International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

IJIEMR Transactions, online available on 28th nov 2021.
Link : https://ijiemr.org/downloads/Volume-10/Issue-11

DOI: 10.48047/IJIEMR/V10/I11/37

Title **ACADEMIC PERFORMANCE PREDICTION BASED ON MULTISOURCE, MULTIFEATURE BEHAVIORAL DATA WITH ML**

volume 10, Issue 11, Pages: 241-245
Paper Authors:**AARE MOHANAPRIYA, V.DINESH BABU**

USE THIS BARCODE TO ACCESS YOUR ONLINE PAPER

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

ACADEMIC PERFORMANCE PREDICTION BASED ON MULTISOURCE, MULTIFEATURE BEHAVIORAL DATA WITH ML

[1]AARE MOHANAPRIYA, [2]V.DINESH BABU
[1]M.Tech Scholar, Dept of CSE, Chebrolu Engineering College, Chebrolu, Andhra Pradesh, India
[2]Associate Professor, Dept of CSE, Chebrolu Engineering College, Chebrolu, Andhra Pradesh, India

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 (i) combine these data to obtain a holistic view of a student, (ii) use these data to accurately predict academic performance, and (iii) use such predictions to promote positive student engagement with the university. To initially alleviate this problem, in this paper, a model named Augmented Education (Augment ED) is proposed. In our study, (1) first, an experiment is conducted based on a real-world campus dataset of college students (N = 156) that aggregates multisource behavioural data covering not only online and offline learning but also behaviours inside and outside of the classroom. Specifically, to gain in-depth insight into the features leading to excellent or poor performance, metrics 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). (2) Second, machine learning-based classification algorithms are developed to predict academic performance. (3) Finally, visualized feedback enabling students (especially at-risk students) to potentially optimize their interactions with the university and achieve a study-life balance is designed. The experiments show that the Augment ED model can predict students' academic performance with high accuracy.

KEY WORDS: Academic Performance Prediction, Behavioural Pattern, Digital Campus, Machine Learning (ML), Long Short-Term Memory (LSTM).

I.INTRODUCTION

In recent times, machine learning (ML) practices have been a big deal in various industries in the world including the educational frontier [1]. The need to automate different tasks such as grading students, improving student retention, testing students, predicting student performance, as well as administrative tasks like material optimizations in the academic facet, has called for the application of machine learning techniques and methods.

Arguably, the most important task in every learning institution is to monitor and improve their student's performance. Early prediction of student [2] performance in the right fashion will improve student retention as well as the testing methods used for the students. This practice will also aid the educators and education policymakers by giving them better information about their students' learning ability as well as how best they can help students who are lagging in a given set [3].

This study focuses on reviewing previous research works on building models to predict student's performance in a learning environment in the last ten (10) years [4]. The authors of this article developed a systematic approach to the review work. This approach is to support the objectives of this study, which are:

1. To identify the existing prediction methods and the tools employed for predicting students performance

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

2. To study and identify the variable type used for the predictive process [5].
3. To identify and study the researchers who employed these learning models to analyze student's performance.

II. RELATED WORK

A. FEATURE EXTRACTION

Feature evaluation plays an important role in designing prediction systems. Features that measure the various behavioural patterns can enhance our understanding of how a student's behaviour changes as the semester progresses. In this part, on the one hand, previous features that quantify students' behavioural patterns are summarized; On the other hand, new features worthy of inclusion are also introduced. In general, behavioural change can be quantified by the following three groups of metrics.

1) BEHAVIORAL CHANGE-LINEAR (BC-LINEAR) Traditionally, behavioural change is mainly quantified by two linear metrics: behavioural slope and behavioural breakpoint. First, the behavioural slope can be captured by computing the slope of the behavioural time series of each student using a linear regression. The value of the slope indicates the direction and strength of the behavioural changes, e.g., a positive slope with a greater absolute value indicates a faster increase in behavioural change. Given a mid-term day during the semester, both the pre-slope and post-slope can be calculated to represent the students' behavioural change during the first and second halves of the semester, respectively.

Second, the behavioural breakpoint can be captured by computing the rate of behavioural changes occurring across the semester. The value of the breakpoint identifies the day during the semester before and after which a student's behavioural patterns differed. Two linear regressions can be used to fit a behavioural time series and then use the Bayesian information criterion (BIC) to select the best breakpoint. If a single regression algorithm is selected, the breakpoint can be set to the last day.

2) BEHAVIORAL CHANGE-NON LINEAR (BC-NONLINEAR)

In recent years, nonlinear metrics have been increasingly applied to time series analysis. Regarding the students' behavioural time series, nonlinear metrics have been used to discover nonlinear behavioural patterns. We consider entropy an example. In entropy is proposed to quantify the regularity/orderliness of students' behaviours, and it was demonstrated that a small entropy value generally leads to high regularity and high academic performance. Another example is entropy calculated based on a Hidden Markov Model (HMM) analysis, which is called HMM-based entropy for simplicity in our study.

HMM-based entropy is proposed to quantify the uncertainty/diversity of students' behaviours, e.g., the uncertainty between the transition of different behaviours and the various activities that behaviour exhibits. In HMM based entropy is evaluated by the following two steps: (i) extracting the hidden states of a behavioural time series by HMM; and (ii) subsequently calculating the HMM based entropy of the extracted hidden states.

B. PREDICTION ALGORITHMS

In general, academic performance prediction can be considered either a regression or a classification problem. A wide variety of algorithms have been used/proposed in literatures to predict academic performance.

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

C. MULTISOURCE AND MULTI FEATURE

It has been verified in many literatures that the predictive power could be improved by multisource data and multi featured fusion. For example, it is demonstrated that the performances of predicting both at-risk students and stock market could be improved by combining multisource data. Similarly, the performances of academic performance prediction are improved by combing traditional diligence features with orderliness (and sleep patterns) features.

III. PROPOSED SYSTEM

In our study, 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 (1) 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.

A. DATA MODULE A flowchart of this module is shown in Fig. 3, which includes the following three parts.

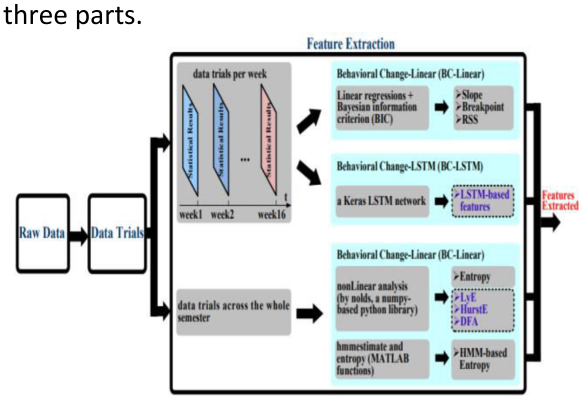

Fig. 1: Flowchart of the data module

1) RAW DATA Permission to access the raw data was granted by the Academic Affairs Office of our university. The raw dataset used in our study was captured from students engaging in the course of "Freshman Seminar" during the fall semester of 2018-2019. The "Freshman Seminar" was chosen for the following reasons: (1) more students were enrolled in this course (N = 156) than other comparable courses, and (2) these 156 students were more active on our self-developed SPOC platform, thus providing abundant valuable behavioural data. Our dataset consists of the following four data sources.

SPOC Data: Two different types of data were collected on the SPOC platform. The first type is log files, which are recorded when a student logs in or out of the system, and the second type is posts on the SPOC discussion forum, which records discussions related to students' learning experience.

Smart Card Data: Similar to most modern universities, in our university, all students have a campus smart card registered under their real name. The usage of this smart card, such as for borrowing books from the library, entering the library, consuming meals in campus cafeterias, shopping on campus, or making an appointment with the school clinic, is captured daily.

WiFi Data: There are approximately 3000 wireless access points at our university, covering most areas of campus. Once a student passes by one of these points, the MAC address of his/her device (e.g., tablet, laptop, or smart phone) can be recorded. In our study, to distinguish among diverse behaviours, the entire campus is divided into several different areas, including a study area and a relaxation/dormitory area.

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

Central Storage Data: other features used in our study, including the students' personal information and academic records, are recorded by the central storage system of our university.

IV. 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. 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 students.

V.REFERENCES

[1] A. Furnham, and J. Monsen, "Personality traits and intelligence predict academic school grades," Learning and Individual Differences, vol. 19, no. 1, pp. 0-33, 2009.

[2] M. A. Conard, "Aptitude is not enough: How personality and behavior predict academic performance," Journal of Research in Personality, vol. 40, no. 3, pp. 339-346, 2006.

[3] T. Chamorropremuzic, and A. Furnham, "Personality predicts academic performance: Evidence from two longitudinal university samples," Journal of Research in Personality, vol. 37, no. 4, pp. 319- 338, 2003.

[4] R. Langford, C. P. Bonell, H. E. Jones, T. Pouliou, S. M. Murphy, and E. Waters, "The WHO health promoting s chool framework for improving the health and well-being of students and their academic achievement," Cochrane Database of Systematic Reviews, vol. 4, no.4, pp. CD008958, 2014.

[5] A. Jones, and K. Issroff, "Learning technologies: Affective and social issues in computer-supported collaborative learning," Computers & Education, vol. 44, no. 4, pp. 395-408, 2005.

[6] D. N. A. G. Van, E. Hartman, J. Smith, and C. Visscher, "Modeling relationships between physical fitness, executive functioning, and academic achievement in primary school children," Psychology of Sport & Exercise, vol. 15, no. 4, pp. 319-325, 2014.

[7] R. Wang, F. Chen, Z. Chen, T. Li, and A. T. Campbell, "StudentLife: Assessing mental health, academic performance and behavioral trends of college students using smartphones," In Proc. of the ACM International Joint Conference on Pervasive & Ubiquitous Computing, Seattle, WA, USA, 2014.

[8] R. Wang, G. Harari, P. Hao, X. Zhou, and A. T. Campbell, "SmartGPA: How smartphones can assess and predict academic performance of college students," In Proc. of the ACM International Joint Conference on Pervasive & Ubiquitous Computing, Osaka, Japan, 2015.

[9] M. T. Trockel, M. D. Barnes, and D. L. Egget, "Health-related variables and academic performance among first-year college students: Implications for sleep and other behaviors," Journal of American College Health, vol. 49, no. 3, pp. 125-131, 2000.

[10] D. M. Hansen, S. D. Herrmann, K. Lambourne, J. Lee, and J. E. Donnelly, "Linear/nonlinear relations of activity and fitness with children's academic achievement," Med Sci Sports Exerc. vol. 46, no. 12, pp. 2279-2285, 2014.

[11] A. K. Porter, K. J. Matthews, D. Salvo, and H. W. Kohl, "Associations of physical

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

activity, sedentary time, and screen time with cardiovascular fitness in United States adolescents: Results from theNHANES national youth fitness survey (NNYFS)," Journal of Physical Activity and Health, pp. 1-21, 2017.
[12] K. N. Aadland, O. Yngvar, A. Eivind, K. S. Bronnick, L. Arne, and G. K. Resaland, "Executive functions do not mediate prospective relations between indices of physical activity and academic performance: the active smarter kids (ask) study," Frontiers in Psychology, vol. 8, pp. 1088, 2017.