iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** VI **Month of publication:** June 2022

DOI: https://doi.org/10.22214/ijraset.2022.43192

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue VI June 2022- Available at www.ijraset.com

A survey on Prediction and Analysis of Students Academic Performance Using Machine Learning Technique

Ashmina Khan[1], Prof. K. N. Hande[2]
[1]*Student, Department of Computer Science & Engineering, PBCOE, Nagpur*
[2]*Assistant Professor, Department Of Computer Science & Engineering, PBCOE Nagpur*

Abstract: ***Digital data trails from disparate sources covering different aspects of student life are stored daily in most modern university campuses. However, it remains challenging to combine these data to obtain a holistic view of a student, use these data to accurately predict academic performance, and use such predictions to promote positive student engagement with the university. In our study, first, an experiment is conducted based on a real-world campus dataset of college students that aggregates multisource behavioural data covering not only online and offline learning but also behaviors inside and outside of the classroom. Specially, to gain in-depth inside into the features leading to excellent or poor performance, matrix measuring the linear and nonlinear behavioural changes (e.g. regularity and stability) of campus lifestyles are estimated; furthermore, features representing dynamic changes in temporal lifestyle patterns are extracted by the means of long short-term memory (LSTM). Second, machine learning based classification algorithms are developed to predict academic performance. Finally visualized feedback enabling students (especially at risk-students) to potentially optimize their interaction with the university and achieve a study-life balanced is designed.***

Keywords: Academic performanceprediction, Behavioural Pattern, Digital Campus,Machine Learning (ML)

I. INTRODUCTION

As an important step to achieving personalized education, academic performance prediction is a key issue in the education data mining field. It has been extensively demonstrated that academic performance can be profoundly affected by the following factors:

1) Student Personality e.g., neuroticism, extraversion, and agreeableness.
2) Personal Status (e.g., gender, age, height, weight, physical fitness, cardiorespiratory fitness, aerobic fitness, stress, mood, mental health, intelligence, and executive functions.)
3) Lifestyle Behaviours (e.g., eating, physical activity, sleep patterns, social tie, and time management).
4) Learning Behaviours (e.g., class attendance, study duration, library entry, and online learning).

For example, investigated the incremental validity of the Big Five personality traits in predictingcollege GPA. Demonstrated that physical fitness in boys and obesity status in girls could be important factors related to academic achievement. Meanwhile, showed that a lifestyle could lead to performance among college students showed that the degree of efforts exerted while working could be strongly related corelated academic performance. Additionally, showed that compared with high- and medium achieving students, low achieving student were less emotionally engaged throughout the semester and tended to express more confusions during the final stage of the semester. By analyzing the effect of the factors influencing academic performance, many systems using data to predict academic performance have been developed in the literature. In a multitask predictive framework that captures intersemester and intermajor correlations and integrates student similarity was built to predict students' academic performance. According to their predicted academic performance, early feedbacks and interventions could be individually applied to at-risk students. In recent years, compared with primary and secondary education (i.e. K12) more and more attentions have been paid to the academic performance prediction for higher education. The reasons contributing to this phenomenon warrant further investigation and might include the following. First, for college students on a modern campus life involves a combination of studying, eating, exercising, socializing, etc. (see Fig. 1). All activities that students engage in (e.g., borrowing a book from the library) leave a digital trail in some database.

© IJRASET: All Rights are Reserved |

II. LITERATURE SURVEY

The educational data-mining has been researched extensively in the past and remains a hot areaof research in data-mining (DM), machine-learning (ML), deep-learning, and big-data. The aim of various forms of study is the development of a predictive framework, which will forecast the marks, grades, institutional ranking and institutional recommendations. Different tools and techniques are used to analyse and visualize the data. Below is some leading-edge work that helped us to explain our proposed methodology. The research explains the use of large-data application in education Big- data methods for learning analysis are used in different ways e.g., system performance prediction, visualization of data, student skills estimation, risk detection, fraud detection, system of course recommendation, grouping of students and collaboration with other students. The predictive analysis of this study focuses on student achievements, behaviour, and skill prediction enhances the usefulness of this work. During the forecast of student performance data-mining procedures were utilized to build a predictor framework for the final- marks reliant on students' achievements. A key element regression model trained, and used to forecast the academic achievement of students. As features, non-courses variables, for instance out-of- class student conduct, who notes, focus in video watching and postschool tutoring were used. The factors which influence the validity of application are examined in a study while working on EDM, two sort methods for data-analysis are used, that is, descriptive and predictive models. Descriptive model utilizes unsupervised-learning techniques to explain and recognizethe structure of mined-data while Predictive model utilizes supervised-learning techniques whichdetermine unknown-values. The objective of a study was to evaluate students at the beginning of an academic session and to forecast achievements through academic history, using collaborative filtering technique. The authorized courses reflecting the students learning are chosen. The information system gathered historical data to identify similar characteristics for students. technique by utilizing historical academic data of students as an input, in order to evaluate students' performance. The study was relying on a factorization of low-range-matrixes and dispersed linear model.

Fig. 1: List of common attributes and methods used in predicting student's performance.

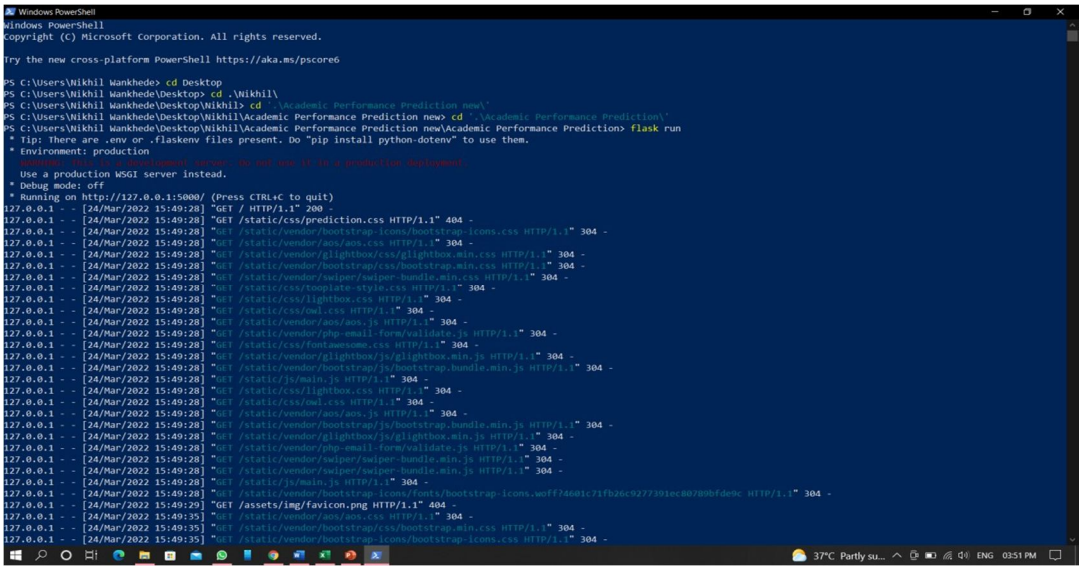
© IJRASET: All Rights are Reserved |

SN	Title of thePaper	Names of the Author	Methods / Techniques / protocols / Algorithm Used	Analysis andObservation
1	Academic Performance Prediction based on multisource, multifeatured behavioural data	Liang Zhao	Decision tree Algorithm, Multisource and Multifeatured.	This student Academic Prediction performance can potentially lead to continual investigation on largescale.
2	A review onPredicting students' Performance using Data Mining Techniques.	Amirah Mohamed Shahiri	Data Mining.	It helps the educational system to monitorthe students' performance in systematic way.
3	Predicting Academic Performance	Arto Hellas	Statistical and Machine learning methods.	Determine thecurrent state ofpredictingstudents' academic performance.
4	Prediction ofStudents Performance by Modellingsmall datasetsize.	Lubna Mahmoud Abu Zohair	Data Mining	It helps the educational system to monitorthe students' performance in systematic way.
5	Students Performance Prediction and Analysis	Ditika Bhanushai	Data Mining, Navies Bayes,ID3 Algorithm.	It reduces the time that iswasted in manual process.

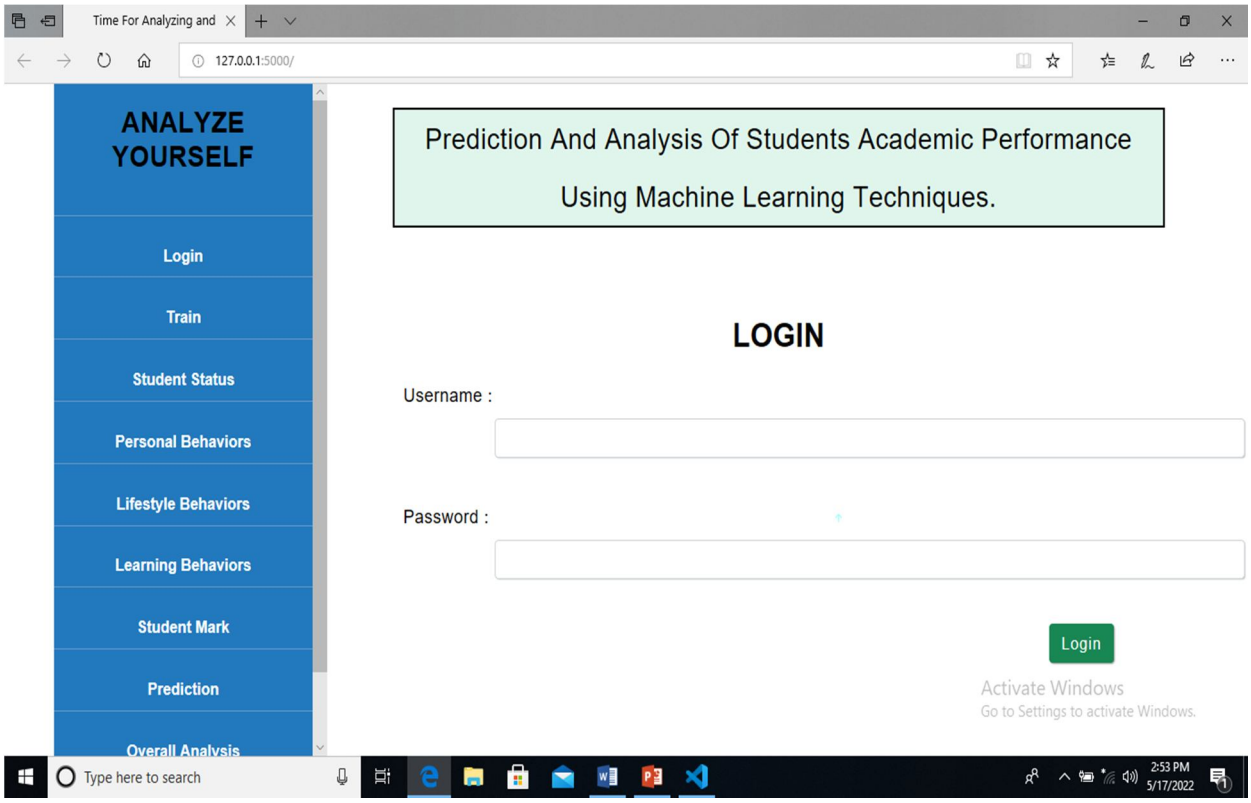
III. IMPLEMENTED WORK

Academic performance prediction is considered as a classification problem. According to the high-low discrimination index proposed by Kelley academic performance is divided into low-, medium-, and high groups. Given a digital campus dataset, according to Fig. 2, the main task is to first extract features from the raw multisource data; then select the features that are strongly correlated with academic performance and use these features to train the classification algorithm; and finally provide visualized feedback based on the prediction results.

Step 1- Open Windows PowerShell then type command flask run

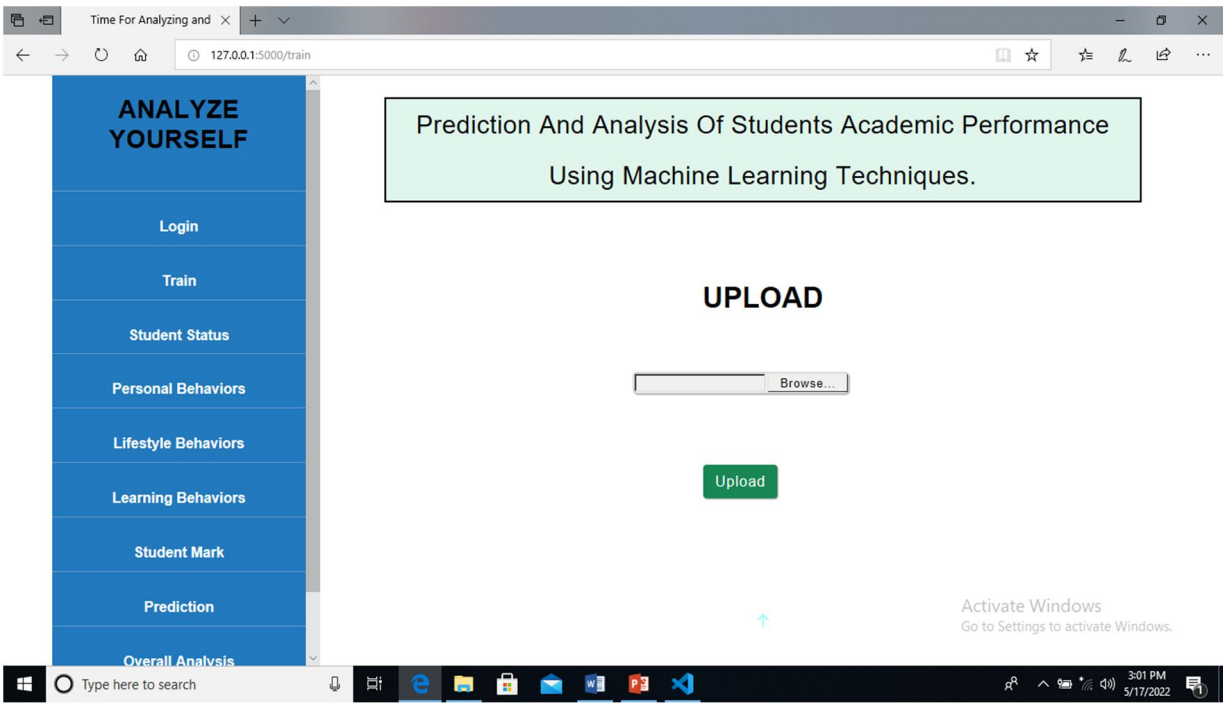


Screenshot 1 : Start Live Server.



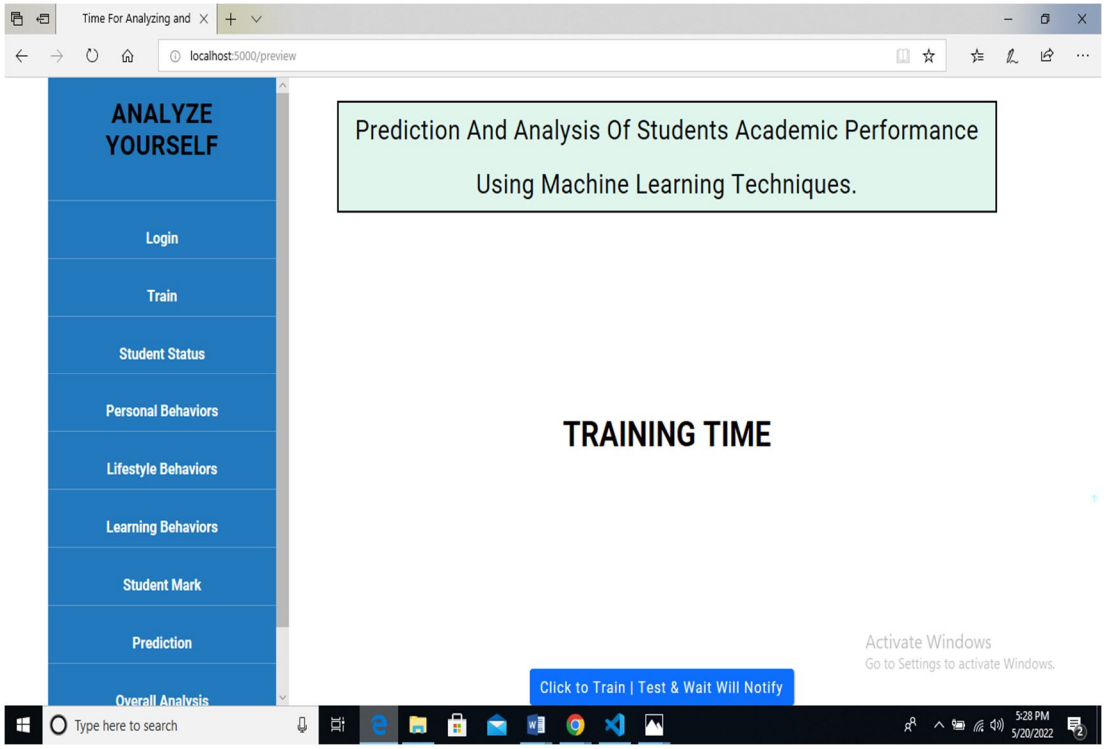
Screenshot 2. : Login Page

Step 3 – Upload CSV file on this project.



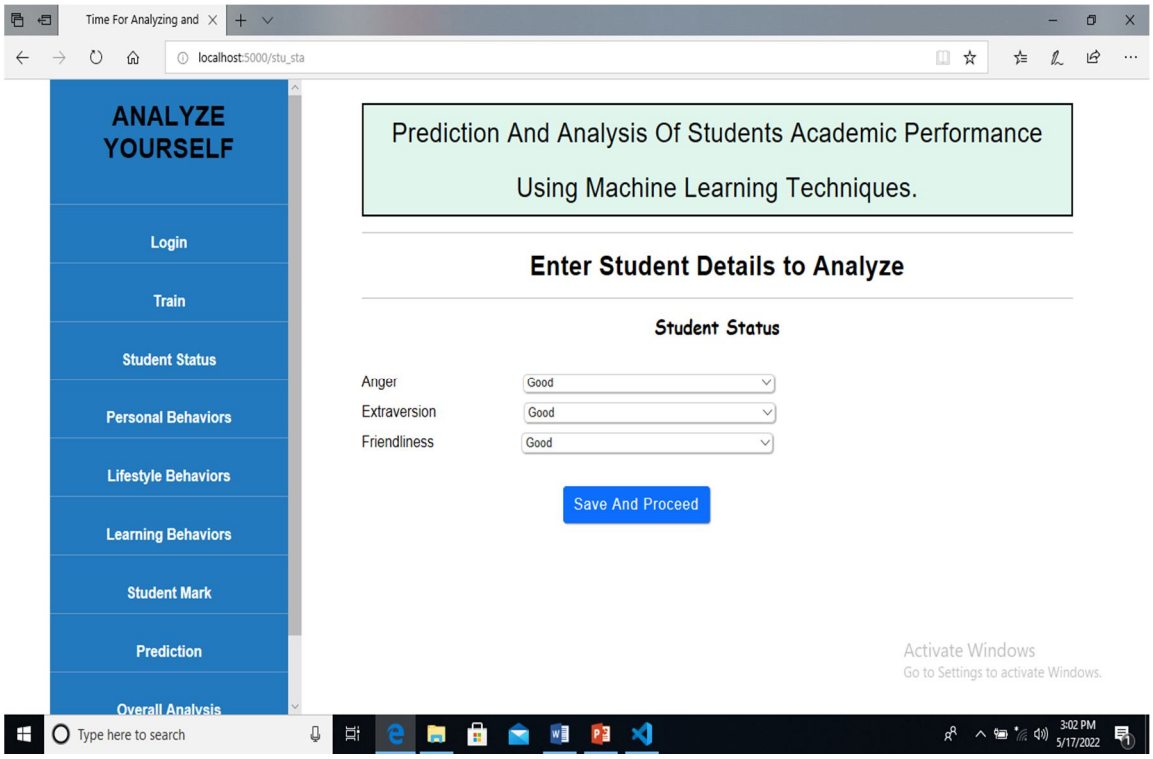
Screenshot 3 : Upload CSV File.

Step 4 – After upload the CSV file to Train and Test.



Screenshot 4 : Training Time

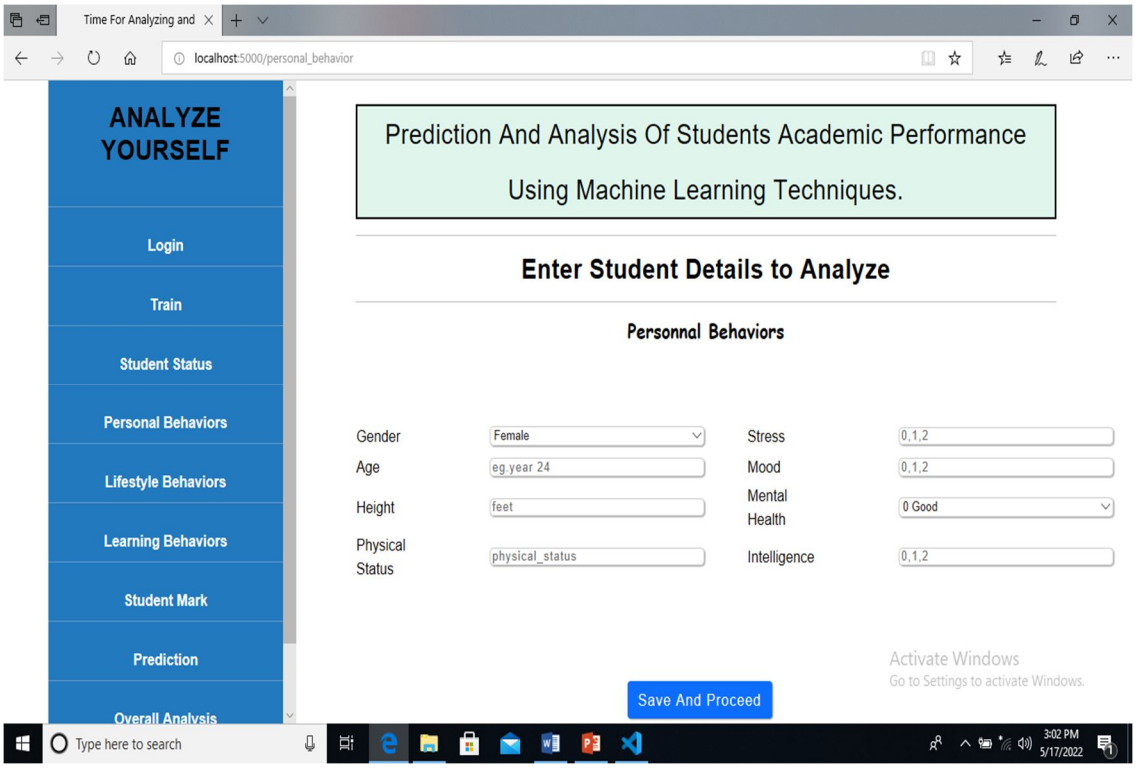
Step 5 – Give some inputs to the Student Personality Page.



Screenshot 5 : Student Personality.

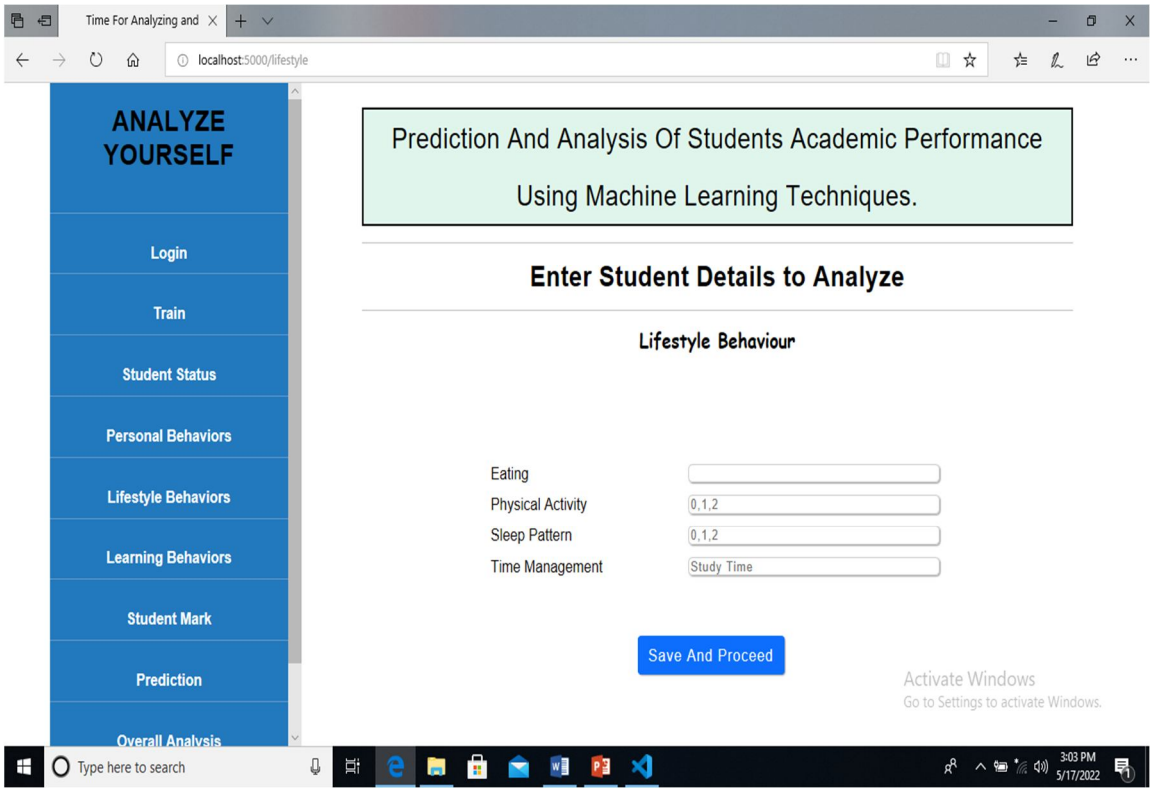
© IJRASET: All Rights are Reserved |

Step 6 – Add Personal Status for Student in 0, 1, 2 Format 0 – Good, 1- Average, 2- Bad.



Screenshot 6: Student Personal Status.

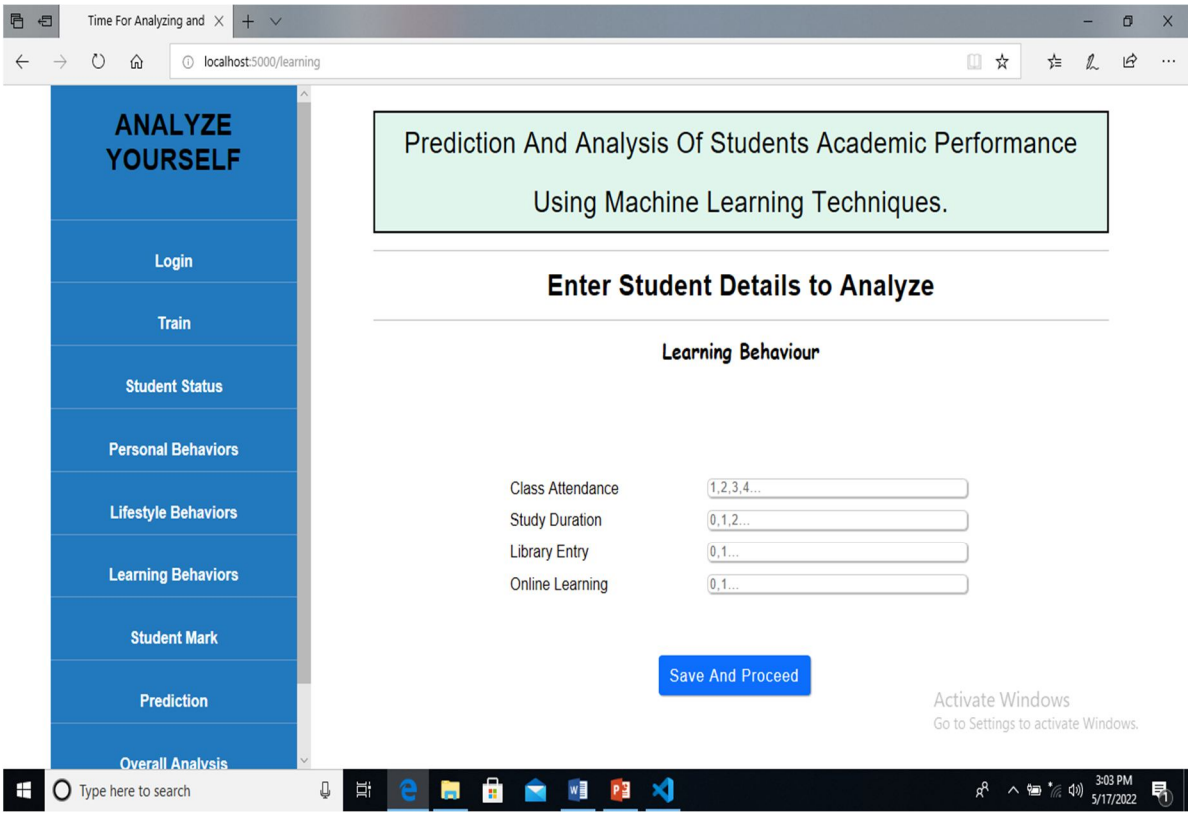
Step 7 – Enter Lifestyle Behaviour for Student.



Screenshot 7: Student Lifestyle Behaviors

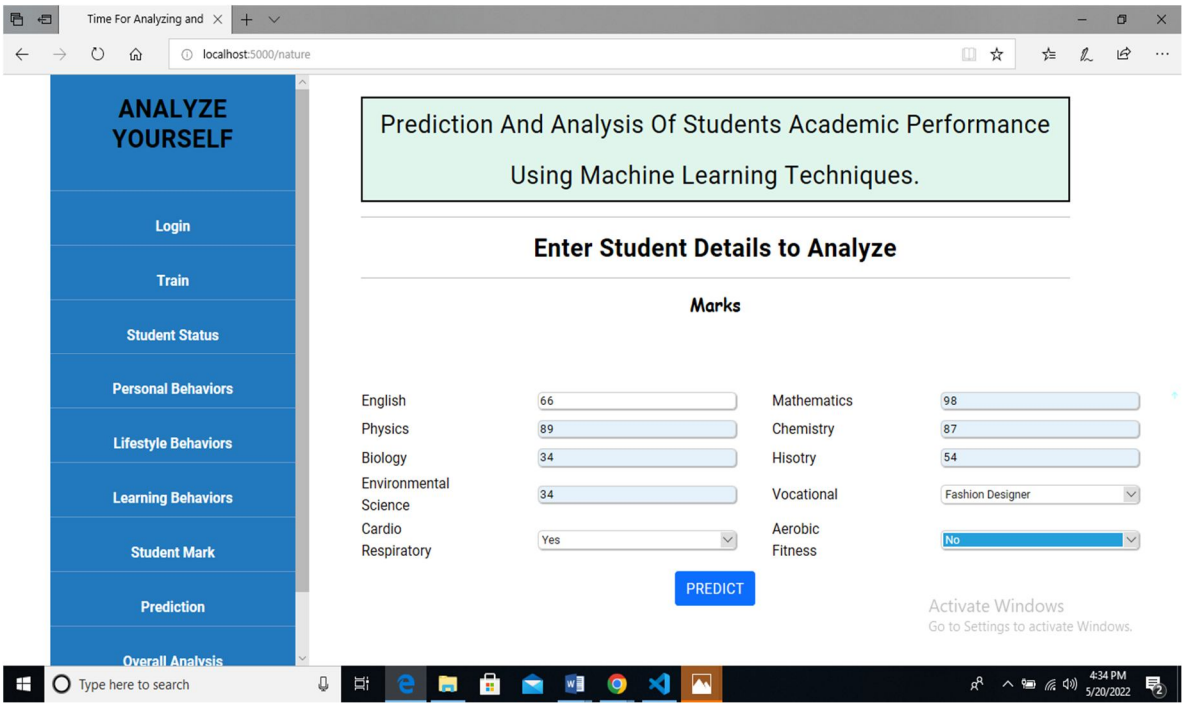
© IJRASET: All Rights are Reserved |

Step 8 – Enter Learning Behaviour for Student.



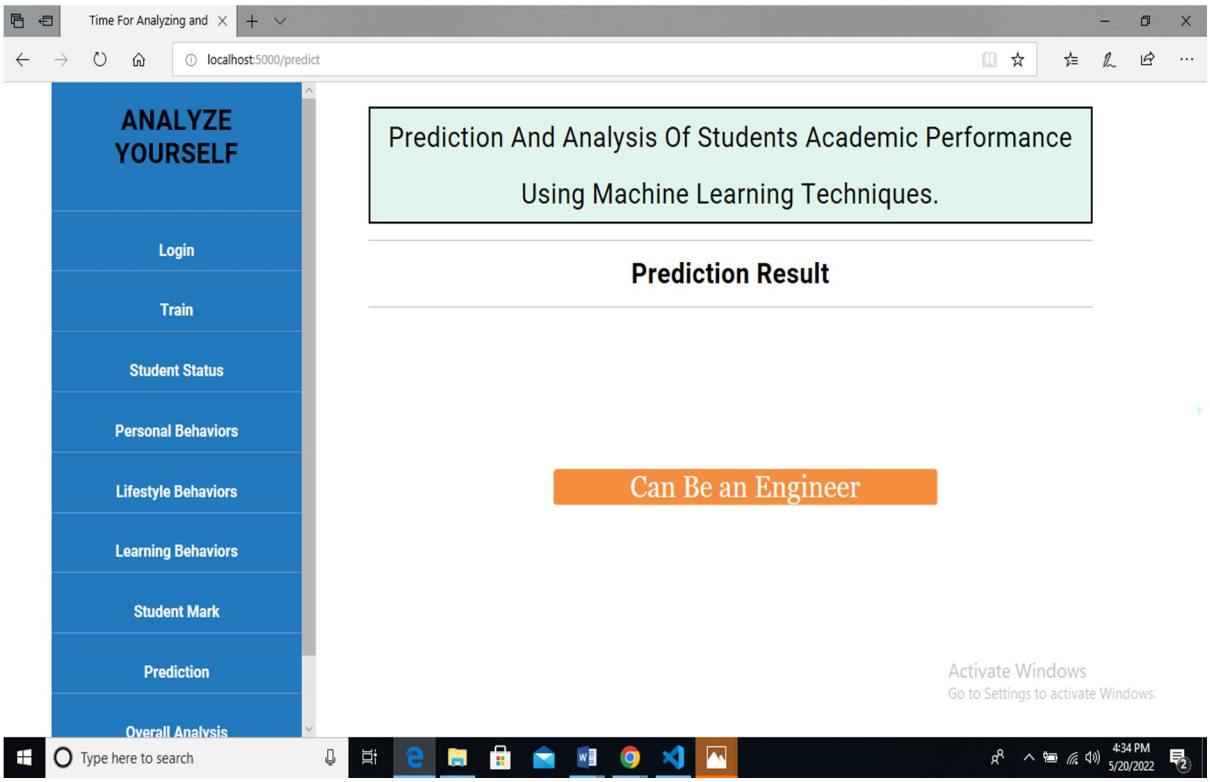
Screenshot 8 : Student Learning Behaviors

Step 9 – Enter the Student Marks for the based on 12th Standard.



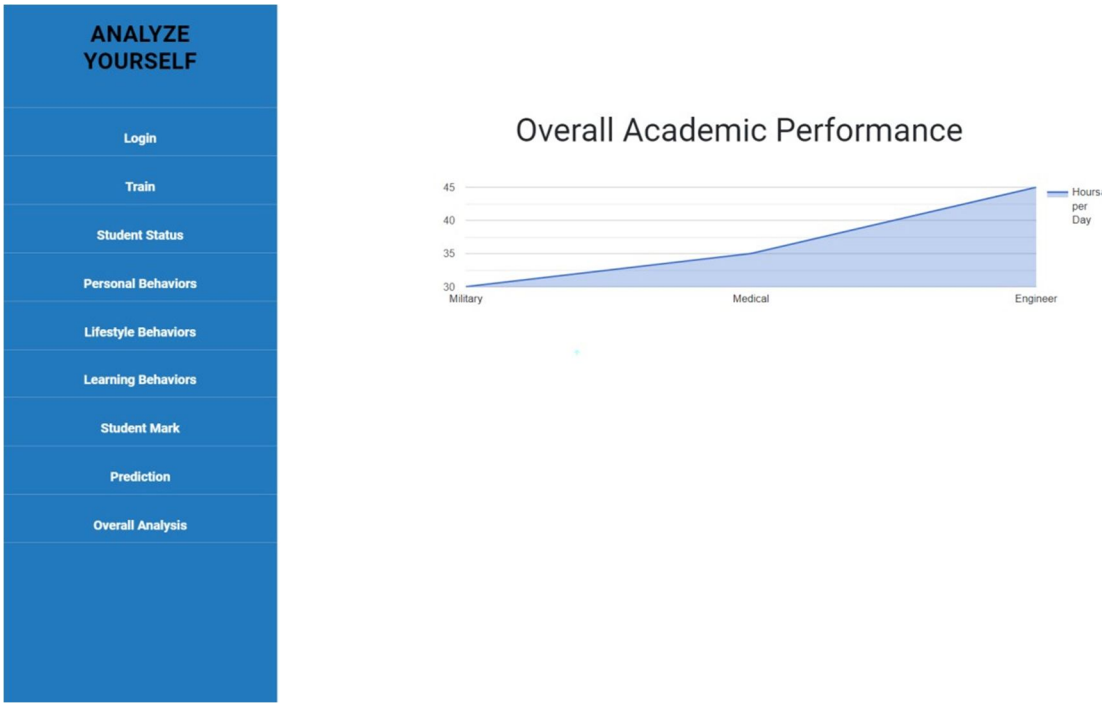
Screenshot 9 : Student Marks.

Step 10 – Prediction Result for which field have to go on the based on Student Marks.



Screenshot 10 : Prediction Page.

Step 11 – Overall Academic Performance chart graph.



Screenshot 13 : Overall Academic Performance.

IV. METHODOLOGY

Academic performance prediction is considered as a classification problem. According to the high-low discrimination index proposed by Kelley academic performance is divided into low-, medium-, and high- groups. Given a digital campus dataset, according to Fig. 2, the main task is to first extract features from the raw multisource data; then select the features that are strongly correlated with academic performance and use these features to train the classification algorithm; and finally provide visualized feedback based on the prediction results.

In this section, the three modules designed in AugmentED (see Fig. 2) are described in details.

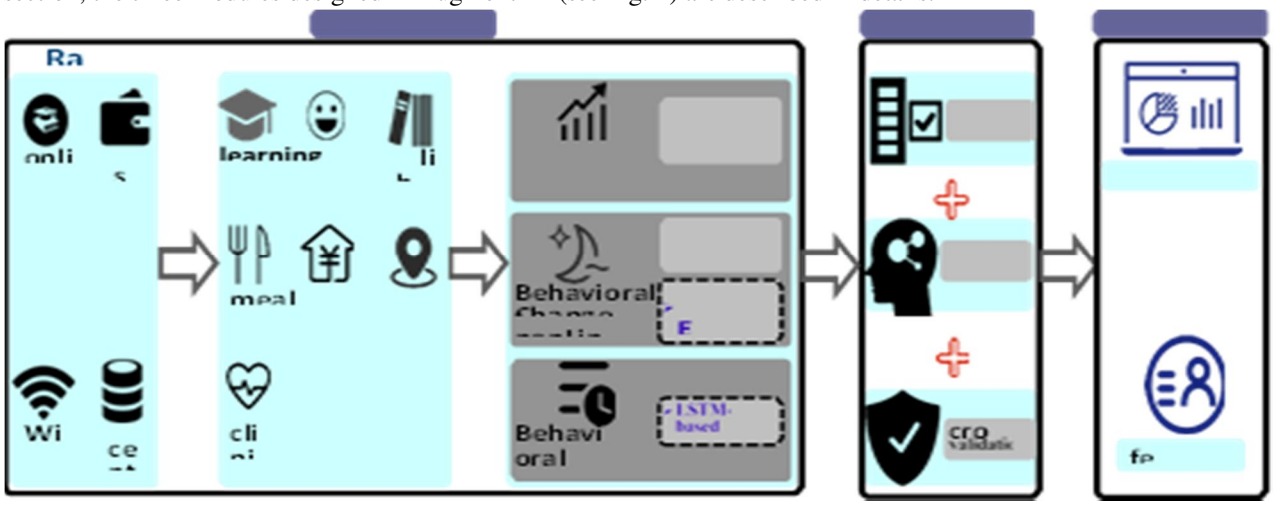


Fig:2 Overview of AugmentED.

In the data module, the features blocked in dashed boxes (including LyE, HurstE, DFA, and LSTM-based features) are proposed in our study, to the best of our knowledge, which is usedfor the first time in student's behavioural analysis.

The main aim of the system is to predict the future performance of the student using certain data of the student such as pervious semester marks, attendance records, etc. After predicting the student performance, the system will also compare the results generated by two classification algorithms and there after determine which of them is more accurate and efficient.

The data to be provided as the input must have the values of the attributes classified into specific variables, for example, the student marks for the previous semester can be classified as good if marks >= 70%, average if 70% > marks >=55% and poor if marks.

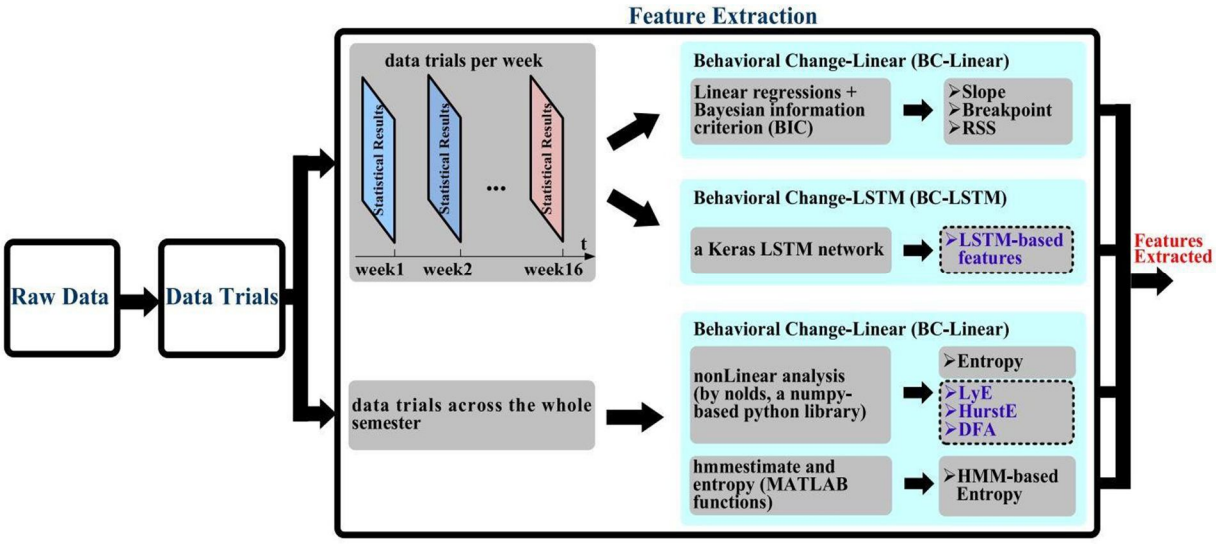


Fig 3: Prediction Module

© IJRASET: All Rights are Reserved |

V. BACKGROUND STUDY

This section introduces the basic concepts of student outcomes and student performance, followed by identifying the research gaps in the literature concerning the prediction of student learning outcomes.

A. *Student Outcomes*

Outcome-based education (OBE) has emerged as a new school of thought in education and has recently enjoyed wide acceptability and adoption. This educational paradigm shifts the focus of the teaching and learning process from the traditional teacher objectives to the so-called student outcomes. In simple terms, student outcomes refer to the knowledge, skills, and values to be attained by the students at the time of graduation or at end of a course. The outcomes, representing the targeted competencies, might be defined and measured at the course level, i.e., course outcomes, or program level, i.e., program outcomes. Essentially, course outcomes enable the accomplishment of program outcomes, and their alignment (i.e., courses to program) is performed in a critical activity referred to as curriculum mapping. Computerized tools were developed to assist in realizing the OBE goals and effectively document the educational assessment activities. Their usefulness could be extended by incorporating intelligent models thatcan prognosticate the attainment of learning outcomes during academic terms. Measuring studentoutcomes in higher education indisputably brings about various benefits, including the establishment of program expectations for students and course instructors, the practical assessment of the quality of courses and programs, and the provision of key success indicators ofthe program, among others. Several quality assessments instruments, e.g., and quality assurance frameworks, e.g., were proposed to realize the outcome-based education philosophy and acquire program accreditations. Moreover, the ability to forecast the attainment of student outcomes adds further invaluable advantages, such as the ability to introduce corrective interventions to the learning processes. However, few works surveyed the intelligent prediction of student outcomes. Furthermore, the factors and attributes that impact educational outcomes are still vague. Studies suggest that these factors range between academic factors, e.g., teaching quality and online engagement and non-academic traits, e.g., family engagement and student motivation. In this work, we aim, through a systematic survey, to understand the landscape of student outcomes prediction using data mining and machine learning, identify the main challenges hindering the prediction of student outcomes, and propose relevant recommendations.

B. *Student Performance*

Albeit, due to the substantial educational shift in teaching and learning, i.e., OBE, student performance remains a significant concern in higher education especially given the low grades and increasing dropout rates even at world-class universities. Previous reviews showed that the cumulative GPA and course assessments are the most used predictors of student performance and success. Indeed, several studies used next-term course grades as the main indicator of student performance, e.g., However, it is not uncommon to measure student performance in otherforms, including dropout rate, student knowledge, post-course outcomes, among other indicators.In our view, student academic performance should not be assessed using assessment grades only.Instead, it should be studied within a broader context, particularly using the student outcomes, which are now guiding the learning process by looking at the cohort performance. Moreover, recent research recommends exploring the prospect of predicting the attainment of student outcomes to infer student performance.

The intelligent techniques employed in learning analyticsto forecast student achievements are generally categorized into supervised learning, unsupervised learning, data mining, and statistical approaches. Each category incorporates a wealth of intelligent algorithms, such as Artificial Neural Networks, Support Vector Machine, K-Nearest Neighbour, and Random Forests. The attributes that predict student performance are surveyed extensively in the literature, leading to a mix of academic (e.g., pre-admission scores and entry qualifications) and non-academic factors (e.g., emotional intelligence and resilience). However, mystery still surrounds the factors that influence the attainment of course and program outcomes.Measurable student outcomes are developed to improve the quality of learning processes and educational programs. Effectively, these outcomes assess what students can perform with what they have learned. The attainment of learning outcomes, both at the course and program level, is performed using direct and indirect assessment methods at the end of the learning process. The direct assessment methods seek to find tangible evidence demonstrating student learning, while the indirect methods rely on the students' reflections on their learning experience. To calculate the attainment rate of outcomes, one should identify a priori the attainment targets and levels andthen properly align student grades to the appropriate attainment level. In our work, we examined the studies that predict the attainment of student outcomes, irrespective of their form.

© IJRASET: All Rights are Reserved |

C. Existing Student Performance Reviews and Literature Gaps

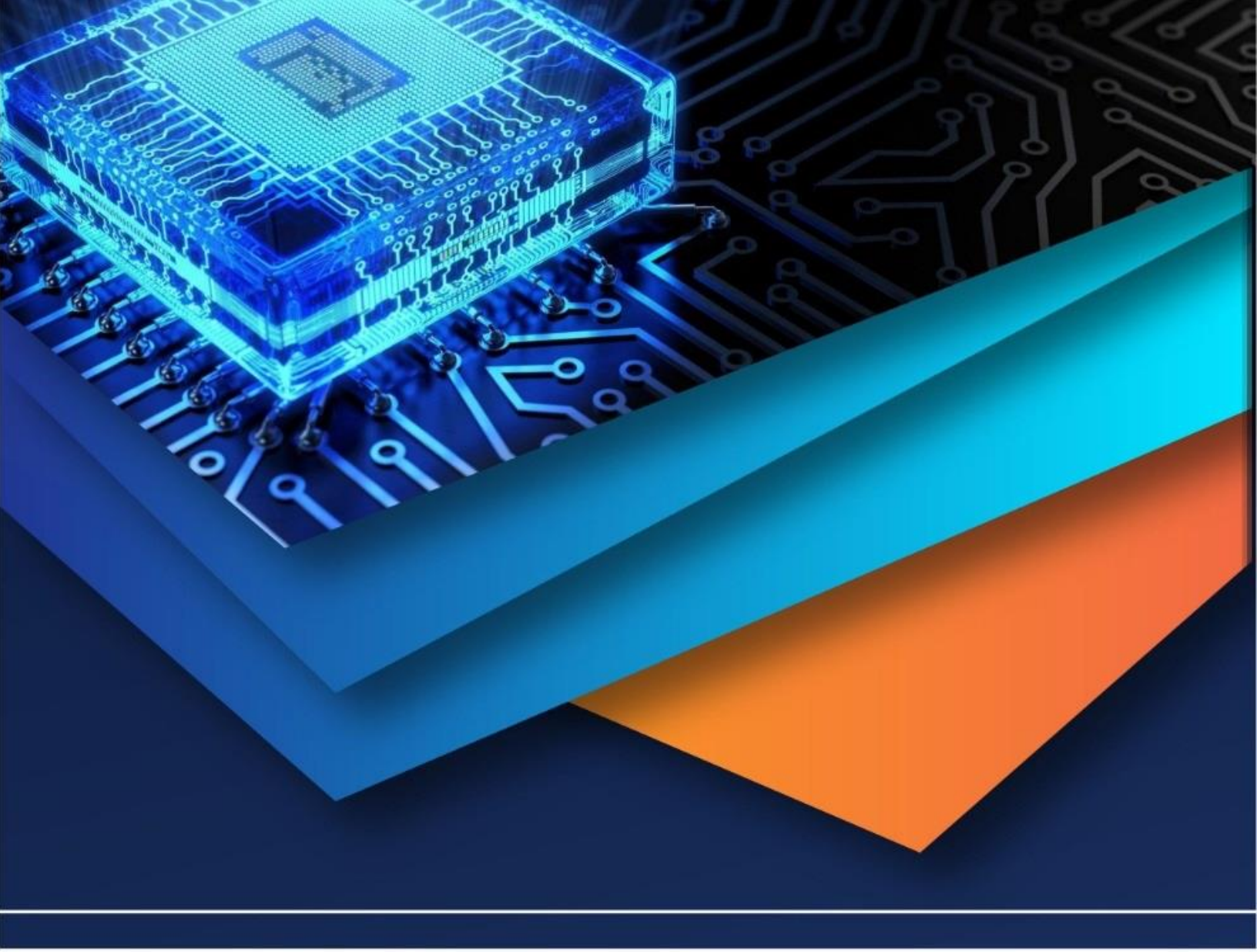
Our extensive review of previous surveys revealed that, to the best of our knowledge, no systematic literature survey was carried out focusing on the prediction of novel approach for student academic performance from the learning outcomes perspective. Summarizes the prominent surveys carried out on the prediction of student performance and emphasizes their focus and weaknesses. Indeed, our search returned numerous surveys on the use of data mining techniques in education (i.e., EDM) to unravel student modelling activities and predict academic performance. These reviews suffered from several limitations, for they were generally broad, did not focus on using student outcomes as an indicator of student performance, suffered from quality issues (e.g., methodologies not thoroughly defined), and were not published in highly indexed venues. Other less relevant surveys published in the field focused on the effects of homework assignments on student performance the impact of using interactive whiteboards on student achievement the predictors of student success in the first year of study and the factors of graduate success. Unlike the above-mentioned surveys, our research opted to conduct a systematic review by implementing a comprehensive review process that allows synthesizing concrete answers to well-defined research questions, in the context of predicting student learningoutcomes.

VI. CONCLUSION

As an important issue in the education data mining field, academic performance prediction has been studied by many researchers. However, due to lack of richness and diversity in both data sources and features, there still exist a lot of challenges in prediction accuracy and interpretability. To initially alleviate this problem, our study aims at developing a robust academic performance prediction model, to gain an in-depth insight into student behavioural patterns and potentially help students to optimize their interactions with the university. In our study, a model named AugmentED is proposed to predict the academic performance of college students. Our contributions in this study are related to three sources. First, regarding data fusion, to the best of our knowledge, this work is the first to capture, analyse and use multisource data covering not only online and offline learning but also campus-life behaviours inside and outside of the classroom for academic performance prediction. Based on these multisource data, a rich profile of a student is obtained. Second, regarding the feature evaluation, behavioural change is evaluated by linear, nonlinear, and deep learning (LSTM) methods respectively, which provides a systematically view of students' behavioural patterns. Specifically, it is the first time that three novel nonlinear metrics (LyE, HurstE, and DFA) and LSTM are applied in students' behavioural time series analysis. Third, our experimental results demonstrate that AugmentED can predict academic performance with quite high accuracy, which help to formulate personalized feedback for at-risk (or unself disciplined) students. However, there are also some limitations in our study. To gain a multisource dataset, we scarified the scale the dataset by only using student-generated data within a single course. This limitation might have a certain negative influence on the generalization of AugmentED. Furthermore, in this study, we mainly focus on behavioural change. Other characteristics/features (e.g., peer effect, sleep) that are worthy of consideration were not evaluated in this study. In conclusion, our study is based on a complete passive daily data capture system that exists in most modern universities. This system can potentially lead to continual investigations on a larger scale. The knowledge obtained in this study can also potentially contribute to related research among K-12 students.

REFERENCES

[1] K. Porter, K. J. Matthews, D. Salvo, and H. W. Kohl, ''Associations of physical activity, sedentary time, and screen time with cardiovascular fitness in United States adolescents: Results from the NHANES national youth fitness survey (NNYFS),'' Oct. 2017.
[2] K. N. Aadland, Y. Ommundsen, E. Aadland, K. S. Brønnick, A. Lervåg, G. K. Resaland, and V. F. Moe, ''Executive functions do not mediate prospective relations between indices ofphysical activity and academic performance: The active smarter kids (ASK) study,'' Jun. 2017.
[3] E. L. Faught, J. P. Ekwaru, D. Gleddie, K. E. Storey, M. Asbridge, and P. J. Veugelers, ''Thecombined impact of diet, physical activity, sleep and screen time on academic achievement: A prospective study of elementary school students in Nova Scotia, Canada,'' Dec. 2017.
[4] V. Kassarnig, E. Mones, A. Bjerre-Nielsen, P. Sapiezynski, D. D. Lassen, and S. Lehmann, ''Academic performance and behavioral patterns,'' Dec. 2018.
[5] N. Morita, T. Nakajima, K. Okita, T. Ishihara, M. Sagawa, and K. Yamatsu, ''Relationships among fitness, obesity, screen time and academic achieve7ment in Japanese adolescents,'' Sep. 2016.
[6] H. Yao, D. Lian, Y. Cao, Y. Wu, and T. Zhou, ''Predicting academic performance for collegestudents: A campus behavior perspective,'' ACM Trans. Intell. Syst. Technol., Feb. 2019.
[7] X. Zhang, G. Sun, Y. Pan, H. Sun, Y. He, and J. Tan, ''Students perforamance modeling based on behavior pattern,'' Oct. 2018.
[8] A. Akram, C. Fu, Y. Li, M. Y. Javed, R. Lin, Y. Jiang, and Y. Tang, ''Predicting students' academic procrastination in blended learning course using homework submission data,'' Nov. 2019.
[9] Z. Wu, W. Lin, P. Liu, J. Chen, and L. Mao, ''Predicting long-term scientific impact basedon multi-field feature extraction,'' Mar. 2019.
[10] Z. Liu, C. Yang, S. Rüdian, S. Liu, L. Zhao, and T. Wang, ''Temporal emotion-aspect modeling for discovering what students are concerned about in online course forums,'' Aug. 2019.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)