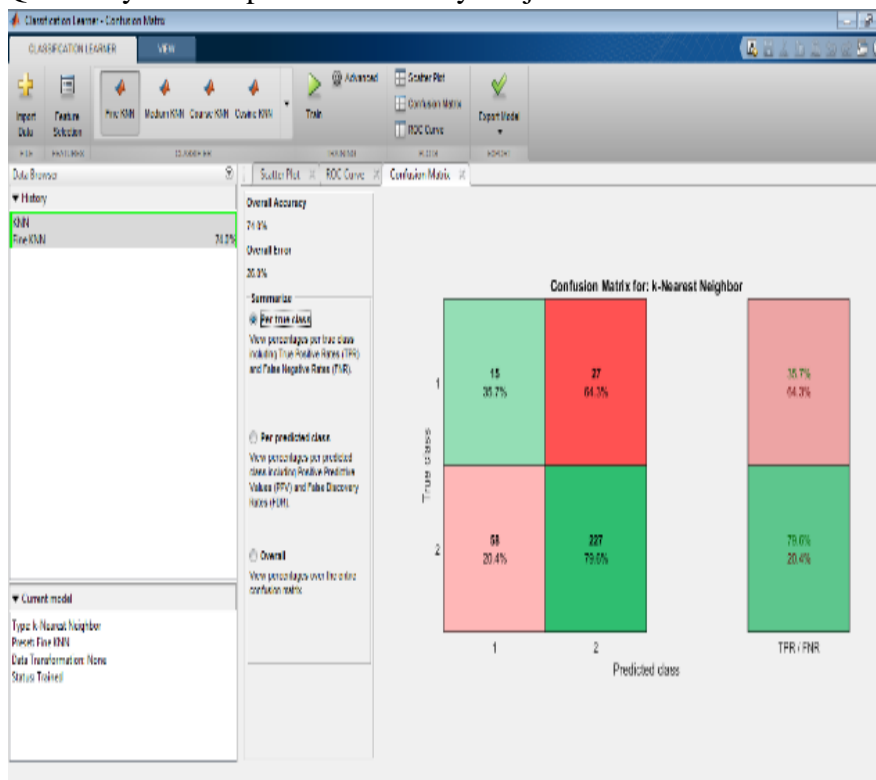


Figure 3: Dependency of do you have problem with any subject parameters classification accuracy is 74.0%.

In this figure Shows accuracy in terms confusion matrix using KNN method when survey question class select training and other survey questions class select a test data.in this survey question find 74% accuracy.

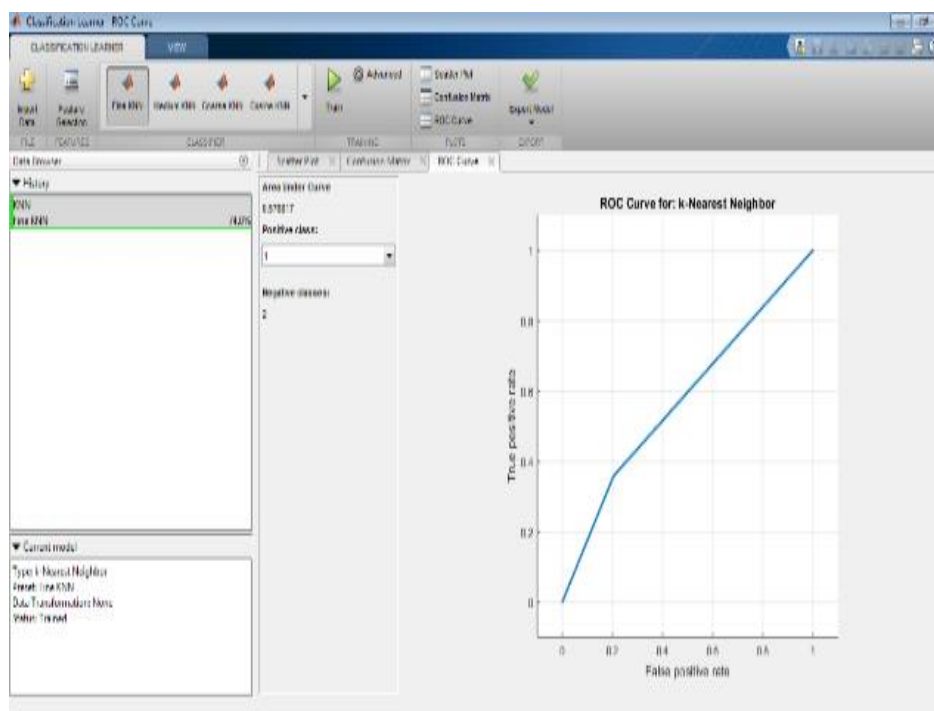


Figure 4: ROC of prediction class.

Q. 3. Do you have friends in school?

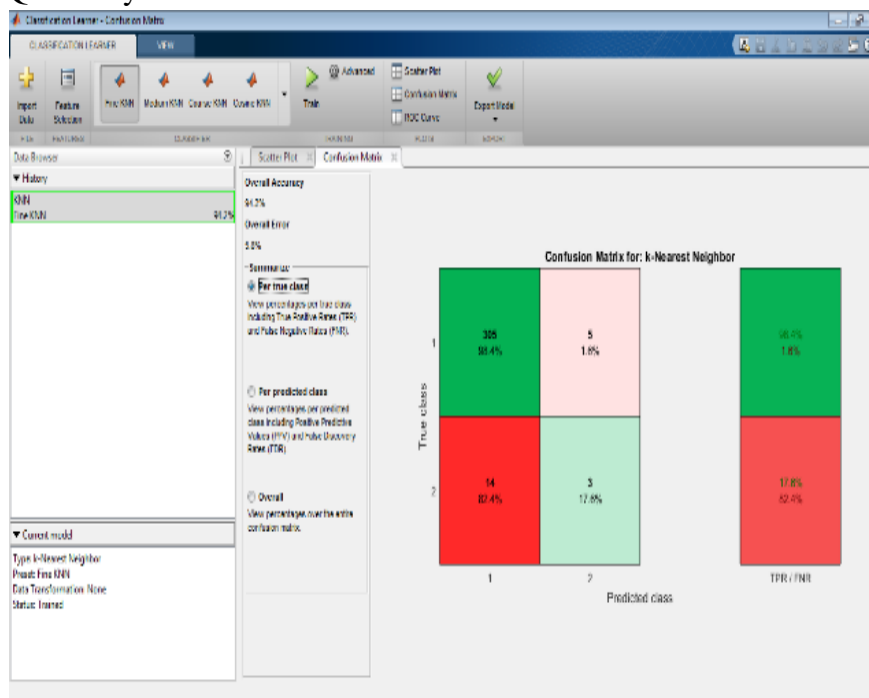


Figure 5: Dependency of Do you have friends in school parameters classification accuracy is 94.2%.

In this figure Shows accuracy in terms confusion matrix using KNN method when survey question class select training and other survey questions class select a test data.in this survey question find 94.2 % accuracy.

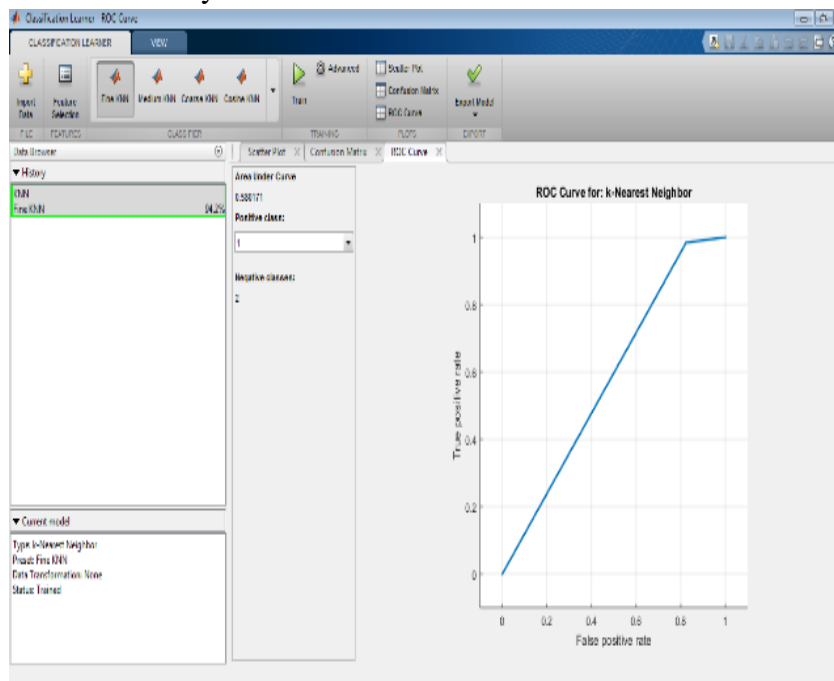


Figure 6: ROC of prediction class.

Q. 4. Do you have any problem with your classmate?

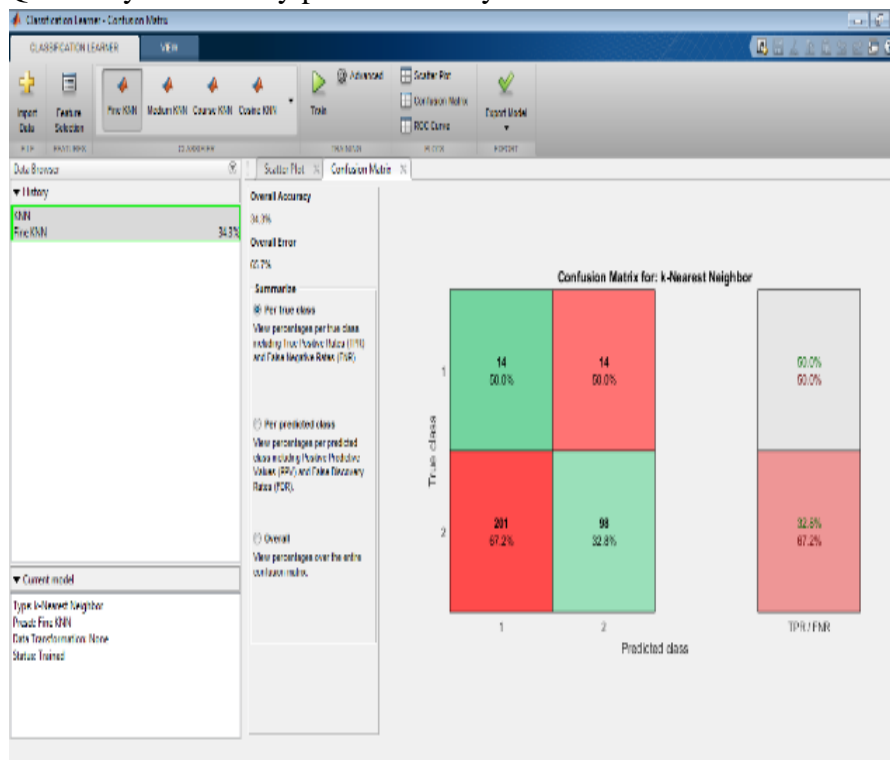


Figure 7: Dependency of Do you have any problem with your classmate parameters classification accuracy is 34.3%.



In this figure Shows accuracy in terms confusion matrix using KNN method when survey question class select training and other survey questions class select a test data.in this survey question find 34.3% accuracy.

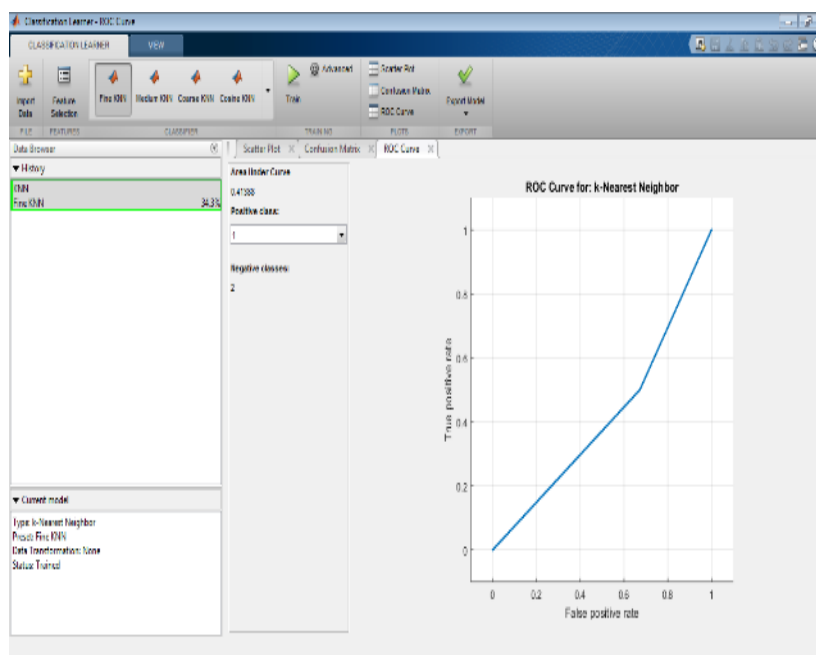


Figure 8: ROC of prediction class.

Q. 5. Do you have drinking/smoking habit?

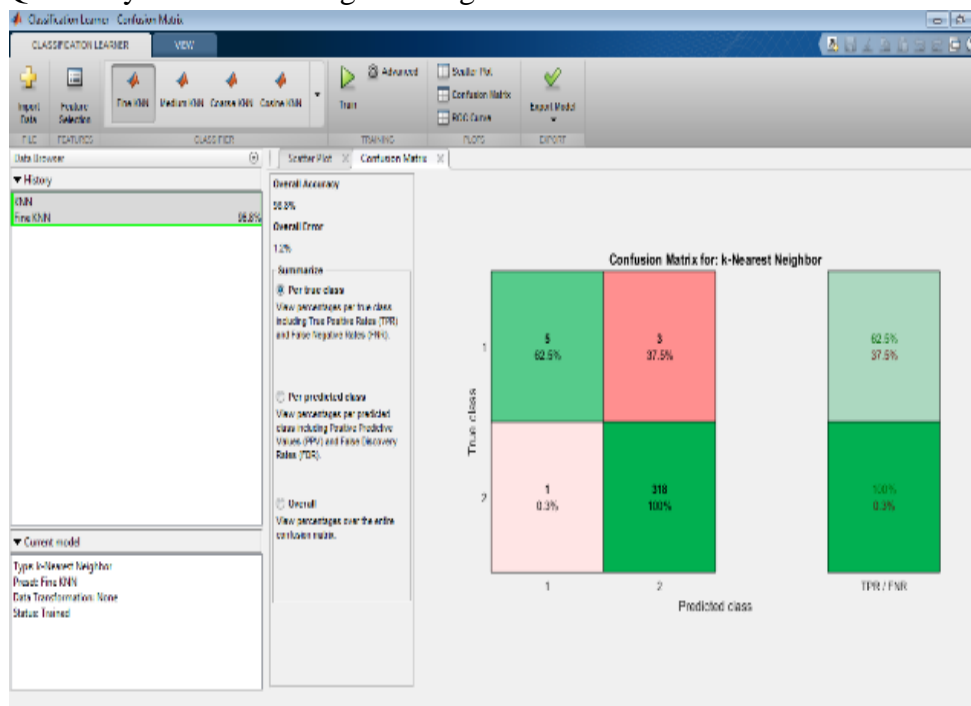


Figure 9: Dependency of Do you have drinking/smoking habit parameters classification accuracy is 98.8%.

In this figure Shows accuracy in terms confusion matrix using KNN method when survey question class select training and other survey questions class select a test data.in this survey question find 98.8% accuracy.

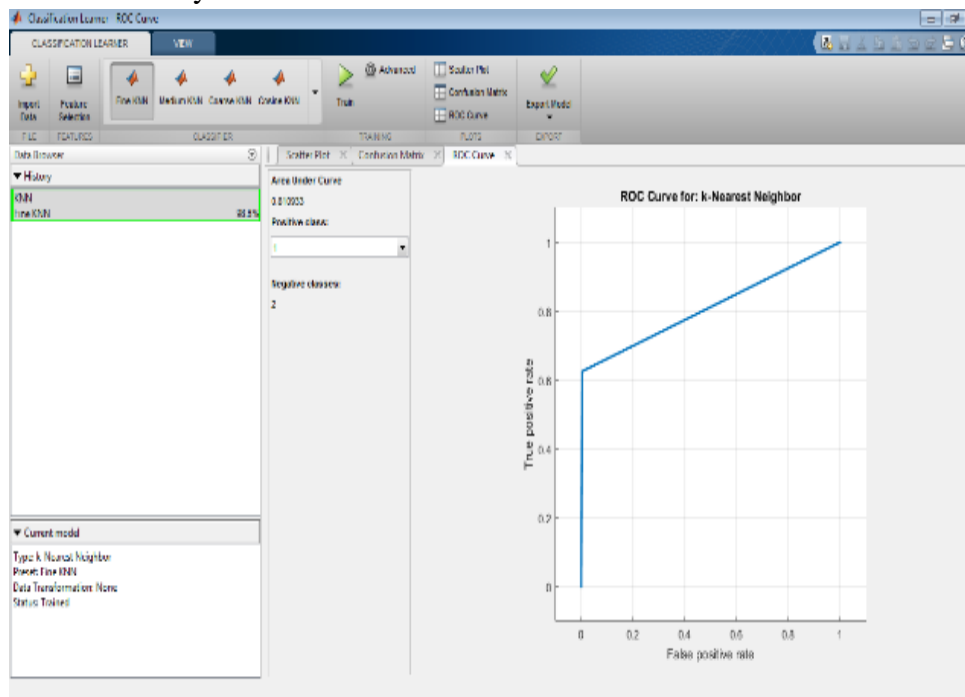


Figure 10: ROC of prediction class.

Q. 6. Do you have any problem with your teachers?

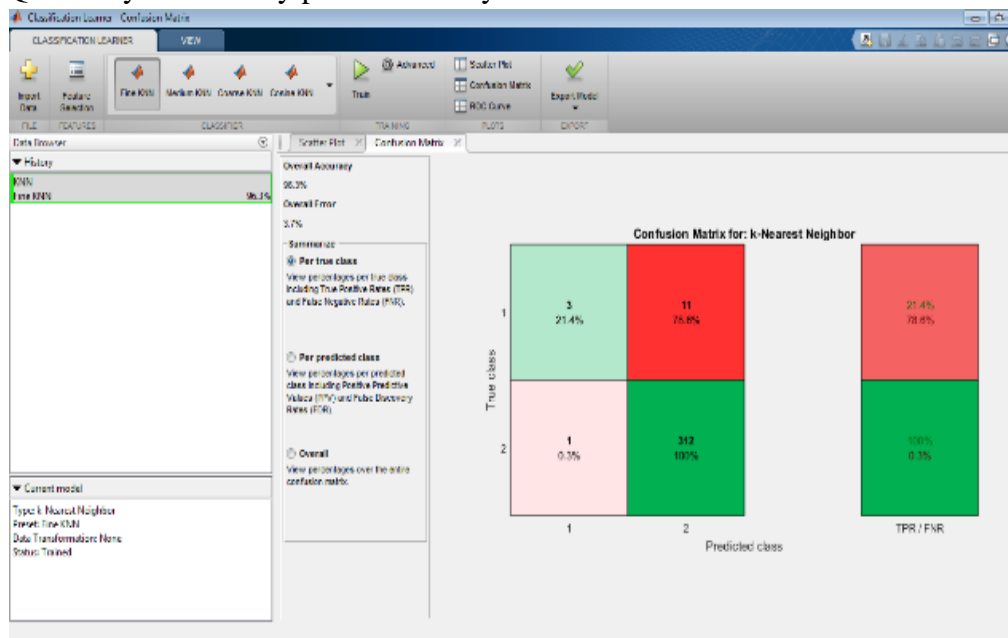


Figure 11: Dependency of Do you have any problem with your teachers parameters classification accuracy is 96.3%.

In this figure Shows accuracy in terms confusion matrix using KNN method when survey question class select training and other survey questions class select a test data.in this survey question find 96.6 % accuracy.

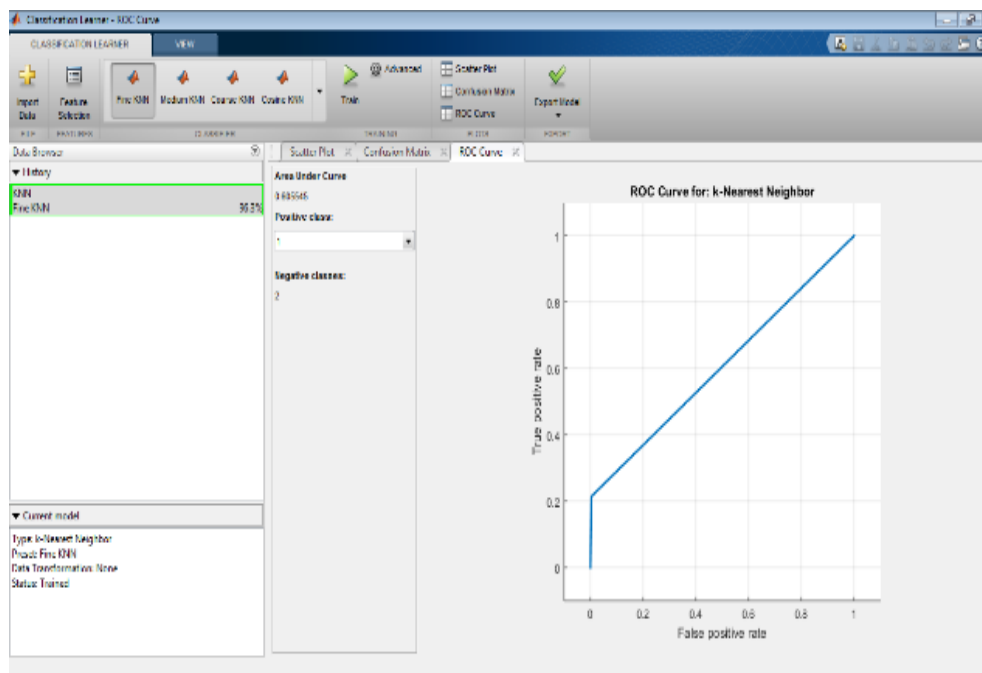


Figure 12: ROC of prediction class.

Q. 7. Do you have any problem with your parents?

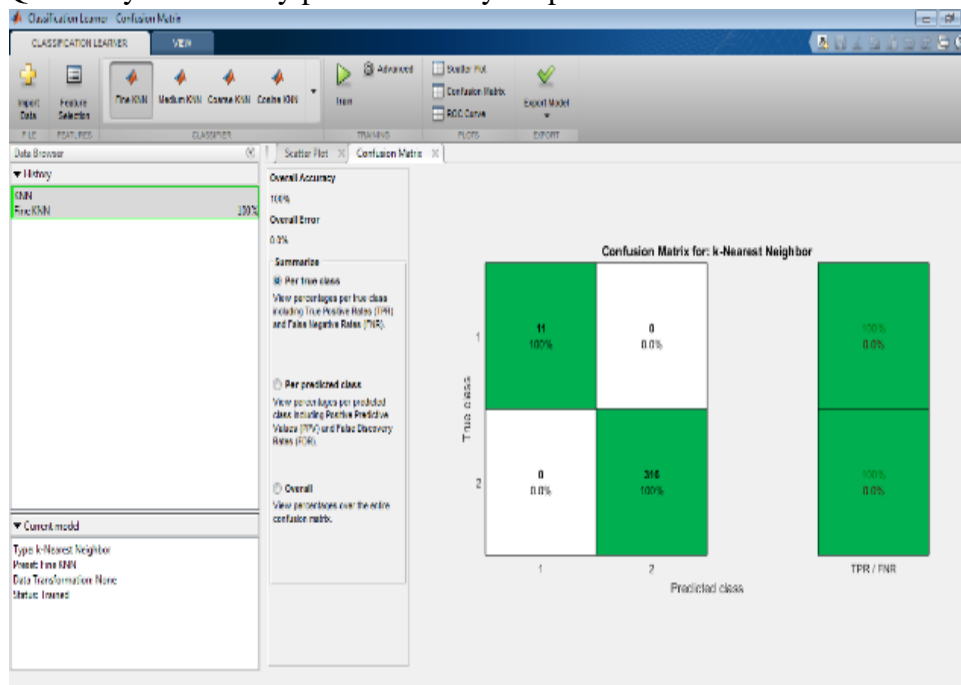


Figure 13: Dependency of Do you have any problem with your parents parameters classification accuracy is 100%.

In this figure Shows accuracy in terms confusion matrix using KNN method when survey question class select training and other survey questions class select a test data.in this survey question find 100 % accuracy.

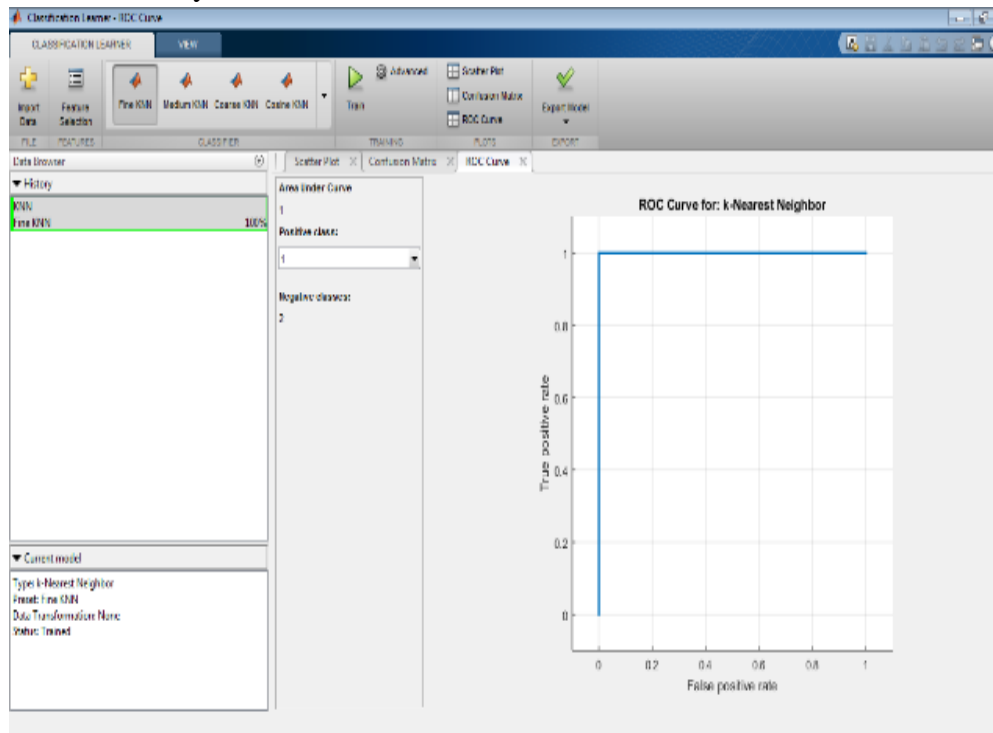


Figure 14: ROC of prediction class.

Q. 8. Any other suggestion/ problem in school or at home

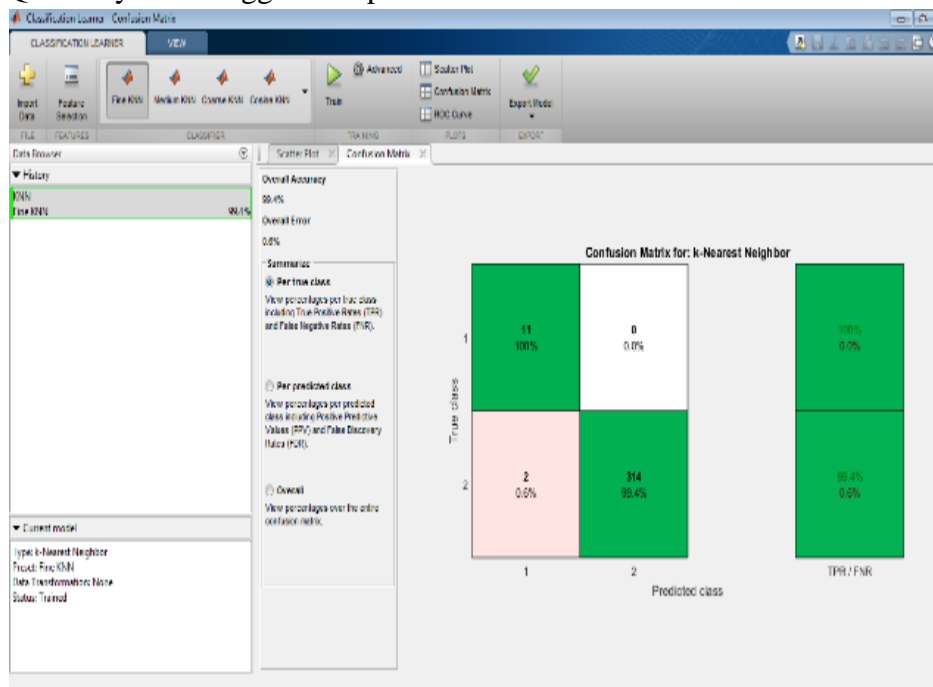


Figure 15: Dependency of Any other suggestion/ problem in school or at home parameters classification accuracy is 99.4%.

In this figure Shows accuracy in terms confusion matrix using KNN method when survey question class select training and other survey questions class select a test data.in this survey question find 99.4 % accuracy.

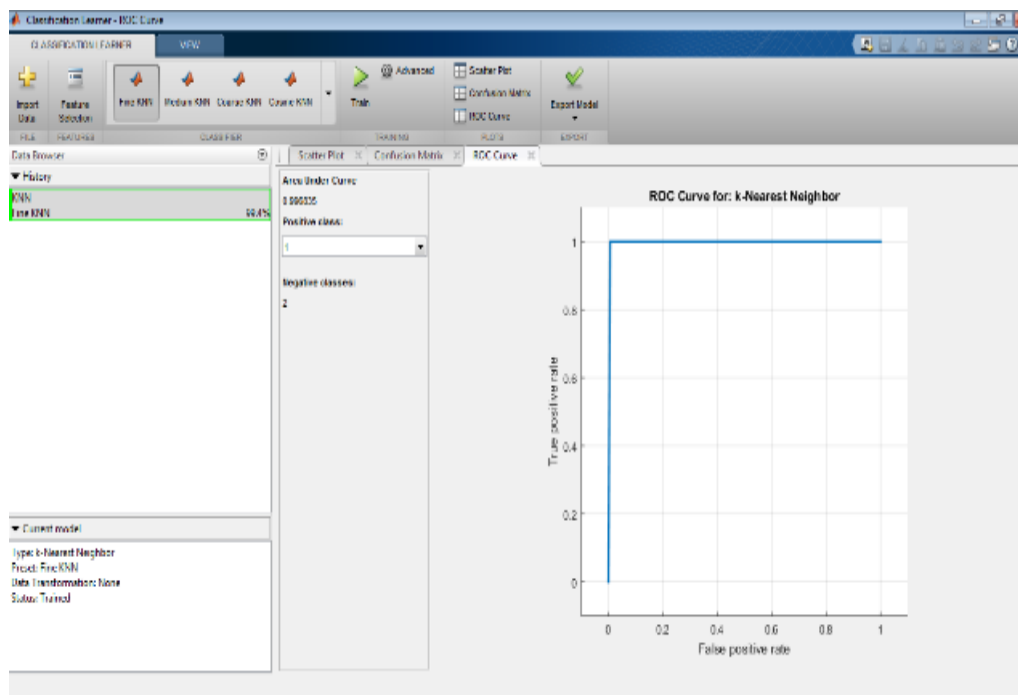


Figure 16: ROC of prediction class.

Table 1:Comparative questions accuracy.

Survey Question	Accuracy
Do you like school work?	93.3
Do you have problem with any subject	74.0
Do you have friends in school	94.2
Do you have any problem with your classmate	34.3
Do you have drinking/smoking habit	98.8
Do you have any problem with your teachers?	96.3
Do you have any problem with your parents	100
Any other suggestion/ problem in school or at home	99.4

## 5. CONCLUSION

The volume of data in educational databases is growing quickly. The classification algorithms are applied to the educational datasets to obtain knowledge about student performance. The study employed KNN major classification algorithms to suggest a model for predicting student 3831

performance. The project's long-term objective is to forecast student behaviour and performance in order to manage all the details for all Educational Institutes. In order to improve accuracy and gather additional data for the forecast, this study will be enhanced in the future by incorporating more data from various years. To improve the field's forecast speed and accuracy, study should be conducted in its place employing a variety of classification and clustering applications.

For example, useful information could include how many students will treat each type equally and what types of courses can be used to attract students. Can student performance be predicted? What factors influence students' performance? The collected stored records can be used to extract, among other things. Databases have grown exponentially as a result of information technology due to the emergence of business intelligence and data mining techniques for this type of information. The database contains valuable information such as trends and patterns that can aid in decision-making and success rate optimization. Data mining is used to extract relevant data from large databases in order to gain knowledge.

## REFERENCES

- [1.]Jawad, H. M., Tout, S., Abualkibash, M., &Xie, Y. (2018, May). The theory of planned behavior and high school students interest in computer science. In 2018 IEEE International Conference on Electro/Information Technology (EIT) (pp. 0764-0769). IEEE.
- [2.]Athani, S. S., Kodli, S. A., Banavasi, M. N., &Hiremath, P. S. (2017, May). Student academic performance and social behavior predictor using data mining techniques. In 2017 International Conference on Computing, Communication and Automation (ICCCA) (pp. 170-174). IEEE.
- [3.]Zeng, T. (2017, October). The research and practice of a five-sided educational data mining framework. In 2017 IEEE 3rd Information Technology and Mechatronics Engineering Conference (ITOEC) (pp. 1050-1053). IEEE.
- [4.]Webb, M. E., Fluck, A., Magenheimer, J., Malyn-Smith, J., Waters, J., Deschênes, M., &Zagami, J. (2021). Machine learning for human learners: opportunities, issues, tensions and threats. *Educational Technology Research and Development*, 69(4), 2109-2130.
- [5.]Tedre, M., Toivonen, T., Kahila, J., Vartiainen, H., Valtonen, T., Jormanainen, I., & Pears, A. (2021). Teaching machine learning in K–12 classroom: Pedagogical and technological trajectories for artificial intelligence education. *IEEE Access*, 9, 110558-110572.
- [6.]Feng, G., Fan, M., & Chen, Y. (2022). Analysis and Prediction of Students' Academic Performance Based on Educational Data Mining. *IEEE Access*, 10, 19558-19571.
- [7.]Feng, G., Fan, M., &Ao, C. (2022). Exploration and Visualization of Learning Behavior Patterns From the Perspective of Educational Process Mining. *IEEE Access*, 10, 65271-65283.
- [8.]Shafiq, D. A., Marjani, M., Habeeb, R. A. A., &Asirvatham, D. (2022). Student Retention Using Educational Data Mining and Predictive Analytics: A Systematic Literature Review. *IEEE Access*.
- [9.]Sharma Y K, Dr. Bharani A., A Systematic Review On Study Behaviour Predictors Using Advanced Machine Learning For School Education *International Journal Of Engineering, Education And Technology*,ISSN 2320-883X,VOLUME 10 ISSUE ,2022.

- [10.] Alsariera, Y. A., Baashar, Y., Alkaws, G., Mustafa, A., Alkahtani, A. A., & Ali, N. A. (2022). Assessment and Evaluation of Different Machine Learning Algorithms for Predicting Student Performance. *Computational Intelligence and Neuroscience*, 2022.
- [11.] Adnan, M., Habib, A., Ashraf, J., Mussadiq, S., Raza, A. A., Abid, M. & Khan, S. U. (2021). Predicting at-risk students at different percentage of course Length for early intervention using machine learning models. *Ieee Access*, 9, 7519-7539.