## A Comparative Study of Academic Performance between Athletic and Non-Athletic Undergraduates in the Faculty of Science, University of Colombo using Machine Learning Techniques

## E. A. I. P. De Silva 1, G. H. S. Karunarathna 2

1,2 Department of Statistics, University of Colombo, Sri Lanka

**Abstract:** The effect of sports activities is a possible factor affecting the academic work of students. University students face this problem majorly when they want to engage in sports activities but have doubts regarding the fact that it may affect their academic activities. The objectives of the study were to identify whether there is an effect of sports activities on university students, factors that affect the academic performance of the students, and factors that lead university students to engage in sports through machine learning techniques. This study was initiated by the survey with a sample of 271 students out of a target population of 2225 through a stratified random sampling technique based on students' academic year.

from the exploratory data analysis, it was revealed that being a sportsperson does not significantly impact the grade point average (GPA) overall, although an influence was noted among 4th-level undergraduates. Moreover, the quantity of sports engaged in exhibited a significant correlation with a student's GPA and the time allocated to academic and sports activities displayed a noteworthy association.

In pursuit of predictive models, machine learning techniques were employed. Logistic regression, KNN, Linear discriminant analysis, Random Forest regressor, Support vector regressor, Gradient boosting, and decision trees were used here. Random Forest classification emerged as the most effective model in determining the factors influencing a student's choice to participate in sports with 76% test accuracy. Notably, the student's academic year was identified as a key determinant. However, none of the applied algorithms could accurately predict the factors linked to the GPA of a sportsperson, suggesting a lack of discernible connection between the variables. Comparing students' opinions, it was ascertained that sports were generally perceived to have favourable impacts on academic endeavours, enhancing overall academic performance. Conversely, non-sportspersons expressed concerns about negative implications, such as time constraints affecting academic achievements.

**Keywords:** Comparative study**,** Athletes, Academic performance, Machine Learning

**Introduction:** The impact of sports participation on academic achievement among university students is a concept that holds theoretical ambiguity. It has been asserted that sports could amplify students' enthusiasm, foster collaboration, and cultivate self-discipline, ultimately leading to positive academic outcomes (Castelli et al., 2007; Chomitz et al., 2009; Trudeau & Shephard, 2008). Conversely, certain studies indicate that sports might curtail students' capacity to attain higher academic accomplishments (Bowen et al., 2011; Paul D. Umbach et al., 2013). Given the disparity in these viewpoints, a significant number of university students have struggled to determine whether to partake in competitive sports alongside their university education.

Upon reviewing earlier research (Castelli et al., 2007; Chomitz et al., 2009; Paul D. Umbach et al., 2013), it became evident that this correlation diverges based on academic environments and geographical locations. To establish a coherent understanding of the connection between sports and academic performance, this study necessitated an examination of both athletes and non-athletes at the University of Colombo. Given the variation in academic backgrounds and sports facilities across different universities, a comprehensive analysis is imperative to grasp the interplay between sports and academic endeavours. Also, Machine learning techniques were not famous among those researchers.

The primary objective of this study is to study the academic performance of students in the University of Colombo's Faculty of Science who engage in sports activities with that of students who abstain from such activities. The secondary objectives encompass evaluating the extent of the contribution of sports and identifying the sport types that are most conducive to enhancing academic performance. When focusing on the target population, the Faculty of Science at the University of Colombo encompasses students enrolled in four major academic programs, each of whom has the opportunity to participate in up to 24 sports activities within the university. To accomplish these objectives, primary data was collected by targeting Faculty of Science, University of Colombo by survey and different types of machine learning techniques were used to compare the performance of those techniques under this research as a novelty contribution.

**Methodology:** A survey was conducted covering the target population of 2225 active students using stratified random sampling. Using the Slovin's formula, 340 were selected as the required sample size to conduct the survey. The survey questionnaire was built, including four main sections: General Information, Sports Information, Academic Information, and Students' Opinions Section.

Information collected in the General section was analysed in descriptive analysis to understand the representativeness of the sample. Furthermore, in the descriptive analysis, the bivariate relationship between the GPA of the sportspersons and the factors related to their sports careers was examined.

Data was pre-processed by identifying duplicates, handling missing values, and creating new variables. Advanced analysis was based on building multivariate classification models to predict whether a student is likely to be an athlete and a predictive model that determines the GPA value of a student. Due to the diversity of responses, machine learning models with higher classification power were used.

Random Forest classifier, Gradient Boosting, K Nearest Neighbour classifier, Decision Trees, Linear Discriminant Analysis, and Support Vector Regressor were used. These were chosen due to their simple yet powerful ability in classification problems. Mean Square Error and average metrics such as accuracy and precision were used for selecting the best models. Apart from model building, understanding the importance of the factors was a main concern when using these techniques.

Opinions of the students collected through Likert scale problems were analysed using averages and medians to get an idea about the overall opinions of the students and determine the differences of opinions between sportspersons and non-sportspersons.

**Results and Discussion:** The dataset could be considered a good representation of the population, accounting for factors such as gender, ethnicity, and other general characteristics. At least one student from each degree program was selected for sampling, although there was a slight bias towards physical science students.

Through bivariate analysis, it has been identified that there is a higher probability of males being sportspersons compared to female students. Additionally, the proportion of students who are sportsmen gradually decreases with the progression of academic years.

The chi-square test, with a significance of 0.669, indicates that there is no significant relationship between GPA and sportsmanship. However, for final-year students, a relationship between GPA and sportsmanship is observed with a significance of 0.048.

In terms of time allocation, a negative correlation is found between sports and academic time, as well as between time spent on sports and academic time. These conclusions are drawn from the chi-square statistic and the phi coefficient.

The number of sports participated, and academic time were found to be correlated when considering the sport-related factors associated with the GPA of a sportsperson.

In the process of model building to identify feature importance and predict whether a student is a sportsman, both machine learning and classical methods exhibit accuracy metrics as shown in Table 1.

Table 1 Accuracy Matrices- Model 1

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Accuracy** | **Precision** | **Recall** | **F1** |
| Logistic Regression | 0.708 | 0.49 | 0.46 | 0.47 |
| Lasso | 0.724 | 0.41 | 0.34 | 0.37 |
| KNN | 0.717 | 0.58 | 0.31 | 0.40 |
| Random Forest | 0.708 | 0.52 | 0.61 | 0.56 |
| Decision Tree | 0.760 | 0.74 | 0.71 | 0.72 |
| SVM | 0.736 | 0.01 | 0.00 | 0.00 |
| Gradient Boosting | 0.704 | 0.38 | 0.24 | 0.29 |
| Linear Discriminant Analysis | 0.727 | 0.404 | 0.23 | 0.293 |

The Decision Tree classifier with the highest F1 score was selected as the best model. The model exhibited a test accuracy of 0.76, with both precision and recall values surpassing 0.7. This suggests that the machine learning model performed well in classification. According to this model, Academic year and Degree type were identified as the most influential factors for engaging in sports.

A second model, Model 2, was constructed to predict the GPA of sportspersons based on characteristics of their sports careers, including the type of sports and time spent on sports. Accuracy metrics obtained from these trained models are presented in Table 2. Although machine learning models demonstrated relatively higher performance compared to other classification models, none of the models achieved an accuracy exceeding 40%. From this, it was concluded that these variables are insufficient to develop a significant predictive model. Additionally, the number of sports students engaged in was found to be negatively correlated with their GPA scores.

Table 2 Accuracy Matrices- Model 2

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Algorithm** | **Accuracy** | **Precision** | **Recall** | **F1** |
| Logistic Regression | 0.27 | 0.13 | 0.20 | 0.210 |
| KNN | 0.23 | 0.125 | 0.18 | 0.144 |
| Random Forest | 0.23 | 0.11 | 0.22 | 0.179 |
| Decision Tree | 0.36 | 0.30 | 0.31 | 0.292 |
| SVC | 0.34 | 0.15 | 0.26 | 0.183 |
| Gradient boosting | 0.28 | 0.21 | 0.26 | 0.258 |
| Linear discriminant Analysis | 0.35 | 0.36 | 0.32 | 0.300 |

Upon analysing student opinions, it became apparent that the opinions of sportspersons tend to be biased towards the benefits of sports, whereas the opinions of non-sportspersons lean towards the negative effects of sports on students. This suggests that misconceptions about sports might be a significant barrier preventing students from participating in sports activities.

**Conclusions:**

Male students are more likely to being an Athlete compared to female students.

Students in higher academic years are less forward to being athletes.

with a significance of 0.669, chi-square test indicates there is no effect of being an athlete towards the GPA of a student in overall.

When considering the effect of academic year, only 4th-year students show a significant relationship with GPA values, with a significance of 0.048 in the chi-square test.

a negative correlation is found between sports and academic time, as well as between time spent on sports and academic time.

Decision tree classifier was selected as the best model for prediction and identification of the factors related to engage in sports activities.

Academic year and Degree type were identified as the most influential factors for engaging in sports activities.

Number of sports students engaged in was found to be negatively correlated with their GPA scores.

Opinions of athletes tend to be biased towards the benefits of sports, whereas the opinions of non-athletes biased towards the negative effects of sports on students.

**References:**

Bowen, W. G., Levin, S. A., Shulman, J. L., Campbell, C. G., Pichler, S. C., & Kurzweil, M. A. (2011). Reclaiming the game: College sports and educational values. *Reclaiming the Game: College Sports and Educational Values*, *1546*, 1–490. https://doi.org/10.1515/9781400888924-012

Castelli, D. M., Hillman, C. H., Buck, S. M., & Erwin, H. E. (2007). Physical fitness and academic achievement in third- and fifth-grade students. *Journal of Sport and Exercise Psychology*, *29*(2), 239–252. https://doi.org/10.1123/jsep.29.2.239

Chomitz, V. R., Slining, M. M., McGowan, R. J., Mitchell, S. E., Dawson, G. F., & Hacker, K. A. (2009). Is there a relationship between physical fitness and academic achievement? Positive results from public school children in the Northeastern United States. *Journal of School Health*, *79*(1), 30–37. https://doi.org/10.1111/j.1746-1561.2008.00371.x

Paul D. Umbach, Palmer, M. M., Kuh, G. D., & Hanna, S. J. (2013). INTERCOLLEGIATE ATHLETES AND

EFFECTIVE EDUCATIONAL PRACTICES: Winning Combination or Losing Effort? *Educational Technology Research and Development,* *57*(5), 613–627. https://doi.org/10.1007/sl

Trudeau, F., & Shephard, R. J. (2008). Physical education, school physical activity, school sports and academic performance. *International Journal of Behavioral Nutrition and Physical Activity*, *5*, 1–12. https://doi.org/10.1186/1479-5868-5-10