“A Comparative Analysis Of Daily Habits On Academic Performance”

1st  Tejas Baravkar

Department of Information Technology , Vishwakarma Institute of Information Tehnology ,

Pune , *Maharashtra , India*

[tejas.22210614@viit.ac.in](mailto:tejas.22210614@viit.ac.in)

4th Anurag Deshpande

Department of Information Technology , Vishwakarma Institute of Information Tehnology , Pune, Maharashtra ,India

[anurag.22211603@viit.ac.in](mailto:anurag.22211603@viit.ac.in)

2nd Rupesh Ambavne

Department of Information Technology , Vishwakarma Institute of Information Tehnology , Pune ,*Maharashtra , India*

[rupesh.22210361@viit.ac.in](mailto:rupesh.22210361@viit.ac.in)

5th Dr. Chaitali Shewale

Department of Information Technology , Vishwakarma Institute of Information Tehnology , Pune ,Maharashtra , India

[chaitali.shewale@viit.ac.in](mailto:chaitali.shewale@viit.ac.in)

3rd Soham Bhagurkar

Department of Information Technology , Vishwakarma Institute of Information Tehnology, Pune ,Maharashtra , India

[soham.22210758@viit.ac.in](mailto:soham.22210758@viit.ac.in)

*Abstract*— The goal of this project is to develop an AI-assisted classroom prediction model that uses students' everyday routines to forecast their overall scoring category. The study examines information from college students, including their academic grades, social media usage, sleep duration, number of days they consume fast food, and study hours. In order to achieve this, two machine learning models—Random Forest and XG Boost—have been developed, and their performances have been compared across the student mark collection category. Determining the relationship between students behaviours and academic achievement is the primary goal of this study. Additionally, an online tool was created where students may input their results and receive recommendations for better practices. This study presents a comparison of methodology models, practical applications, and the usage of artificial intelligence to improve student performance and well-being.

Additionally, an online tool was created where students may input their results and receive recommendations for better practices. This study presents a comparison of methodology models, practical applications, and the usage of artificial intelligence to improve student performance and well-being.

Keywords— Predictive Model, Analysis, Habits, XG Boost , Marks

1. Introduction

Theoretical Background

The contemporary learning systems deeply emphasize on the level of academic performance that students exhibit. Many factors come into play when it comes to this kind of accomplishment — some natural, others associated with the environment such as a daily routine. A lifestyle impacts study outcomes on a significant scale, as with sleeping routine; consumption of fast foods and network interactions in the local flu clinics at private hours. The big data and machine learning surge allows us to dive deeply into these complex factors, unearthing trends that classical statistics left dark.

The primary goal of this study is to classify and analyse student behaviours correlated with academic success across various marks category students over a long period. We used a dataset showing marks of college students with respect to their hours of sleep, fast food intake, social media involvement and the number of study hours. The two methods include the well-known algorithms, Random Forest and Decision Tree to handle complicated data for accurate predictions from machine learning.

Work well — In addition to providing a model, we even made an online system where students could enter their marks and get recommendations on how to shape their daily schedule for better performance.

The purpose of this application is to aid students with the data about their lifestyle in such a way that they are able to engage in the kinds of behaviors which are likely to optimize academic performance.

In this paper, evaluating the usefulness of machine learning methods for predicting academic performance based on lifestyle variables and presented a relevant system designed to assist students.[1]

1. Objective

This study focus mainly on three areas: Developing and validating the machine learning techniques that would be used to predict the academic performance of student separated by dream, fast food eating habit, sleeping duration and social network time consumption and their own studying level. In this regard, the study centers around:

1. Investigating how students’ habits affect their educational success

2. Developing two machine learning models of Random Forest, XG Boost and checking their accuracy to predict the marks categories affected by these facts.

3. Designing an interface, where interested individuals could type in their scores to be directed towards personalized advice regarding habits that can have a negative impact on academic performance.

4. Presenting useful evidence based arguments to assist students in selection of habits that would most benefit their academic performance.

Therefore Industrial bases goals to show how with the right data in place, the relevant solutions can result in meaningful improvement to the academic outcomes of learners.[2]

1. **Literature Review**

*Academic Performance and Sleep*

It is well known that getting enough sleep is essential for maximizing learning for everyone. Curcio, Ferrara, and De Gennaro (2006) compiled a meta-analysis of several research examining the connection between sleep and performance both within and outside of the lab. Their findings demonstrate that sleep does, in fact, have a good impact on academic performance and that kids who get more and better sleep are better able to retain information, which is essential for succeeding in academic competitions. [3]According to a study by Hershner & Chervin (2014), getting too little sleep impairs focus and productivity, which rarely results in tardiness in class and poor grades.

*Self-Study and Time Management*

Besides, self-study besides effective time management aimed at study also affects academic performance. In Zimmerman's self-regulated learning study (2002), it was demonstrated that students who exert effort in self-study, organize themselves, and take care of time perform better than the others. Similarly, Nonis and Hudson (2010), in their longitudinal study noted that the number of hours spent in solo study was a good predictor of GPA, and self-disciplined students, more often than not, had the highest grades.[4]

Of course, even recent studies, as the one done by Komarraju et al. (2013) also revealed, that students who are intended to self-study and not going to be exposed to other forms of media such as social media nor involve in multitasking for pure hours in order to perform well academically. The finding also reveals that it's not just the quantity of studying but the quality of the time devoted.

*Physical Activity and Cognitive Function*

Academic research confirms the relationship between physical activity and cognition, by which means academic performance. For instance, Hillman, Eckerson, and Kramer prepared a full review in 2008 to discuss the role of regular physical exercises in memory, executive functioning, and cognitive functioning in its entirety. According to them, the ability to concentrate and solve problems, very important for education, is enhanced in students who perform moderate physical activity.

To prove this, Trudeau and Shephard (2008) undertook the influence of physical exercise on school achievement by children in a retrospective study; in the researchers indicated that the children who participate in regular physical exercises usually perform well in school. In the same context, Donnelly et al. (2016) has also reported better class management and academic scores for those students who adopted the physical activity schedule on a daily basis.[5]

*Social Media and Academic Distraction*

While social media becomes a tool for students, splitting their study time between studies by using the social media tool, studies about the impact of social media on studies have remained mixed. For example, Junco in 2012 found that poor academic performance strongly correlated with students' failure to regulate their use of websites such as Facebook and Twitter, which it found to be very distracting during "study" time. Although while the study subjects depended on these social media in order to share ideas, such that resourcing used for education was shared, the researcher concluded that the social media played the role of an enhancing tool to the education system.

However, this study was against a research done by Kirschner and Karpinski (2010) where questionnaires shown that students who spent most of their time on online social networks spend less time on academic activities and even have poor GPA. In the same line, other studies which were conducted later including in Paul, Baker, and Cochran (2012) do reveal that limiting the use of social networking during lecturing/studying sessions benefits the students' performance.[6]

*Fast Food Consumption and Cognitive Health*

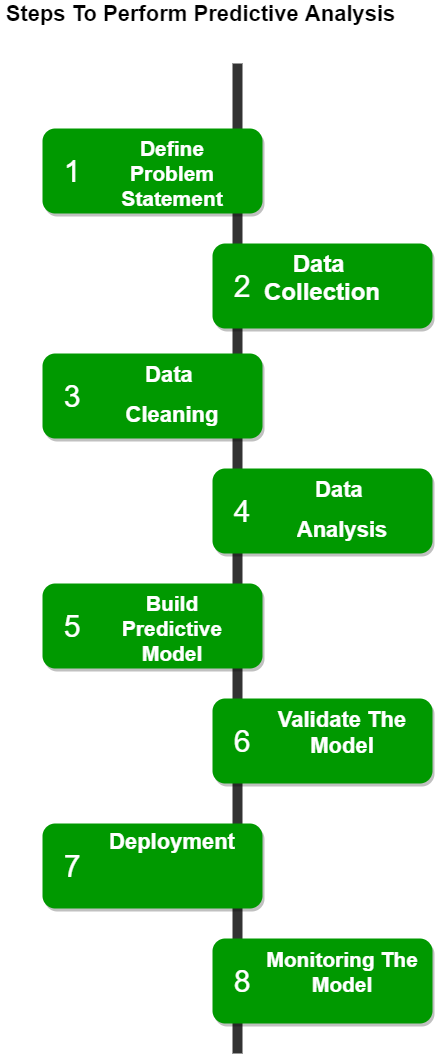
Florence, Asbridge, and Veugelers (2008) state that children or adolescents who consume fast food in high amounts tend to perform poorly in her that is schooling. It is being seen again that the use of a diet that is rich in sugar and fats impairs a person's ability to remember and focus.[7]

Moreover, the study of Burrows, Golley, and Khambalia (2017) indicated that students who consumed less fast food and healthier foods within the healthy diets had better grades compared to the students without healthy diets, thus it is possible to conclude that ideal nutrition plays its role in achievement. That implies that good nutrition will be important if measures are to be advanced to help arrest decline of cognitive signs and improve academic performance.[8]

This literature review helps to show quite clearly certain practices-sleeping, studying and exercising, using social media, and diet-affect academic performance. Although each behavior has its measurable degree of contribution, optimal performance does not appear to happen without healthy sleeping practices along with mothering self-studies, some sporting activities, and eating right. To a certain extent, social networking sites and fast foods negatively relate to one's performance. This comparative analysis therefore portrays the need to apportion appropriate daily schedules to the learners for enhancement of academic success.[9]

**[IV] Methodology**

This research paper established a model with the basic objective of explaining the linkage between five specific behaviors and academic performance.

****

*Fig.(1)*

As has been reported elsewhere, detailed attention is always given to survey questionnaires in this analysis while commenting on the association in the articles. Analysis of the questionnaire revealed five specific lifestyle habits. The Five questionnaire statements are: Lifestyle Habits H-Sleep per Night, H- Self Study, Days Exercising 7/V, Days Fast Food/V, Daily Social Media In total, 100 of these respondents were Bachelor Degree students at British universities.

The questionnaire was carried out with both close ended questions and tick boxes with the objective of either obtaining statistics or words. The questions adopted in the current research were developed for the evaluation of the incidences of some e.g. these behaviors and their part in enhancing the performance of the learners. Thus, some of the interrogative questions that were raised against me were rather explanatory in nature regarding an individual's behavior while others were pretty vague and involved an entire system of such action.

Infill of the questionnaire was mainly on Google Forms, and this was shared with the students and also sent to the college WhatsApp group. This methodology was suitable and relevant in eliciting responses. Upon completion of the data collection process, the information gathered was input into Google Forms for direct tabular output and then exported to Excel for further analysis and processing of the data.

To minimize the bias of free-form response scale, the questionnaire was prepared in a structured manner. Aligned with the same, draft versions of the questionnaire were drawn up based on the main research questions proposed above and were forwarded to the respective focus group of students. The draft of the questionnaire was tested by some experts who provided feedback on logic of understanding.[10]

**[V] Data Analysis And visualization**

Any research work, which bases itself on data, has data visuals as one of its focal features due to the need to represent both simple and complex aggregated data in a more visual, interpretable form. Inclined towards that, the current study utilizes visualizations to examine and depict the correlation between students’ performance (i.e., marks) and their daily routines including but not limited to sleeping, eating junk food, surfing social media, and studying. Looking at the graphical forms of data, one can spot some correlations and trends that are not so easily seen when looking at numbers from a spreadsheet. This helps with understanding the variables that help in achieving academic success.

The visualizations in this study are conceived as: Presenting Trends and Relationships - For example, when we display students' marks in relation to the hours spent sleeping, the frequency of fast food consumption, the time spent on social media, and the number of hours spent studying, we can evaluate the relationship of these variables. For example, one might use a scatter plot or line graph to let graduate students know if sleeping more or studying more hours helps one achieve more marks or if eating more junk food reduces one academic performance.[11]

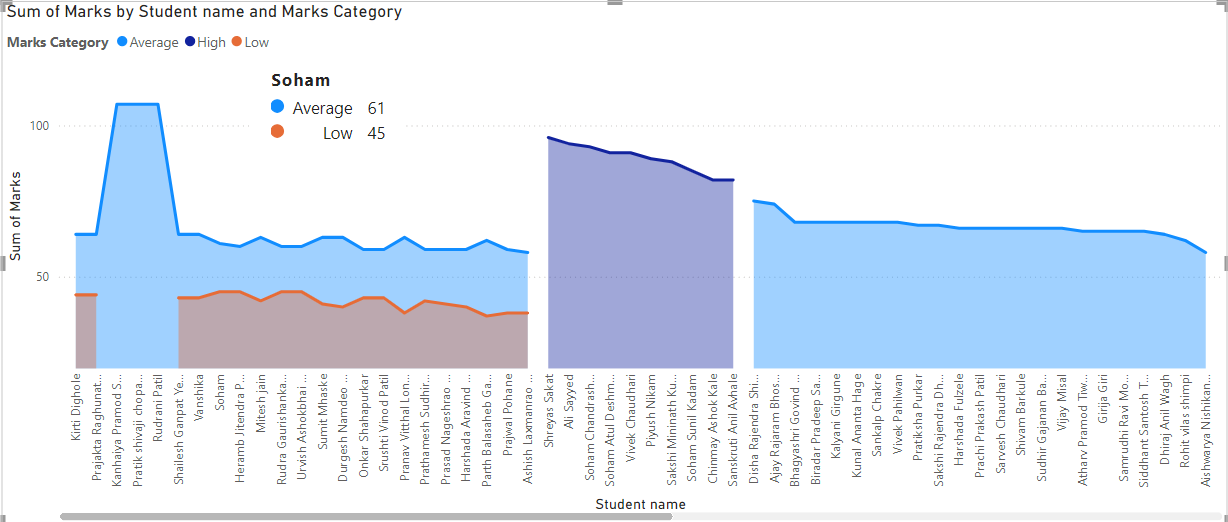
Enhancing Differentiation: Using bar charts and box plots make it easier to show the differences in scores between the different classes of habits e.g. scores between low fast food eaters and high fast food eaters. This emphasizes the presence of differences in academic success, based on different habits.

Enhancing the visualization of Extreme Values: These images depend on the data, for example, box plots and scatter plots, which help see outliers in the data, that is, individuals whose behaviors are far from what the typical student does and investigate if they affect performance.

Here's Another graphical representation of our dataset in Power BI

Students Categorization in to the categories of High , Average and Low.

Figure 2 is a stacked area plot, bearing the title, "Sum of Marks According to Student Name and Marks Category." The x-axis consists of the students' names, with the y-axis being the overall mark scored.



##### *Fig.(2)*

The chart has three classifications of marks;

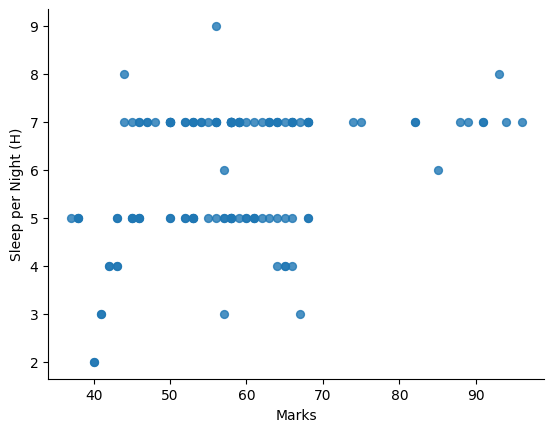
• Low (shown in orange): This classification has thinner slices at the bottom, which represents students whose marks were classified as ‘Low.’

• Average (depicted in light blue): This section is most of the time, the largest cover for the pupils forming the range of this chart.

• High (shown in dark blue): Certain sections mainly in the middle and right parts represent upper-level scores.

In general, the chart enables one to distinguish the students from one another according to their score, whether low, average, or high, illustrating their total scores. Graphs depicting relations between students marks and their habits :

1.Marks vs Sleep Per Night (H)



*Fig. (3)*

Figure (3) is a scatter graph showing the association between Marks (the x-axis) and Hours of Sleep (per night in hours) (the y-axis) in detail. Here's the breakdown:

- On the x-axis, the scope of the variable marked by Marks is nearly between 40 and 100.

- The y-axis is for the concept of hours of sleep per night, which musters between 2 and 9 hours as the extremes.

Every single point on the graph indicates one data entry that is the marks obtained by a student and the number of hours he/she sleeps a night. Along the following:

- Quite a number of data points are seen among the category of students who slept for between 5 and 8 hours.

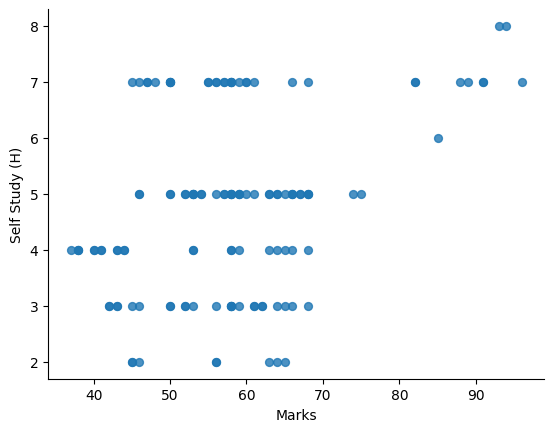
- A diverse set of marks appears to exist for each sleeping pattern which included students scoring from 40 to 90 marks irrespective of the hours spent in sleep.

- Its also noted that students sleeping for below 5 hours tend to secure quite lower scores between the range of 40 to 60 bands.

- It can be observed that students who attained higher marks (around 80 to 90) covered a variety of sleeping patterns including 6 to 8 hours sleep.

In the aggregate, the data plotted suggest scores depend on sleep to some extent, but removed is any semblance of a linear relationship between one’s sleep duration and the resultant score achieved.

2. Marks vs Self Study Hours (H)



*Fig. (4)*

This visual representation is a scatter diagram on Marks (which is on x-axis) Vs Self Study (H) which is on the y-axis. This is further explained as follows:

- The marks are represented along x axis which is approximately around 40 to 100.

- The self-study hours are shown along y-axis which range from 2 to 8 hours in a day.

Fitting a line is simple, as every point on this plot is a student, basing on the number of self-study hours to the marks attained.

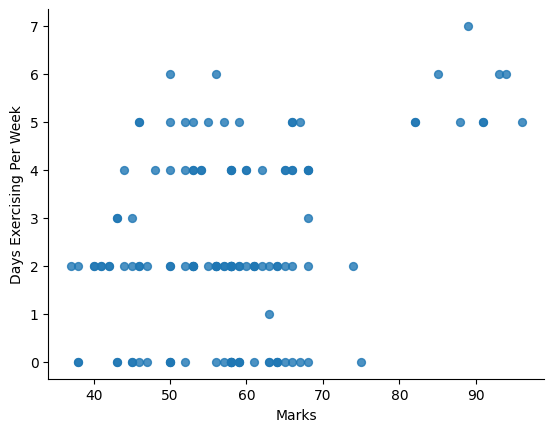
- There are also quite a number of students who self-study for 3 to 6 Hours and components of their marks are less than 50-70.

- The other category are the students who can manage to self-study for 7-8 hours where a majority of the points lie between the range of 70-90 marks.

- Students in the bottom range of above 50 40 50 marks are able to self-study for anything from two hours to nearly six hours as their self-study looks wider.

For the most part the graph depicts that there might be some related increase in regards to higher self-study time and improvement in marks, though they give the most marks where will study for 7 i.e. 8 hours a day on average.

3. Marks vs Days Exercising per week



*Fig. (5)*

This scatterplot demonstrates the association between two variables determining the X-axis ‘Marks’ and Y-axis ‘Days exercising per week’.

- X-axis (Marks):The distribution of marks on the y-axis seems to fall between 35 and 95.

- Y-axis (Days Exercising Per Week): The range on the y axis for number of days of students per week, exercises between 0 to 7 days inclusively.

Key observations :

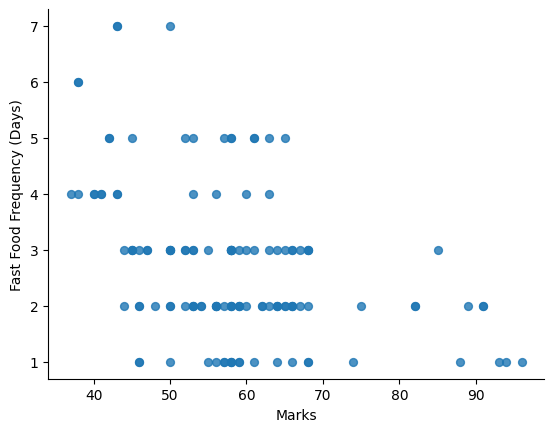
- Many students scored between 40 to 60 marks and therefore did not engage in exercise for any more than four days.

- Some students who score higher marks (above 80) tend to practice exercise frequently, even five to seven days a week.

- There is a low-performing cluster of students who do little exercises or none at all and there is a separate d high exercise days and high scores cluster.

The pattern presents a notion that there exists a relationship between the marks achieved and the number of exercise days in a week, if at all marks above a certain level are considered.

4. Marks vs Fast Food Frequency (Weekly)



*Fig. (6)*

The provided image is a scatter diagram with respect to the "Marks" and "Fast Food Frequency (Days)" on the X and Y axis respectively.

- X-Axis (Marks): The range of marks is between approximately 35 to 95.

- Y-Axis (Fast Food Frequency): The number of fast food servings against the number of days per week for fast food intake from 1 day to all days of the week.

*Key observations:*

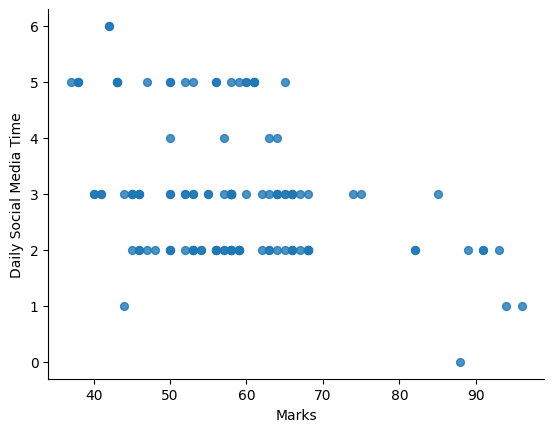
- Students who scored over 40 but less than 60 years have a high prevalence of fast food consumption and could eat it from 1-7 times a day.

- For students scoring higher than 80, fast food consumption is relatively low and indicates an average of between 1 and 3 days per week of having fast food on average.

- There is a majority of students who have lower marks and so tend to eat more fast food often, while those with higher marks do not have that tendency as they eat fast food less often.

The present plot indicates that there is an inversely proportional relationship between the two variables fast food intake and marks scored where it can be deduced that the more a student eats fast food the lower the marks they score and vice versa.

5. Marks vs Daily Social Media Time (H)



*Fig. (7)*

The diagram is a scatterplot demonstrating the relationship between Social media time in hours on a daily basis and Marks obtained by Students in Esade. The x-axis corresponds to Marks, while the y-axis represents Social medial hours spent per day. Each point on the graph signifies an individual value in the data, the positioning of which is determined by the Marks and Social media time.

Following the plot, here are a few remarks:

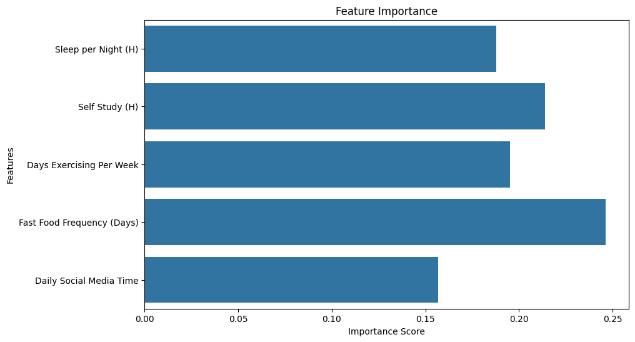
A small yet existent negative relationship could be detected; that is, students who engages more in social as per the self-reported hours usually do worse in terms of the scores. This aspect is quite apparent that the relationship is not very close since there is a lot of crowding on the graph.

Most of the values plot around the y axis on values 2 and 3 which means that most of the students spend an average of 2 to 3 hours on social media daily.

There are a few anomalies in the scatter such as the mark extremely low and the daily social media activity extremely high. These outliers are extremely high or extremely low in the context and may be interesting as well in terms of the given problem and its overall aspect ordinate values in other aspects.

These visualizations collectively provide an intuitive understanding of how various daily habits influence academic performance, helping to form a clear narrative about the lifestyle factors that contribute to or detract from students' success.

Following we have a **Feature Importance** map



*Fig. (8)*

The image communicates the Feature Importance plot that measures the proportion of the roles of five factors with regard to the lifestyle habits in estimating academic performance through an XG Boost model. The features include:

1. Fast Food Frequency (Days): This has the highest importance coefficient, meaning that the frequency by which students take fast foods has the strongest impact towards students’ performance given the studied features.

2. Self Study (H): The second index is the amount of time students spend on self-study, indicating that studying time considerably affects the result.

3. Sleep per Night (H): The hours of sleep per night are also significantly associated on the same predictors underlining the need for the adequate number of hours of sleep to ensure a good performance on the academic tasks.

4. Days Exercising Per Week: However, the number of days students exercise per week but a bit less influential than the three mentioned above contributed to students’ academic performance may be due to the benefits of exercise in students, including physically and mentally.

5. Daily Social Media Time: This feature has the lowest importance which mean the number of hours that students spend in social media significantly has lesser contribution on students’ performance.

This plot shows that life style factors; which include fast foods, self study, and sleep are the chief variables that concerns academic performance as posited by the above model.

**[VI] Results And Discussions**

As part of our research work we advanced and fitted two prediction systems aimed at assessing the influence of students’ lifestyle on their performance in school. Pursuant to this aim, data from 100 university students were collected and five habits were studied in detail: Hours with Closed Eyes a Day (H), Time spent on Self Study, Days When Exercises are Practiced a Week, Fast Food Habit (Days) and Minutes or Hours Using Social Media a Day.

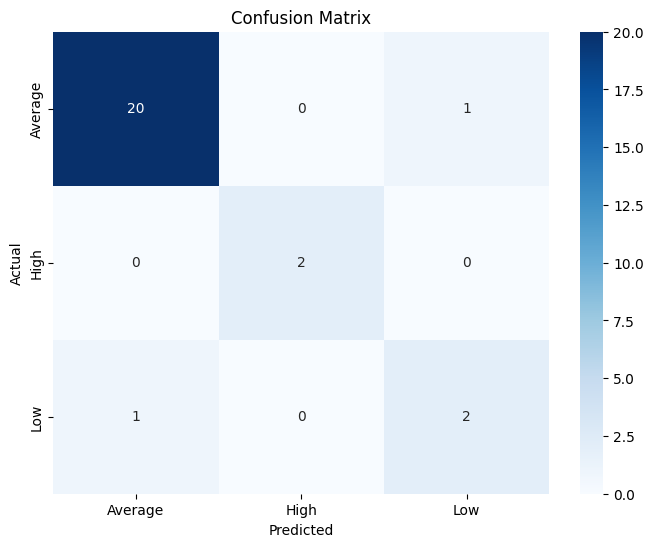
The obtained responses were used to train two machine learning algorithms solutions namely: XG Boost and Random Forest. The two models therefore were aimed at predicting the academic achievement of students with respect to their habits. The XG Boost model was able to outperform other models, with an accuracy rate of 92%, in predicting academic performance, while the Random Forest model was able to predict academic performance at an accuracy of 80%.

Thanks to these models we were able to discover the habits the most contributing to academic success Fast Food Frequency and Self Study came out as the strongest shifts as can be seen from the feature importance chart. Such precision means that our models possess a certain level of accuracy when predicting within which category (low, average, high) of marks a student’s lifestyle habits will fall into which is instrumental in bettering learning outcomes through changing habits.

*Let’s compare both the models :*

We have developed confusion matrix for both the models Understanding and Interpretation of the Confusion Matrices.

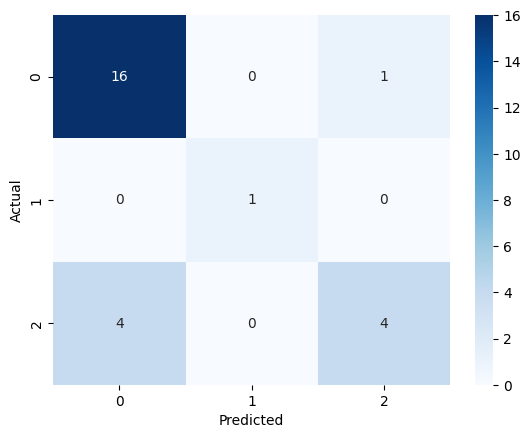
1. *XG Boost Model*



*Fig. (9)*

As shown in this matrix model, it contains three classes namely average, high, and low. The elements on the diagonal (20,2,2) are indicative of the correct predictions made, the other elements (off-diagonal elements) are errors which denote the misclassified instances. It appears that the category of Average had more precision and recall for the predictions made using the XG Boost model, whereas a small error was made in predicting the Low and High.

1. *Random Forest Model*

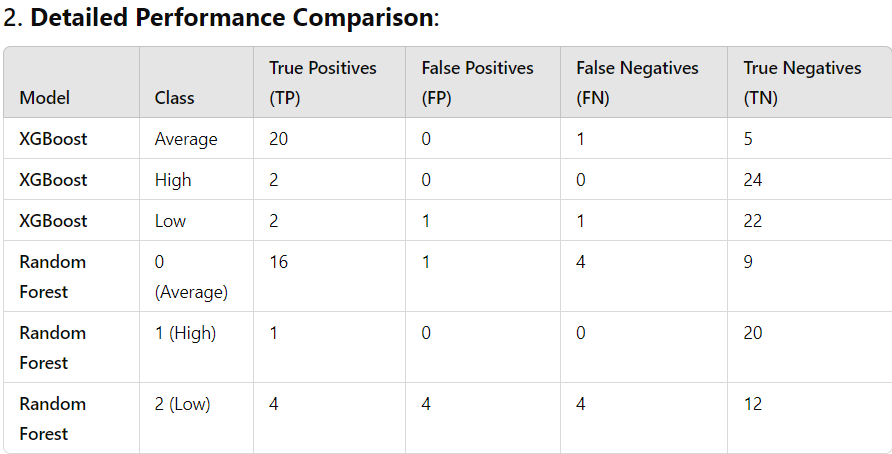


*Fig. (10)*

Like XG Boost, Random Forest also has three classes: 0 (Average), 1 (High) and 2 (Low). Appropriate predications are suggested in (16, 1, 4) and misclassifications are observed in other parts of the matrix (2D). Contrarily, Random Forest appears to have issues with class 2 (Low), attributing to the more errors in its predictions.

1. *In-depth Comparison of Performances:*

Model Class True Positives (TP) False Positives (FP) False Negatives (FN) True Negatives (TN)



*Fig. (11)*

1. *Performance Metrics:*

Accuracy: The ratio of the total number of correctly predicted observations to all predicted observations.

**XG Boost**



**Random Forest**



**[VII] Conclusion**

From the comparison made in the confusion matrix above, XG Boost outperforms Random Forest with more correct classification of all classes. It show higher precision and high recall than using all word and there are less misclassification especially for the data that belong to Average class.

Even so, the performance tested on Random Forest proved relatively satisfactory, but the classifier has more confusion with more instances placed in the Low class on the scale of misclassification.

Based on this analysis, this researcher submits that XG Boost is the one which will be preferred model if more accuracy is desired in this context inter-swap differential particularly when trying to disentangle between the Average, and Low resolutions.

**[VIII] References**

[1] O. F.Y, A. J.E.T, A. O, H. J. O, O. O, and A. J, “Supervised Machine Learning Algorithms: Classification and Comparison,” *International Journal of Computer Trends and Technology*, vol. 48, no. 3, pp. 128–138, Jun. 2017, doi: 10.14445/22312803/IJCTT-V48P126.

[2] H. Jafari, A. Aghaei, and A. Khatony, “Relationship between study habits and academic achievement in students of medical sciences in kermanshah-iran,” *Adv Med Educ Pract*, vol. 10, pp. 637–643, 2019, doi: 10.2147/AMEP.S208874.

[3] G. Curcio, M. Ferrara, and L. De Gennaro, “Sleep loss, learning capacity and academic performance,” Oct. 2006. doi: 10.1016/j.smrv.2005.11.001.

[4] B. J. Zimmerman, “Becoming a self-regulated learner: An overview,” 2002, *Ohio State University Press*. doi: 10.1207/s15430421tip4102\_2.

[5] F. Trudeau and R. J. Shephard, “Physical education, school physical activity, school sports and academic performance,” Feb. 25, 2008. doi: 10.1186/1479-5868-5-10.

[6] P. A. Kirschner and A. C. Karpinski, “Facebook® and academic performance,” *Comput Human Behav*, vol. 26, no. 6, pp. 1237–1245, Nov. 2010, doi: 10.1016/j.chb.2010.03.024.

[7] M. D. Florence, M. Asbridge, and P. J. Veugelers, “Diet quality and academic performance: Research article,” *Journal of School Health*, vol. 78, no. 4, pp. 209–215, Apr. 2008, doi: 10.1111/j.1746-1561.2008.00288.x.

[8] R. K. Golley *et al.*, “Validity of short food questionnaire items to measure intake in children and adolescents: a systematic review,” Feb. 01, 2017, *Blackwell Publishing Ltd*. doi: 10.1111/jhn.12399.

[9] H. W. . Kohl and H. D. . Cook, *Educating the student body : taking physical activity and physical education to school*. National Academies Press, 2013.

[10] B. Sánchez-Hernando *et al.*, “A comparative study of life skills, lifestyle habits and academic performance in health promoting and non-health promoting schools in the Autonomous Community of Aragon, Spain,” *PeerJ*, vol. 10, Mar. 2022, doi: 10.7717/peerj.13041.

[11] O. Llaha, “Data Visualization: A brief review.” [Online].Available: <https://www.researchgate.net/publication/359384184>

[12] Ackerman , D. S. and Gross , B. L. 2003 . Is time pressure all bad? Measuring between free time availability and student performance perceptions . *Marketing Education Review* , 12 : 21 – 32

[13] Barling , J. and Charbonneau , D. 1992 . Disentangling the relationship between the achievement striving and impatience-irritability dimensions of Type A behavior, performance, and health . *Journal of Organizational Behavior* , 13 : 369 – 377

[14] Clarke C, Mullin M, McGrath D, Farrelly N. University students and study habits. *Irish Journal of Psychological Medicine*. 2024;41(2):179-188. doi:10.1017/ipm.2021.28

[15] Rezaie Looyeh H, Seyed Fazelpour S F, Reza Masoule S, Chehrzad M M, Kazem Nejad Leili E. The Relationship between the Study Habits and the Academic Performance of Medical Sciences Students . JHNM 2017; 27 (2) :65-73

**.**