A Comparative Study of Academic Performance of Sportspersons and Non-Sport Persons in the University of Colombo Faculty of Science

By

E. A. I. P. De Silva

This dissertation is submitted as a partial fulfilment of the B.Sc. Honours in Statistics

Department of Statistics,

University of Colombo, Sri Lanka.

March 2023

Abstract

The effect of sports activities is a possible factor affecting the academic works 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.

This is a survey-based study conducted at the University of Colombo Faculty of Science with the aim of understanding the relationship between academic activities and sports activities of a university undergraduate. As the objectives of the study, identify whether there is an effect of sports activities on university students, factors that affects the academic performance of the students, and factors that lead university students to engage in sports were considered. The study was carried out on a representative sample of 271 students from a target population of 2225 students. For the selection of the sample, a stratified random sample based on the academic year of students was used.

A descriptive study was carried out to identify the data and to identify the relationships between the variables. Through the chi-square test of independence, it was discovered that being a sportsperson has no significant impact on the grade point average, but the number of sports has a significant effect on the GPA value of a student. Even though, 4^{th} level undergraduates have a significant effect of sports activities on their GPA value. Through pairwise comparisons of variables using the Kruskal-Wallis test and chi-square test of independence, it could be identified that time spend on academic activities has a significant association with engaging in a sport and time spent on sports activities.

By analyzing through multiple modeling techniques, Random Forest classification was selected as the best model to identify the factors that determine an undergraduate to be a sportsperson. The academic year of a student was identified as the most significant factor.

The median and of the responses were considered for comparing undergraduates' opinions. It has been identified that students believe sports have many positive effects on academic activities and that sports activities improve a student's overall academic performance. Non-sports persons strongly believe there are negative effects of sports such as time consumption, which affect the academic performance of a student.

Declaration

This thesis is my original work and has not been submitted previously for a degree at this or any other university/institute. To the best of my knowledge, it does not contain any material published or written by another person, except as acknowledged in the text.

Candidate: E. A. I. P. De Silva Signature ……..…...

Date ……....

This is to certify that this dissertation is based on the work carried out by Mr. E. A. I. P. De Silva, under my supervision. This dissertation has been prepared according to the format stipulated and is of an acceptable standard.

Supervisor: Dr. Hasani Karunarathne Signature ……..…...

Senior lecturer / lecturer,

Department of Statistics, Date ……....

University of Colombo.

Coordinator: Dr. Gayan Dharmarathne Signature ……..…...

Senior lecturer / lecturer,

Department of Statistics, Date ……....

University of Colombo.

Acknowledgment

I would like to offer my sincere gratitude to all the people who have given me enormous support in several ways.

Foremost, I would like to express my sincere gratitude to my advisor, Dr. Hasani Karunarathne, for the continuous support of my final year research and for her patience, motivation, enthusiasm, and immense knowledge. Her guidance helped me in all phases of the research and writing of this thesis.

Besides my advisor, I would like to thank all the professors who imparted knowledge to me in the Faculty of Science at the University of Colombo. The knowledge I gained from them was a huge support in making this project a success.

Table of contents

List of Figures

List of Tables

Chapter 1

Introduction

This chapter focuses on giving a brief description of the study. It contains an outline of the thesis as well as information on the study's background, goals, and importance. The purpose of this chapter is to provide a firm background of the thesis.

1.1 Background

The impact of sports participation on academic achievement as a university student is a theoretically equivocal concept. It has been claimed that sports could enhance students' enthusiasm and promote collaboration and self-discipline, resulting in favourable academic spillovers (Castelli et al., 2007; Chomitz et al., 2009; Trudeau & Shephard, 2008). Even though, some believe that sports diminish the ability of students to achieve higher academic achievements (Bowen et al., 2011; Paul D. Umbach et al., 2013). With the differences in these opinions, most university students haven’t been able to decide whether to engage in sports activities with a competitive university education or not.

To get a clear identification of the relationship between sports and academic performance, this study had to be based on information from sportspersons and non-sportspersons at the University of Colombo. Since academic backgrounds and sports facilities vary from university to university, it is important to study them to understand the relationship between sports and academic activities.

1.1.1 Sports Activities at University Level

University education, which is the most competitive and advanced education system of any country, also cooperates with sports activities on many levels to build students' physical abilities and provide a good leisure time activity. In Sri Lanka, universities provide students with the chance to engage in sports by providing them with the necessary equipment, training courts, and sports advisors. They also encourage the students for engage in sports activities by holding sports competitions at university, national, and international levels. The first inter-university sports championship was organized in the 1950s. Since then, inter-university games have been conducted yearly for

several sports to determine the winner. Every three years, Sri Lanka holds the University Games Championship, which has been the most competitive sports championship among university sports held on different university premises. The latest Sri Lanka University Games (SLUG) meet has been held at University of Ruhuna in 2019. The University of Colombo has been the champion of university games many times since the beginning of the SLUG championship. The earliest win for the University of Colombo was the 2010 SLUG championship. (Sri Lanka University Games- SLUG | *Sri Lanka Universities Sports Association*)

Inter-faculty sports meets are also held annually separately at the university level. The Freshers Championship and the Inter-Faculty Sports Championship are the major sports meets at the University of Colombo. The main aim of these sports events is to encourage the students to participate in sports activities, identify their sports skills, and prepare them for national and international sports events. The Faculty of Science at the University of Colombo has been a leading faculty of sports for a long time. Even with the tough academic activities, these students have been able to compete and win sports awards at SLUG and other national and international levels. (Handbook - Faculty of Science - University of Colombo 2021/2022)

1.1.2 University of Colombo, Faculty of Science – Education

The University of Colombo is the leading university in Sri Lanka based on its education and all other factors. Among science faculties in Sri Lankan universities, the Faculty of Science, University of Colombo (FOS UOC) occupies a pivotal position. Presently, FOS UOC consists of 7 departments, which are Chemistry, Nuclear Science, Mathematics, Physics, Plant Sciences, Statistics, Zoology and Environment Sciences. (Handbook - Faculty of Science - University of Colombo 2021/2022)

Students in FOS UOC study for their first 2 years based on the subject streams they have selected when applying to the university. Four such streams are available in FOS UOC.

- Physical Science
- Biological Science
- Biochemistry and Molecular Biology
- Industrial Statistics and Mathematical Finance

At the end of two academic years, students are given the chance to select their degree program. Intending to provide its students with the best learning experience, FOS UOC grants students multiple degree programs related to various educational aspects.

There are three types of degree programs offered by FOS UOC to its students. They are,

- BSc degree program (three-year duration)

This program is designed for students who wish to complete their degree program and enter the job market. Students follow this degree program under the same subject Strems they have been following for the past two years.

- BSc Honours degree program, 4-year duration (research-oriented)

Students have to get selected for these degree programs based on their results in the first two years. These degree programs are focused on giving students the skills they need to work in academia.

- BSc Honours degree program 04-year duration (industrial oriented)
  This degree program is based on the skills required for a career in industry. According to the requirements of the industry, the university offers this degree program covering a variety of themes.

Universities supply students with sufficient resources to improve their academic activities, including infrastructure and qualified lecturers. FOS UOC bestows multiple awards on students upon completion of their degree programme in recognition of their academic achievements. (Handbook - Faculty of Science - University of Colombo 2018/2019)

1.1.3 University of Colombo, Faculty of Science – Sports Activities

The Department of Physical Education at UOC supports students with their sports activities. This department offers students physical fitness programs and motivational programs. For the support of sportspersons, the university supplies sports equipment and courts for 24 sports. (Handbook - Faculty of Science - University of Colombo 2021/2022)

Sports activities

- Badminton
- Carrom
- Chess
- Karate
- Scrabble
- Swimming
- Table Tennis
- Taekwondo
- Baseball
- Basketball
- Beach volleyball
- Cricket
- Elle
- Football
- Hockey
- Tennis
- Rugby
- Track and Field
- Volleyball
- Netball
- Road Race
- Rowing
- Weightlifting
- Wrestling

1.2 Objectives

In this study, the objectives are,

Primary Objective

- Compare the academic performance of students in UOC FOS who participate in sports activities and the students who do not participate in sports activities.

Secondary Objectives

- Determine the contribution and most favorable sports types that enhance academic performance.
- Identify the factors that motivate sportspersons toward sports activities.

1.3 Significance of the Study

- Through the comparison of sportspersons and non-sportspersons regarding their academic progress in the university, it can be recognized whether sports are good or bad for the university's academic activities.
- By recognizing the above fact, students who started academic activities in university can decide whether they should or shouldn’t engage in sports activities.
- Students can recognize which sports work well with academic activities and how much effort should be given towards those sports.
- From the results, The University can motivate students to engage in the correct types of sports and include sports activities that improve academic performance for university students.
- This type of study has not been done based on Sri Lankan university education.

1.4 Organization of the Thesis

This thesis is divided into seven chapters. The first chapter contained background information with an overall description of the study, its significance, and the research objectives of the study.

Chapter 2 – Literature Review

Provides a review of past studies associated with the objectives of this study regarding the findings and methodologies.

Chapter 3 – Theory and Methodology

Includes details about the methodology used in the study and a brief explanation of the theories used in the project.

Chapter 4 - Data

demonstrates how to calculate the sample size, collect data, and prepare data for analysis. This chapter also includes the variable description.

Chapter 5 – Results

This section consists of results from exploratory data analysis and advanced analysis.

Under exploratory data analysis, visualization of data and numerical tests on data were carried out to understand the characteristics of the variables and the association between variables.

Advanced analysis includes model building to identify relationships among response variables and predictor variables.

Chapter 6 – Discussion and Conclusion

This Chapter includes the overall discussion and conclusions of the study.

After this chapter will be the list of references and the appendices, which will have the questionnaire and all the relevant programming codes used in this study.

Chapter 2

Literature Review

Relationship between the academic activities and sports is a conversational topic in society. Much research has been done to identify this relationship worldwide. When referring to literature, many researchers have tried to identify the effect of sports activities on academic activities by looking at various perspectives of the problem. Some researchers have been able to identify the significance of these positive and negative factors that affect academic activities using various experiments and techniques while others have focused on the overall effect of sports on academic activities.

2.1 Positive Effects of Sports on Academic Activities

At the end of their studies, most researchers found that sports could improve the academic performance of students significantly. These studies have been conducted on students from various age groups using various analysis techniques.

When considering such studies, Trudeau & Shephard (2008) conducted a systematic review of literature based on quasi-experimental studies to identify the sport-academic relationship. Through these cross-sectional studies, they discovered that, by reducing the academic time of primary school students and allocating that to Physical education, primary school students performed better in mathematics subjects in comparison to students who maintained regular time for academic activities at the end of the 5-year time frame. Ahamed et al., (2007) conducted a 16-month cluster randomized control trial in Canadian schools by increasing the physical education time by 45 minutes. By analyzing the Canadian achievement test scores using independent sample t-tests, the author showed that the learning efficiency of these school children has improved with the involvement of sports.

Castelli et al., (2007); Chomitz et al., (2009); Coe et al., (2006); Dyson et al., (2018) also implied that physical fitness is associated to obtain higher grades and therefore schools should not reduce the time allocated for sports and physical education sessions to increase the academic time. Through a research-based study on New Zealand schools, Dyson et

al., (2018) discovered that more than 20 minutes is good to allocate for physical education sessions. The author has come to this conclusion using an exploratory factor analysis model based on survey data. Chomitz et al., (2009) have used multivariate logistic regression analysis to access the probability of passing the academic tests when coming to the above conclusion.

According to the research carried out based on secondary schools in the metropolis, Sani et al., (2019) stated that sports give improved health benefits and improved academic grades to students. The authors also stated that students who engage in sports are more likely to attend classes and complete academic work, leading to higher grades. This study was conducted through a survey containing a Likert scale questionnaire focused on the experiences on sports of both students and teachers in sports.

When considering the literature on factors that lead to improved academic activities, some researchers have found sports and physical education improve specific abilities of students which leads them to perform better in academic work.

Development of cognitive function

The development of cognitive function in the brain has a direct influence on the academic progress of students. Some researchers have identified sports increase the cognitive functions of students with regular physical activities. In a study based on students who participated in after-school physical activities, students who regularly engaged in those activities demonstrated significant improvements in cognitive performance compared to students who do not (Chaddock-Heyman et al., 2014; Chaddock et al., 2011). In a similar study, it was discovered that engaging in 10 minutes of physical activity per day can increase brain functions responsible for cognitive control and attention (Voss et al., 2016). The authors have come to these conclusions based on cross-sectional studies supported by some randomized, controlled trials. Geda et al., (2010) imply that not only physical activity boosts cognitive brain functions but also a slower rate of cognitive decline as you get older. These conclusions are based on the comparison of odds ratios for age groups and the amount of physical exercise.

Improvement of time management

Sports require a level of discipline and time management skills. Researchers have experimented with the fact that these skills have been used in academic activities as well.

In a study of high school athletes, athletes showed better time management skills than non-athletes (Darling et al., 2005). And also, through univariate general linear models, the authors have identified that demographic variables such as ethnicity and age are significant factors that lead to participation in extracurricular activities.

Improvement of mental health

According to medical experts, exercise helps to reduce stress, anxiety, and depression which improves mental health and impacts academic performance positively. Buckworth et al., (2007) discovered that college students who engaged in physical activities daily reported higher levels of self-esteem and lower levels of depression than non-athletes. Through a study of high school sportspersons, Eime et al., (2013) achieved similar results that indicate improvements in mental health. In his study, authors used a systematic review of 14 electronic databases to come to this conclusion.

Improvement of teamwork and social skills

Teamwork and group projects are part of the educational system at any level. Since sportspersons get to practice social skills and develop their teamwork within their sports activities, it has been claimed that sportspersons are good at group projects and collaborations in academic work (Buunk et al., 2010; Darling et al., 2005). Buunk et al., (2010) conducted an experimental project based on schools in the Netherlands including 921 students. At the end of the study, a Likert scale questionnaire was used to collect and compare the differences between groups. Through mean comparison, Buunk et al., (2010) have identified that students who engage in sports activities show improved social skills, the author also implies that sportspersons get a higher level of social support when conducting academic activities.

2.2 Negative Effects of Sports on Academic Activities.

Some researchers strongly suggest that sportspersons underperform in academic activities compared to non-sportspersons (Bowen et al., 2011). Through a comparison of 33 colleges that conduct athletics and others that do not, Bowen et al., (2011) have identified that the athletes underperform in their academic activities. Also, authors have mentioned that athletic recruitment in educational institutions is a problematic fact due to this, and most of the time selected athletes perform worse in academic activities over time.

Studying the academic performance of athletes, Umbach et al. (2013) have stated that athletes have lower grades compared to non-athletes in schools. The author has come to this conclusion by observing a series of hierarchical linear models.

Both Simons et al., (1999) & Bowen et al., (2011) have discussed the effect of recruiting students to schools based on their previous athletic records. Even though both of these researchers have concluded that sports are affecting academic activities negatively, Simons et al., (1999) have mentioned that when setting preschool differences in academic activities are under control, the academic performance of school students does not differ significantly.

When referring to the main factors that could cause negative effects on academic activities, the time required for athletics and physical exhaustion can be identified.

Time constraints

Since most sports require higher practice time which could be spent on academic activities such as studying or attending lectures, the academic performance of a student can be negatively affected. When considering the literature on this topic, a study of college athletes found that they reported spending less time on academic work than non-athletes (Pascarella et al., 2004).

Physical exhaustion

Over-exhaustion in sports can lead to physical exhaustion, which can negatively impact concentration and performance in academic work. In a study of university athletes, Beasley et al., (2003), reported that athletes feel more tired and less alert during academic work compared to non-athletes. The author has come to this conclusion by using hierarchical and stepwise regression analysis techniques on survey data.

2.3 Effect of Sport Activities on Academic Activities

When considering the types of sports, some researchers have tried to identify which sports do not negatively affect academic activities and which sports affect academic activities positively. Among the three levels of physical activity, which are vigorous, moderate, and light (World Health Organization), Coe et al., (2006) indicated that higher grades in academic activities can be achieved only by increasing the time spent on vigorous physical activities. He also stated that moderate physical activities do not have a significant effect on academic performance. Even though this is the case, Ahamed et al.,

(2007) stated that Most of the schools have allocated insufficient time for sports and spend more than half the time on non-vigorous activities.

Since the education systems are significantly different according to the regions, it was identified that this relationship could be different in different countries. In Sri Lanka, studies carried out on this topic are pretty rare. Pamudini & Manel (2022) have identified that students who engage in sports activities have better knowledge and skills compared to others. This has led them to score higher grades in subjects like science and mathematics. The authors came to these conclusions based on analysis of survey data on the governmental schools around Gampaha District using an independent sample t-test and a chi-square test. The authors suggested that improving sports in schools is essential for the development of education in Sri Lanka. No advanced analytic techniques have been used in this study.

In summary, the effects of sports on academic performance have been different according to the region, education level, subject streams, types of sports, and many other factors. Therefore, it is crucial to analyse the related population to come to a conclusion about the relationship between academic activities and sports. Conducting the research project on the faculty level will allow for an accurate conclusion on how the academic performance of University of Colombo Faculty of Science students is related to sports activities. Also, the results of this research will coincide with the relationship between sports and academic activities in other science faculties in Sri Lanka and countries that have a similar educational background to Sri Lanka.

Chapter 3

Theory and Methodology

This chapter focuses on the basic theories and methodologies used to achieve the objectives of this study. Theories and methodologies behind survey planning, preliminary analysis, and advanced analysis will be discussed in this section.

3.1 Identify Research Background

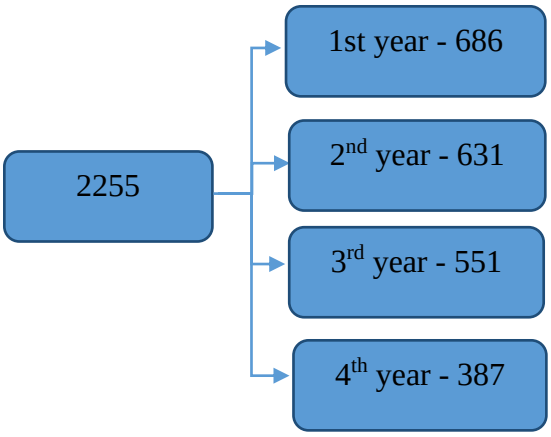
It was identified that the link between sports and academic activities is considered by many researchers worldwide(Ahamed et al., 2007; Chomitz et al., 2009; Paul D. Umbach et al., 2013; Simons et al., 1999). Even though some studies show similar results to each other, there are studies that show completely opposite conclusions on the relationship between sports and academic activities. At the end of the background study, it was found that different results about this problem may have been caused by differences in education systems, regions, etc. To get a clear idea about the specific relationship between sports and academic activities at the Faculty of Science, University of Colombo (FOS, UOC), it is crucial to study the data gathered from FOS, UOC students.

Study was aimed to conduct on the primary data gathered through a survey in Faculty of Science, University of Colombo.

3.2 Survey Plan

Target population

- All students in UOC FOS are considered as the target population.
- The current target population of the faculty of Science, university of Colombo consists of 2255 students from four academic levels.



Sampling technique

Academic activities and effort have a significant difference when comparing academic years. Since this study aimed to compare the academic performance of students, it was necessary to include the best representation of students from each academic year in the sample. Considering this, stratified sampling according to the academic year of students was used in the survey.

For determine the sample size within each stratum proportional allocation was used by taking the proportion of students in each academic year as reference. By choosing people at random for the survey, it was hoped that the number of athletes would be better represented.

For the selected sample size, survey questioner was sent through university emails via Information Technology Unit, UOC FOS.

Sample size calculation

A survey's sample size can be chosen using a variety of methods, including the power analysis formula, the margin of error formula, the Bayesian sample size, the Solvins formula, etc. Since it's easy to use and does not require prior knowledge about the population, Solvin's formula was selected to use for the sample size calculation for the survey.

Solvin's Formula:

$$n = \frac{N}{1 + Ne^2}$$

$N = Population\ Size$ (2255)
$n = Sample\ Size$
$e = Error\ limit$ (0.05)

After deciding the overall size of the sample, the number of students from each academic year to include in the survey was determined by proportional allocation.

The equation for proportional Allocation:

$$n_h = \frac{N_h}{N} \times n$$

$N = Population\ Size$
$n = sample\ Size$
$N_h = Size\ of\ the\ Stratum$
$n = Size\ of\ the\ sample\ Stratum$

3.3 Questionnaire Design

The questionnaire was designed to collect all the data necessary to understand the relationship between sports and academic activities among the FOS, UOC students. For the sake of comparing academic activities between sportspersons and non-sportspersons, all the details except sports information were collected from all students who were selected for the survey. The details related to the sports are collected from only sportspersons by using the filter questions in the questionnaire.

The questionnaire consists of five main sections.

Section 01 - General information

Under this section, demographic factors related to both athletes and non-athletes have been captured. The aim of this information is to identify whether there is a connection between these demographic variables and the fact that students decide to engage in sports activities. The section included a total of 8 questions.

Gender, district of residence, and ethnicity are collected through three multiple-choice questions. To capture all the degree programs granted by the university, five multiple-choice questions were included to facilitate the respondents.

Section 02 – Whether a sportsman or not

This section consists of a single question for respondents. If the students identify themselves as sportspersons, they can mark it as "yes" or otherwise "no."

Section 03 - Information on sports carrier

If a student identified himself as a sportsperson, they were directed to this section to capture the necessary information regarding their sports career. This section included six multiple-choice questions and a check box where the person could mark all the sports they have been playing. Details such as how long they have been a sportsman and how much time they spent on sports activities were gathered through this section.

Section 04 - Information on academic activities

3 multiple choice questions were included in this section. The questions include the most basic factors that can be used to measure the academic progress of a university student. For the comparison of academic performance, this information were made compulsory for each student in the sample.

Section 05 - Opinions of undergraduates on sports activities

To check the opinions of the undergraduates, 10 questions with a 5-point Likert scale were used. The objective of this section was to determine the differences in impressions between sportspersons and non-sportspersons on sports activities. It was expected that these opinions are the major factor that leads students to engage in sports activities.

The survey was sent through the Information Technology Unit at FOS, UOC, to 441 randomly selected students in FOS, UOC, according to the sampling scheme.

To increase the response rate further, regular announcements were made through WhatsApp and personal meetups to find out and encourage students who were randomly selected to answer the survey.

3.4 Exploratory data analysis techniques

For the identification and visualization of patterns in the data, exploratory data analysis was used. For the visualization of univariate and bivariate characteristics, pie charts and bar charts were used since almost all the variables in the analysis were either categorical or discrete. To identify the relationships among variables, hypothesis tests were applied that are designed to detect associations within the variables.

Kruskal Wallis Test

Kruskal Wallis test measures the association between a numerical dependent variable and categorical independent variable with more than two levels. It is the non-parametric alternative to the one-way ANOVA and an extension version of the Mann-Whitney U test.

Hypotheses

H_0: Medians are same for each group

H_1: Medians are different

Test Statistic (H statistic)

$$H = \left[\frac{12}{n(n+1)}\sum_{j=1}^{c}\frac{R_j^2}{n_j}\right] - 3(n+1)$$

$$n = sum\ of\ group\ sizes\ for\ all\ groups$$

$$c = number\ of\ groups$$

$$R_j = sum\ of\ ranks\ in\ the\ j^{th}\ group$$

$$n_j = size\ of\ the\ j^{th} group$$

Rejection criteria

reject H_0 if, H statistic > χ^2_{c-1}

Assumptions

1. The independent variable should be with two or more levels.
2. The dependent variable should be in ordinal scale, ratio scale or interval scale.
3. The observations should be independent.
4. All groups should have the same shape distributions.

Chi-square test of independence

If there is a statistically significant correlation between two categorical variables, it can be found using the chi-square test of independence. The test is based on comparing the variable frequencies that were observed to those that were anticipated, assuming that the variables are independent.

By creating a contingency table that shows the joint distribution of the two categorical variables, the test is carried out. In each cell of the contingency table, the observed frequencies are noted. Using the marginal sums of the contingency table and the assumption that the variables are independent, the expected frequencies are then calculated.

Hypotheses

H_0: 2 variables are independent

H_1: 2 variables are not independent

Test Statistic

$$\chi^2 = \sum_{i=1}^{n}\sum_{j=1}^{m}\frac{(O_i - E_i)^2}{E_i}$$

$$O_i = Observed\ frequency$$

$$E_i = Expected\ frequency$$

n = Number of categories in variable 1

m = Number of categories in variable 2

Rejection criteria

reject H_0 if, $\chi^2 > \chi^2_{(n-1)(m-1)}$

Assumptions

1. The observations used in the test must be independent of each other.
2. The sample size must be large enough to ensure that the expected frequencies for each category are at least 5.
3. The data used in the test should be categorical or count data. If the data is continuous, it should be grouped into categories before conducting the test

3.5 Methodology for Advanced Analysis

Advanced analysis was used to achieve two secondary objectives of the study which are,

- Determine the contribution and most favorable sports types that enhance academic performance.
- Identify the factors that motivate sportspersons towards sports activities.

Considering the above 2 objectives, the main concern of the advanced analysis is to detect related variables toward academic performance and being a sportsperson. Feature Selection algorithms were proposed to achieve that objective. Prior to that, it was vital to check whether these response variables could be forecasted with good accuracy using the predicted variables in the dataset.

For model building and feature selection classification models, which are machine learning algorithms, better performance was needed.

Machine learning algorithms

Machine learning algorithms are a type of statistical modelling that uses computational methods to learn patterns and relationships from data without being explicitly programmed.

Machine learning algorithms can be broadly classified into three categories:

I. Supervised learning

   In supervised learning, the algorithm learns from labelled examples, where each example consists of a pair of input and output data. The objective is to discover a mapping function between inputs and outputs.

II. Unsupervised learning

   In unsupervised learning, the algorithm learns from unlabelled data, where there are no pre-defined output values. The goal is to discover patterns or relationships in the given data.

III. Reinforcement Learning

In reinforcement learning, the algorithm interacts with its environment and learns through trial and error. The objective is to discover a policy that maximises a reward signal.

In the analysis supervised machine learning was used since both input and output variables were present within the dataset.

1. **K-Nearest Neighbors (KNN)**

   KNN is a non-parametric algorithm in the field of machine learning, used for classification and regression tasks. The algorithm works by finding the K closest data points to the target point and then classifying the target based on the majority class or regression value among its K-nearest neighbors. In the classification task, KNN assigns a class to the target point by selecting the majority class from the K nearest neighbors.

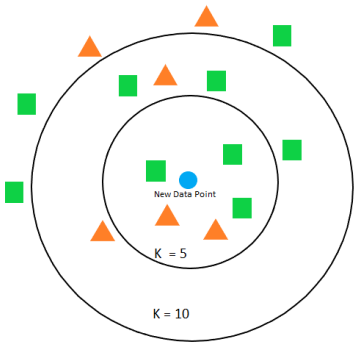


Figure 3-1 KNN Diagram

Source: https://debuggercafe.com/an-introduction-to-k-nearest-neighbors-in-machine-learning/

KNN is a simple algorithm that requires no prior training, making it easy to implement and understand. However, the algorithm can be computationally expensive, especially when dealing with large datasets, as it requires computing distances between the target point and all the other data points. Overall, KNN is a powerful algorithm for classification and regression tasks, particularly in cases where the data is non-linear or when there is no clear separation between classes.

2. Decision trees

Decision trees are a popular machine learning algorithm used for classification. The algorithm works by recursively splitting the data into subsets based on the most informative features, creating a tree-like structure where the leaf nodes represent the final prediction.

In a decision tree, each node represents a decision based on the value of a specific feature. The algorithm decides which feature to split on by selecting the feature that provides the most information gain, which measures the reduction in entropy or impurity of the dataset after the split. Decision trees have many advantages over other machine learning algorithms.

They are easy to understand and interpret, can handle both categorical and numerical data, and can identify important features in the dataset.

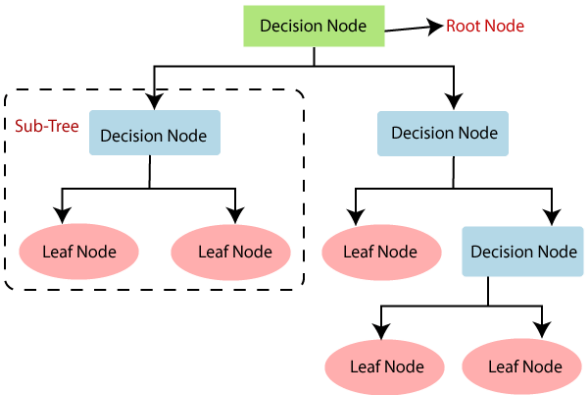


Figure 3-2 Decision Tree Diagram

Source: https://www.javatpoint.com/machine-learning-decision-tree-

3. Random Forest

Random Forest works by generating multiple decision trees, where each tree is built using a random subset of the training data and a random subset of features at each split. This randomness helps to reduce the risk of overfitting and improve the overall accuracy of the model. During prediction, the output of each decision tree is combined through a voting mechanism to produce the final prediction.

Random Forest has many advantages over other machine learning algorithms. It can handle both categorical and numerical data, can handle missing values, and can identify important features in the dataset.

Additionally, it can also handle high-dimensional data and noisy data, making it suitable for a wide range of applications.

However, Random Forest also has some limitations. It can be computationally expensive, especially when dealing with large datasets, and it may not perform well on imbalanced datasets or datasets with strong class dependencies.

4. Linear Discriminant Analysis (LDA)

LDA is a statistical method used for classification tasks in machine learning. LDA aims to find a linear combination of features that best separates the classes in the data by maximizing the between-class variance and minimizing the within-class variance. The algorithm works by first calculating the mean and covariance matrix of each class in the dataset. It then projects the data onto a lower-dimensional space by finding a set of linear discriminants that maximize the ratio of the between-class variance to the within-class variance. Even though LDA assumes that the data follows a normal distribution and that the classes have equal covariance matrices. In practice, these assumptions may not hold, and thus, LDA may not be the optimal algorithm for all datasets. LDA has many advantages over other classification algorithms, such as being computationally efficient and having a low risk of overfitting. It can handle high-dimensional data and is robust to outliers.

5. Support Vector Machines (SVM)

Support Vector Machines (SVMs) are a popular machine learning algorithm used for classification and regression tasks. SVMs work by finding the hyperplane that maximally separates the data into different classes. In SVM, the margin is the distance between the hyperplane and the closest data points from each class, and the goal is to maximize this margin while minimizing the classification error. If the data is not linearly separable, SVM can use a kernel function to map the data into a higher-dimensional space where a linear boundary can be found. SVM has many advantages over other machine learning algorithms. It can handle high-dimensional data, is robust to overfitting, and can be used for both classification and regression tasks.

6. **Gradient Boosting**

Gradient Boosting is a machine learning algorithm that works by building an ensemble of weak predictive models, such as decision trees, and combining them to create a stronger model. The algorithm trains each model in a sequence, with each subsequent model trying to correct the errors of the previous one. The algorithm works by first fitting a simple model to the data, such as a decision tree with a small number of nodes. It then calculates the error of this model and uses it to adjust the weights of the training data. The algorithm then trains a new model on the adjusted data and adds it to the ensemble. This process is repeated, with each new model trying to reduce the residual error of the previous models.

Gradient Boosting can capture complex nonlinear relationships in the data and has been shown to perform well in many machine learning tasks, such as classification, regression, and ranking.

Evaluation metrics

Mean Squared Error (MSE)

MSE is a commonly used metric for evaluating the performance of a predictive model in regression tasks. The MSE is calculated by taking the average of the squared differences between the predicted and actual values over the entire dataset.

Mathematically, it is defined as:

$$MSE = \frac{\sum_i^2 (y_i - \hat{y_i})^2}{n}$$

A smaller MSE value indicates a better fit of the model to the data. MSE also has some limitations. It can be sensitive to outliers in the data and may not always be the best metric to use in certain situations. Additionally, the MSE does not provide any information about the direction or magnitude of the errors.

Accuracy

Accuracy is a common metric used in machine learning to evaluate the performance of a classification model. It is the ratio of the number of correct predictions to the total number of predictions made by the model, expressed as a percentage.

Mathematically, accuracy can be defined as:

$$Accuracy = \frac{Number\ of\ correct\ predictions}{Total\ number\ of\ predictions}$$

Precision

Precision is a metric used in machine learning to evaluate the performance of a classification model, particularly in cases where the focus is on minimizing false positives. It measures the proportion of true positive predictions among all the positive predictions made by the model.

Mathematically, precision can be defined as:

$$Precision = \frac{True\ positive}{True\ positives + false\ positives}$$

In other words, precision is the number of true positive predictions divided by the total number of positive predictions made by the model. A high precision score indicates that the model is making relatively few false positive predictions.

Recall

Recall is a metric used in machine learning to evaluate the performance of a classification model, particularly in cases where the focus is on minimizing false negatives. It measures the proportion of true positive predictions among all the instances that are actually positive.

Mathematically, recall can be defined as:

$$Recall = \frac{True\ positive}{True\ positives + False\ negatives}$$

In other words, recall is the number of true positive predictions divided by the total number of positive instances in the dataset. A high recall score indicates that the model is correctly identifying a large proportion of the positive instances.

F1 Score

F1 score is a commonly used metric in machine learning for evaluating the performance of a classification model. It is a harmonic mean of precision and recall and is particularly useful in cases where there is an imbalance between the number of positive and negative instances in the dataset.

The F1 score can be calculated as:

F1 score = 2 * (precision * recall) / (precision + recall)

$$F1\ Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

The F1 score provides a balance between precision and recall, and a high F1 score indicates that the model is correctly identifying a large proportion of the positive instances while making relatively few false positive predictions.

<u>Correlation</u>

Phi coefficient

The phi coefficient, also known as the phi correlation coefficient, is a statistical measure used to determine the association between two categorical variables. It ranges from -1 to +1, where -1 indicates a perfect negative association, 0 indicates no association, and +1 indicates a perfect positive association.

$$\phi = \frac{(ad - bc)}{\sqrt{(a+b)(c+d)(a+c)(b+d)}}$$

$$a = True\ positive$$

$$b = false\ positive$$

$$c = False\ negative$$

$$d = True\ negative$$

Chapter 4

Data

This chapter includes an overview of the collected data for the study. A description of the targeted sample, variables collected through the survey, and data pre-processing for the analysis are discussed here.

4.1 Sampling Design

Sample size calculation

By comparison of characteristics of the known sample techniques, Solvin's formula was selected to use for the sample size calculation for the survey.

$$n = \frac{N}{1 + Ne^2}$$

$$n = \frac{2255}{1 + (2255) \times 0.05^2}$$

$$n = \frac{N}{1 + Ne^2}$$

$$n = 339.7 \approx 340$$

$$N\ (Population\ Size) = 2255$$

$$e\ (Error\ limit) = 0.05$$

Since the response rate for the online surveys was quite low according to the details gathered from previous surveys done within the science faculty, 440 students were selected as the final count of the students as participants in the survey by assuming a 10% non-response rate.

According to this, the number of students from each academic year to include in the survey was determined by proportional allocation according to table 4-1.

Table 4-1 Sampling Design

Academic Year	Population Size	Population Proportion	Sample size
1st year	686	0.304	134
2nd year	631	0.280	123
3rd year	551	0.244	108
4th year	387	0.172	76

4.2 Questionnaire Design

Total number of questions in the survey is 28. Survey includes multiple choice questions (MCQ), checkboxes and a series of Likert scale questions. Covering all aspects related to research problem, questions were included under each section as in table 4-2

Table 4-2 Questionnaire design

Section	Number of questions	Question type
General Information	8	MCQ
Being a Sportsman or not	1	MCQ
information on academic activities	7	MCQ, checkbox
information on sports	3	MCQ
Student opinions	10	Likert scale

General information

This section was consisted of 4 questions directed towards demographic information of the students. Another 4 other questions were included about the degree program of the students. Due to the variations of the degree programs among students, degree programs of the undergraduates were captured through these 4 questions. Each respondent of the questionnaire has to answer at most 2 questions out of these four. For the analysis answers of these questions were merged to capture the degree program of a student.

Information on academic activities

This section included 3 multiple choice questions which are related to academic progress of the students. Since response for these three questions were marked as necessary, none of the variables computed using them included missing values.

Information on sports career

This section was available only for the students who claimed to be sportspersons.

6 MCQ questions were directed to measure their contribution and progress of sports activities.

Using the data collected by the question ‘What are the sports you are engaged in?’, all the sports played by each sportsperson were captured. Using the sports engaged by each sportsperson, 2 new variables were created.

1. Number of sports
2. Type of sports

   According to the Centers for Disease Control and Prevention (CDC) and the World Health Organization (WHO), all physical exercises can be categorized into 3 levels considering their intensity.

   i. Light Activities

      Activities that require the least amount of effort.

      Ex: golf, yoga, chess, carrom etc.

   ii. Moderate Activities

      Require some kind of effort to make the person breath harder

      Ex: Table Tennis, Cycling, Swimming, etc.

   iii. Vigorous Activities

      Activities that require higher efforts for practicing. These activities elevate heart rate and require higher muscle strength

      Ex: Football, Volleyball, Martial Arts, etc.

According to the sports they were engaged in each sportsman was categorized into one of these three levels. When assigning values to the variable for students who engaged in more than one sport activity, the sport at the highest intensity level that they engaged in was considered when designing the level.

Opinions of the students on academic vs sports relationship

For 10 of Likert scale questions, response of all the students were marked necessary. Their responses were recorded from sting to numeric values according to the slandered order of Likert scale charts.

5–point Likert scale	1 - Strongly disagree 2 - Disagree 3 - Neutral 4 – Agree 5 – Strongly Agree

4.3 Variable Description

From the 441 random students selected for the sample, only 271 students responded to the survey. Considering the answers to the questionnaire, the dataset for analysis consisted of 29 variables. Names of the variables and their characteristics are shown in table 4-3.

Table 4-3 Variable Description

Variable name	**Description**	**Scale**	**Categories**
Gender	gender	Nominal	Male/Female
District	District of residence	Nominal	Ampara/ Anuradhapura/ Badulla/ Batticaloa/ Colombo/ Galle/ Gampaha/ hambanthota/ Jaffna/ Kalutara/ Kandy/ Kegalle/ Kilinochchi/ Kurunegala/ Mannar/ Matale/ Matara/ Monaragala/ Mullativu/ Nuwara Eliya/ Polonnaruwa/ Puttalam/ Rathnapura/ Trincomalee/ Vavuniya
Ethnicity	Ethnicity	Nominal	Sinhala/ Tamil/ Muslim/ Burghers
Academic_year	Current academic year of the student	Ordinal	1/ 2/ 3/ 4
Subject_stream	Subject stream followed by 1st & 2nd year students	Nominal	Physical Science/ Biological Science/ Industrial Statistics and Mathematical Finance/ Biochemistry and Molecular Biology
Degree_program	Type of degree program followed by 3rd and 4th year students	Nominal	3-year General degree program/ Honours degree - Research oriented/ Honours degree - Industrial oriented
Degree_research	If engaged in a research-oriented degree program, name of the degree	Nominal	23 research-oriented degree programs offered by FOS, UOC
Degree_Industrial	If engaged in an industrial-oriented degree program, name	Nominal	9 industrial-oriented degree programs offered by FOS, UOC

	of the degree		
Sportsman_yes	Whether student is a sportsperson or not	Nominal	Yes/ no
Sport_places	Places that students engage in sport activities	Nominal	University/ Outside/ Both
Sports_names	Names of the sport activities	Nominal	25 sports granted by the university
S_num	Number of sports done by a student	Discrete	
S_type	Type of the sport according to the intensity level	Ordinal	light/ moderate/ vigorous
Sports_howlong	How long a student has been participated in sports activities within the university	Ordinal	0-1 years/ 1-2 years/ 2-3 years/ More than 3 years
Sport_hours	Time spent on sports activities in hours (per week)	Ordinal	0 - 5 hours/ 5-10 hours/ 10-15 hours/ 15-20 hours/ More than 20 hours
Pre_sport	Have the sportspersons engaged in sports before came to the university or not	Nominal	Yes/ No
Awards	Whether students could achive any awards or wining for their sports activityes	Nominal	Yes/ No
Awards_level	If awards were won, level of the awads	Ordinal	Inter faculty freshers/ Inter faculty/ Inter University/ District/ Provincial/ All Island
GPA	Current Grade Point Average Value.	Ordinal	0.0 - 2.0/ 2.0 - 2.5/ 2.5 - 3.0/ 3.0 - 3.3/ 3.3 - 3.5/ 3.5 - 4.0
Academic_time	Time spent on academic works in hours	Ordinal	less than 2 hours/ 2- 5 hours/ 5 - 8 hours/ 8 - 12 hours/ More than 12 hours

Diploma	Whether students engaged in external academic programs other than university degree program	Nominal	Yes/ No
O_1	Sports can cause negative performance in academic activities due to harmful effects on general health.		Strongly agree/ Agree/ Neutral/ Disagree/ Strongly disagree
O_2	Sports can cause fatal injuries which affects the studies.		
O_3	Sports might have improved immunity to diseases which affects the studies.		
O_4	Sports participation improves self-esteem which improves their confidence toward academic work.		
O_5	Sports reduce depression that can enhance academic performance		
O_6	Participation in sports activities is time-consuming for academic work.		
O_7	Sports participation negatively affects attendance at lectures		
O_8	Participation in sports negatively affects the completion of assignments.		
O_9	Achievements from sports will help to find better career paths		
O_10	Engaging in sports activities affect positively on Academic performance of a student.		

4.3.1 Data transformation for advanced analysis - Part 1

For identification of the factors that motivate sportspersons towards sport activities, model building and feature selection methods were carried out in the analysis. For this analysis, four variables were included.

I. Gender
II. Ethnicity
III. District
IV. Academic_year
V. Degree_program

Degree_program variable was created by concatenating subject_stream and degree_program variables. When creating the degree_program, difference in 1st and 2nd year degree programs vs 3rd and 4th year degree programs was neglected.

4.3.2 Data transformation for advanced analysis - Part 2

For the identification of the factors contributing to the GPA value of a sportsperson, machine learning techniques were applied for model building and feature selection. For this, data gathered under the 'sports information' section was used. Since these data were only available for sportspersons, the dataset for the analysis consisted of only 71 students who were sportspersons.

Out of 9 variables regarding sport activities 7 variables were chosen for the analysis

Name of the sports variable was dropped after converting the variable to number of sports and type of the sports variable. Variable 'Awards' was also removed after combining it within the Awards_level variable.

I. Sport_places
II. S_num
III. S_type
IV. Sports_howlong
V. Sport_hours
VI. Pre_sport
VII. Awards_level

Chapter 5

Results

The number of respondents for the survey is 271. In this chapter, the data collected from these respondents have been analyzed to answer the research questions of the project. As the first step, univariate analysis has been carried out to understand the variables we are dealing with and to check whether the sample represents the target population. For the visualization and descriptive analysis, IBM's statistics package was used.

5.1 Exploratory Data Analysis

Among the 271 respondents, there are 139 males and 132 females. According to the student records in FOS, the proportion of male to female population within the Faculty of Science is similar to this which indicates the sample is a proper representation of the population considering gender status. (Figure 5-1)

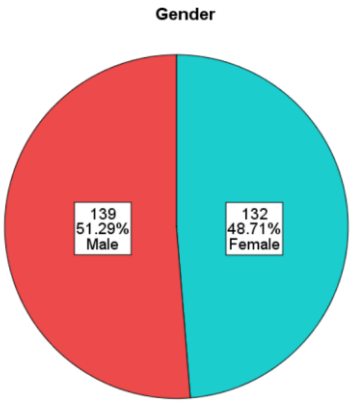


Figure 5-1Gender of Students

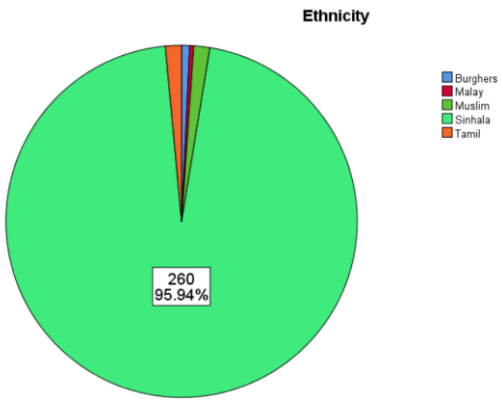


Figure 5-2 Ethnicity of Students

When considering Figure 5-2, over 95% of the sample who answered the questionnaire was Sinhalese. This might be because the lower number of students belongs to other ethnicities within the faculty or a biased sample towards Sinhala people.

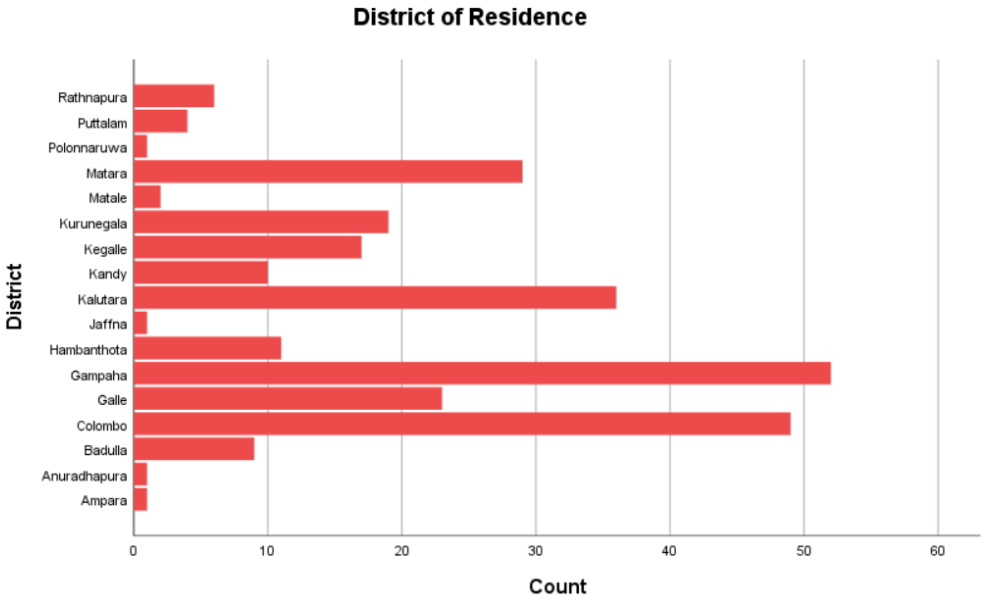


Figure 5-3 District of Residence

The sample consists of students from 17 districts out of 25 and the majority is from the Western Province. According to the chart in Figure 5-3, Colombo and Gampaha have the highest count of respondents.

When considering the academic year, the majority of the students who responded to the survey were level 1 students. When considering all 4 years distribution of science faculty students according to their academic year is shown in table 5-1.

Table 5-1 Distribution of academic year in the sample

Academic level	**No. of respondents in the sample**	**Sample proportion**	**Population proportion**
1	76	28%	30%
2	65	24%	28%
3	61	22.5%	24%
4	69	25.5%	17%

According to table 5-1, it can be identified that the number of respondents from each year is approximately equal. According to the population proportion, the number of students is inversely proportional to the academic year. When compared with the sample, the first, second, and third years follow the same pattern as the population size. Even though 4th-year students have an increased rate of responding that exceeds levels 1 and 2.

When considering the educational backgrounds of respondents, physical science students have greatly contributed to the study among the students who are in 1st and 2nd academic years. When considering the UOC FOS student population, the number of biology students and the number of physical science students are approximately equal. And also, Industrial statistics and Microbiology students have an approximately equal number of students which is less than half of the students in physical science or Biology subject Streams. Considering these facts, the collected sample of students seems biased towards the data from physical science and Industrial statistics students.

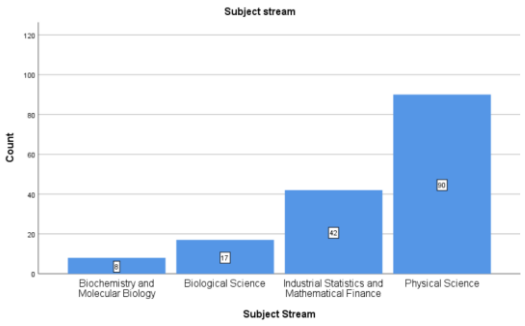


Figure 5-4 Subject Stream

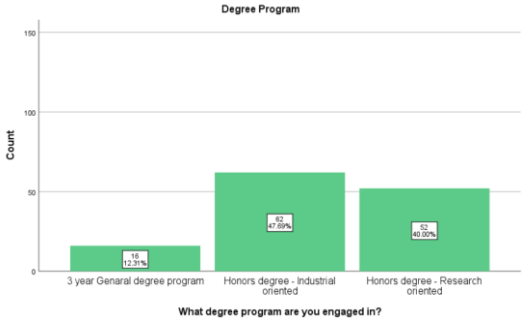


Figure 5-5 Subject Stream

According to figure 5-5, out of 3rd and 4th-year students, more than 85% follow honours degree programs. According to the UOC FOS academic status, this percentage is accurate considering the population. When considering these 85% honours students, the counts of students from each degree program are shown in table 5-2 and table 5-3.

Table 5-2 Students count of each Degree Program – Industrial Oriented

Degree program	**Sample count**
Applied Statistics	16
Business & Environment	10
Electronics & IT	9
Finance & Insurance	10
IT & Management	9
Molecular biology & Biotechnology	1
Science & Management	7

Table 5-3 Students count of each Degree Program – research Oriented

Degree program	**Sample count**
Applied Mathematics	1
Biochemistry & Molecular Biology	1
Bioinformatics	1
Chemistry	3
Computational Chemistry	4
Computational Mathematics	2
Computational Physics	3
Engineering Physics	3
Finance, Business and computational Mathematics	2
Immunology and Integrative Molecular Biology	1
Industrial Statistics	4
Mathematical Finance	2
Mathematics	2
Mathematics & Statistics with Computer Science	3
Nuclear Medical Science	1
Physics	5
Plant Bio technology	1
Plant Science	1
Statistics	6
Statistics with Computer Science	6

According to table 5-2 and table 5-3, it can be identified that there is at least one student from each degree program within the sample. Even though some degree programs have a higher or lower count, a selected sample could be identified as a reasonable one considering undergraduates from each department have a presence within.

When identifying the academic performance of a university student, the grade point average (GPA) value, which indicates the average score of all the examinations, is a highly important measure. Hence, the GPA value was considered the main response variable that determines the academic performance of a student.

Distribution among GPA values of the science faculty undergraduates is shown in the figure 5-6.

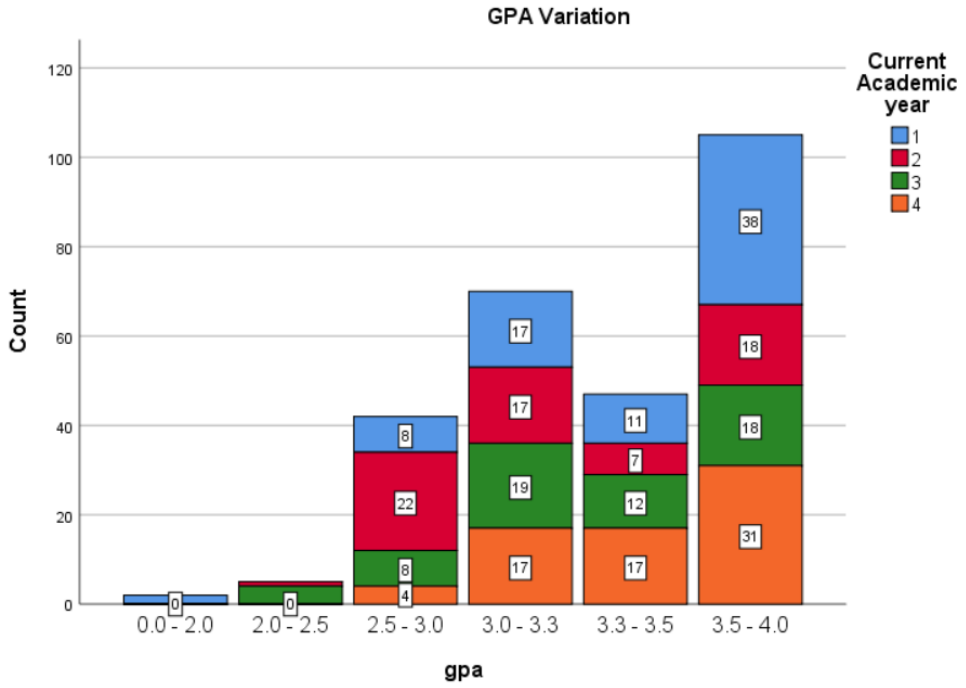


Figure 5-6 Distribution of GPA value

Among the 6 groups that were introduced to represent the GPA values, it can be identified that almost all the students in FOS, UOC have achieved good GPA scores. 40.2% of the undergraduates have scored above 3.5 GPA while another 17.2% have scored above 3.3 GPA.

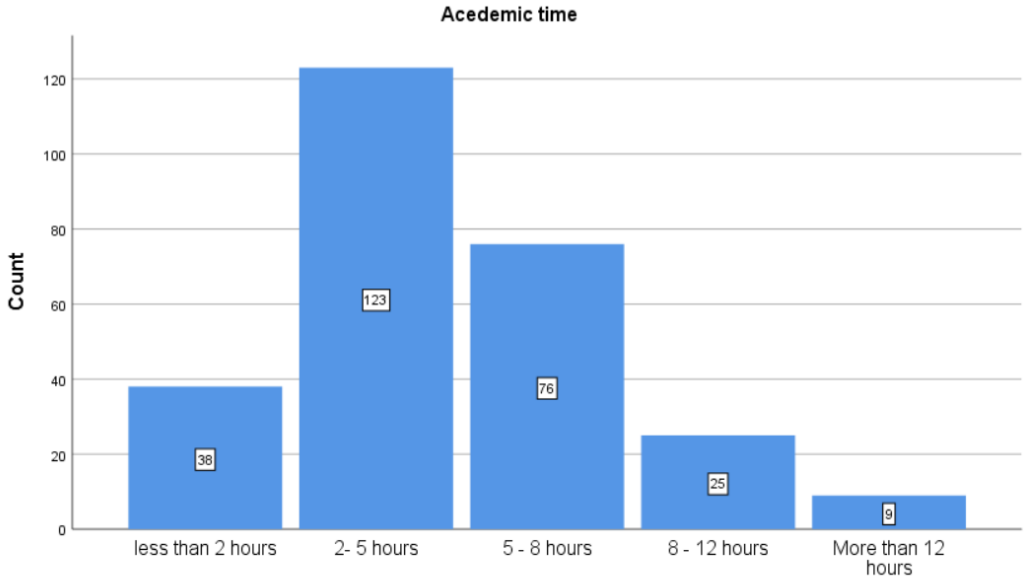


Figure 5-7 Distribution of academic time

When considering the contribution of a student towards academic activities, the time spent on academic activities can also be taken as a good measure. By observing the above graph in figure 5-7, it can be identified that the majority of students spent 2-5 hours on their academic activities per day. There are 4.6% of students who spend less than 2 hours per day on academic activities, which can be an issue when considering the toughness of university education. Even though approximately 10% of students spent 8-12 hours and There are 3.4% of students are committed to academic activities by spending more than 12 hours per day on academic activities.

Also, there is 16.24% of students engage in additional academic programs such as diplomas, external degree programs, etc.

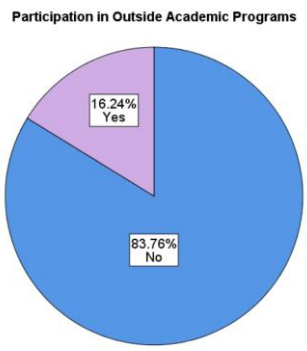


Figure 5-8 Participation in external academic programs

Sportspersons

Out of 271 respondents, 71 students claimed that they are currently engaged in sports activities parallel to their academic work. That indicates approximately 26% of UOC FOS students are sportspersons.

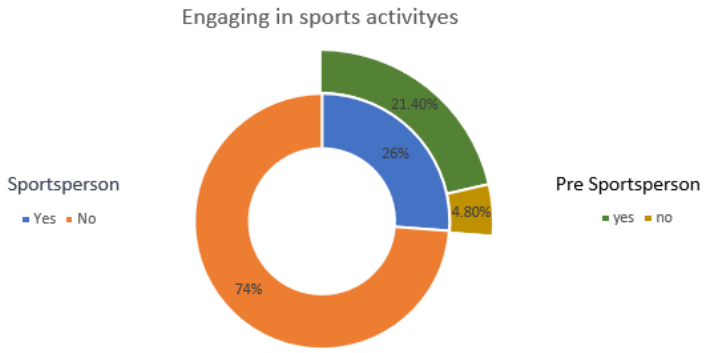


Figure 5-9 Sportspersons vs pre-sportspersons

According to figure 5-9 it can be identified that the majority of students have been engaging in sports activities before coming to the university. But, 18.3% of sportspersons started their sports careers after they came to the university. These new sportsmen are an indication of the encouragement given by the university to the undergraduates to engage in sports activities parallel to their academic activities.

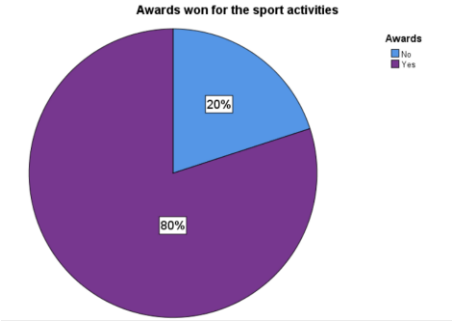


Figure 5-11 Awards won for sports

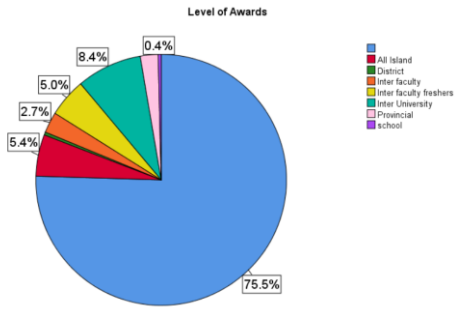


Figure 5-10 Maximum Award level achieved

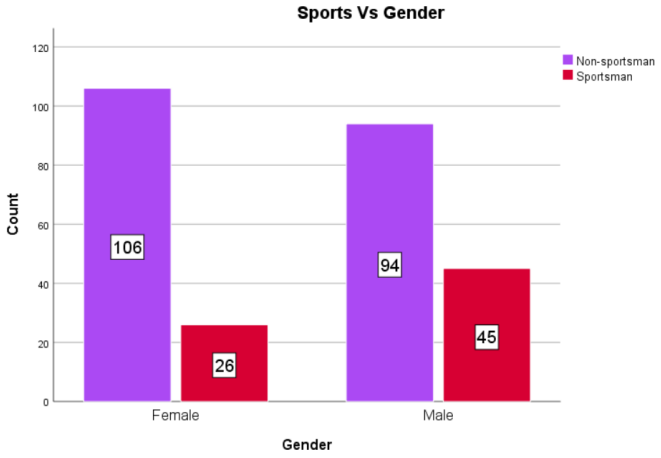
It was also observed that, 80% of sportsman have won awards for their academic careers within the university. According to figure 5-11 out of award winners, 16.1% have achieved awards at the university level and 5.4% at all Island levels. This shows a significant chance for a university student to achieve All-island awards.

5.1.1 Bivariate Analysis

Association of Demographic variables and sport participation

This section is dedicated to identifying the significant factors that lead a university undergraduate to engage in sports activities. Variables that were collected through the survey that are assumed to be relevant to sportsmanship are gender, academic year, ethnicity, and degree program. These factors were selected according to the literature and public opinions of the society. Other than visual representation, analysis techniques such as the independent sample t-test and the Chi-square test of association were used in the analysis.

Association with Gender.



Chi-Square test results

H_0: No Association
H_1: Association

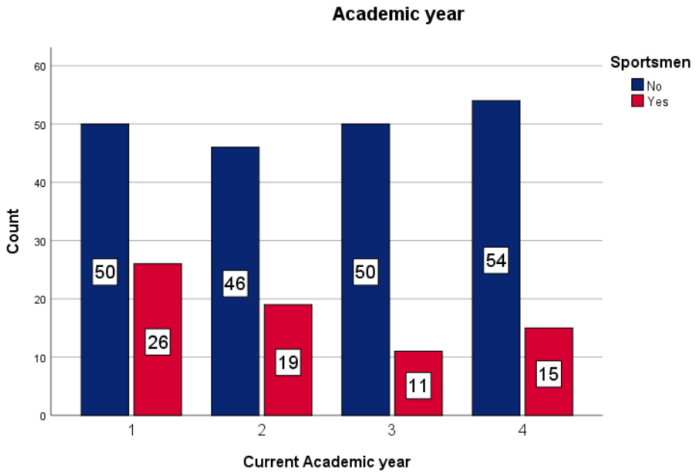
Test Value: 5.627
d.f.: 1
Significance: 0.018

Figure 5-12 Sports Vs Gender

Figure 5-12 represents the distribution of the count of sports and non-sports people. Proportion of male students who were also sportspersons is higher than the proportion of female sportspersons. This indicates male students are more likely to engage in sports activities when coming to the university. Using a study based in 37 countries, it has been proven that a male is 3.7 times more likely to engage in sports activities than a non-competitive activity. (Deaner et al., 2016)

When considering the significance of gender for a person to engage in a sports activity, the chi-square test of independence suggests that there is a significant relationship between gender and sports engagement. Overall, it can be concluded that being a male significantly increases the chance for a student to be a sportsman.

Association with Academic year



Chi-Square test results

H_0: No Association
H_1: Association

Test Value: 5.646
df: 3
Significance: 0.13

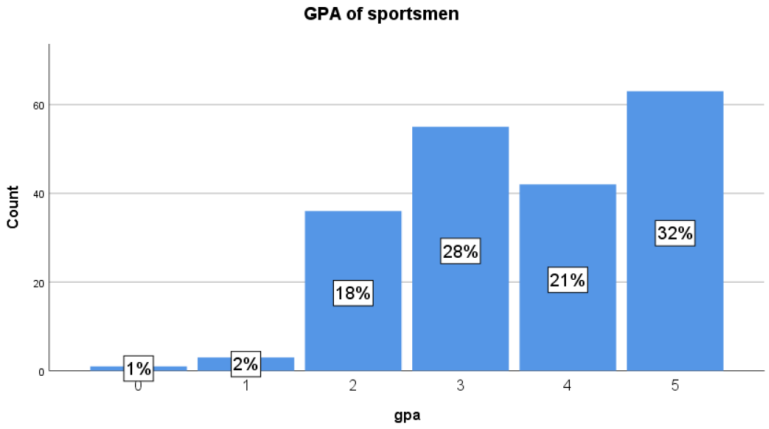
Figure 5-13 Association of Academic year

The bar chart in Figure 5-13 indicates the distribution of sportspersons vs. non-sportspersons according to their academic year. 1st and 2nd-year participation in sports is more than 30% of the number of students in each year. Compared to this, 3rd and 4th-year students have lower participation in sports. This might be caused by the increase in the academic workload of the final academic years, which causes students to leave their academic work. The COVID pandemic, which will lead to online education in 2021–2022, has led to a reduction in sports practices in every university in the country. Considering the above fact, it is expected to have a lower number of athletes among 2nd and 3rd-year students. By observing the above graphs, it can be seen that even though the pandemic and country lockdown have affected the sports activities of university students, they have recovered in a short period of time.

Even though these differences can be observed, when checking the similarity of proportions using Pearson's chi-square test for independent samples, it indicates there is no significant difference among the proportion of sportspersons overall.

Association of sports activities and academic activities

Sports vs GPA



GPA

0 : 0.0 - 2.0

1 : 2.0 - 2.5

2 : 2.5 - 3.0

3 : 3.0 - 3.3

4 : 3.3 – 3.5

5 : 3.5 – 4.0

Figure 5-14 GPA distribution of sportspersons

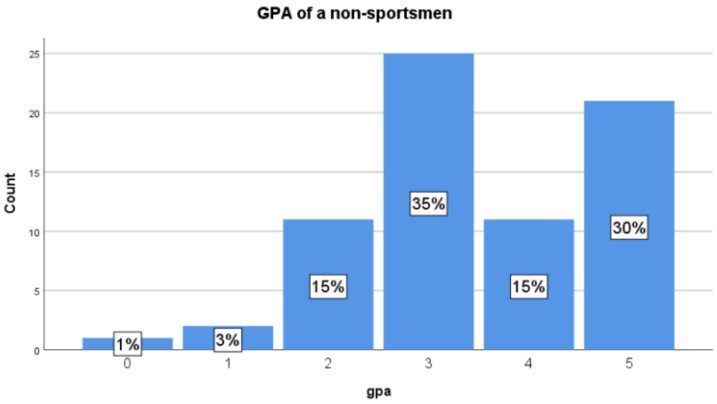


Figure 5-15 GPA distribution of non-sportspersons

By observing variation of GPA values among sportspersons and non-sportspersons, according to Figure 5-14 and Figure 5-15, all students have been performed well at their academic works. Both graphs in figure 5-14 and figure 5-15 have skewed left towards higher values. Even though both charts are approximately similar in visual appearance, percentage of sportspersons who score over 3.3 GPA is 53% while percentage of non-sportspersons who score over 3.3 is 45%. According to this it can be identified that sportspersons have been comparatively good at achieving high GPA values compared to non-sportspersons.

Table 5-4 chi square test results: sports vs GPA

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)
Chi-Square	3.199[a]	5	.669
Likelihood Ratio	3.092	5	.686
Linear-by-Linear Association	.552	1	.458
N of Valid Cases	271		

.

According to Table 5-4, with a significance value of 0.669, the Chi square test of independence suggests that there is no significant relationship between GPA and being a sportsperson overall. Since the GPA values of students vary according to the academic year as shown in table 6, the association between GPA and being a sportsman was analyzed separately using chi-square test results.

Table 5-5 chi square test results: sports vs GPA (Considering academic year)

Academic year	Chi square test value	d.f	significance
1	4.241	4	0.374
2	0.624	4	0.960
3	4.222	4	0.377
4	7.904	4	0.048

When considering the chi-square test results in Table 5-5, 1st, 2nd, and 3rd year students do not show a significant difference between GPA and doing sports activities. Even though it can be observed that 4th year students have a significant relationship between GPA and their engagement in sports activities, the most possible reason for this is the deference in academic works of a final-year student compared to those at other levels. Further analysis of this will be carried out under the advanced analysis section.

Other than GPA, there are factors such as time spent on academic activities that can be used to measure the academic performance of a student. Other than doing a sport activity, factors such as the number of sports, type of sports, and time spent on those sports can also be related to the academic performance of a student. To identify these, a series of comparison tests were used to compare the above academic progress measures with sport-related factors.

Chi-square test of independence

To observe the relationship between nominal vs ordinal categorical variables, chi-square test of independence was used.

H_0: There is no relationship between 2 categorical variables

H_1: There is a relationship between 2 categorical variables

Table 5-6 Chi-square test results

variables	Test statistic	p-value	conclusion
Sport vs academic time	9.651	0.047	Correlated
Sport type vs GPA	7.164	0.724	Independent
Sport type vs Academic time	9.674	0.290	Independent
Sport hours vs academic time	9.682	0.021	Correlated

Based on the results from Table 5-6, it was identified that sports are correlated with time spent on academic activities. As the probable reason for this could be the time spent on sports activities forced students to reduce academic time considerably. Even though, as previously identified, sportspersons somehow have been able to maintain a GPA value similar to that of non-sportspersons. Type of the sport does not show any significant relationship to the academic activities of a sportsperson. Also, the time spent on academic

activities has a correlation towards time spend on academic activities. This can be considered as a possible relationship since sportspersons have to balance their time between academic activities and sports.

For the selected 2 related variables Phi coefficient of correlation have been calculated to measure the direction and magnitude of the relationship. According to Table 5-7 time spent on academic activities have a moderate negative relationship with the time spent on sports activities.

Table 5-7 Phi Coefficient for Categorical variables

Variables	Phi coefficient	Relationship
Sport vs academic time	-0.112	Week negative relationship
Sport hours vs academic time	-0.360	Moderate negative relationship

Independent sample Kruskal-Wallis test

H_0: There is no correlation between two variables

H_1: There is a correlation between two variables

Table 5-8 Kruskal-Wallis test results

variables	p-value	conclusion
No of sports vs GPA	0.177	Uncorrelated
No of sports vs Academic time	0.038	Correlated
Sport hours vs GPA	0.841	Uncorrelated

Number of sports does not affect the academic activities of a FOS, UOC undergraduate in overall. As a reason for this it can be assumed that some sportspersons have time management issues when they engaged in multiple sports activities

5.2 Likert Scale Data Analysis

Likert-scale questions in the research questionnaire were based on the most debatable opinions of the society on the relationship between sports and academic activities. Since the sample used in the survey is identified as a good representation of the Faculty of Science, University of Colombo (FOS, UOC), an analysis of the overall opinions of the sample, aside from the fact of whether or not participants are engaged in sports, could possibly interpret the general idea of FOS, UOC, on this relationship. Conclusions on these were based on the median of the responses.

Table 5-9 Likert Scale Analysis- Overall

Var. Name	Variable Description	median	mean	s.d	Conclusion
O1	Sports can cause negative performance in academic activities due to harmful effects on general health	2	2.13	0.955	Disagree
O2	Sports can cause fatal injuries which affects the studies.	3	3.12	0.963	Neutral
O3	Sports might have improved immunity to diseases which affects the studies.	4	3.40	1.002	Agree
O4	Sports participation improves self-esteem which improves their confidence toward academic work.	4	4.03	0.883	Agree
O5	Sports reduce depression that can enhance academic performance.	4	4.11	0.855	Agree
O6	Participation in sports activities is time-consuming for academic work.	3	3.18	0.903	Neutral
O7	Sports participation negatively affects attendance at lectures.	3	3.07	0.894	Neutral
O8	Participation in sports negatively affects the completion of assignments.	3	2.96	0.930	Neutral
O9	Achievements from sports will help to find better carrier paths.	4	3.94	0.907	Agree
O10	Do you think engaging in sports activities affect positively on Academic performance of a student?	4	3.84	0.909	Agree

When considering the overall opinion of University of Colombo, Faculty of Science students, the majority disagree with the notion that sports can have a negative effect on general health that may cause drawbacks in academic activities.

According to table 5-9, for the fact that sports might cause fatal injuries, the responses indicate that the opinions of the students are neutral. This may be due to the unawareness of such events in sports.

And also, for the other factors of sports that might negatively affect academic activities, such as time required for sport activities, lower attendance to the lectures, and being unable to complete assignments in time, have a majority vote for neutral. This indicates that even though the majority of students do not engage in sports activities, they also do not have a good understanding of the negative effects of sports.

But students have agreed with the fact that sports can improve mental health by improving self-esteem and reducing depression through sports. They also agree on the fact that sports might improve immunity for diseases in students, which can be a support to studying by staying healthy.

According to the students, they also have the opinion that achievements from sports activities might be helpful in finding better career paths in the future. Since every student is focused on their future career, this is a crucial factor that could affect students willingness to engage in sports.

Table 5-10 Likert Scale Analysis- Sportspersons vs Non-Sportspersons

	Non-sports		sports		Test Scores	
	Mean	s.d	mean	s.d	Test stat	significance
O1	2.16	0.932	2.04	1.020	3.238	0.519
O2	3.08	0.942	3.23	1.017	6.219	0.183
O3	3.37	1.023	3.49	0.939	1.281	0.865
O4	3.97	0.885	4.21	0.860	7.029	0.134
O5	4.05	0.813	4.30	0.947	14.933	0.005
O6	3.25	0.861	2.99	0.993	9.203	0.044
O7	3.16	0.871	2.82	0.915	12.839	0.12
O8	3.07	0.871	2.66	1.027	19.380	0.001
O9	3.86	0.924	4.15	0.822	7.517	0.111
O10	3.73	0.850	4.15	0.995	30.134	0.000

According to Table 5-10, the opinions of sportspersons and non-sportspersons do not differ significantly in most aspects.

Even though both sportspersons and non-sportspersons agree that ‘sports reduce depression, which helps to improve academic performance’, sportspersons have a significantly higher opinion on that.

Similarly, even though sportspersons also agree that sports are time-consuming, non-sportspersons have a higher opinion towards the time-consuming nature of sport activities.

When considering the completion of assignments, while sportspersons state sport does not affect the completion of the assignments, non-sportspersons believe sports have a negative effect on the completion of the assignments.

Considering sportspersons have the experience of the time they spent on sport activities and the completion of assignments while engaging in sports, it can be said that non-sport persons have a misleading opinion of these negative effects.

In overall both sportspersons and non-sportspersons states, the effect of sports might be positive towards academic works. This positive feedback from sportspersons is significantly higher compared to non-sportspersons.

5.3 Summary of the descriptive Analysis

- Sample can be considered as a good representation of the target population. Only the degree type of the students has biased towards physical science and Industrial statistics students. 26% of the students in the sample was identified as sportspersons.
- Gender shows is significant impact on sports, indicating males are more likely to engage in sports activities.
- Considering the GPA value of a students, Sports does not have a significant impact on academic activities. Even though, when comparing against academic year, 4^{th} year students have a significant effect on academic activities while students in other academic levels do not.
- Engaging in a sport, Amount of time spent on sports and number of sports have affected on the academic time spent.
- Opinion of the students in overall says sports are positively affected on the academics while Sportspersons highly confident on the positive effects of sports and non-sportspersons have strong belief on some negative effects of the sports.

5.4 Advanced Analysis

This chapter includes the analysis carried out to achieve secondary objectives of the study. For identifying factors associated with being a sportsman and factors associated with GPA among sportspersons are discussed in this chapter. Built-in Python Libraries were used in this part of the analysis.

5.4.1 Model 1 -Factors associated with being a sportsperson

Being a sportsman is a choice an individual makes according to their opinions. For a university student, a load of academic works depending on their degree programs may be the factors that lead to this decision. Other than that, demographic factors such as age, ethnicity, and residence were identified as possible factors. To identify whether these factors are associated with being a sportsperson, classification models were fitted to the data. Advance machine learning techniques were used for the best performance in building classification models.

As the initial dataset to build the model, variables that were collected under the "general information section of the questionnaire were used. Six variables were selected for the model building.

Response variable: Sports_yes

Predictor variables:

```
0   Gender       271 non-null    object
1   District     271 non-null    object
2   Ethnicity    271 non-null    object
3   Ac_year      271 non-null    int64
4   Degree_Type  271 non-null    object
5   sports_yes   271 non-null    int64
```

As District variable contains 17 unique categories, this variable was transformed into a new variable province which contains only 9 variables.

To check whether these variables can create a successful predictive model, seven classification techniques were used, and their accuracy and f1 scores were compared against each other.

When considering multicollinearity between the factors using Cramer’s V statistic. None of the variables showed significant correlations among them. Even though variable

academic year showed Cramer's V statistic of 0.423 between academic year and Degree type which was the highest correlation recorded among factor variables.

Table 5-11 Training accuracies of models – Model 1

	Accuracy	**Precision**	**Recall**	**F1**
Logistic Regression	0.708	0.30	0.11	0.156
Lasso	0.724	0.21	0.02	0.037
KNN	0.717	0.33	0.21	0.255
Random Forest	0.708	0.42	0.29	0.305
Decision Tree	0.713	0.47	0.33	0.401
SVM	0.736	0.0	0.0	0.0
Gradient Boosting	0.704	0.38	0.24	0.290
Linear Discriminant Analysis	0.727	0.404	0.23	0.340

According to the results in table 5-11, it could be observed that each of these models has an accuracy of over 70%. Even though the precision and recall values are less than 50% for each model. To select the best model out of them, the F1 score was used. The highest f1 score was achieved by the decision tree algorithm, while Random Forest and Linear Discriminant Analysis also performed with a closer f1 scores.

Identification of the most important factors for students to be sportsmen was one of the main objectives of this study. Achieving this objective was done using feature importance algorithms.

Considering these factors, the decision tree algorithm which has the highest F1 scorewas selected as the best classification technique for prediction and identifying variables of importance for being a sportsperson.

Decision Tree

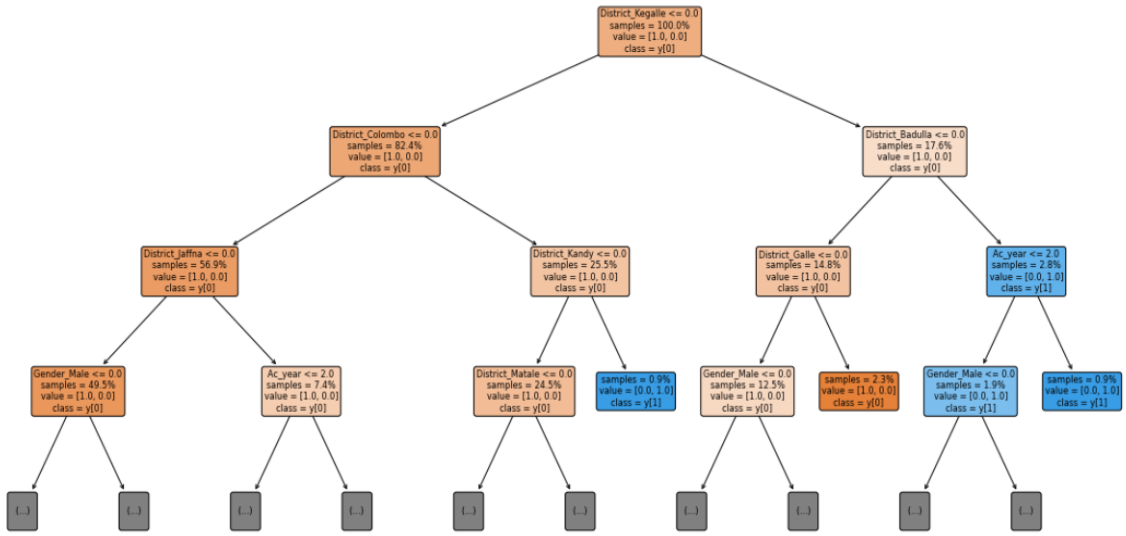


Figure 5-16 Decision Tree – Model 1

Confusion matrix

Table 5-12 Confusion Matrix -Model 1

		Actual	
		1	**0**
Predicted	**1**	36	5
	0	8	6

Accuracy = 0.76

Precision = 0.878

Recall = 0.818

The accuracy, precision, and recall values of the model are high, so it can be concluded that it has performed well in the classification.

Feature importance towards engaging in sports activities were captured through the model.

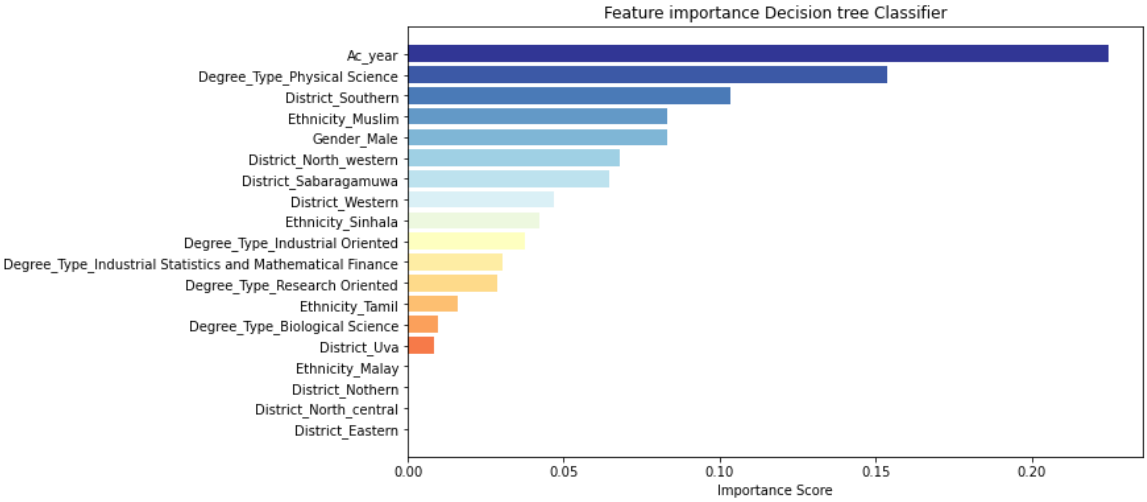


Figure 5-17 Feature Importance Chart – Model 1

Table 5-13 Feature importance - coefficients

Feature	**Importance**
Ac_year	0.2160
Degree_type_Physical_Science	0.1538
District_Southern	0.1099
Gender_Male	0.0914
Ethnicity_Muslim	0.083
District_North_western	0.0679
District_Sabaragamuwa	0.0579
Ethnicity_Sinhala	0.0579
District_Western	0.0579
Degree_type_Industrial_Oriented	0.0469
Industrial_Statistics_and_Mathamatical_Finance	0.0375
Degree_type_Research_Oriented	0.0304
Degree_type_Biological_Science	0.0286
District_Uva	0.0095

When considering the feature importance towards sports participation, academic year is the most important variable. Gender has played a greater role in sport participation, which

could also be identified through the explanatory data analysis. Other than these, the district of residence has also been an important factor in response. From all the districts, Kalutara, Galle, Hambanthota, Kandy, Colombo, and Kegalle show significant importance towards engaging in a sport. From degree types, Physical Science, Biological Science, and Industrial Statistics degree programs play a greater role.

Even though these factors were identified as being significant towards engaging in a sports activity, it must be taken into account that differences in sample sizes might affect the importance score.

5.4.2 Model 2 - Factors associated with the GPA of a sportsperson.

The GPA of a student might be affected by the level of effort and time spent on their sports career. This analysis tested factors that were assumed to be related to GPA value. For the prediction of the GPA of a student, machine learning models were fitted since these algorithms showed higher performance compared to the classical models in use. For this multi-level classification, which includes six categories in the response variables, six machine learning models were trained, and their f1 scores and accuracy were compared with each other.

Table 5-14 Training Accuracies of models – 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

The best f1 score of the models was detected through linear discriminant analysis. Since Decision tree algorithm also follows a closer f1 score of 0.29 both of these models were considered validating against the test data.

Table 5-15 Testing Accuracies of models – Model 2

Model	MSE	Test accuracy	F1
Decision tree	2.16	0.26	0.25
Linear discriminant Analysis	4.46	0.26	0.22

Considering the lower mean squared error and higher F1 score, the decision tree algorithm was selected as the best classifier in this problem.

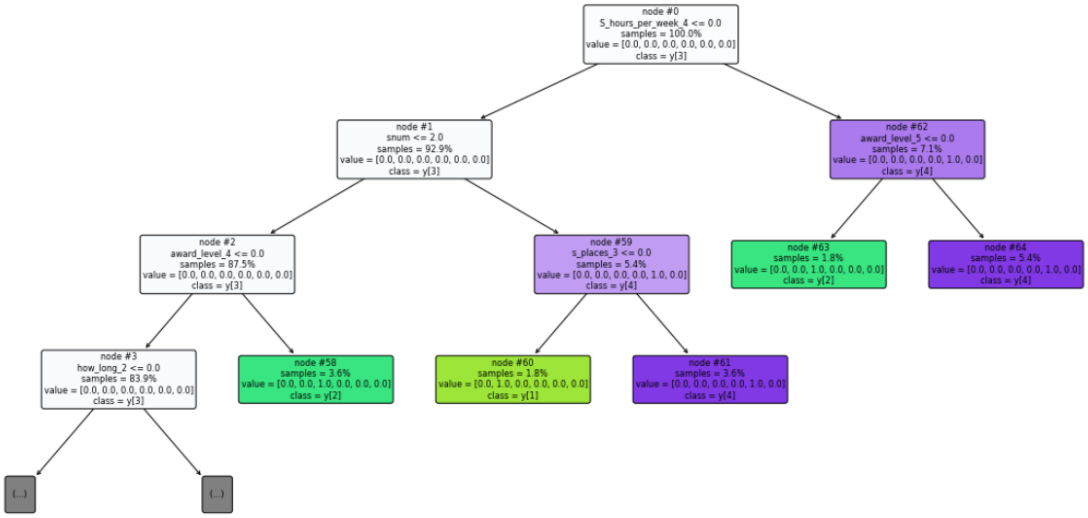


Figure 5-18 Decision Tree - GPA of a Sportsperson

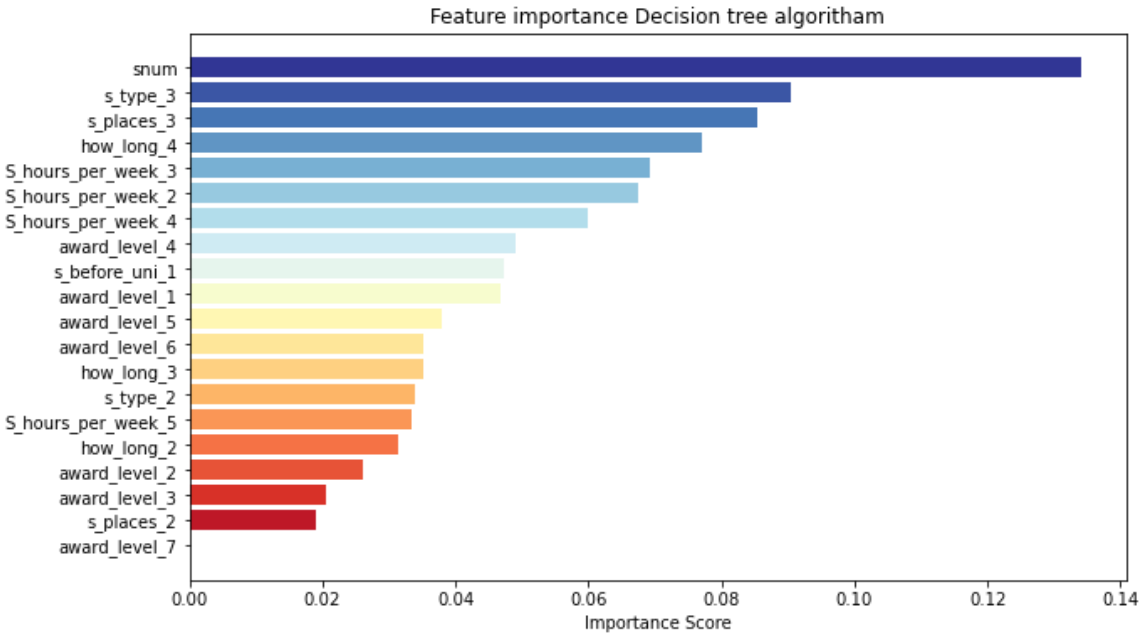


Figure 5-19 Feature Importance Chart – Model 2

Considering feature importance, the number of sports is the most important variable included in the model. The type of sports considered when using the intensity level of the sports was also identified as an important factor for the GPA. Other than that, students who were engaged in sports for more than 2 years (among 3rd and 4th year students) had a significant effect on their GPA scores. The next important factor in determining the GPA value is the number of hours spent on their sports activities. Students who spent more than 15 hours per week have significantly lower GPA scores.

5.5 Summary of the Advanced Analysis

- When considering feature importance towards sports participation, academic year and Gender plays a significant role. Considering Degree programs, doing a physical science degree program is an important factor that decide a student to be a sportsperson. Residence province such as Southern province, North Western province and Sabaragamuwa also have a significant impact in engaging in sport activities.
- Number of sports have a high impact on GPA value of a sportsperson. Even though sport type was considered as an insignificant variable towards academic performance in overall, engaging in a vigorous sport activity also affected on the GPA of a student.

Chapter 6

Discussion and Conclusion

6.1 General discussion

This chapter discusses the major findings and conclusions that can be drawn from the overall analysis of this study. Furthermore, this will give some suggestions for further studies. The objective of this study is to identify the association between academic progress and sports participation.

Exploratory Data Analysis

Considering gender and district of residence, the collected sample was a good representation of the population. Even though the sample consists of a very low number of students from ethnicity groups other than Sinhalese, it was identified that the number of non-Sinhalese people in the FOS and UOC is lower. When considering the academic year, the fourth-year representation of the sample was quite large compared to the expected proportion of the fourth year, which is 17%. Resampling the 4th year's student population by 21 could accomplish this. Since the number of athletes that could be included in that proportion of students is important for model building, resampling wasn't used to balance the dataset.

When the degree programs of the students were taken into account, the responses from the students in biological science were lower than expected. Considering 3rd and 4th year students participation in the sample, it was approximately similar to the expected values. Also, at least one student from each degree program has contributed to the study. In general, it can be concluded that the sample can be considered a good representation of the target population.

When factors that affect academic performance were looked at, it was found that FOS and UOC undergraduates do better than the average. Mode of the GPA categories was the GPA range of 3.5–4.0, and time spent on academic activities daily was more than 2 hours for more than 75% of the sample.

The number of sportspersons in the sample was 71, which indicates that more than 15% of the students in FOS at UOC are engaged in at least one sport activity. Also, 4.8% of the

students have started their sports careers after coming to the university. For a student to begin a sports career from the beginning, it can be assumed that the university has provided excellent sports facilities to students.

Association of Demographic variables and sports participation

Through Chi-square test of independence, it was identified that gender is significantly associated with the sports participation of students with a significance of 0.018. Even though a decrease in sport participation can be observed through the bar chart in Figure 5-12, the Chi square test of association suggests there is no significant association between academic year and sports participation.

Using the visualizations and test results regarding the association between sports activities and academic activities, it was identified that engaging in sports has no significant effect on the GPA value. Both athletes and non-athletes have been able to achieve similar GPA scores. Even though effort towards sports activities remains the same for any year within the university, academic activities change significantly under each year. Considering this fact, a chi-square test was used to check the connection between sport participation and GPA value for each year. It was identified that even though sports activities in the first three years don't affect GPA values of the students, they have a significant effect on the fourth-year undergraduates.

By comparison of every possible pair between 2 academic measures, which were GPA and academic time, with the sport-related variables, which were sport type, number of sports, and time spent in hours. Engaging in a sport vs. academic time has weakly negative correlation between the variables while and time spent on academic activities vs. time spent on sports, has a significantly high negative correlation among them.

Likert Scale Analysis

In overall students agree that sports can improve immunity, improve self-esteem and reduce depression which helps to improve academic performance. Also, Students believe that Sports helps to find better carrier paths in the future. With a mean value of 3.84 students agree with the fact that sports effect positively on the academic performance of a student.

While the opinions of sportspersons and non-sportspersons do not differ significantly on most factors, sportspersons have a higher opinion of the fact that "sports help reduce

depression" compared to non-sportsmen. Similarly, even though both groups agree that sports improve academic activities, sportspersons have a higher opinion on that. On the other hand, non-sportspersons believe academic activities are time-consuming and might have a negative effect compared to sports.

Advanced Analysis

Factors associated with being a sportsperson

Considering accuracy and the F1 score the decision tree algorithm was selected as the best model for prediction to identify the importance of the feature of being a sportsperson. The decision tree predicted the response with an accuracy of 0.745 and an f1 score of 0.42, which can be considered as good prediction accuracy.

Even though it was identified that the model was trained on too few predictor variables. By adding more variables related to being a sportsperson, we may have increased the performance of the model.

Through feature importance the academic year was selected as the most important variable for sports participation. Gender also has a high importance score, indicating it is one of the most important factors in sport engagement. This factor was also identified through exploratory analysis. Other than this degree program, there was also another important factor towards sport engagement.

Factors Associated with the GPA of a sportsperson

Through the comparison of six machine learning algorithms, the decision tree was selected as the best classification technique to predict the GPA of a sportsperson. With a 0.25 f1 score, this model indicates that even though it performed better among the used algorithms, the prediction strength of the algorithm is quite low. This could happen due to the lack of data for the algorithm. By increasing the sample size, a better model could have been created.

Through the feature importance number of sports, type of sport was identified as the most important variable. Students who spent more than 2 years on sports activities within the university also play an important role in predicting their GPA values.

6.2 Conclusions

- Considering the prevalence of sport activities, 26% of the University students in FOS at UOC are engaged in any kind of sport, which can be taken as a good involvement rate for the sport activities.
- In overall there is no significant relationship between participating in a sport activity and GPA value.
- Sports have a significant effect on the time spent on academic activities. Student who can balance the sports with their academic time can overcome negative effect of sports on academic works.
- Final-year students have a significant negative impact on sports activities.
- Academic year and gender were identified as the most significant factors that influence students' participation in sports activities.
- The number of sports they are engaged in have significantly affected their GPA scores.
- Considering the opinions of students, FOS, and UOC undergraduates believe, the impact of sports on academic progress is positive while sportspersons much strongly believe in that fact.
- Considering the negative effects of sports, such as high time consumption, non-sportspersons have a higher expectation that they might affect academic activities negatively. These negative opportunities might also be a reason for non-sportspersons to refrain from engaging in sports activities.

6.3 Limitations of the study

The research was based on a simple survey that included the most necessary questions to answer the research questions. The number of variables used in the analysis was not sufficient to conduct a good statistical model.

GPA, which is the response variable, was measured as a categorical variable as a support for the respondents. The accuracy of the prediction of GPA values has been affected by this.

The sample of the sportspersons, which was 71, was quite low to perform a good prediction model on their GPA values.

6.4 Suggestions for future works

The accuracy of the predicted model can be increased by increasing the sample size.

Additional questions can be asked of the respondents during the data collection process to include more predictor variables for model training.

Other feature selection methods could be carried out for the variable selection.

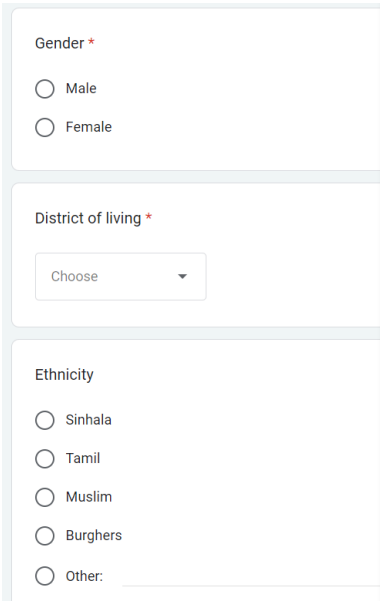
Expanding the study to whole university can be used to determine the sports versus academic relationship in a university student in overall.

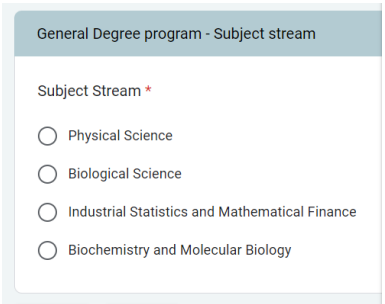
Chapter 7

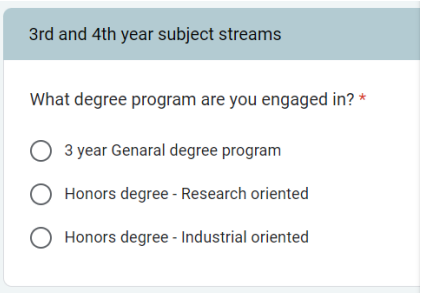
Appendices

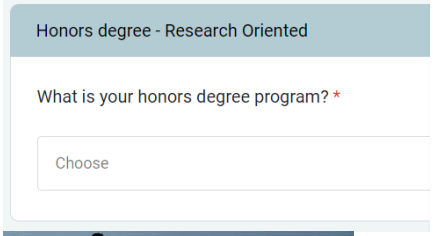
Appendix A : Survey Questionnaire

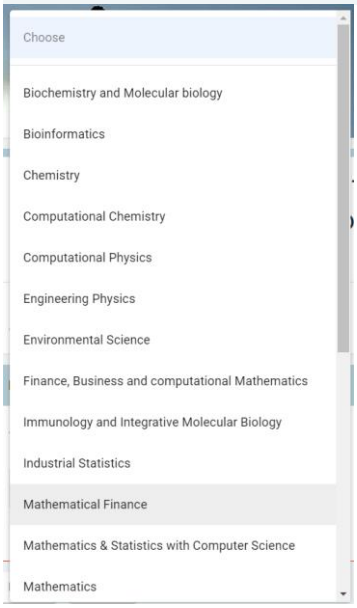
Section 01

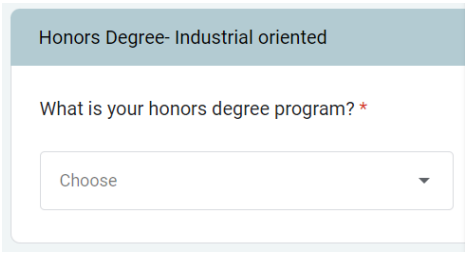


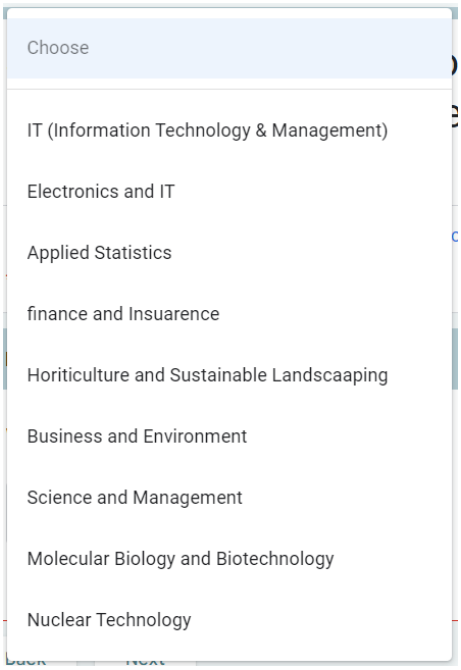




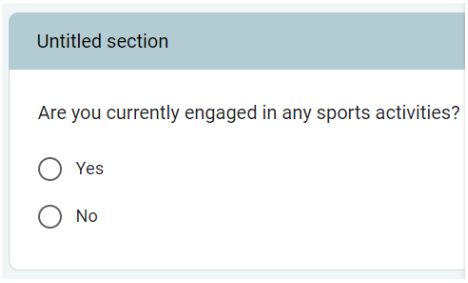




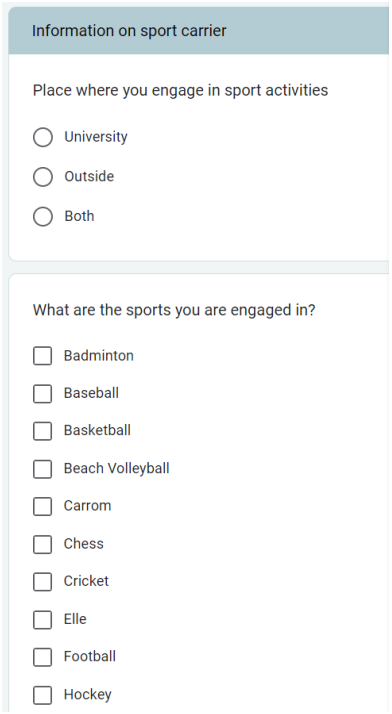


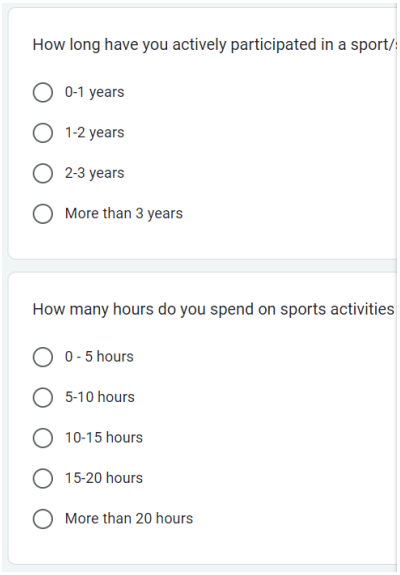


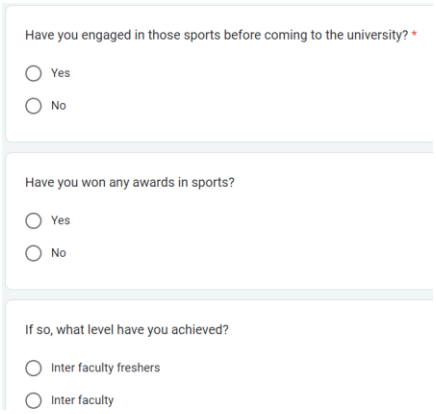
Section 02



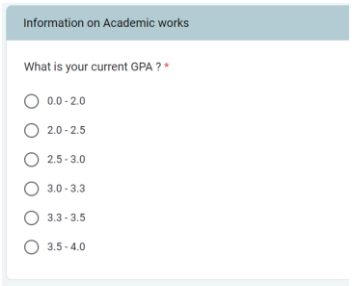
Section 03

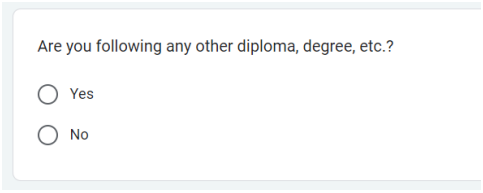




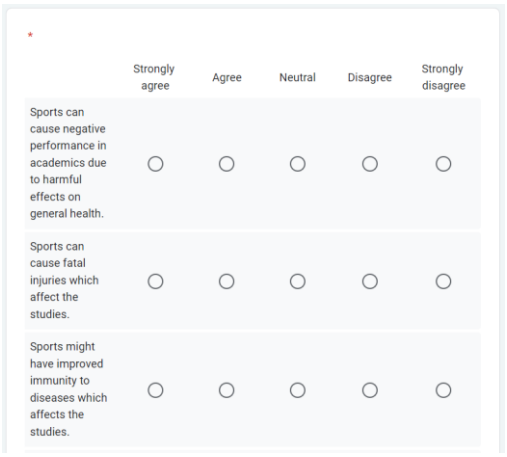


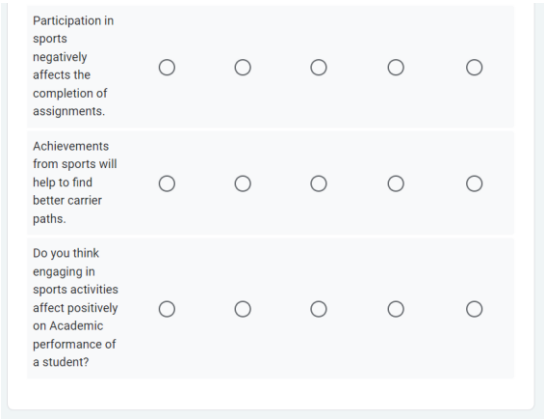
Section 04

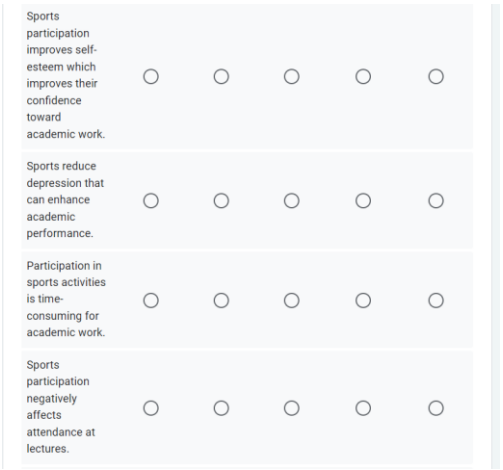




Section 05







Appendix B : Images of the dataset

A	B	C	D	E	F	G	H	I	J
Timestamp	Gender	District of living	Ethnicity	Current Academic ye	What degree program	Subject Stream	What is your honors (	What is your honors (	Are you currently eng
1/10/2023 14:36:42	Male	Gampaha	Sinhala	4	Honors degree - Research oriented		Statistics		Yes
1/10/2023 15:19:27	Female	Matara	Sinhala	2		Physical Science			No
1/10/2023 18:52:42	Female	Kalutara	Sinhala	1		Biological Science			No
1/10/2023 19:59:31	Male	Kalutara	Sinhala	3	Honors degree - Industrial oriented			Applied Statistics	No
1/10/2023 22:28:29	Male	Colombo	Tamil	1		Biological Science			Yes
1/11/2023 10:50:31	Male	Matara	Sinhala	2		Physical Science			Yes
1/11/2023 12:15:29	Female	Gampaha	Sinhala	2		Physical Science			No
1/11/2023 19:03:36	Female	Colombo	Sinhala	2		Physical Science			No
1/11/2023 23:59:12	Female	Colombo	Sinhala	1		Biological Science			Yes
1/12/2023 0:22:44	Female	Colombo	Sinhala	1		Biological Science			Yes
1/12/2023 14:25:10	Female	Kalutara	Sinhala	4	Honors degree - Research oriented		Statistics		Yes
1/12/2023 15:12:11	Female	Rathnapura	Sinhala	4	Honors degree - Research oriented		Industrial Statistics		No
1/13/2023 1:52:30	Female	Colombo	Sinhala	1		Biochemistry and Molecular Biology			Yes
1/13/2023 20:29:20	Male	Colombo	Sinhala	1		Biochemistry and Molecular Biology			Yes
1/13/2023 22:50:47	Male	Kalutara	Sinhala	2		Physical Science			Yes
1/13/2023 22:52:45	Male	Matara	Sinhala	2		Industrial Statistics and Mathematical Finance			No
1/13/2023 22:53:29	Male	Galle	Sinhala	3	Honors degree - Industrial oriented			Business and Enviro	No
1/13/2023 22:54:25	Male	Badulla	Sinhala	4	Honors degree - Industrial oriented			IT (Information Techr	No
1/13/2023 22:57:54	Male	Kurunegala	Sinhala	4	Honors degree - Research oriented		Mathematics & Statistics with Computer Sc		Yes
1/13/2023 22:59:14	Female	Gampaha	Sinhala	2		Industrial Statistics and Mathematical Finance			No
1/13/2023 23:01:31	Male	Kurunegala	Sinhala	2		Industrial Statistics and Mathematical Finance			Yes

K	L	M	N	O	P	Q	R	S	T
Place where you enga	What are the sports y	How long have you a	How many hours do y	Have you engaged in	Have you won any aw	If so, what level have	What is your current C	How long do you spe	Are you following any
University	Carrom	0-1 years	5-10 hours	No	Yes	Inter faculty	3.5 - 4.0	less than 2 hours	No
							3.5 - 4.0	5 - 8 hours	No
							3.5 - 4.0	2- 5 hours	No
							3.5 - 4.0	less than 2 hours	No
University	Cricket	0-1 years	5-10 hours	Yes	No		3.0 - 3.3	5 - 8 hours	No
University	Table Tennis	0-1 years	5-10 hours	Yes	No		3.5 - 4.0	5 - 8 hours	No
							2.5 - 3.0	2- 5 hours	No
							3.5 - 4.0	5 - 8 hours	No
University	Hockey	0-1 years	5-10 hours	Yes		Inter University	3.0 - 3.3	8 - 12 hours	Yes
Outside	Swimming	0-1 years	0 - 5 hours	Yes	No		3.5 - 4.0	2- 5 hours	Yes
Both	Volleyball	More than 3 years	0 - 5 hours	Yes	Yes	Inter faculty freshers	3.5 - 4.0	5 - 8 hours	No
							3.3 - 3.5	2- 5 hours	No
University	Cricket, Volleyball	0-1 years	0 - 5 hours	Yes	Yes	Inter faculty freshers	3.5 - 4.0	More than 12 hours	No
University	Baseball	0-1 years	10-15 hours	Yes	Yes	Inter University	3.0 - 3.3	8 - 12 hours	No
Outside	Cricket, Martial arts	1-2 years	10-15 hours	Yes	Yes	All Island	2.5 - 3.0	less than 2 hours	No
							3.5 - 4.0	2- 5 hours	No
							3.0 - 3.3	2- 5 hours	Yes
							3.5 - 4.0	2- 5 hours	No
University	Rugby	2-3 years	15-20 hours	Yes	Yes	Inter University	3.0 - 3.3	less than 2 hours	No
							3.5 - 4.0	8 - 12 hours	No
University	Taekwondo	1-2 years	5-10 hours	No	Yes	Inter University	2.5 - 3.0	2- 5 hours	No

U	V	W	X	Y	Z	AA	AB	AC	AD
[Sports can cause ne	[Sports can cause fa	[Sports might have ir	[Sports participation	[Sports reduce depre	[Participation in spor	[Sports participation	[Participation in spor	[Achievements from	[Do you think engagir
Disagree	Disagree	Neutral	Agree	Neutral	Disagree	Disagree	Disagree	Agree	Agree
Neutral	Neutral	Agree	Agree	Agree	Agree	Neutral	Neutral	Disagree	Strongly agree
Disagree	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral
Disagree	Neutral	Neutral	Agree	Agree	Agree	Agree	Agree	Agree	Neutral
Disagree	Neutral	Neutral	Agree	Agree	Neutral	Neutral	Neutral	Agree	Agree
Disagree	Agree	Agree	Agree	Strongly agree	Disagree	Disagree	Disagree	Neutral	Strongly agree
Strongly disagree	Agree	Strongly agree	Strongly agree	Neutral	Neutral	Neutral	Neutral	Agree	Strongly agree
Disagree	Agree	Agree	Agree	Agree	Agree	Agree	Neutral	Agree	Agree
Neutral	Agree	Neutral	Strongly agree	Strongly agree	Neutral	Neutral	Strongly disagree	Strongly agree	Strongly agree
Disagree	Agree	Neutral	Agree	Agree	Neutral	Neutral	Neutral	Agree	Agree
Strongly disagree	Agree	Neutral	Strongly agree	Strongly agree	Disagree	Disagree	Disagree	Strongly agree	Strongly agree
Strongly disagree	Disagree	Agree	Agree	Agree	Neutral	Neutral	Neutral	Agree	Agree
Disagree	Agree	Agree	Agree	Agree	Neutral	Neutral	Neutral	Agree	Neutral
Neutral	Agree	Agree	Agree	Strongly agree	Neutral	Neutral	Disagree	Agree	Agree
Strongly disagree	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral
Disagree	Agree	Agree	Agree	Agree	Neutral	Agree	Agree	Agree	Agree
Disagree	Neutral	Strongly disagree	Agree	Agree	Agree	Neutral	Agree	Agree	Agree
Disagree	Agree	Strongly agree	Strongly agree	Strongly agree	Agree	Agree	Neutral	Strongly agree	Agree
Disagree	Disagree	Agree	Agree	Strongly agree	Strongly agree	Disagree	Strongly disagree	Strongly agree	Strongly agree
Disagree	Agree	Agree	Agree	Strongly agree	Neutral	Agree	Agree	Agree	Agree
Disagree	Disagree	Agree	Agree	Agree	Neutral	Disagree	Neutral	Agree	Neutral

Appendix C : Python Codes for advanced Analysis

```
import numpy as np
import pandas as pd
import sklearn
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.model_selection import train_test_split

from sklearn.model_selection import GridSearchCV,
RepeatedStratifiedKFold, RandomizedSearchCV, cross_val_score,
train_test_split
from sklearn.metrics import classification_report, confusion_matrix,
accuracy_score
from sklearn.neighbors import KNeighborsClassifier
from sklearn.svm import SVC
from sklearn.ensemble import RandomForestClassifier,
GradientBoostingClassifier
from sklearn.tree import DecisionTreeClassifier
from sklearn.linear_model import LogisticRegression
from sklearn.discriminant_analysis import LinearDiscriminantAnalysis
from xgboost import XGBClassifier


from sklearn.tree import DecisionTreeClassifier, plot_tree
import matplotlib.pyplot as plt
from sklearn.metrics import f1_score



df = pd.read_excel("Research_sport.xlsx")
df1=df.drop(['sports_names','deploma','academic_time','awards'],axis=1)
df1.info()

#creating dummy variables
for i in ['Gender', 'District','Ethnicity','Degree_Type']:
    tempdf= pd.get_dummies(df1[i],drop_first=True,prefix=i)
    df1= pd.concat([df1,tempdf],axis=1)
    df1.drop(i,axis=1,inplace=True)

df1.head()

x=df1.drop('sports_yes',axis=1)
y=df1['sports_yes']

#train test split
from sklearn.model_selection import train_test_split
x_train,x_test,y_train,y_test=
train_test_split(x,y,random_state=0,test_size=0.2,stratify=df1['sports_y
es'])
train_df= pd.concat([x_train,y_train],axis=1)
test_df= pd.concat([x_test,y_test],axis=1)

x1=train_df.drop('sports_yes',axis=1)
y1=train_df['sports_yes']

x2=test_df.drop('sports_yes',axis=1)
y2=test_df['sports_yes']

#model training
models = [
```

```
    KNeighborsClassifier(),
    LogisticRegression(),
    RandomForestClassifier(),
    DecisionTreeClassifier(),
    SVC(),
    GradientBoostingClassifier(),
    LinearDiscriminantAnalysis(),
    XGBClassifier()
]

names = [
    "KNeighborsClassifier",
    "LogisticRegression",
    "RandomForestClassifier",
    "DecisionTreeClassifier",
    "SVC",
    "GradientBoostingClassifier",
    "LinearDiscriminantAnalysis",
    "XGBClassifier"
]

for model, name in zip(models, names):
    print(name)
    for score in ["accuracy", "precision", "recall","f1"]:
        print(score," : ",cross_val_score(model, x1, y1 ,scoring=score,
cv=5,error_score='raise').mean())
    print('\n')


# Create a Decision Tree model and fit it to the training data
#dt = DecisionTreeClassifier(max_depth=8)
dt = DecisionTreeClassifier()
dt.fit(x1, y1)



# Calculate the feature importances and max_depth=10plot them in a bar
chart
importances = dt.feature_importances_
plt.bar(x1.columns, importances)
plt.xticks(rotation=90)
plt.ylabel("Importance")
plt.show()


#ranforest prediction
y_pred = dt.predict(x2)

y_pred
plt.figure(figsize=(20,10))
plot_tree(dt, filled=True, feature_names=x1.columns, max_depth=3,
fontsize=8, label='all', proportion=True,
          impurity=False, rounded=True, precision=0, node_ids=False,
class_names=True)
plt.show()

accuracy = accuracy_score(y2, y_pred)
mse = mean_squared_error(y2, y_pred)
precision = precision_score(y2, y_pred, average='weighted')
recall = recall_score(y2, y_pred, average='weighted')
f1=f1_score(y2, y_pred)
```

```
print("Accuracy:", accuracy)
print("Mean Squared Error:", mse)
print("Precision:", precision)
print("Recall:", recall)
print("f1:",f1)

#Decision tree
cm = confusion_matrix(y_test, y_pred)

print("Contingency Table:")
print(cm)

rf = RandomForestClassifier()
rf.fit(x1, y1)
y_pred2 = rf.predict(x2)


importances = dt.feature_importances_
sorted_idx = np.argsort(importances)

# Plot feature importance graph
cmap = plt.cm.get_cmap('RdYlBu')
colors = [cmap(i) for i in np.linspace(0, 1, x1.shape[1])]

fig, ax = plt.subplots(figsize=(10, 6))
ax.barh(range(x1.shape[1]), importances[sorted_idx],color=colors)
ax.set_yticks(range(x1.shape[1]))
ax.set_yticklabels(x1.columns[sorted_idx])
ax.set_xlabel("Importance Score")
plt.title("Feature importance Random Forest Regressor")
plt.show()
```

References

Ahamed, Y., Macdonald, H., Reed, K., Naylor, P. J., Liu-Ambrose, T., & Mckay, H. (2007). School-based physical activity does not compromise children's academic performance. *Medicine and Science in Sports and Exercise*, *39*(2), 371–376. https://doi.org/10.1249/01.mss.0000241654.45500.8e

Beasley, M., Thompson, T., & Davidson, J. (2003). Risilience in response to life stress: The effects of coping style and cognitive hardiness. *Personality and Individual Differences*, *34*(1), 77–95. https://doi.org/10.1016/S0191-8869(02)00027-2

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

Buckworth, J., Lee, R. E., Regan, G., Schneider, L. K., & DiClemente, C. C. (2007). Decomposing intrinsic and extrinsic motivation for exercise: Application to stages of motivational readiness. *Psychology of Sport and Exercise*, *8*(4), 441–461. https://doi.org/10.1016/j.psychsport.2006.06.007

Buunk, B. P., Kuyper, H., & Zee, Y. G. Van Der. (2010). *Basic and Applied Social Psychology Affective Response to Social Comparison in the Classroom Affective Response to Social Comparison in the Classroom*. *May 2013*, 37–41. https://doi.org/10.1207/s15324834basp2703

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

Chaddock-Heyman, L., Hillman, C. H., Cohen, N. J., Kramer, A. F., Chaddock-Heyman, L., Hillman, C. H., Cohen, N. J., & Kramer, A. F. (2014). The relation of childhood physical activity to brain health, cognition, and scholastic achievement: III. The importance of physical activity and aerobic fitness for cognitive control and memory in children. *Monographs of the Society for Research in Child Development*, *79*(4), 25–50. http://www.ncbi.nlm.nih.gov/pubmed/25387414%5Cnhttp://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=2014-48859-004&site=ehost-live%5Cnhttp://lchaddo2@illinois.edu

Chaddock, L., Pontifex, M. B., Hillman, C. H., & Kramer, A. F. (2011). A review of the relation of aerobic fitness and physical activity to brain structure and function in children. *Journal of the International Neuropsychological Society*, *17*(6), 975–985. https://doi.org/10.1017/S1355617711000567

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

Coe, D. P., Pivarnik, J. M., Womack, C. J., Reeves, M. J., & Malina, R. M. (2006). Effect of physical education and activity levels on academic achievement in children. *Medicine and Science in Sports and Exercise*, *38*(8), 1515–1519. https://doi.org/10.1249/01.mss.0000227537.13175.1b

Darling, N., Caldwell, L. L., & Smith, R. (2005). Participation in school-based extracurricular activities and adolescent adjustment. *Journal of Leisure Research*, *37*(1), 51–76. https://doi.org/10.1080/00222216.2005.11950040

Deaner, R. O., Balish, S. M., & Lombardo, M. P. (2016). Sex differences in sports interest and motivation: An evolutionary perspective. *Evolutionary Behavioral Sciences*, *10*(2), 73–97. https://doi.org/10.1037/ebs0000049

Dyson, B., Cowan, J., Gordon, B., Powell, D., & Shulruf, B. (2018). Physical education in Aotearoa New Zealand primary schools: Teachers' perceptions and policy implications. *European Physical Education Review*, *24*(4), 467–486. https://doi.org/10.1177/1356336X17698083

Eime, R. M., Young, J. A., Harvey, J. T., Charity, M. J., & Payne, W. R. (2013). A systematic review of the psychological and social benefits of participation in sport for adults: Informing development of a conceptual model of health through sport. *International Journal of Behavioral Nutrition and Physical Activity*, *10*. https://doi.org/10.1186/1479-5868-10-135

Geda, Y. E., Roberts, R. O., & Knopman, D. S. (2010). *Physical Exercise, Aging, and Mild Cognitive Impairment*. *67*(1), 80–86.

Handbook - Faculty of Science - University of Colombo

https://science.cmb.ac.lk/wp-content/uploads/2020/01/ughandbook2019_2020.pdf

Pamudini, K., & Manel, D. P. . (2022). *Study on the Effects of Sports on Ed- ucation Intentions of Grade Eleven Students in Government Schools : With reference to Gampaha District*.

Pascarella, E. T., Pierson, C. T., Wolniak, G. C., & Terenzini, P. T. (2004). First-generation college students: Additional evidence on college experiences and outcomes. *Journal of Higher Education*, *75*(3), 249–284. https://doi.org/10.1080/00221546.2004.11772256

Umbach, P. D., 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

Sani, M. U., Danlami, A. I., & Amos, L. (2019). *Influence of Sport Participation on Students ' Academic Performance in Public Secondary Schools in Dutsin-Ma Metropolis of Katsina State Nigeria*. *2665*, 105–110. https://doi.org/10.21276/jaep.2019.3.3.6

Simons, H. D., Van Rheenen, D., & Covington, M. V. (1999). Academic motivation and the student athlete. *Journal of College Student Development*, *40*(2), 151–162.

Sri Lanka University Games- SLUG. (2015, March 9). Sri Lanka Universities Sports

Association. http://slusa.lk/compititions/sri-lanka-university-games-slug/

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

Voss, M. W., Weng, T. B., Burzynska, A. Z., Wong, C. N., Cooke, G. E., Clark, R., Fanning, J., Awick, E., Gothe, N. P., Olson, E. A., McAuley, E., & Kramer, A. F. (2016). Fitness, but not physical activity, is related to functional integrity of brain networks associated with aging. *NeuroImage, 131*, 113–125. https://doi.org/10.1016/j.neuroimage.2015.10.044