

UTILIZING NUTRITIONAL AND LIFESTYLE DATA FOR PREDICTING STUDENT ACADEMIC PERFORMANCE: A MACHINE LEARNING APPROACH

Mohammed S. Salih ^{a,*}, Soran A. Pasha ^a

^a Kalar Technical College, Garmian Polytechnic University, Kalar, Kurdistan Region, Iraq.

(mohammed.sarwat, soran.pasha)@gpu.edu.iq

Received: 13 Mar., 2024 / Accepted: 10 July., 2024 / Published: 14 Aug., 2024. <https://doi.org/10.25271/sjuoz.2024.12.3.1288>

ABSTRACT:

Nutrition and lifestyle factors have an enormous impact on students' academic performance. However, there is a shortage of machine learning models to predict students' academic performance based on their nutrition and lifestyle. This paper intends to fill those gaps based on an extensive dataset of various attributes, underlining the capabilities of advanced machine learning models in uncovering the complex relationship between nutrition, lifestyle factors and student's academic performance. A cross-sectional study was conducted in Kalar Technical College, Garmian Polytechnic University in Kurdistan region - Iraq, that involved 500 undergraduate students whose ages range from 18 to 22 years old; the dataset contains demographic characteristics, dietary intake, physical activity, and anthropometric measurements. Various Techniques, tools, and machine learning algorithms such as logistic regression and decision tree classifiers were employed using Python's Scikit-Learn library; finally, Pre-processing of the data was carried out to ensure its suitability for analysis. The machine learning model that the authors developed showed promising prediction results. While the logistic regression model had an accuracy of 70.4%, the decision tree model excelled with an accuracy of 98.55%. Furthermore, exercise, BMI, and dietary intake notably impacted students' academic performance.

KEYWORDS: Machine Learning, student performance prediction, nutritional factors, lifestyle factors, predictive modeling

1. INTRODUCTION

Undergraduate students encounter numerous challenges, such as adapting to new lifestyles due to transitioning to university life away from home and family (Perusse-Lachance et al., 2010). Such drastic changes in their lives can affect their nutrition and overall shape their dietary intake and eating behaviour during their undergraduate tenure (Perusse-Lachance et al., 2010).

The University environment boosts unhealthy eating habits that are characterized by skipping or irregular meals, unhealthy snacks, and junk food, especially for students living in dormitories (Fedewa et al., 2014). Studies have identified harmful eating behaviours that are prevalent among university students, including but not limited to immoderate consumption of soft drinks, saturated fat, late-night snacks, and skipping breakfast (Yahia et al., 2008; Al-Rethaiaa et al., 2010). These patterns of unhealthy lifestyle during their university years may lead to diet-related chronic diseases in later years, like obesity, diabetes, Gastrointestinal, Osteoporosis, and hyperlipidemia problems (Smith et al., 2012; Fateh et al., 2023).

Recently, Machine learning models have made huge progress in various fields, exhibiting their ability in tasks such as image recognition, text interpretation, and healthcare applications (Peng & Gulshan, 2016). For example, deep-learning models have surpassed human ophthalmologists in the identification of diabetic retinopathy by using image analysis (Peng & Gulshan, 2016). This impressive success is largely due to improved computing infrastructure and access to large amounts of training data.

Despite these advances in the field of machine learning, data collection remains a major bottleneck. The considerable amount of time required for data preparation, such as collection, preprocessing, analysis, visualization, and feature engineering,

still presents a clear challenge. Our study aims to use machine learning models to predict students' academic performance based on their nutritional and lifestyle factors to address the shortage of research in this field.

What makes our study novel lies in the way the authors combined various nutrition and lifestyle factors to see how they affect students' academic performance using machine learning models. In contrast to other studies that incorporated just one or two factors, the authors of this study encompassed all those factors together for a complete view. Furthermore, our focus on students in the Kurdistan region of Iraq gives fresh insights into their eating and lifestyle habits, which haven't been studied much before. Our decision tree model was very accurate, getting 98.55%. This shows how powerful machine learning can be in understanding and predicting how well students do in school. It can also be used to recommend healthy lifestyles and diets so that students do better academically.

Literature Review:

The literature review suggests the lifestyle changes that occur during university years have a major effect on eating habits and dietary intake, which may lead to lifelong health problems (Perusse-Lachance et al., 2010; Smith et al., 2012, Fateh et al., 2023).

On the other hand, researchers continue in their effort to find the relationship between academic performance and obesity, fitness, and exercise. Researchers suggest that there is a positive correlation between childhood obesity and academic performance (Li & O'Connell, 2012). Additionally, they also pointed out that boys who are obese got lower marks on school tests in contrast to those who are normal or overweight weight (Torrijos-Niño et al., 2014). Their view is supported by our finding of a positive correlation between BMI and academic achievement. Furthermore, studies indicating that exercise might enhance memory and cognitive function provide support for the

* Corresponding author

This is an open access under a CC BY-NC-SA 4.0 license (<https://creativecommons.org/licenses/by-nc-sa/4.0/>)

positive relationship between exercise and academic achievement. (Mandolesi et al., 2018) Meanwhile, machine-learning models made huge advancements in various domains (Peng & Gulshan, 2016). Thus, using it to predict students' academic performance based on nutritional and lifestyle factors is still an unexplored area. This study aims to bridge the gap through the use of machine learning models to comprehend the intricate associations between nutrition, lifestyle, and academic performance among university students.

2. METHODOLOGY

Study Design and Participants:

The cross-sectional study was conducted from February 2022 to April 2023 at Kalar Technical College, which included 500 university students aged 18-22 years from Garmian Polytechnic University/Kalar Technical College, Kurdistan Region, Iraq. Students were excluded if they had ever used an antibiotic, an anti-acid H2 blocker, a proton pump inhibitor (PPI), a bismuth compound, or an NSAID within the previous four weeks. These qualification requirements led to the recruitment of 500 consecutive individuals for this research.

Data Collection and Academic Performance Metrics:

Comprehensive data encompassing demographics, health status, dietary patterns, and academic records was collected. Academic performance metrics, vital for predictive modelling, were obtained from student's cumulative grade point averages (CGPAs) for the preceding academic year. Categorization of academic performance facilitated precise prediction models across different performance levels.

Demographic Data Collection:

Collected information included age, sex, marital status, place of residence, income status, chronic diseases, food allergies, drug usage, stomach issues, heartburn, and the number of diseases obtained through a demographic questionnaire.

Physical Activity Assessment:

The level of physical activity was evaluated using the validated International Physical Activity Questionnaire (IPAQ) and classified based on Metabolic Equivalents (METs); the recorded quantities were displayed and divided into three groups (very low: <600, low: 600-3000, and moderate and high > 3000 MET-min/week) (Kwak et al., 2011).

Anthropometry Measurements:

The participants wore little clothes and no shoes while having their weight recorded using the InBody 770 device (Inbody Co, Seoul, Korea). The participant was asked to stand without shoes while having their height measured using an automated stadiometer, model BSM 370 (Biospace Co., Seoul, Korea), with an accuracy of 0.1 cm.

Dietary Assessment:

A food frequency questionnaire (FFQ) was used to obtain and measure long-term dietary intake. The seven food categories in the FFQ were modified to fit the Iraqi diet (Al Khalidi et al., 2021). These food categories included (Dairy products, Meat, Poultry, Vegetables, Fruits, Grains, and sweets). By choosing from one of four options (every day, "3-4 times a week", once a week, and monthly), frequency was evaluated.

Research Process:

Chart 1 illustrates the research process for this paper. This is a general explanation, and we go into detail in the later sections

of the methodology. The chat starts with the authors collecting data from the participating students. After that, there was an extensive data preprocessing phase, handling missing values and normalization, which was followed by data engineering and model development using the two mentioned algorithms and utilizing various ML techniques in our model. This was followed by model evaluation and, finally, interpretation.

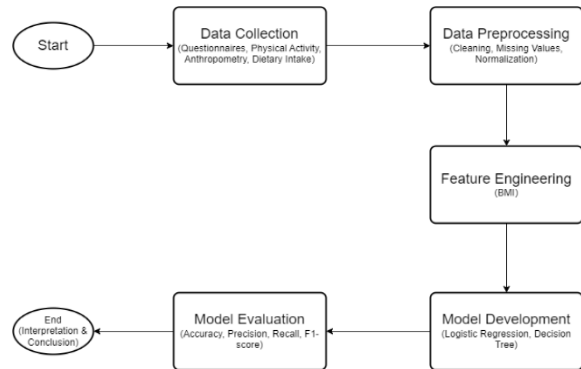


Chart 1: Research Process

Data Preprocessing and Feature Engineering:

Preprocessing involved extensive data cleaning, handling missing values, and addressing inconsistencies. Techniques such as one-hot encoding for categorical variables, derivation of Body Mass Index (BMI) from anthropometric data, and normalization using StandardScaler were employed to ensure model compatibility.

Model Development, Evaluation, and Machine Learning Tools:

The study leveraged Python's scikit-learn library to implement Logistic Regression and Decision Tree classifiers, demonstrating their effectiveness in predictive analysis. To counter class imbalances, Gaussian noise augmentation was incorporated, enhancing model robustness, particularly in predicting the performance levels of minority classes. Models were evaluated using accuracy, precision, recall, and F1-score metrics.

Logistic Regression and Decision Tree Classifiers:

Using a given data set of independent variables, logistic regression calculates the probability that an event will occur. This type of statistical model, often referred to as the logit model, is widely used for prediction and classification analytics. With probability as the outcome, the dependent variable's range is 0 to 1. The odds in logistic regression are calculated by dividing the probability of failure by the likelihood of success, and this is done using a logit transformation. This is also called the log odds or the natural logarithm of odds (Schober & Vetter, 2021).

Meanwhile, decision tree is a non-parametric supervised learning approach that is used for both regression and classification applications (Charbuty & Abdulazeez, 2021). With a root node, branches, internal nodes, and leaf nodes, it has a hierarchical tree structure. A decision tree begins with a root node that has no incoming branches, as shown in Chart 2. The internal nodes, sometimes referred to as decision nodes, receive input from the outward branches of the root node. Both types of nodes perform assessments based on available attributes to create homogeneous subsets, which are represented by leaf nodes or

terminal nodes. Every conceivable result in the dataset is represented by the leaf nodes.

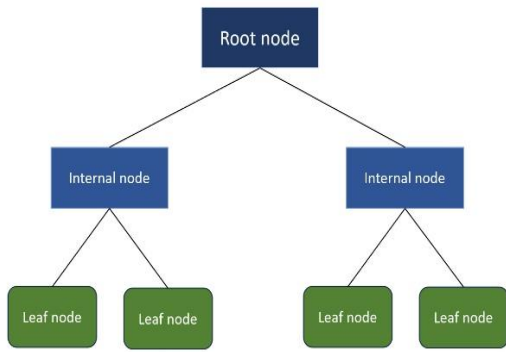


Chart 2: Decision Tree Classifier

Model Interpretation, Deployment:

Interpretive and post-modelling techniques such as feature importance ratings and partial dependence plots were used to understand the relationships between dietary and lifestyle factors and predicted outcomes. In addition, deployment strategies, including continuous improvement and monitoring, were used for practical implementation.

Result and Discussion:

This research paper aims to achieve two goals. First, developing a machine learning model to predict students' academic performance based on nutritional and lifestyle factors. And second, finding the association between eating habits, exercise, BMI, and academic performance.

To achieve those goals, the authors collected data from 500 Kalar Technical College students with a wide range of academic grades and diverse lifestyles and nutritional habits. Those data were used in the development of the machine learning models. The authors used two distinct algorithms to find the most suitable for our data and find the relationship between academic performance and the factors mentioned above.

Figure 1 illustrates the correlation between students' academic performance and their dietary habits. The analysis revealed a positive correlation between dairy intake and academic performance, suggesting that students with a healthy consumption of dairy tended to achieve higher academic scores. Conversely, a negative correlation was observed between fruit intake and student performance. Similar patterns were identified for meat and vegetables, with the former displaying a positive correlation and the latter demonstrating a negative association with academic performance.

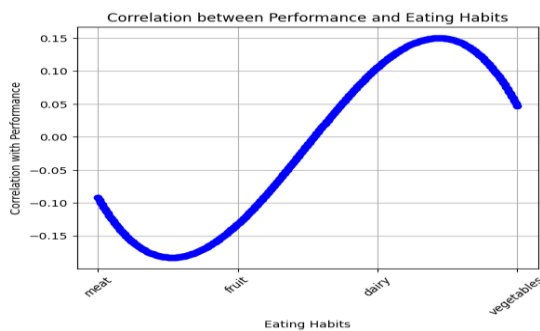


Figure 1: correlation between performance and eating habits

Figure 2 displays a negative correlation between students' academic scores and their Body Mass Index (BMI). The correlation coefficient of -0.063 indicates a weak negative relationship between performance and BMI. This implies that as

BMI increases, there is a slight tendency for academic performance to decrease and vice versa.

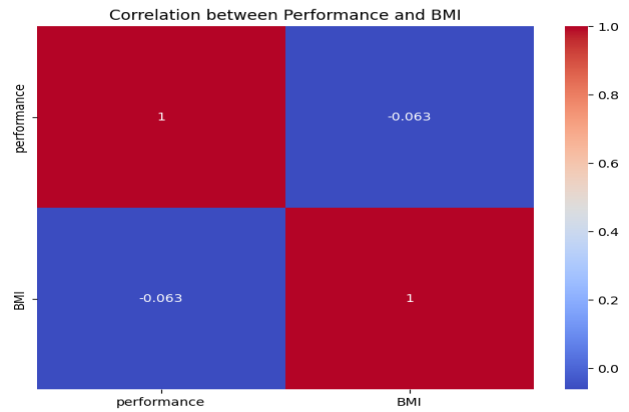


Figure 2: Correlation between performance and BMI

Moreover, a positive correlation was identified between exercise frequency and higher academic scores, as depicted in Figure 3. Students engaging in regular exercise tended to achieve higher academic scores, while those who did not exercise regularly displayed comparatively lower scores.

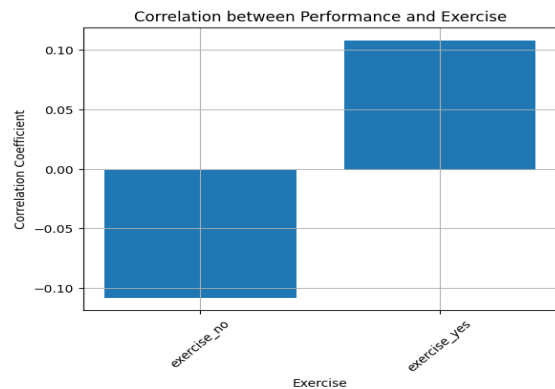


Figure 3: correlation between performance and

Furthermore, while the Logistic Regression model exhibited an accuracy of 70.4% in predicting student performance across diverse categories, the Decision Tree model notably outperformed, boasting an accuracy of 98.55%. Moreover, detailed performance metrics such as precision, recall, and F1 scores are presented in Table 1.

Table 1: Performance Metrics (Precision, Recall, and F1-scores)

Metrics (weighted avg)	Logistic Regression	Decision Tree
Accuracy	70.4%	98.55%
Precision	70%	98%
Recall	70%	98%
F1-score	70%	98%

These metrics highlight the substantial difference in performance between the Logistic Regression and Decision Tree models, with the Decision Tree model demonstrating significantly higher accuracy and overall predictive capability compared to the Logistic Regression model.

Our study breaks new ground by pioneering the use of a machine learning model to predict student academic performance based on nutritional and lifestyle factors in the Kalar District, Kurdistan Region, Iraq. Our model with the decision tree

classifier achieved an excellent accuracy of 98.55%; the authors chose it due to its simplicity and its high accuracy in similar tasks previously; its impressive accuracy also outperforms other similar models, showing its potential as a reliable tool for predictive analysis. Furthermore, our analysis shows a positive correlation between dairy/vegetable consumption, lower BIM, regular exercise, and academic performance.

The simplicity and ease of use of the decision tree classifier make it popular for prediction tasks. Our paper shows its utilization in six other studies previously where the highest accuracy achieved was by (Altujjar et al., 2016), and the lowest was by (Islam et al., 2019) as shown in Table 2.

Table 2: Comparing Authors Results with Other

Decision Tree Metrics	Authors	Altujjar et al.	Islam et al.
Accuracy	98.55%	98.2	64%

It is paramount to acknowledge the multifaceted effect of diet on humans, which affects physical and mental health, physical effort, and cognitive function. Poor sleep patterns, often stemming from a lack of parental guidance, can lead students towards readily available, less healthy food choices, making good nutrition particularly important for academic success (Deliens et al., 2014). Nutritional deficiencies can also affect students' thinking, concentration, behaviour, and overall wellbeing. (Belot & James, 2011) Research supports this, demonstrating that students participating in a junk food ban initiative achieved higher scores than those who didn't.

Our study highlights a clear link between student performance and eating habits. The authors observed a positive association between dairy consumption and academic performance, suggesting that students with healthy dairy intake generally fared better than those who consumed more fruit, which yielded the opposite effect. Similar patterns emerged for meat and vegetables, with the former showing a positive association and the latter a negative one. Elsayead and Said (Abd & Said, 2020) found a similar link between dietary status and academic success in Saudi Arabia, where university students face less pressure to consume specific foods due to financial support. This aligns with the broader observation that physical health is generally linked to academic success.

Interestingly, our study found that students with lower BMIs performed better academically than those with higher BMIs (Anderson & Good, 2017; de Almeida Santana et al., 2017). While some research associates obesity with negative academic impacts (Anderson & Good, 2017), others suggest that overweight children may perform better due to attempts to compensate for negative self-perception (de Almeida Santana et al., 2017). Furthermore, the authors observed a positive correlation between exercise and academic scores, with exercising students achieving higher results than those who were less active. While these findings warrant further investigation, particularly given the inconclusive nature of the influence of exercise on adult cognition (as highlighted by various studies), they offer valuable insights for future research.

A key strength of our study lies in its pioneering nature, being the first of its kind to develop a machine learning model for predicting student academic performance based on nutritional and lifestyle factors within our region.

Limitations and Future Directions:

Acknowledging the limitations inherent in the study, including dataset size and diversity, the authors propose future research avenues that could explore ensemble models or more advanced neural networks. These avenues may further enhance the accuracy and applicability of predictive models in forecasting student performance.

CONCLUSION

Our study underscores the successful utilization of advanced machine learning techniques, notably the Decision Tree classifier, in accurately predicting student performance based on lifestyle and nutritional parameters. These models exhibit robustness and high predictive accuracy, laying the foundation for their implementation within educational frameworks.

Declarations

Ethics approval and consent to participate

The Ethics Committee of Garmian Polytechnic University, Kalar Technical College, approved the study. All methods were carried out in accordance with relevant guidelines and

Table 2: Comparing Authors Results with Other Researchers

regulations. All the participants were provided oral and written informed consent. All methods were carried out according to relevant guidelines and regulations. This study was conducted by the Declaration of Helsinki.

Consent for publication

Not applicable.

Availability of data and materials

The data analyzed in the study are available from the corresponding author upon reasonable request.

Competing interests

The authors declare no conflicts of interest.

Funding Sources

This research was supported by Garmian Polytechnic University.

Authors' contribution

Mohammed Sarwat designed the study. Mohammad Sarwat Developed the model and analyzed the data. Mohammad Sarwat and Soran Pasha prepared the draft of the manuscript.

Acknowledgements

The authors extend our gratitude to the research team for their support and collaboration. Our families deserve heartfelt thanks for their unwavering support and understanding. The authors also acknowledge the GPU deputy for technical assistance.

Finally, our appreciation goes to all study participants for their invaluable contributions.

Thank you all for your support and involvement.

REFERENCE

- Abd Elaleim Elsayead, M., & Said, A. S. (2020). A Study of The Relationship Between Nutritional Status and Scholastic Achievement Among Primary School Students in Wadi Eldawasir City in Kingdom of Saudi Arabia. *Journal of medical & pharmaceutical Sciences*, 4(1).
- Al Khalidi, N. M., Kadhim, Z. G., & Almousawi, H. Y. (2021). Dietary patterns in adult patients with Non-Alcoholic Fatty Liver Disease in Iraq. *Medical Science*, 25(115), 2292–2301.
- Al-Rethaiaa, A. S., Fahmy, A. E. A., & Al-Shwaiyat, N. M. (2010). Obesity and eating habits among college students in Saudi Arabia: a cross-sectional study. *Nutrition journal*, 9, 1-10.
- Altujjar, Y., Altamimi, W., Al-Turaiki, I., & Al-Razgan, M. (2016). Predicting critical courses affecting students performance: a case study. *Procedia Computer Science*, 82, 65-71.

- Anderson, A. S., & Good, D. J. (2017). Increased body weight affects academic performance in university students. *Preventive medicine reports*, 5, 220-223.
- Belot, M., & James, J. (2011). Healthy school meals and educational outcomes. *Journal of Health Economics*, 30(3), 489-504.
- Charbuty, B., & Abdulazeez, A. (2021). Classification based on decision tree algorithm for machine learning. *Journal of Applied Science and Technology Trends*, 2(01), 20–28.
- Deliens, T., Clarys, P., De Bourdeaudhuij, I., & Deforche, B. (2014). Determinants of eating behaviour in university students: a qualitative study using focus group discussions. *BMC Public Health*, 14, 1-12.
- de Almeida Santana, C. C., Farah, B. Q., De Azevedo, L. B., Hill, J. O., Gunnarsdottir, T., Botero, J. P., ... & Do Prado, W. L. (2017). Associations between cardiorespiratory fitness and overweight with academic performance in 12-year-old Brazilian children. *Pediatric Exercise Science*, 29(2), 220-227.
- Fateh, H. L., Kamari, N., M. Ali, A., Moludi, J., & Rezayaeian, S. (2023). Association between diet quality and BMI with side effects of Pfizer-BioNTech COVID-19 vaccine and SARS-CoV-2 immunoglobulin G titers. *Nutrition & Food Science*, 53(4), 738-751.
- Fedewa, M. V., Das, B. M., Evans, E. M., & Dishman, R. K. (2014). Change in weight and adiposity in college students: a systematic review and meta-analysis. *American Journal of Preventive Medicine*, 47(5), 641–652.
- Islam, R., Sazid, M. T., Mahmud, S. R., Ferdous, C. N., Reza, R., & Hossain, S. A. (2019, May). Parametric study of student learning in IT using data mining to improve academic performance. In 2019 Joint 8th International Conference on Informatics, Electronics & Vision (ICIEV) and 2019 3rd International Conference on Imaging, Vision & Pattern Recognition (icIVPR) (pp. 286-290). IEEE.
- Kwak, L., Proper, K. I., Hagströmer, M., & Sjöström, M. (2011). The repeatability and validity of questionnaires assessing occupational physical activity—a systematic review. *Scandinavian Journal of work, environment & health*, 6-29.
- Yahia, N., Achkar, A., Abdallah, A., & Rizk, S. (2008). Eating habits and obesity among Lebanese university students. *Nutrition journal*, 7, 1-6.
- Li, J., & O'Connell, A. A. (2012). Obesity, high-calorie food intake, and academic achievement trends among US school children. *The Journal of Educational Research*, 105(6), 391–403.
- Mandolesi, L., Polverino, A., Montuori, S., Foti, F., Ferraioli, G., Sorrentino, P., & Sorrentino, G. (2018). Effects of physical exercise on cognitive functioning and wellbeing: biological and psychological benefits. *Frontiers in Psychology*, 9, 347071.
- Peng, L., & Gulshan, V. (2016). Deep learning for detection of diabetic eye disease. *Google Research Blog*.
- Perusse-Lachance, E., Tremblay, A., & Drapeau, V. (2010). Lifestyle factors and other health measures in a Canadian university community. *Applied Physiology, Nutrition, and Metabolism*, 35(4), 498-506.
- Schober, P., & Vetter, T. R. (2021). Logistic regression in medical research. *Anesthesia & Analgesia*, 132(2), 365-366.
- Smith, M. L., Dickerson, J. B., Sosa, E. T., J McKyer, E. L., & Ory, M. G. (2012). College students' perceived disease risk versus actual prevalence rates. *American Journal of Health Behavior*, 36(1), 96–106.
- Torrijos-Niño, C., Martínez-Vizcaíno, V., Pardo-Guijarro, M. J., García-Prieto, J. C., Arias-Palencia, N. M., & Sánchez-López, M. (2014). Physical fitness, obesity, and academic achievement in schoolchildren. *The Journal of Pediatrics*, 165(1), 104-109.