An International Multidisciplinary Journal
अनन्यशास्त्रम्™
(A Unique Treatise of Knowledge)
ISSN:3049-3927 (Online)

ANANYAŚĀSTRAM:
An International Multidisciplinary Journal
(A Unique Treatise of Knowledge)
Peer-Reviewed | Refereed | Open Access Journal
DOI: https://doi.org/10.64328/aimjv01i030043

Volume: 01 | Issue: 03 | December: 2025
Review Article

Predictive analytics model through ML for enhancing Holistic Education and wellness under NEP-2020

Dr Bhapkar H.R.[1], Pranjal Mahesh Dalvi[2], Vedanti Swapnil Dalvi[3], Rohit Raskar[4] and * Dr Pratibha Vijay Jadhav

Abstract

Applying statistical models with statistical approaches has developed a powerful tool for advancement of education systems. The Neuro NSGA_MHED dataset is used to explore predictive frameworks which captures all multi-dimensional variables like health, lifestyle, learning outcomes. The relationships were identified by carrying statistical analysis which defines the relationships between wellness indicators and educational attributes. Also some machine learning methods are applied to classify the categories of learners and predict their performance. Results proves that the ensemble models performed for only single learners with random forest but it achieves robust predictive accuracy. Results gives confirmation that if we combine statistical methods with machine learning methods it helps to develop systems easily. Having large datasets and analysing them can derive useful insights and built more strategies for progress of institutions.

Keywords: Descriptive and inferential methods, Machine Learning methods and Clustering algorithms etc.

Introduction

The Indian Higher Education has undergone a policy which mainly focuses on progress of organization in terms of physical, emotional, and wellness education .This guarantees that student are prepared for long-time not only for academics. The NEP-2020 highlights the importance about student centric and also inspires the

[1] Associate Professor and Department of Mathematics, Central University of Kashmir, Tulmulla, J&K, India. Email: hrbhapkar@gmail.com@gmail.com, https://orcid.org/0000-0002-8814-5698.

[2] PG Student, Department of Applied Science and Humanities, MIT School of Engg and Science, MIT-ADT University, Pune. India. Email: pranjaldalvi06@gmail.com

[3] PG Student, Department of Applied Science and Humanities, MIT School of Engg and Science, MIT-ADT University, Pune. India. Email: vedantidalvi626@gmail.com

[4] Assistant Professor, Department of Applied Science and Humanities, MIT School of Engg. and Science, MIT-ADT University, Pune. India. Email: Rohit.Raskar@mituniversity.edu.in,

*Coresponding Author: Assistant Professor, Department of Applied Science and Humanities, MIT School of Engg and Science, MIT-ADT University, Pune. India, Email: pratibhajadhav1@gmail.com, https://orcid.org/0000-0002-6940-9563

incorporation of advanced, evidence-based practices into teaching and institutional structure to foster holistic growth.

The data-driven health and education initiatives have received more attention on a national and international levels in framing parallel with policy making. There are now more opportunities for predictive analytics and machine learning (ML) applications to support decision-making, personalise learning, and improve student outcomes due to the growing availability of large-scale, multidimensional data from educational environments [2][4]. Techniques such as Educational Data Mining (EDM) and some learning techniques had played a major role for accessing student involvement and identifying students who are at risk[3],[6].

Some of the wellness indicators such as stress levels, physical activity, sleep quality and mental health are identified as critical determinants of academic achievements and student retention [7], [12],[8].

This collaboration of policy vision under NEP-2020 and technological advancements in predictive analytics provides good opportunity to build new integrated frameworks for holistic development of students.

Problem Statement:

As there is large scale data about education and wellness most of approaches can only focus on one factor at a time either on academic or wellness separately which may cause to separated insights. It becomes easier to use machine learning methods with statistical methods together because they can identify the complex relationships and patterns between them. Such type of integration is important because it leads to development of the educational systems according to NEP-2020.

Objectives:

The following objectives are:

- To identify key wellness determinants—including stress, lifestyle, engagement, and cognitive attributes—that significantly influence academic outcomes.
- To develop predictive models capable of forecasting student performance trends and wellness categories, thereby supporting personalized educational interventions.

The main goal is to combine all methodological gaps by creating a data-driven framework.

Significance of the Study:

It plays a major role in contribution to the development of predictive analytics as it provides evidence based data with is focused on academics and wellness. The findings also aligns with:

- Some of the personal individual strategies which mainly focus on well-being of the learners and individual needs of learners for development.

- Making good decisions based on the data and evidence improves student's development and it is also focused on the strategies of NEP-2020.
- It leads to advancement of research by applying statistical techniques with machine learning models.

The identification of gap between education analytics and wellness framework research provides a scalable approach to higher institutions which helps them for success.

Review of Literature:

Approaches of holistic education and NEP_2020:

These approaches determines various methods which can improve learning methods and achieve stability. It highlights the importance of flexible, multidisciplinary educational framework .Similarly these policy directions have led to integration of learning ,and technology based education to enhance student development[13].These policies motivate the research field for using machine learning and predictive analytics.

Wellness and Academic Outcomes: Studies on How Stress, Lifestyle, and Thinking Affect Learning:

The significance of literature have establishes links between the wellness indicators like stress, lifestyle, physical activity and mental health .Although ,[7] has emphasized the student involvement and well-being are some critical predictors of retention and success.[12] and [8] highlighted how the physical activities and mental wellness effects outcomes .By combining these wellness dimensions into educational analytics offers deep insights about the learning performance and patterns but such combinations are limited in educational predictive modelling frameworks.

Machine Learning in Education: Predictive Modelling, Clustering, and Personalization Approaches:

Early work in educational data mining(EDM) has proved the potential of applying data analytics to large educational datasets to understand the learning patterns and performance[3],[4].Some of the predictive modelling techniques like decision trees, regression, neural networks and ensemble methods have been applied to classify students, forecast academic outcomes and identify risk learners[6],[5].Random forest[9] and deep learning architectures(Good Fellow et al.,2016)are such machine learning models which have shown robust predictive capabilities across various domains. In the Indian context, [2] and [8] has highlighted the growing role of ML in institutional planning and curriculum personalization .Recently, Jadhav et al(2024) presented the power of integrating ai with statistical analysis to identify key factors influencing wellness and learning outcomes.

Research Gap

The studies and analysis is done can either focus on academic or wellness indicators. Some higher education have combined machine learning techniques with statistical models to improve the overall performance of the organizations. Some of the attributes like lifestyle, cognitive, and educational are considered as multi-variate relationships which is not explored in deep in large scale datasets.

This research gap is noticed when the machine learning techniques are applied with statistical methods. This combination have played a major role in study.

Research Methodology

Dataset Description:

The dataset used for the study is the NeuroNSGA_MHED dataset which has the data that is collected by physical form from learners. Dataset consists of stress levels, sleeping hours, average heart rate and daily step counts also some of behavior attributes like screen time, sentiment scores. Measures denoting academic performance includes learning outcomes, assignment completion. Lifestyle measures are attendance, activity. The existing dataset is multidimensional dataset which includes all these metrics.

Data Preprocessing:

A Structured preprocessing pipeline was implemented to ensure the quality of data and consistency using python (pandas, scikit-learn).

Handling Missing Values:

Numerical variables were filled using the median strategy.

Categorical variables were filled using the most frequent category.

The data was encoded and categorical data was converted into numerical data.

Normalization and Standardization was done because it makes the numeric features more standardized by z-score scaling method and it helps to compare it with other measurement units, which is important in clustering and distance based algorithms.

Target Variable Construction was done by using median split method so that it can derive binary academic performance.

Statistical Analysis:

In this step both descriptive and statistical analyses were conducted for determining the relationships between wellness indicators and academic performance. Here, descriptive statistics include mean, standard deviation, median and distribution analysis was conducted on numerical variables to understand patterns. Also, correlation analysis was performed like person correlation to understand associations between wellness and academic variables and through that matrices were generated. Multivariate Analysis of variance (MANOVA) was performed by using wellness indicators as dependent variables and academic as independent variable. This ensures that groups that ae defined by academic scores are different from multiple dimensions.

This analysis helps for feature selection and model interpretation.

Machine Learning Models:

Supervised and unsupervised learning approaches are used here to build predictive frameworks.

Supervised Classification:

Random Forest Classification

Logistic Regression model

Multi-Layer Perceptron which is for complex relationships.

The machine learning models was trained in such a way that it predict all the outcomes by undertaking wellness and behavioral predictors. Through the prediction performance was evaluated using accuracy, precision, recall, F1-score.

Unsupervised Clustering:

K-means clustering and agglomerative hierarchical clustering are some of the unsupervised clustering methods which were applied on wellness-related numeric features to define learners into wellness profiles.

Experimental Setup:

The dataset was sperated into 80% training and 20% testing subsets according to the target variable so it maintains balance. Also, stratified k-fold cross validation was done for robust model evaluation and Hyper parameter for Random Forest using gridsearchCV. All these analysis was done in python language by using python libraries such as pandas, numpy, scikit-learn. By split of data it can leads to accuracy of analysis.

Ethical Considerations:

The participants were aware there personalized data will be used educational purpose. Their data was collected before the analysis was started. The data was collected in structured way by respecting the participant by seeking their permission.

Results / Findings

Descriptive Statistical Analysis:

Descriptive analysis was performed to calculate the population of students by categorizing them by their wellness habits, academic engagements and overall performance. It is also performed for examining the distribution and variability of academic variables and based on this mean and standard deviation is calculated

It is reported that average sleeping hours of student is 7.1 hours and their mean stress level is 4.8 and standard deviation is 13.5.Average study time of student is approximately 4.5 hours and its standard deviation is 1.8.

Correlation Analysis:

Correlation matrix of a person was generated to explore more linear combinations between wellness indicators and academic performance. Sleeping hours and content engagement were positively related with academic scores($r=0.38$ and $r=0.42$, respectively<0.01).Stress levels of the students are discovered as weak but significantly negative($r=--0.21$, $p<0.05$).Variables like screen time showed an important association with both engagement and performance.

Because of these correlations it is derived that wellness behaviours are closely related to academic outcomes a previous supports it'.

MANOVA Results:

Multivariate analysis of variance (MANOVA) was performed by using wellness indicators like sleeping hours, stress level, mental health keyword frequency and content engagement as dependent variables

And academic performance variables as independent such as high vs low, median split.

The MANOVA yielded significant multivariate effects (Wilks' $\lambda = 0.0253$, $F(4, 2995) = 28877.64$, $p < 0.001$) specifies that the learners which have higher academic performance have significant difference across wellness indicators. Univariate tests were conducted which results that sleeping hours, stress levels and engagement scores has major contribution for this group differences. It specifies that this wellness dimensions simultaneously influence learning outcomes which supports NEP-2020 holistic approach

Classification Model Performance:

Random forest, logistic regression and neural Network are three supervised learning models which are trained to predict academic performance classes using wellness and behavioural variables:

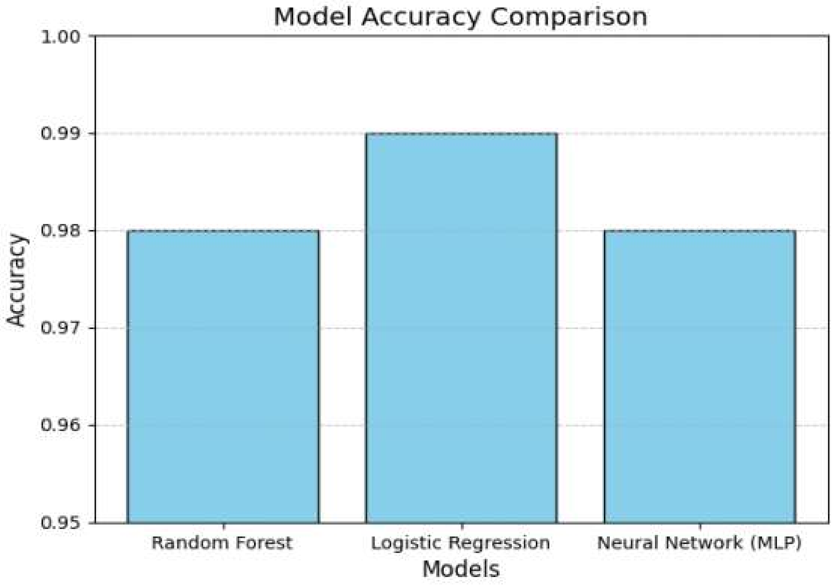

Fig 1.Accuracy comparision of models.

Cross-validation (5-fold Stratified) confirmed stable performance across folds (mean accuracy ≈ 0.98–0.99, SD ≈ 0.005). Logistic Regression performed surprisingly well, indicating that wellness variables linearly separate academic groups effectively. Random Forest and MLP showed comparable performance, suggesting the presence of both linear and non-linear patterns in the data.

Feature importance analysis from the Random Forest model indicated that content engagement, sleep hours, stress level, and resilience score were among the top predictors of academic outcomes. This aligns with international research highlighting the role of engagement and well-being in academic achievement [4],[2].

Clustering and Learner Profiling:

Unsupervised clustering is done using K-means and three learner segments are identified:

Table 2.Proportion of wellness Profile cluster wise

Cluster	Wellness Profile	Proportion
1	High sleep, low stress, high engagement	34%
2	Moderate sleep and engagement, average stress	41%

3	Low sleep, high stress, low engagement	25%

Clustering centroids showed separations in sleep, stress and engagement which reveals that cluster1 has highest proportion of highest performing students that is 82% and cluster 3has lowest that is 29.

Discussion / Analysis:

The combination of statistical methods and machine learning models gives a comprehensive analysis of the relationships between wellness behaviours and academic outcomes. Statistical analysis has determined some significant relations and group differences whereas predictive models showed higher accuracy in classifying learners based on wellness features.

Clustering Analysis highlights students are not homogeneous in their behaviour and their patterns. These insights can give you evidence based information and also engagement focused curriculum.

With easy adaptive learning ways.

These findings are always stable with their previous studies describing the impact of stress, sleep and engagement of academic success. Importantly, the present study extends the prior work by providing predictive framework.

Conclusion

This study presents a framework that combines the statistical methods and machine learning techniques which identifies the relationships between the wellness indicators and academic behaviours .This is done by using Neuro NSGA_dataset. Whole analysis is done revealing significant correlations between wellness indicators. Multivariate analysis is done which defines that groups which have highest performing students have the highest wellness profiles aligning with the holistic educational vision in NEP-2020.

Predictive modelling ensures that variables like wellness and behavioural can classify learners into academic performance categories by using machine learning models like logistic regression, random forest and neural networks which can give higher accuracy. Clustering analysis provides actionable insights.

These findings can highlight the potential of the strategies which combines academic wellness and behavioural dimensions.

Future Directions:

Future research focuses on expanding the datasets and creating useful dashboards that demonstrate day-wise planning and progress. Social, economic , physical are variables which will help in future to improve accuracy.

Reference

Government of India. (2020). National Education Policy 2020. Ministry of Human Resource Development, Government of India. Retrieved from https://www.education.gov.in/

Kumar, V., & Singh, A. (2021). Machine learning in education: Applications and opportunities. International Journal of Advanced Computer Science and Applications, 12(4), 245–252. https://doi.org/10.14569/IJACSA.2021.0120432

Baker, R. S., & Inventado, P. S. (2014). Educational data mining and learning analytics. In J. A. Larusson & B. White (Eds.), Learning Analytics: From Research to Practice (pp. 61–75). Springer. https://doi.org/10.1007/978-1-4614-3305-7_4

Romero, C., & Ventura, S. (2020). Educational data mining and learning analytics: An updated survey. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 10(3), e1355. https://doi.org/10.1002/widm.1355

Paigude, S., Pangarkar, S. C., Majajan, R. A., Jadhav, P. V., Shirkande, S. T., & Shelke, N. Occupational health in the digital age: Implications for remote work environments. South Eastern European Journal of Public Health, 97-110.

Bhattacharya, S., & Nath, A. (2020). Application of predictive analytics in Indian higher education: A review. Journal of Emerging Technologies and Innovative Research, 7(5), 195–202.

Nair, S., & Bhatia, M. P. S. (2022). Predictive modeling in higher education using machine learning techniques. Procedia Computer Science, 192, 2635–2644. https://doi.org/10.1016/j.procs.2021.09.040

Jadhav, P. V., & Patil, V. V. (2022). Application of Decision Tree for Developing Accurate Prediction Models.

Tinto, V. (2017). Through the eyes of students. Journal of College Student Retention: Research, Theory & Practice, 19(3), 254–269. https://doi.org/10.1177/1521025115621917

Jadhav, P. V., Patil, V., & Gore, S. (2020). Classification of categorical outcome variable based on logistic regression and tree algorithm. Int J Recent Technol Eng, 8(5), 4685-90.

Singh, A., & Sharma, R. (2021). Integrating wellness frameworks in higher education: A data-driven approach. Indian Journal of Educational Research and Innovation, 16(1), 33–41.

Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32. https://doi.org/10.1023/A:1010933404324

Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. MIT Press.

Han, J., Pei, J., & Kamber, M. (2011). Data mining: Concepts and techniques (3rd ed.). Morgan Kaufmann.

World Health Organization. (2020). Mental health and psychosocial considerations during the COVID-19 outbreak. WHO. Retrieved from https://www.who.int/

UNESCO. (2021). Reimagining our futures together: A new social contract for education. UNESCO Publishing.

Mishra, S., & Gupta, D. (2023). Statistical approaches for wellness analytics in higher education. Indian Journal of Applied Statistics and Data Science, 9(1), 76–88.

Kuri, M., Jadhav, P., Patil, S., Goure, P., Chandre, P., & Kamble, P. (2025, July). Modelling and Classifying Sleep Disorders with Machine Learning Algorithms. In International Conference on ICT for Sustainable Development (pp. 78-92). Cham: Springer Nature Switzerland.