SMARTINTERNZ REPORT

PROJECT TITLE:

STUDENT’S ADAPTIBILITY PREDICTION

IN ONLINE EDUCATION

TEAM MEMBER:

GOVIND AGARWAL

KRITIKA JINDAL

JASNEET SINGH BHATTI

MANAV JAIN

INDEX:

I. INTRODUCTION:

OVERVIEW:

The field of education has undergone a significant transformation with the rise of online learning platforms. Online education offers unprecedented flexibility, accessibility, and personalized learning experiences. However, the transition from traditional classroom settings to online environments presents unique challenges for students. Understanding and predicting student adaptability in the context of online education is crucial for designing effective interventions and providing targeted support.

The advent of online education has opened doors to learning opportunities for students worldwide. However, not all students thrive in the virtual learning environment. Some may struggle to adapt to the independent and self-directed nature of online learning, leading to reduced engagement, lower academic performance, and potential dropouts. By predicting student adaptability levels, educational institutions can proactively identify students who may face challenges and implement targeted support mechanisms to ensure their success.

This project's objectives include exploring the factors impacting student adaptability in online education, gathering and analyzing relevant data, and employing machine learning techniques to develop a predictive model. By examining various factors such as demographic information, academic performance, engagement metrics, and self-reported adaptability measures, we aim to uncover hidden patterns and correlations that influence student adaptability.

PURPOSE:

The Purpose of this project is to examine the adaptability level of students in online education. The project will explore the factors that in influence students’ ability to adapt to online education.

By understanding and forecasting students' adaptability, educational institutions can design targeted interventions and provide appropriate support to enhance the online learning experience. Leveraging data-driven methodologies, we aim to identify key factors that influence student adaptability and develop an accurate predictive model.

Also the project will have social impact such as personalised learning, support and intervention, equity and inclusion, academic success and retention. Along with these social impacts the project will have various business impact such as improved student satisfaction, enhanced reputation, cost optimization and continuous improvement. Overall, the social and business impacts of this project revolve around creating a more inclusive and effective online education environment, enhancing student success and satisfaction, and optimizing resource allocation for educational institutions.

II. LITERATURE REVIEW:

EXISTING SOLUTION:

1. Adaptability and High School Students' Online Learning During COVID-19: A Job Demands-Resources Perspective

Authors: Andrew J. Martin, Rebecca J. Collie and Robin P. Nagy

The above study examined, examined how adaptation aided high school students in navigating their online learning, during a COVID-19 period that involved totally or partially remote online instruction. The study utilised information from a sample of 1,548 Australian high school students in nine schools and the Job Demands-Resources hypothesis. Adaptability was found to be significantly associated with higher levels of online learning self-efficacy and with gains in later achievement, independent of the effects of online learning demands, online and parental learning support, and background characteristics. These results confirm the significance of adaptability as a personal resource that can support students in their online learning, including during times of remote instruction, like during COVID-19. The findings from the study gave educators and organizations ideas on how to enhance students' adaptability in online learning more effectively. It was advised to promote self-regulated learning abilities, create social contact, give accessible resources and services, and provide technical training and support. Finally, the study acknowledged its limitation regarding its specific focus on the COVID-19 pandemic context and potential biases in self-reported survey data.

2. Adaptability to a Sudden Transition to Online Learning During the COVID-19 Pandemic: Understanding the Challenges for Students

Authors: Avi Besser, Gordon L. Flett, Virgil Zeigler-Hill

In the above case study, researchers delved into the relationships between pandemic adaptability, personality, along with the different types of learning methods including affective, cognitive, and behavioral. The survey collected data through online surveys, data from a sample of 1,217 college students from Israel who had switched to synchronous online during the COVID-19 epidemic was gathered and utilized for the study. The study found that students' response to the online condition, owing to the epidemic, were consistently unfavorable, according to comparisons of those responses to more conventional face-to-face learning settings. The ability to adjust to the epidemic was also generally linked to more favourable responses across several measures. The results of this study demonstrate the importance of flexibility as well as the major difficulties faced by college students who had to quickly adapt to the COVID-19 pandemic-related disruptions and uncertainties in their learning and living environments.

PROPOSED SOLUTION:

The solution proposed in the project is to develop a machine learning model using available data about student adaptability in online education and then integrate the developed model with a website in order to predict the adaptability using user inputs. In order to achieve this the data was collected from Kaggle and then it was pre-processed to feed into the model. The best performing model was then selected for hyper-parameter tuning and was saved using pickle. For the development of front-end of the website html, CSS and JavaScript were used and backend was developed using Flask framework.

III. THEORETICAL ANALYSIS:

BLOCK DIAGRAM:

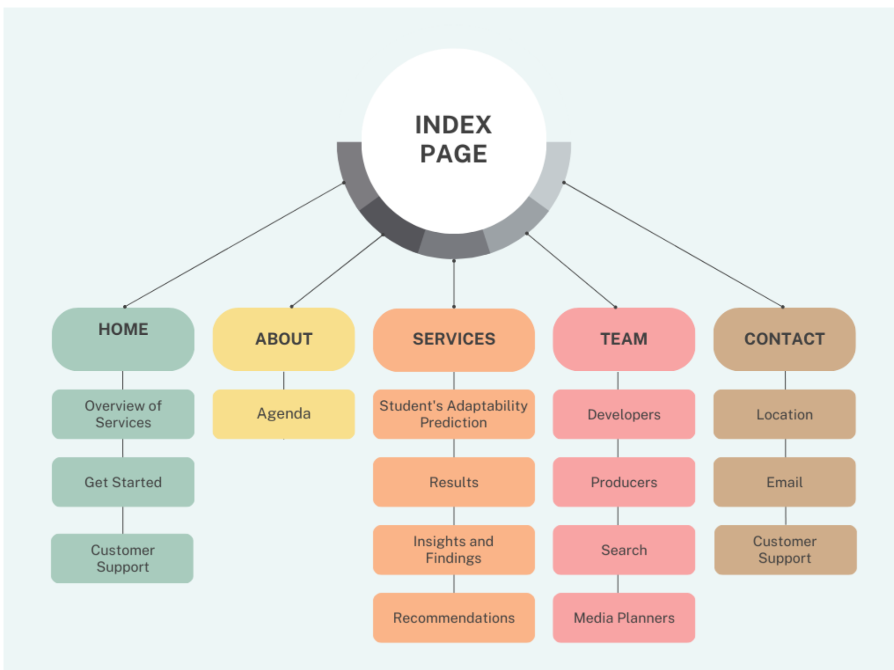


Fig. 1

HARDWARE/SOFTWARE DESIGNING:

3.1 System Requirements

Software Requirements:

- Microsoft Windows 7/8/10 or Linux.

- VS Code or any other text editor.
- Chrome or any other browser.

Hardware Requirements:

- Intel Processor 2.0 GHz or above.
- 2 GB RAM or more

IV. EXPERIMENTAL INVESTIGATIONS:

DATASET DESCRIPTION:

The dataset has been taken from Kaggle named “Student Adaptability Level in Online Education”. The dataset consists of 14 features and has 1205 instances in total. All the features in the dataset are Categorical type. The target variable is divided into 3 classes with the values ‘Moderate’ with 625 instances, ‘Low’ with 480 instances, ‘High’ with 100 instances.

VISUALIZATIONS:

Python libraries matplotlib and seaborn were utilized for visualizing the dataset’s features. In the project count plot is employed to visualize the variables, as it depicts the count of observations in each category concerning the feature used to analyze the pattern between each feature’s contribution to the student’s adaptability classes.

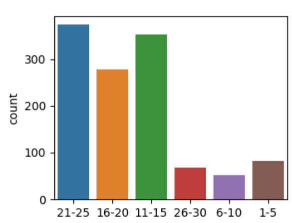


Fig -2

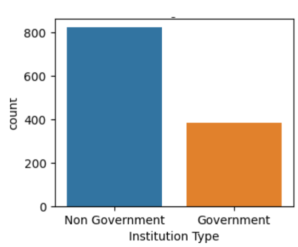


Fig -3

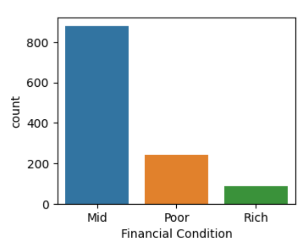


Fig - 4

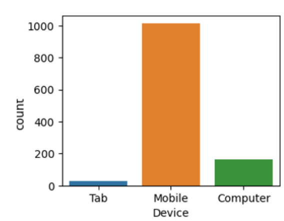


Fig - 5

It can be observed from Fig-2 that most of the data is ranged between the age group from 11 to 25 and very minority of the data is in the age ranged from 1 to 10 and 25 above. From Fig-3 it can be inferred that most of the data consists of non-Government institution type. Fig-4 shows that the dataset consists the majority of the instances from the middle class financial condition and less data from the Poor and rich sections.

From Fig-5 it can be deduced that the majority of the data is collected from the students who use mobile devices for online classes as the majority of the data is of Mobile devices.

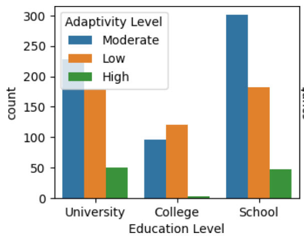


Fig-6

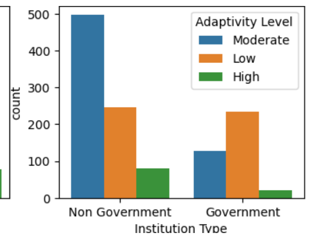


Fig-7

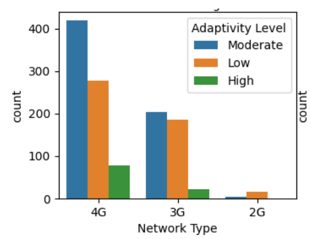


Fig-8

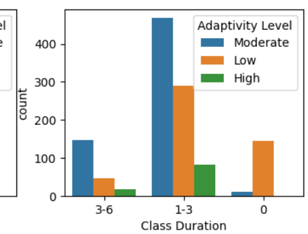


Fig-9

From Fig-6 it can be observed that very less students from colleges have high level of adaptivity towards online education as compared to Universities and Schools. Fig-7 depicts that in non-government institutions the adaptivity level of majority of the students is moderate. From Fig-8 it can be noticed that most of the student with a 4G network have moderate adaptivity towards online education. From Fig-9 it can be observed that if the class duration is between 1 to 3 hours then the adaptivity level is moderate for the majority of the students.

SMOTE:

The observed frequencies of a target variable are significantly varied over its range of potential values, the data is said to be imbalanced. SMOTE is an algorithm that adds synthetic data points to the actual data points to accomplish data augmentation. The minority examples that are near the feature space are chosen. Then, a new sample is drawn at a location along the line that is drawn between the examples in the features space. The dataset before using the SMOTE technique had a significant imbalance where ‘High’ was the minority class with only 100 instances, whereas ‘Low’ and ‘Moderate’ were the majority class with an instance count of 480 and 625 respectively. SMOTE reduced the class imbalance by increasing the minority class i.e. ‘High’ instances to 528 and reducing the majority class instances of ‘Low’ and ‘Moderate’ with the instance count of 451 and 462 respectively.

V. FLOWCHART:

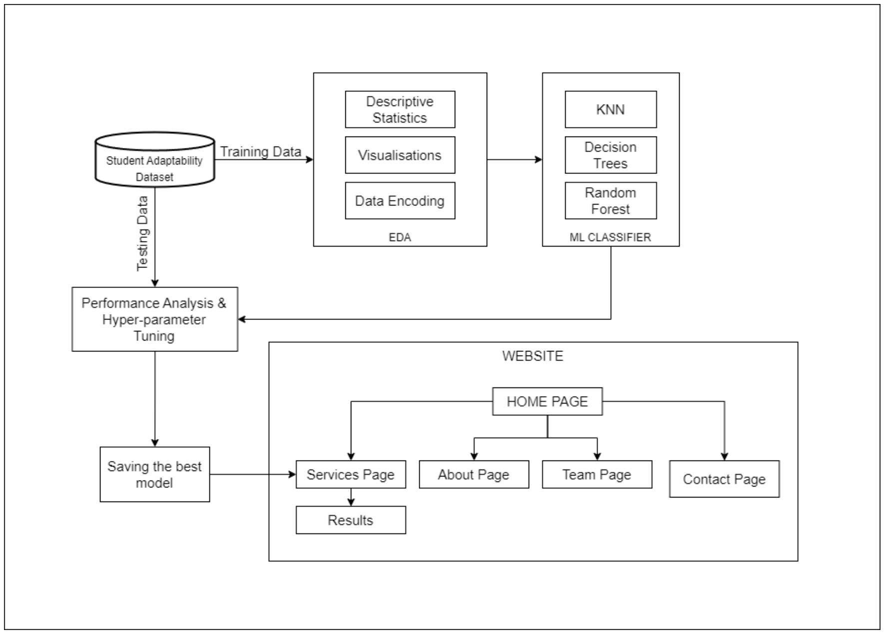


Fig. 10

VI. RESULTS:

Confusion Matrix:

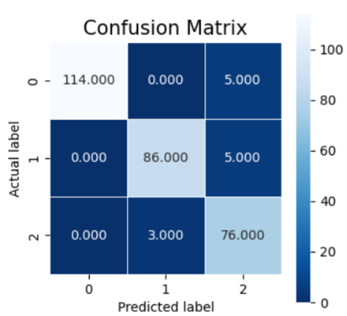


Fig-11 KNN Classifier

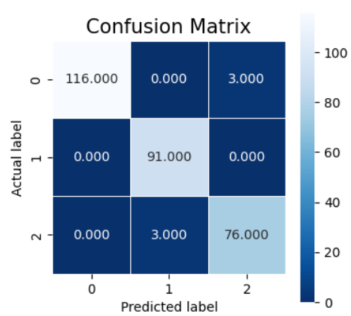


Fig – 12 Decision Trees Classifier

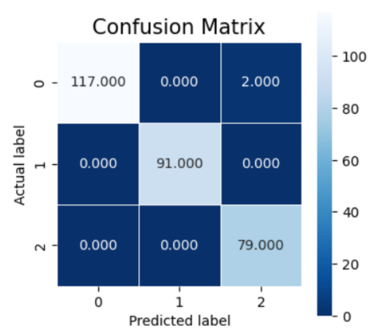


Fig – 13 Random Forests Classifier

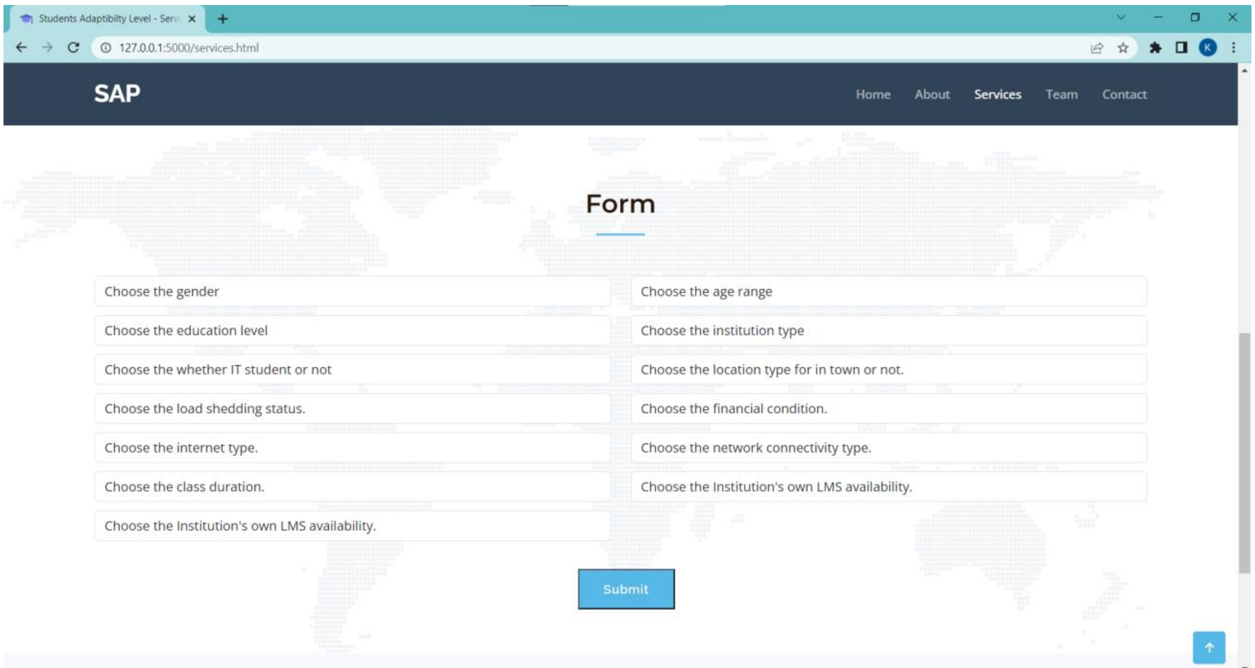
S. No	Models	Accuracy on Train (%)	Accuracy on Test (%)	Precision	Recall	F1 Score
1	**KNN Classifier**	97.82	95.50	0.955	0.955	0.955
2	**Decision Trees Classifier**	98.26	97.92	0.979	0.979	0.979
3	**Random Forests Classifier**	100	99.30	0.993	0.993	0.993

Table I: Performance Analysis of all the ML Algorithms

The experimental results of the performance metrics of the different classifiers used in the model are shown in Table I. Scikit-learn was used for model evaluation. The results show that the Random Forests classifier outperformed the other algorithms with an accuracy of 99.3%, the highest precision rate of 0.993, a recall of 0.993 and an F1 - score of 0.993. It can be noted from the Fig-13 that the true positive rate for Random Forests is highest for all classes i.e for moderate it is classifying the maximum true values with a true positive rate of 79 which is higher than the decision trees and KNN classifier, for 'Low' class a true positive rate of 91 which is higher than the true positive rate of KNN which is 86 and for 'High' it has a true positive rate of 117 which is higher than KNN and Decision Trees Classifier.

FINAL PRODUCT:

WEBSITE SCREENSHOTS:



VII. ADVANTAGES AND DISADVANTAGES:

Advantages of the Project:

1. Personalized Support: By predicting student adaptability levels, this project enables educational institutions to provide targeted support and interventions tailored to individual students' needs. This personalized approach can enhance student engagement, motivation, and overall learning outcomes in online education.
2. Early Intervention: The project allows for the early identification of students who may struggle with adapting to online education. Institutions can intervene promptly, offering resources and support to help these students overcome challenges and improve their chances of success.
3. Improved Retention Rates: Understanding and addressing student adaptability can positively impact student retention rates in online education. By providing the necessary support and interventions, institutions can help students stay engaged, persist in their studies, and ultimately graduate.
4. Resource Optimization: The predictive model developed in this project can assist in resource allocation by identifying students who require additional support. This optimization of resources ensures that interventions are targeted where they are most needed, maximizing their effectiveness and efficiency.
5. Enhanced Learning Experience: By identifying factors that influence student adaptability, institutions can make informed decisions to improve the overall online learning experience. This may involve optimizing course design, instructional strategies, and the integration of technology to better meet students' needs.

Disadvantages and Limitations of the Project:

1. Complexity of Adaptability: Student adaptability is a multifaceted construct influenced by various factors, some of which may be challenging to measure accurately. The project may face limitations in capturing the entirety of student adaptability, potentially leading to incomplete predictions or oversimplification of the concept.
2. Ethical Considerations: Collecting and analyzing student data for predictive purposes raises ethical concerns regarding privacy, data security, and informed consent. It is crucial to ensure compliance with relevant regulations and ethical guidelines to protect the privacy and rights of the students involved.
3. Generalizability: The predictive model developed in this project may have limitations in terms of generalizability across different populations or contexts. Factors influencing student adaptability can

vary across demographics, cultural backgrounds, and educational settings, potentially impacting the model's performance when applied to diverse populations.

4. Human Factors: The project primarily focuses on data-driven predictions and may not fully account for the nuances of individual student experiences or subjective factors that influence adaptability. It is important to consider the human element, such as social support, personal circumstances, and intrinsic motivation, which may not be captured by the data used in the analysis.
5. Continuous Model Updating: As student adaptability is influenced by evolving factors and changing educational landscapes, the predictive model needs to be regularly updated and refined to maintain its accuracy and relevance. This requires ongoing data collection, analysis, and model adaptation, which may pose logistical challenges.

VIII. APPLICATIONS:

1. Personalized Learning Interventions: The project's predictive model can be utilized to provide personalized learning interventions tailored to individual students' adaptability levels. Educational institutions can develop targeted resources, adaptive learning strategies, and support programs to help students overcome specific challenges and maximize their learning potential.
2. Early Warning System: The predictive model can serve as an early warning system to identify students who may be at risk of struggling with adaptability in online education. Institutions can proactively intervene by offering guidance, mentorship, or counseling to help these students navigate the challenges and improve their chances of success.
3. Course Design and Instructional Strategies: Insights gained from the project can inform the design and delivery of online courses. Institutions can optimize course structures, instructional strategies, and assessments to enhance student adaptability and engagement. For example, incorporating interactive elements, multimedia resources, and collaborative activities can promote active learning and adaptability.
4. Student Support Services: The project's outcomes can guide the development of targeted support services for students in online education. Institutions can establish dedicated support teams, virtual communities, or mentoring programs to address specific adaptability challenges, foster a sense of belonging, and provide guidance throughout the online learning journey.
5. Program Evaluation and Improvement: The project's findings can contribute to program evaluation and continuous improvement efforts in online education. Institutions can assess the effectiveness of their online programs by examining the relationship between student adaptability, engagement, and academic performance. This information can guide future program enhancements and modifications.

6. Student Success and Retention Strategies: By identifying factors that influence student adaptability, institutions can develop comprehensive student success and retention strategies. Understanding which adaptability-related factors contribute to student attrition or success can enable institutions to implement targeted interventions and create a supportive online learning environment that promotes student persistence and completion.

IX. CONCLUSION:

In this system, we implemented a machine learning model with web framework development using Flask to classify the adaptability level of students in online education. The project dataset is acquired from Kaggle. Prior to the development of the model, the dataset was cleaned and preprocessed using descriptive statistics analysis, visualisations for understanding the patterns, one hot encoding data transformation, and SMOTE technique to address the issue of class imbalance in the dataset. After training several models, including KNN, Decision Trees, and Random Forests, the performance analysis revealed that the Random Forest algorithm outperformed all the others with an accuracy of 99.3% and the highest precision, recall, and F1-score values of each 0.993. After this hyper-parameter tuning was performed on the best performing model i.e Random Forests, and it was found that the default parameters were equally performing well as compared to after parameter optimization.

X. FUTURE SCOPE:

Refinement of Predictive Model: The predictive model developed in this project can be further refined and optimized to improve its accuracy and performance. This can involve incorporating additional variables, exploring more advanced machine learning algorithms, or integrating emerging techniques such as deep learning or natural language processing. The continuous refinement of the model will enhance its predictive capabilities and applicability.

Longitudinal Analysis: Conducting longitudinal studies can provide insights into how student adaptability evolves over time in online education. By tracking and analyzing adaptability levels at different stages of a student's educational journey, institutions can gain a deeper understanding of the factors influencing adaptability and identify potential patterns or trends.

Comparative Analysis: Comparing student adaptability levels between different types of online education programs, institutions, or learning modalities can shed light on the effectiveness of specific approaches.

This comparative analysis can help identify best practices, successful interventions, and areas for improvement, contributing to the overall advancement of online education.

Intervention Effectiveness Assessment: Evaluating the effectiveness of specific interventions or support mechanisms designed to enhance student adaptability is an important future scope. Institutions can conduct intervention studies, measure the impact of targeted support programs, and determine which strategies yield the most positive outcomes in terms of improving student adaptability and overall learning performance.

Integration with Learning Analytics: Integrating student adaptability predictions with learning analytics data can provide a more comprehensive understanding of students' learning experiences. By examining the relationship between adaptability levels, engagement metrics, and academic performance, institutions can uncover deeper insights into the dynamics of online learning and inform data-driven decision-making processes.

XI. BIBLIOGRAPHY:

[1] Martin, A. J., Collie, R. J., & Nagy, R. (2021). Adaptability and High School Students' Online Learning During COVID-19: A Job Demands-Resources Perspective. *Frontiers in Psychology, 12*. https://doi.org/10.3389/fpsyg.2021.702163

[2] Besser, Avi, et al. "Adaptability to a Sudden Transition to Online Learning during the COVID-19 Pandemic: Understanding the Challenges for Students." *Scholarship of Teaching and Learning in Psychology*, vol. 8, no. 2, 19 Oct. 2020, web.a.ebscohost.com/ehost/detail/detail?vid=0&sid=ef211af2-8a01-437b-a00e-bde5d39c7d67%40sessionmgr4006&bdata=JnNpdGU9ZWhvc3QtbGl2ZSZzY29wZT1zaXRl#AN=2020-77530-001&db=pdh, https://doi.org/10.1037/stl0000198.

[3] Scikit-learn. "3.2.4.3.1. Sklearn.ensemble.RandomForestClassifier — Scikit-Learn 0.20.3 Documentation." *Scikit-Learn.org*, 2018, scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html.

[4] scikit-learn. "1.10. Decision Trees — Scikit-Learn 0.22 Documentation." *Scikit-Learn.org*, 2009, scikit-learn.org/stable/modules/tree.html.

[5] scikit-learn developers. "Sklearn.neighbors.KNeighborsClassifier — Scikit-Learn 0.22.1 Documentation." *Scikit-Learn.org*, 2019, scikit-learn.org/stable/modules/generated/sklearn.neighbors.KNeighborsClassifier.html.

[6] "Sklearn.metrics.accuracy_score — Scikit-Learn 0.24.1 Documentation." *Scikit-Learn.org*, scikit-learn.org/stable/modules/generated/sklearn.metrics.accuracy_score.html#sklearn.metrics.accuracy_score.

[7] "Sklearn.metrics.confusion_matrix — Scikit-Learn 0.21.3 Documentation." *Scikit-Learn.org*, 2019, scikit-learn.org/stable/modules/generated/sklearn.metrics.confusion_matrix.html#sklearn.metrics.confusion_matrix.

[8] "Sklearn.metrics.f1_score." *Scikit-Learn*, scikit-learn.org/stable/modules/generated/sklearn.metrics.f1_score.html#sklearn.metrics.f1_score.

[9] "Sklearn.metrics.precision_score — Scikit-Learn 0.24.1 Documentation." *Scikit-Learn.org*, scikit-learn.org/stable/modules/generated/sklearn.metrics.precision_score.html#sklearn.metrics.precision_score.

[10] "Sklearn.metrics.recall_score." *Scikit-Learn*, scikit-learn.org/stable/modules/generated/sklearn.metrics.recall_score.html#sklearn.metrics.recall_score.

[11] "Students Adaptability Level in Online Education." *Www.kaggle.com*, www.kaggle.com/datasets/mdmahmudulhasansuzan/students-adaptability-level-in-online-education.

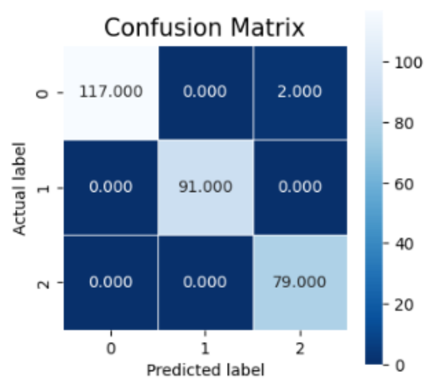
APPENDIX:

Screenshot of the best performing ML model with output:

```
In [31]: #fit the model on train data
rf_clf2=RandomForestClassifier(n_estimators=20, min_samples_split=2, min_samples_leaf=1, max_depth= 50).fit(X_train2,Y_train2)

#predict on test
test_preds_rf = rf_clf2.predict(X_test2)
#accuracy on test
print("Model accuracy on train is: ", accuracy_score(Y_train2, rf_clf2.predict(X_train2)))
print("Model accuracy on test is: ", accuracy_score(Y_test2, test_preds_rf))
print('-'*50)
print("Model precision score on test is: ", round(precision_score(Y_test2, test_preds_rf,  average='micro'),3))
print('-'*50)
print("Model recall score on test is: ", round(recall_score(Y_test2, test_preds_rf,  average='micro'),3))
print('-'*50)
print("Model F1 score on test is: ", round(f1_score(Y_test2, test_preds_rf,  average='micro'),3))
print('-'*50)
cm = confusion_matrix(Y_test2, test_preds_rf)
plt.figure(figsize=(4,4))
sns.heatmap(cm, annot=True, fmt=".3f", linewidths=.5, square = True, cmap = 'Blues_r');
plt.ylabel('Actual label');
plt.xlabel('Predicted label');
plt.title('Confusion Matrix', size = 15);
plt.show()
```

```
Model accuracy on train is:  1.0
Model accuracy on test is:  0.9930795847750865
--------------------------------------------------
Model precision score on test is:  0.993
--------------------------------------------------
Model recall score on test is:  0.993
--------------------------------------------------
Model F1 score on test is:  0.993
--------------------------------------------------
```



GitHub Link:

https://github.com/govind01a/Student-Adaptability-Level

Video Link:

https://drive.google.com/file/d/1_7ufwDMyzLMjScVgXFLuzy-rszILyGZq/view?usp=sharing