STUDENT PERFORMANCE PREDICTION BASED ON PERSONAL FACTORS BY USING DASHBOARD

BY

TAN CHUNG HAO

A REPORT

SUBMITTED TO

Universiti Tunku Abdul Rahman

in partial fulfillment of the requirements

for the degree of

BACHELOR OF INFORMATION SYSTEMS (HONOURS) BUSINESS INFORMATION SYSTEMS

Faculty of Information and Communication Technology

(Kampar Campus)

JUNE 2024

UNIVERSITI TUNKU ABDUL RAHMAN

REPORT STATUS DECLARATION FORM

Title: STUDENT PERFORMANCE PREDICTION BASED ON PERSONAL FACTORS BY USING DASHBOARD

Academic Session: JUNE 2024

I TAN CHUNG HAO

(CAPITAL LETTER)

declare that I allow this Final Year Project Report to be kept in Universiti Tunku Abdul Rahman Library subject to the regulations as follows:

1. The dissertation is a property of the Library.
2. The Library is allowed to make copies of this dissertation for academic purposes.

Verified by,

(Author's signature)

(Supervisor's signature)

Address:
270, Lorong 11, Taman Intan Mas,
36000 Teluk Intan,
Perak.

Dr Shakiroh Binti Khamis

Supervisor's name

Date: 12/9/2024

Date: 13/9/2024

FACULTY OF INFORMATION AND COMMUNICATION TECHNOLOGY

UNIVERSITI TUNKU ABDUL RAHMAN

Date: 12/9/2024

SUBMISSION OF FINAL YEAR PROJECT /DISSERTATION/THESIS

It is hereby certified that **TAN CHUNG HAO** (ID No: **20ACB05504**) has completed this final year project/ dissertation/ thesis* entitled "STUDENT PERFORMANCE PREDICTION BASED ON PERSONAL FACTORS BY USING DASHBOARD" under the supervision of Dr Shakiroh Binti Khamis (Supervisor) from the Department of Digital Economy Technology, Faculty of Information and Communication Technology, and _____ (Co-Supervisor)* from the Department of Digital Economy Technology, Faculty of Information and Communication Technology.

I understand that University will upload softcopy of my final year project / dissertation/ thesis* in pdf format into UTAR Institutional Repository, which may be made accessible to UTAR community and public.

Yours truly,

TAN CHUNG HAO

_____

(*Student Name*)

*Delete whichever not applicable

DECLARATION OF ORIGINALITY

I declare that this report entitled "**STUDENT PERFORMANCE PREDICTION BASED ON PERSONAL FACTORS BY USING DASHBOARD**" is my own work except as cited in the references. The report has not been accepted for any degree and is not being submitted concurrently in candidature for any degree or other award.

Signature : 

Name : TAN CHUNG HAO

Date : 12/9/2024

ACKNOWLEDGEMENTS

I would like to express my sincere thanks and appreciation to my supervisors, Ts Dr Shakiroh binti Khamis who has given me this bright opportunity to engage in an Data Analyst project. It is my first step to establish a career in the Data Analyst field. A million thanks to you.

Additionally, we would like to extend our sincere appreciation to University Tunku Abdul Rahman for the excellent facilities that were made available to us. By making use of the library's resources, including databases and references, we were able to conduct the research with greater efficiency. Additionally, we would like to extend our gratitude to our colleagues who provided any assistance during the course of this research. We express our sincere gratitude for the valuable insights, opinions, and support provided by our acquaintances. Aside from that, we would also like to extend our deepest gratitude to our families for their encouragement and support throughout the project's execution.

ABSTRACT

Within the field of education, this research initiative operates at the interface of methods of statistical and/or logical analysis used to describe, illustrate, summarize, and assess data. In order to improve instructional techniques, it handles the problem of comprehending and forecasting student performance based on individual factors. Conventional teaching methods frequently ignore the variety of individual circumstances affecting a student's performance. In order to provide educators with a data-driven tool to support targeted interventions and enhance overall student outcomes, this project aims to discover, model, and forecast these aspects. A six-step data science lifecycle, beginning with business understanding and ending with model deployment, is used in the study process. To build an extensive dataset, information is gathered from educational institutions and through surveys. An interactive dashboard is designed to provide instructors with insights, and machine learning methods are utilized in the development of the prediction model. The project starts with a thorough literature review to establish the theoretical basis. Following data comprehension, preprocessing, and modeling, the predictive model is created. The created prediction model shows excellent generalizability and accuracy in a variety of educational contexts. The model's efficacy in recognizing and forecasting student performance based on individual factors is demonstrated by evaluation measures like the SME (Mean Square Error). The successful implementation of an interactive dashboard, which provides educators with a useful tool for data-driven decision-making, marks the research's conclusion. By addressing the highlighted problem and providing a systematic way to analysing and forecasting student performance, the initiative contributes to the progress of educational practices.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

MLR	Multi Linear Regression
HE	Home Environment
SI	Student Interest
SH	Study Habit

Chapter 1
Introduction

This research investigates the individual factors that influence the academic performance of students. The objective of our research is to use a dashboard to present an analysis of the personal factors that influence students' academic performance. This chapter will encompass the following: problem statement and motivation, research objective, project scope and direction, research contributions, and report organization. Ultimately, the study's overview offers a brief overview covering Chapters 1 through Chapter 6.

1.1 Problem Statement and Motivation

The main issue that this research will look into is the difficulties educators have while trying to comprehend and deal with the various aspects of personal factors that affect students' academic performance. Even in cases where data is readily available, tools for systematic analysis and forecasting to spot patterns and trends are frequently lacking. The three main issues that need to be resolved; the first problem is identifying the most crucial personal factor that affects academic performance. It's important to recognize and comprehend the different personal aspects that have a big influence on a student's performance. This covers socioeconomic background, study habits, and study interests.

Furthermore, this problem will lead to the predictive modeling aspect. It is difficult to create predictive models that are both accurate and trustworthy. Since the educational data is dynamic, this study attempts to examine how well machine learning algorithms predict student performance based on personal factors.

Additionally, the difficulty lies in delivering the insights understandably, even with strong prediction models. Due to this, there is need a for user-friendly predictive modelling to interpret it. It will benefit educators to make informed decisions based on data, and an interactive dashboard will be created to effectively express the forecasts.

This project's motivation comes from the expanding demand for tailored and focused interventions in education to address various personal factors that influence the academic success of students. Thus, the diversity in the learning profile is the first motivation to do this project as students have distinct learning styles, different study habits, and a range of socioeconomic backgrounds, understanding and accounting for this variation is critical for educators in building an inclusive educational environment that caters to each student's particular needs.

Besides that, enhancing educator decision-making is also the motivation to do this project. It can be difficult for educators to derive valuable insights from the deluge of data they encounter. By understandably providing complicated data, the creation of an interactive dashboard would empower teachers and give them the ability to make well-informed decisions regarding ways to teach, and support for students, and interventions.

Furthermore, this project gives motivation to get used to new technologies. It is necessary to utilize technology to the advantage of both teachers and pupils since it will always play a crucial part in education. This research is in line with the larger trend of using data analytics and machine learning in education to shift towards a more data-informed approach.

1.2 Objectives

This study investigates the relationship between personal factors, intending to address and predict the issue of students performing poorly academically. This research presents an idea that examines the association between personal factors such as Student Habits, Home Environment, and Student Interest among students.

1.2 General Objective

- To predict student performance based on personal factors by using dashboards.

1.2.1 Specific Objective

i. To determine which factors, have the impact on a student's academic performance.

ii. To develop the predictive model based on the identified personal factors by utilizing machine learning methods.

iii. To create an interactive dashboard that explains the model's predictions to educators in an efficient manner.

1.3 Project Scope and Direction

The scope of the project includes identifying the personal factors that affect the student's performance. The project's main goal is to investigate how personality factors may affect a student's performance. These could include study interests, study habits, socioeconomic status, and other traits that can be measured and examined individually. As a result, the sample for this research comprises enthusiastic undergraduates from various faculties and academic years enrolled at UTAR in Kampar, Perak, who are enrolled full-time. Next is to create the interactive dashboard. The interactive dashboard's design and implementation will be restricted to showcasing the predictive model's results. The dashboard is designed to make it easy for educators to understand and apply the insights produced by the model.

1.4 Contributions

The goal of this project's research is to solve the stated issues and motives to significantly advance the field of education. The first contribution is the identification of influential personal factors toward academic performance. By methodically identifying and evaluating the personal factors that significantly affect student performance, the study aims to contribute to providing insights into which aspects play a critical influence on academic success by utilizing machine learning algorithms and data analytics.

In addition, the study that addresses the lack of utilization of advanced data visualization tools in the context of education is extremely valuable because it can improve the Decision-Making of educators. Educators working with huge, dynamic, and complicated datasets must implement powerful data visualization technologies. These solutions, like Power BI, include dynamic and interactive capabilities that can greatly improve the capacity to identify key patterns and trends in the data. I may equip decision-makers with visual insights and help them make more strategic and well-informed decisions by implementing these technologies without hesitation. This will allow decision-makers to gain a deeper grasp of their data to improve student academic performance.

Furthermore, this study contributes significantly to resolving the insufficiency of using advanced analytics tools for in-depth analysis of personal factors that affect student academic performance in the education field. The advanced analytics tools are important because they enhance the capabilities for analysis. Employing technologies such as Python enables enterprises to analyze a variety of datasets in a more sophisticated manner, revealing subtle patterns, preferences, and trends that could otherwise go unnoticed.

Moreover, predictive model development and validation also contribute to this project. The goal of the study is to provide a prediction model that can accurately forecast student performance by utilizing the personal factors that have been uncovered. With the use of historical data, the model will be thoroughly built and verified, giving educators a trustworthy tool for spotting possible problems in the lives of students early on.

In addition, this project also contributes by becoming a framework for future research. It is the goal of the project to establish a platform for further study in the fields of education, data analytics, and machine learning. The methodology and findings can be used to guide future research into personalized learning, early intervention strategies, and the use of technology in educational practices.

1.5 Report Organization

Chapter 1: Introduction

Chapter 1 comprises the research objective, problem statement and motivation, project scope and direction, and contribution of this study. It offers a comprehensive outline of the entire research investigation.

Chapter 2: Literature Review

The literature evaluations in this section evaluate all secondary data sources, such as journal articles, magazine articles, and websites. This part comprises the literature review, theoretical foundation, and theoretical framework that will serve as the basis for the formulation of hypotheses and further analysis.

Chapter 3: Research Methodology

Methodology, covered in Chapter 3, includes the data science life cycle, sampling design, data collection, research instrument, procedure for data analysis, data processing, questionnaire design, and questionnaire measurement scale.

Chapter 4: Data Analysis and Result

This chapter is dedicated to data analysis, which involves the provision and interpretation of statistical results obtained from reliability tests. Subsequently, contrast the outcome with the formulated conclusions and examine the fundamental thesis.

Chapter 5: Development

This chapter will provide a comprehensive explanation of the technical procedures required to construct both the predictive model and the dashboard. Subsequently, it will elaborate on the interactive dashboard's development process, which encompasses the crucial design decisions and the tools employed to establish an intuitive user interface that enables educators to comprehend the predictions.

Chapter 6: Conclusion and Future Work

The final chapter will provide a comprehensive summary of the study, including the personal factors that were influential in predicting student performance. Additionally, it will address the manner in which the interface can assist educators in making well-informed decisions by utilizing the predictions. The chapter will address the study's limitations and propose areas for future research, such as the exploration of additional factors or the enhancement of the dashboard's capabilities.

Chapter 2
Literature Review

2.1 An analysis of previous empirical research

2.1.1 Academic Performance

Academic performance of students, which is one of the most important and foundational goals of education, is comprised of the knowledge obtained by students and assessed by instructors via grading systems, in addition to educational goals established jointly by students and instructors to be achieved within a specified timeframe [4]. The principal objective of academic institutions is to assist students in attaining academic excellence through the facilitation of enhanced academic performance. The importance of a student's academic success is great because a strong academic record is thought to be a requirement for landing excellent employment, advancing in one's career, and ultimately leading a quality life. Due to the vague nature of "Academic Performance," a variety of tools can be utilized to measure it while keeping in mind the particular viewpoint that is being taken into consideration. The grade point average, or "GPA," has been the most widely used standard for assessing academic success. It is used to assess how well students perform within a given semester [4].

2.1.2 Home Environment

Students' psychological, emotional, social, and financial well-being are greatly influenced by their family environment. A person's home environment has an impact on them because their parents are the ones who socialize them from an early age [1]. This is because a child's family background and environment have an impact on both his academic achievement and his response to life's circumstances. Put simply, students who come from families with a higher level of education are more likely to achieve academic success and acquire more quickly than those whose parents don't have a formal education. Moreover, empirical evidence indicates that a significant proportion of students encountered considerable financial and sociological challenges within their families, which in turn influenced their academic performance by reducing their drive to seek knowledge [2]. Students raised by alcoholics frequently endure stress, feelings of insecurity, rage, and frustration, which negatively impact their academic

achievement. Thus, the research conducted discovered that compared to individuals from non-alcoholic homes, individuals from alcoholic families experience higher levels of anxiety and depression [2]. It is reasonable to assume that a student's academic performance in school may be influenced by the type of family, the educational background of the parents, and their socioeconomic status such as income [8].

2.1.3 Student Interest

Numerous psychological studies have demonstrated the relationship between interest and a wide range of other factors, including personal life, aesthetics, careers, motivation, reasoning, information processing, emotion, and recreational pursuits. [1]. Therefore, interest also will affect student performance. The interest in learning is likely a highly influential psychological characteristic, accompanied by a strong emotional response to knowledge and a strong sense of being captivated, enthralled, invigorated, and energized. This leads to a faster and more accurate cognitive processing of information, as well as the most effective application of psychomotor traits such as self-regulatory skills, self-discipline, exerting greater effort and intelligence with optimal persistence. [1] According to the researcher, a good interest in a subject translates into a high degree of demand for achievement and, eventually, a high level of performance. This observation highlights the relationship between attitude and academic success. Stated differently, there is a guarantee of high achievement if students approach a subject with positivity. This supports the idea that students' interest has a significant role in their academic success [2].

2.1.4 Study Habits

Students' academic success is significantly influenced by their study habits [9]. A large number of students may not attend class regularly and pay little to no attention to their academics. [2] Indeed, it is possible that some of them may neglect their obligation to attend class in order to engage in frivolous pursuits such as video club presence, theft, slander, harassment, and other detrimental group activities. Poor study habits are the main cause of underachievement in schools, according to research; even bright students can perform below expectations if they have poor study habits [2]. A significant correlation has been demonstrated by research to exist between effective study techniques, increased academic engagement, and improved academic achievement [4]. A few study habit characteristics that affect students' academic achievement are homework, time management, reading and note-taking, and educator consultation [10].

2.1.5 Multiple Linear Regression for Analysis on Academic Performance

Multiple linear regression (MLR) is a frequently used predictive analysis tool in education that is based on the multivariate statistical technique [5]. The number of variables has a considerable impact on the MLR method's performance. To evaluate the significance of the association between the response and each term in the model and to invalidate the null hypothesis, compare the term's p-value to the preset significance level. There is no correlation between the term and the response, which is the null hypothesis. Typically, a significance level of 0.05 (represented as α or alpha) is acceptable. A 5% chance of assuming there is a link when none exists is indicated by a significance level of 0.05.

P-value $\leqslant$ α : There is statistical significance in the association.

A correlation between the response variable and the term is considered statistically significant when the p-value is equal to or less than the significance level.

P-value > α: The association is not statistically significant.

Determining a statistically significant relationship between the term and the response variable is not possible if the p-value surpasses the predetermined significance level.

Regression models' goodness of fit has been analyzed using the Q-Q plot method, the MSE, and R^2 metrics [5]. Nevertheless, these metrics are unable to assess the regression model's predictive performance. For evaluating prediction error, the mean absolute percentage error (MAPE), which quantifies the percentage of prediction error in a prediction model, is calculated. The prediction error of the prediction model decreases with decreasing MAPE values. With academic performance as a dependent variable, the study interest, study habit, and home environment as the independent variable, form a multiple linear regression to study the significant factors on a student’s academic performance.

2.1.6 Strengths and Weaknesses of MLR

MLR is a powerful concept for deep learning. The strength is the capacity to identify the impacts of each variable, account for confounding variables, and understand complicated relationships [5]. By simultaneously assessing the impacts of several variables, this statistical strategy improves predicted accuracy and model robustness while also accepting non-linear

interactions. In addition, implementing multiple regression is not too difficult, especially when software tools are used. Furthermore, multiple regression may produce extremely accurate predictions if its presumptions are met.

However, these techniques also have several weaknesses. One of the biggest disadvantages of MLR is the assumptions required. A few conditions must be fulfilled for multiple regression to work, including linearity and error independence. Regression analysis may yield deceptive results if these presumptions are not true. Moreover, it also has limitations in handling the missing data and outliers. It is important to note the model's resilience because it can become susceptible to outliers and extreme values when simultaneously observing the impacts of variables. Besides that, when a model has too many independent variables, it might become overfit, which can result in excessively complicated and erratic predictions.

2.2 Literature Review Summary Table

Author	Objective	Number of Variables	Methodology	Sample Domian
M. Dev (2016) [1]	1. Investigation of the individual aspects influencing the academic performance of students.	4 variables: - Intelligent - Gender - Home Environment - Interest	- A descriptive research design implemented as an ex-post-facto technique. - The chosen sample underwent administration of the Generalised Mental Ability Test, Home Environment Inventory, and Multiphasic Interest Inventor.	- 110 students from the class VIII in KenderyaVidyalayas of Delhi.

A. M. Adeyemi and S. B. Adeyemi (2014) [2]	1. To assess the academic performance of students in federal, state, and private institutes of education in that region of the country. 2. The study also encompasses a comparative analysis of personal variables and their impact on students at the researched schools of education.	6 variables: - Home Environment -Students' Interest - Study Habit - Peer Influence - Students' Perception of Course - Parents' Support	- Ex-post facto analysis employing a survey approach. - Multiple regression design - 4-point likert scale questionnaires - SPSS Software	- The research population comprised 1,900 instructors and 12,420 NCE students (at 200 Level and 300 Level) from federal, state, and private NCE-awarding institutions in South Western Nigeria, namely Ekiti, Lagos, Ogun, Ondo, Osun, and Oyo States.
N. Arora and N. Singh (2017) [3]	1. To investigate the determinants influencing students's academic grades. 2. To construct a predictive model for the academic achievement of students using specified data.	5 variables: - Teaching Effectiveness - Distraction Factors - Study Habits - Personality Traits - Family Environment	- 5-point Likert scale survey method - Multiple regression analysis - SPSS Software	- 120 students has been collected from engineering students of Gurugram.

S. J. H. Yang, O. H. T. Lu, A. Y. Q. Huang, J. Huang, H. Ogata, and A. J. Q. Lin (2018) [5]	1. To validate whether MLR is suitable for predict student performance.	- MSE - R^2 - pMSE - pMAPC	- Multiple Linear Regression - Principal Component Analysis	- 58 university freshmen from Northern Taiwan.
Obeta, A. (2014) [8]	1. To determine the home environmental elements influencing the academic success of students.	- 4 variables: - Untimely enrollment of the student in a reputable educational institution - No availability of sufficient educational resources when return from school - Parents/students relationship - Socio-economic status	- 20 items structured questionnaire with 4-point likert scale - Using frequency table, percentage and mean.	- For JSS111 and SS111 students in secondary schools in Abia State, as well as their parents and the principals of the schools.

R. A. Kusurkar, O. T. Cate, C. M. P. Vos, P. Westers, and G. Croiset (2012) [9]	1. To investigate the correlations among motivation, study strategy, study effort, and academic achievement among medical students.	- 5 variables: - Autonomous motivation - Controlled motivation - Relative autonomous motivation - Good study strategies - Study effort	- Electronic questionnaire - SPSS 15.0 - Multiple Regression	- 464 of academic students enrolled in years 2 to 6 at the VU University Medical Center Amsterdam.
Tus, J (2020) [10]	1. To assess the impact of study attitudes and study habits on students' academic achievement	6 variables: - Delay Avoidance - Work Methods - Teacher Approval - Educational Acceptance - Study Habits - Study Attitudes	- Applied Survey of Study Habits and Attitudes comprising 100 items over four scales. - Multiple linear regression - IBM SPSS 20.0	- The study population consisted of 130 of senior high school students attending a Catholic School in Bulacan, Philippines.

2.3 Proposed Conceptual Framework

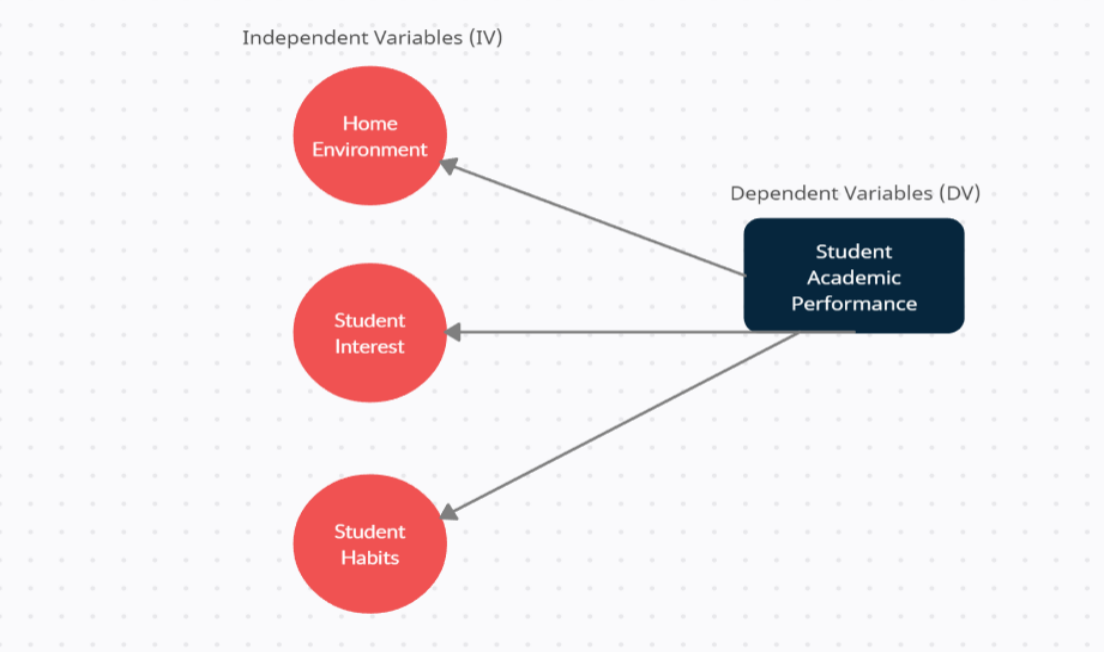


Figure 2.1 Proposed Conceptual Framework

Source: Developed for Study.

We constructed an innovative conceptual framework model by expanding upon previous models published in previous research. The main aim of the proposed conceptual framework was to clarify the correlation between the three separate independent variables (home environment, student interest and student habits) and the dependent variable (student academic performance). This study aimed to develop hypotheses that investigate the relationships between academic performance and three independent variables: home environment, student interest and student habits because mostly research paper will contain these three personal factors for research.

2.4 Hypotheses Development

2.4.1 Relationship between home environment and academic performance among UTAR students.

Based on several earlier studies, the hypothesis on the connection between students' academic performance and home environment has been developed. The majority of students were found to have significant social and financial issues at home, which had an impact on their academic development [2]. The general academic success of students is greatly influenced by the

education of their parents [3]. Academic achievement of students is significantly correlated with home environments [2].

Therefore, the following hypothesis is proposed:

H_0: The home environment does not affect the academic performance among UTAR students.

H_1: The home environment affects the academic performance among UTAR students.

2.4.2 Relationship between student's interest and academic performance among UTAR students.

Based on several earlier studies, the hypothesis on the connection between students' academic performance and student interest has been developed. The degree to which a situation can or will satisfy one's requirements and objectives determines one's interest response in it. It stated that there is a favorable correlation between attitude and academic success and that having a positive attitude towards a subject is supposed to transfer into having a high need for achievement and, eventually, performing at a high level [2]. Put differently, there's an assurance of good achievement if students have a favorable attitude towards a given subject. This supports the idea that students' academic success is significantly influenced by their level of interest [2].

Therefore, the following hypothesis is proposed:

H_0: The student's interest does not affect the academic performance among UTAR students.

H_1: The student's interest affects the academic performance among UTAR students.

2.4.3 Relationship between study habits and academic performance among UTAR students.

Based on several earlier studies, the hypothesis on the connection between students' academic performance and study habits has been developed. It stated that bad study habits are a big reason why students perform poorly in school; even exceptionally talented students can perform poorly if they have bad study habits [2]. Academic achievement and students' motivation to attend classes, take notes, interact with teachers, learn time management skills, and follow the schedule, are strongly correlated [4]. A significant correlation has been

demonstrated by research to exist between effective study techniques, increased academic engagement, and enhanced academic performance [4].

Therefore, the following hypothesis is proposed:

H_0: The study habits do not affect the academic performance among UTAR students.

H_1: The study habits affect the academic performance among UTAR students.

2.5 Conclusion

Chapter two concluded with mostly review of the literature from previous researchers about academic performance and the three independent factors such as home environment, student's interest, and study habits with MLR for data analysis purpose. In order to address the research questions, hypotheses were generated and evaluated in order to figure out the correlation between the independent and dependent variables. The research approach will be covered in more detail in the upcoming chapter.

Chapter 3
Research Methodology

The methods section of a research proposal comprises detailed information about the research methodology. The project plan consists of the activities that are intended to be carried out to accomplish the project's objectives, while the study design encompasses the methodology and methods that are intended to be employed.

3.1 Methodology

The methodology employed in this research is the data science life cycle. The Data Science Lifecycle is an all-encompassing manual that offers precise guidance on the implementation of analytical techniques, such as machine learning, to extract insights and predictions from data with the ultimate objective of accomplishing a business objective. Throughout the process, numerous steps are performed, including business understanding, exploratory data analysis, data preparation, modeling, model evaluation, and development.

3.1.1 Business Understanding

In the data science process, identifying problems is one of the key phases required to discover a specific goal around which all subsequent processes will be developed. Identify and list the issue that needs to be resolved, after that start with a thorough literature review to gain an understanding of the current state of research. In summary, since the business purpose will determine the aim of our investigation, it is critical to comprehend it early on. After that, it is compulsory to identify the solution used, whether is supervised, unsupervised, or reinforcement learning. After identifying the solution, come to make the selection of a algorithm is dependent upon the characteristics of the problem, the size and type of the data, and the particular demands of the application environment. Each algorithm possesses unique attributes and exhibits varying levels of effectiveness in different scenarios, so, careful selection is crucial to attain the best outcomes.

3.1.2 Data Understanding

The next step in the data science lifecycle is data understanding, firstly involves gathering unprocessed data from appropriate sources. The collected data may come in an unstructured or structured format. The data collection techniques could include web scraping, weblogs, surveys, social media, online repositories, data streamed from online sources via APIs, and data included in Excel files or any other source. The objective is to collect all the essential facts that could directly influence our research or model. After this is a data description, grasping the fundamental organization, structure, and substance of the data. This entails analyzing data types, tabular names, row counts, and verifying the presence of missing values. The objective is to familiarize oneself with the data and detect any first patterns or irregularities. Move forward is undertaking a more comprehensive analysis of the data to comprehend its attributes. This includes: Statistical Summary, Data Visualization, Identifying Outliers, and Understanding Relationships. Statistical summary refers to the computation of statistical measures such as mean, median, mode, standard deviation, and distribution of the data. Data visualisation is the creation of charts, histograms, and scatter plots to visually represent correlations, trends, and patterns in data. Outlier identification refers to the detection of any abnormal data points that may distort the analysis. Lastly, an analysis of relationships involves examining the correlations and interactions among several variables.

3.1.3 Data Preprocessing

The next step in the data science lifecycle is data preprocessing, first involves handling missing values. Identifying and addressing missing data points involves eliminating rows or columns with missing values, filling them with statistical parameters such as mean, median, or mode, or employing more sophisticated approaches like interpolation or predictive modelling. After that, identifying and eliminating duplicate data to mitigate selection bias in the model.

3.1.4 Modelling

One of the main stages of data processing, modeling data is sometimes referred to as the "heart" of data analysis. It also entails a variety of activities, such as determining which mode type is best and figuring out if the issue is one of clustering, regression, or classification. Selecting the algorithms to use them requires analysis of the model family. It must be done carefully because it is crucial to get relevant insights from the produced data.

3.1.5 Evaluating

In this instance, the model undergoes evaluation to see if it is ready for deployment. The model is tested on a set of unknown data and assessed using a carefully considered set of evaluation measures. We also need to confirm that the model accurately reflects reality. Like humans, every data science solution, including machine learning models, must be able to change over time, improve with new data, and adjust to new assessment metrics.

3.1.6 Model Development

In the data science life cycle, this is the final stage. The model is now ready to be deployed in the appropriate format and channel following a thorough evaluation procedure. As a result, machine learning models must be documented prior to deployment.

3.2 Data Collection Method

Throughout the entire study, data were gathered using secondary as well as primary data. Reliable information and a deeper comprehension of the variables are provided by the data gathered using these two approaches. The primary source of data is acquired by way of the targeted participants, but data from journals and articles is considered secondary.

3.2.1 Primary Data

First-hand information obtained by the researchers themselves is referred to as primary data and is known as real-time data [6]. Since it takes less money and additional skill, online questionnaires are used to gather primary data. Undergraduate students fill out questionnaires to provide primary data.

3.2.2 Secondary Data

Data that has already been gathered by somebody else is referred to as secondary data and is known as past data [6]. To have a deeper understanding of the variables influencing undergraduate academic success, secondary data is employed to evaluate the literature. The most obvious kind of secondary data employed in this study are journal articles.

3.3 Sampling Design

3.3.1 Target Population

This study examines the personal factors that influence the academic performance of university students. As a result, the population comprises undergraduates enrolled full-time at UTAR, representing various faculties and study years. Questionnaires are provided to undergraduates of UTAR in Kampar, Perak.

3.3.2 Sampling Technique

There are two distinct categories of sampling methods: probability sampling and non-probability sampling [7]. Since there would be no sampling frame of target respondents in this study due to the uncertain identity of the respondents, non-probability sampling is chosen. Non-probability sampling methods are characterised by their lack of objectivity compared to probability techniques. They involve the researcher selecting participants from a target population or having them self-select to participate in the study, rather than ensuring that every member of the population is invited to participate [7]. Therefore, convenience sampling was used in this study due to time and resource constraints.

3.4 Questionnaire Design

Target respondents are given online questionnaires to complete for this study. Essentially, the questions were adapted from earlier studies on the same topics. To make it easier for the respondent to read and complete the questionnaire, it is designed in English, a worldwide language. The goal of this questionnaire will be briefly explained to the respondents at the start of the questionnaire. The questionnaire is separated into three sections which are Section A – C. The inquiry in Section A obtains respondents' demographic data; Section B evaluates their ideas on personal factors (home environment, student interest, study habits) that affect academic performance, which are mentioned above as the independent variable; and Section C collects their most recent GPA score (Academic performance) as the dependent variable.

3.4.1 Adaptation of a Questionnaire from Publicly Existing Research

Variable	Adopted from
Home Environment	- Obeta, A (2014) [8]
Student Interest	- R. A. Kusurkar, O. T. Cate, C. M. P. Vos, P. Westers, and G. Croiset (2012) [9]
Study Habits	- Tus, J (2020) [10]

Table 3.1: Summary of Questionnaire Adaptation

3.5 Constructs and Measurement

The constructs that are analyzed in this research include the factors and academic performance of students [4] which home environment [1][2][8], student interest [1][2], study habits [2][4][9][10].

Two sections of an online questionnaire are utilized to gather primary data. The demographic data of the target respondents is required for the first component of the questionnaire, which uses nominal and ordinal scales. Using the 5-point Likert scale at interval measurement, the dependent and independent variables are measured in the second section of the questionnaire. As a result, respondents must select one of five options: On a scale of 1 (strongly disagree) to 5 (strongly agree), indicate the level of agreement. Informing the queries will be previous journals.

3.5.1 Examples of Questionnaire Topics Used in Actual Research

Factor	Sample Item
Home Environment	1. Poor Parents/student relationship in the home. 2. Socio-economic status of the student's family has influence on the students' academic performance. 3. Adequate Provision of the student's educational needs, teaching, and supervision of the students work at home by parents can enhance the student's academic performance. 4. Educated and high income earner parents can provide the educational needs of the student.

	5. Family size and Parental attitudes (interest) towards the student's academic work/activities enhances student's academic performance. 6. The academic level of the student's parents and positive attitude towards education can enhances the student's academics performances.
Student Interest	1. I find that at times studying gives me a feeling of deep personal satisfaction. 2. I feel that virtually any topic can be highly interesting once I get into it. 3. I find most new topics interesting and often spend extra time trying to obtain more information about them. 4. I find that studying academic topics can at times be as exciting as a good novel or movie. 5. I work hard at my studies because I find the material interesting. 6. I spend a lot of my free time finding out more about interesting topics which have been discussed in different classes. 7. I make a point of looking at most of the suggested readings that go with the lectures.
Study Habits	1. When my assigned homework is extra-long or unusually hard, I either quit or study only the easier parts of the lesson. 2. If I have to be absent from class, I make up missed lessons without being reminded by the teacher. 3. I do not bother to correct errors on the papers of my teachers have graded and returned to me. 4. It takes a long time for me to get warmed up to the job of studying. 5. I am unable to study well because I get restless, moody, or have the blues. 6. I put off doing written assignments until the last minute.

	7. I do poorly on tests because I find it hard to think clearly and plan my study within a short period of time.
Academic Performance	1. What is your latest academic performance (GPA)?

Table 3.2: Sample of Questionnaire From Research Paper

3.6 Data Analysis Techniques

In this study, data analysis was performed using Google Form Analytics, Microsoft Excel, and SPSS software. The model assessment outcome was generated based on the questionnaire results. All primary data are summarized by using appropriate descriptive and inferential statistics.

3.6.1 Descriptive Analysis

Descriptive analysis summarises quantitative data using tables and graphics. Data summarization often involves computing the mean and standard deviation. Descriptive statistics are typically used to examine independent variables' averages and standard deviations, whereas frequency distributions are used to summarise respondents' demographics. Using descriptive analysis, the characteristics of the respondents were examined about their age, gender, and academic performance.

3.6.2 Inferential Analysis

Inferential statistics are utilised to ascertain the correlation between two variables, evaluate hypotheses, and draw conclusions at the population level. The hypothesis is tested in this study using Pearson Correlation and Multi Linear Regression (MLR) Analysis.

3.6.2.1 Multi Linear Regression (MLR)

With the help of information from independent or predictor variables, multiple regression analysis (MR) is used to determine how variable a dependent or criteria variable is [5]. Since all of the variables in this study are measured using interval data, multilinear regression (MLR) analysis is utilized to examine the relationships between the personal factors that have an impact on student performance. multilinear regression enables the investigation of the effects

of two or more independent variables on a single interval scale dependent variable at the same time. Because of this, multilinear regression is used to assess how home environment, student interest, and study habits affect academic performance.

The multiple regression analysis model is formulated as below:

$$y = B_0 + B_1x_1 + ... + B_nX_n + E$$

y = dependent variable

xi = independent variable

B = the slope for the independent variable

n = number of independent variables in the equation

E = Error

3.6.2.2 Pearson Correlation

In addition, a Pearson correlation analysis was carried out to identify cases of multicollinearity, or the presence of a significant relationship between two independent variables. Multicollinearity may result in more complex results from research, making it more challenging for researchers to interpret the data [11]. The meanings of the correlation coefficients are provided in the table below.

Size of Correlation	Interpretation
0.00 to 0.30	Negligible correlation
0.30 to 0.50	Low correlation
0.50 to 0.70	Moderate correlation
0.70 to 0.90	High correlation
0.90 to 1.00	Very high correlation

Table 3.3: Interpretation of Pearson Correlation Coefficient

3.7 Conclusion

This chapter concluded with a thorough explanation of the techniques applied to the data gathered. In addition to going over the methods in depth, more explanations of each method used were also given. In the subsequent chapters, the structure and analysis of the data gathered from the pertinent target respondents will be elaborated.

CHAPTER 4

Research Framework for Data Analysis

This chapter is the the connection between implementation and theory. This section establishes a link between the established methodologies and theoretical framework of the preceding chapters and the tangible components of the project. This serves as a transitional stage between the project's "planning" and "execution" phases. Besides, this chapter also demonstrates the initial findings and detailling experimental setup. For projects involving experimental work, this chapter may include thorough descriptions of the setup, calibration details, and preliminary testing performed to confirm that everything works well. In this section, the research framework is explained in detail based on my study.

4.1 Business Understanding

This stage is to identify and list the issue that needs to be resolved which is predicting student performance based on personal factors. The literature review is conducted based on my topic. From the 50 research papers are collected, 30 articles are related to my study, but only 15 of the research papers are suitable for references. Researchers investigate personal factors such as study habits, home environments, and student interests, which have been identified as influencing elements of academic achievement. This conclusion was reached after conducting a comprehensive literature review. Additionally, it came to my attention that the researcher was on the verge of employing GPA or CGPA as metrics for academic performance. It was subsequently discovered that the predominant predictive analysis instrument employed by the majority of researchers was multiple linear regression. In the end, the null hypothesis was formulated in line with the researcher's paper. The statistical significance of the association between the response and each term in the model is determined by comparing the term's p-value to my significance level (p value < 0.05) to reject or accept the null hypothesis.

4.2 Data Understanding

After making the hypothesis, now is the time to collect the data. To begin, it was imperative to ascertain the specific data category that was necessary for the purpose of analysis. Datasets on student demographics, academic performance, and other pertinent personal factors such as home environment, student interest, and study habits are collected in this process. Undergraduate UTAR students fill out my online questionnaires to provide primary data. The questionnaire is designed in English, which is a worldwide language, convenient for other international country students to fill in. The questionnaire is separated into three sections, the first section collects students' demographic data such as age, gender, etc, the second section collects their ideas about will personal factors can affect academic performance, and lastly collects their latest GPA. Responders must select one of five options: With 1 (strongly disagree) to 5 (strongly agree), indicate the level of agreement. The dependent and independent variables are subsequently assessed in the second section of the questionnaire using an interval 5-point Likert scale. The questions that were asked in the online questionnaire will be adapted from past journals as approval of reliability. A total of 108 questionnaires are given to Kampar's undergraduate students at UTAR. All of the surveys are successfully collected because they are collected on the spot, and this indicates a 100% response rate. 7 of the have 4 were missing values and SPSS determined 3 value that they were outliers.

	Time stamp	Email	age	Gender	Faculty	HE1	HE2	HE3	HE4	HE5	HE6	HE	SI1	SI2	SI3	SI4	SI5	SI6	SI7	SI	SH1	SH2	SH3	SH4	SH5	SH6	SH7	SH	GPA
88	1...	yzl...	18...	Male	Facul...	5	5	5	5	5	4	4.833	5	3	4	3	4	4	5	4.000	4	2	2	3	4	2	4	3.000	2.000
89	1...	xia...	18...	Male	Facul...	4	4	5	4	5	3	4.167	4	3	5	3	4	2	4	3.571	2	4	2	2	4	2	2	2.571	4.000
90	1...	pin...	18...	Male	Facul...	4	4	4	3	4	4	3.833	2	2	3	2	1	2	3	2.143	2	3	2	4	4	4	4	3.286	2.670
91	1...	ywb...	18...	Male	Facul...	3	3	5	2	3	2	3.000	2	4	3	2	3	3	2	2.714	5	5	3	4	2	4	3	3.714	3.500
92	1...	vinc...	18...	Male	Facul...	3	2	3	4	3	4	3.167	4	2	4	3	4	2	3	3.143	2	3	4	3	2	2	2	2.571	4.000
93	1...	suu...	22...	Fem...	Facul...	1	3	4	3	3	3	2.833	4	2	2	1	2	1	2	2.000	3	3	4	3	2	4	2	3.000	3.740
94	1...	qiu...	22...	Fem...	Facul...	2	4	5	4	5	4	4.000	2	3	2	4	2	2	1	2.286	2	5	2	3	3	2	5	3.143	.
95	1...	hus...	22...	Male	Facul...	2	5	2	3	5	4	3.500	2	2	3	1	1	1	2	1.714	4	2	3	5	5	4	4	3.857	2.880
96	1...	yun...	18...	Male	Facul...	4	2	4	5	5	5	4.167	2	2	3	4	5	4	2	3.143	3	2	2	4	4	5	5	3.571	2.950
97	1...	che...	18...	Male	Facul...	5	3	5	5	5	2	4.167	2	5	3	5	2	5	5	3.857	4	4	1	3	5	2	3	3.143	3.600
98	1...	liew...	22...	Fem...	Facul...	5	4	4	4	4	4	4.167	3	3	3	2	2	2	2	2.429	2	4	1	3	3	2	2	2.429	3.500
99	1...	ngy...	18...	Fem...	Facul...	3	2	4	4	4	3	3.333	4	2	3	3	4	3	3	3.143	2	5	1	2	2	2	1	2.143	3.600
100	1...	yan...	18...	Fem...	Foun...	3	3	4	4	4	4	3.667	3	3	3	2	4	3	2	2.857	2	3	2	4	4	2	2	2.714	3.570
101	1...	wiki...	18...	Fem...	Facul...	3	3	4	4	4	4	3.667	4	4	4	3	3	3	2	3.286	4	3	2	3	3	3	3	3.000	3.330
102	1...	sea...	22...	Male	FMHS	3	4	4	4	4	4	3.833	3	4	4	3	4	4	4	3.714	3	4	3	3	3	3	3	3.143	3.330
103	1...	mei...	22...	Fem...	Facul...	4	3	3	3	2	4	3.167	2	3	4	5	4	5	4	3.857	4	2	2	3	3	2	5	3.000	2.500
104	1...	dsf...	22...	Male	Facul...	1	2	2	4	4	5	3.000	3	4	3	3	3	4	1	3.000	2	2	1	4	4	4	4	3.000	2.930
105	1...	yon...	18...	Fem...	Foun...	1	2	3	4	4	4	3.000	3	3	3	3	3	2	2	2.714	3	2	3	4	4	4	2	3.143	3.650
106	1...	mo...	18...	Male	Facul...	2	4	4	4	4	2	3.333	3	1	4	1	3	1	2	2.143	2	4	2	5	1	5	1	2.857	3.560
107	1...	rog...	18...	Male	Facul...	4	3	3	2	2	3	2.833	2	3	2	3	2	2	3	2.429	3	3	3	3	2	2	3	2.714	3.330
108	1...	zhl...	22...	Male	Facul...	1	4	3	3	4	1	2.667	1	5	1	3	3	5	4	3.143	1	5	4	2	3	4	5	3.429	4.000
109																													
110																													

Figure 4.1: Evidence of proving 108 Respondents is Collected

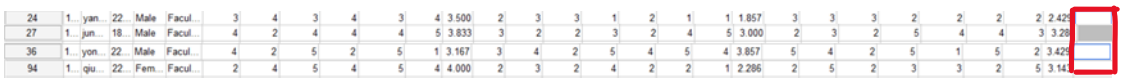


Figure 4.2: Total of respondents of Missing Value in GPA

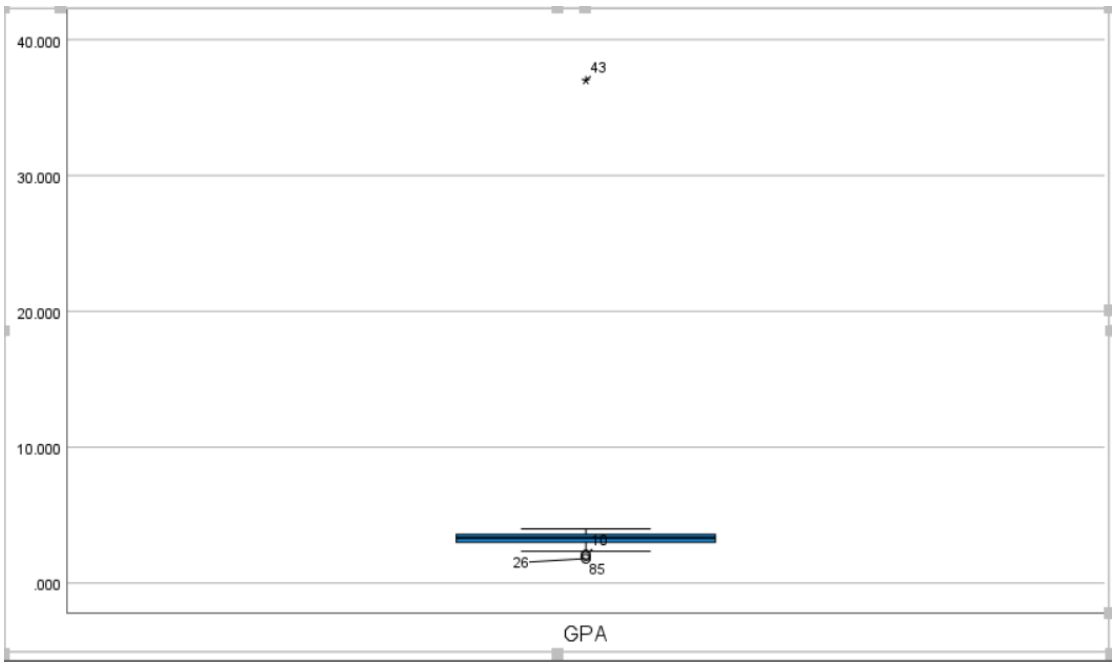


Figure 4.3: Number of respondents that are outliers in GPA

4.2.1 Descriptive Analysis

This study provides a descriptive examination of the respondent's demographic profile and the construct's central tendency assessment. The target respondents' demographic variable is their statistical data. Stated differently, the variables that are measured include age, gender, and faculty of respondents' main study location. Respondents are classified based on the characteristics needed for the tests as demographic information is provided, which facilitates the elimination process and visualization process.

Age	Frequency	Percent	Cumulative Frequency	Cumulative Percentage
18-21	66	65.3	66	65.3
22-25	33	32.7	99	98.0
More than 26	2	2.0	101	100
Total	101	100		

Table 4.1: Age of the Respondents

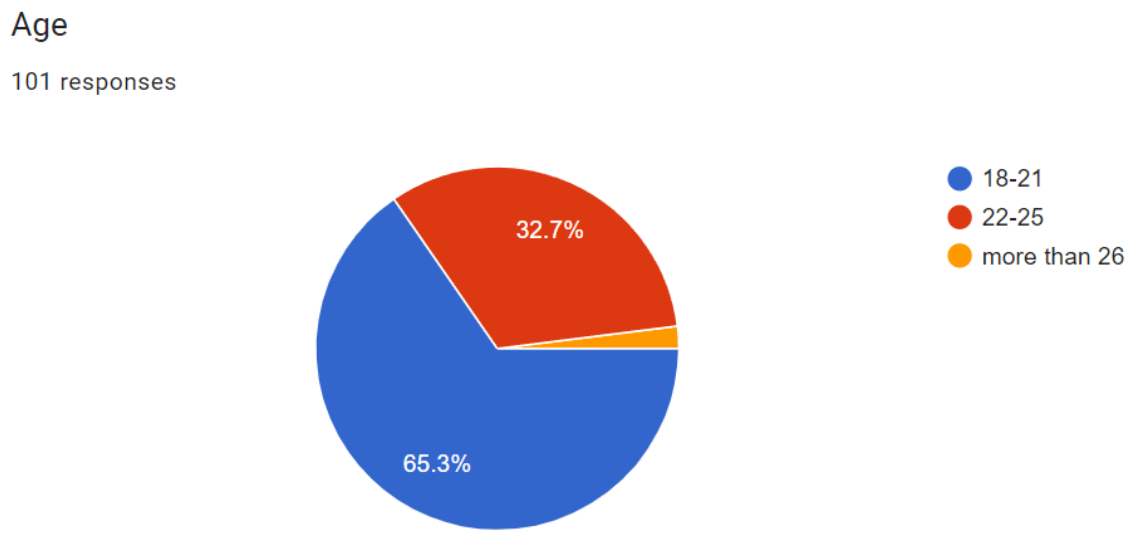


Figure 4.4: Age of the Respondents

The age distribution of the respondents is displayed in Figure 4.4. The greatest frequency of respondents—65.30% (66 respondents)—are between the ages of 18 and 21. The next largest frequency—32.70% (33 respondents)—are between the ages of 22 and 25. Approximately 2.00% (2 respondents) of the total respondents are older than 26.

Gender	Frequency	Percent	Cumulative Frequency	Cumulative Percentage
Male	58	57.4	58	57.4
Female	43	42.6	101	100
Total	101	100		

Table 4.2: Gender of the Respondents

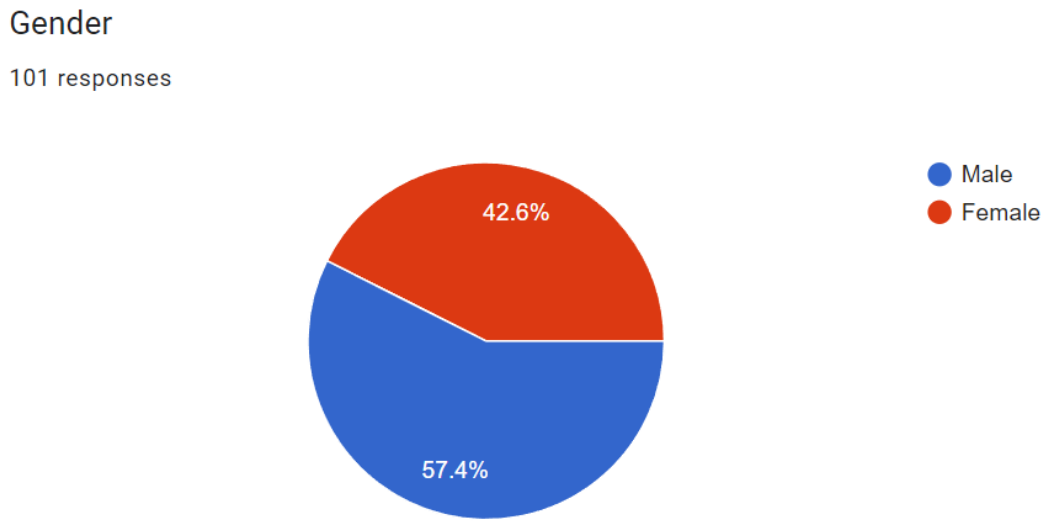


Figure 4.5: Gender of the Respondents

The gender frequency of the respondents is illustrated in Figure 4.5. The pie chart illustrates that the male gender comprises the majority of respondents (57.40% or 58 participants), while the female gender comprises 42.60% or 43 participants.

Faculty	Frequency	Percent	Cumulative Frequency	Cumulative Percentage
FICT	35	34.7	35	34.7
FBF	25	24.8	60	59.5
CFS	19	18.8	79	78.3
FSC	9	8.9	88	87.2
FAS	7	6.8	95	94.0
FEGT	3	3.0	98	97.0
ICS	2	2.0	100	99.0
FHMS	1	1.0	101	100
Total	101	100		

Table 4.3: Faculty Studies of the Respondents

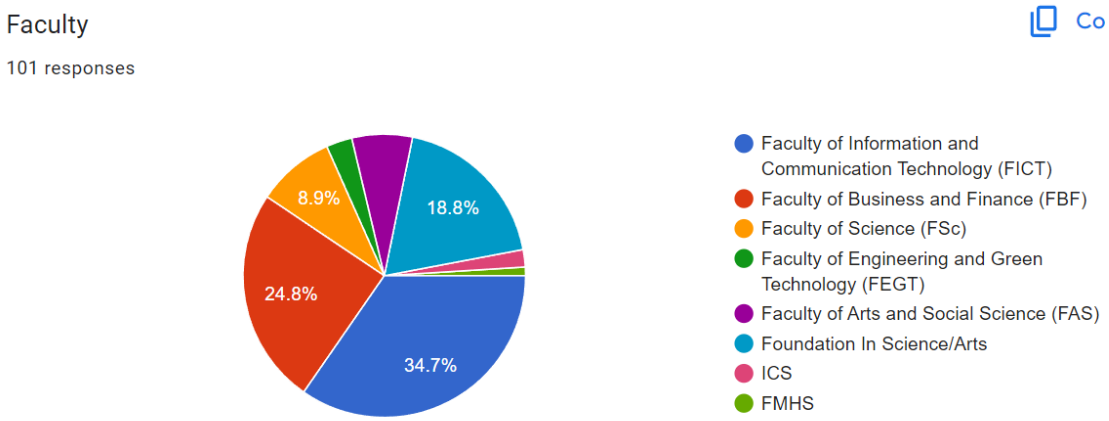


Figure 4.6: Faculty Studies of the Respondents

According to Table 4.3 and Figure 4.6, the majority of respondents (34.7% and 35 out of 101, respectively) come from the Faculty of Information and Communication Technology (FICT). In addition, 10 respondents from the Centre for Foundation Studies (CFS) and 29 respondents from the Faculty of Business and Finance (FBF) made up 18.8% and 24.8% of all respondents, respectively. On the other side, the Faculty of Science (FSC) received 9 responses out of 103, accounting for 8.9% of all responses. Meanwhile, the Faculty of Engineering and Green Technology (FEGT) and the Faculty of Arts and Social Science (FAS) received 3 and 7 replies, accounting for 3.0% and 6.8% of total respondents, respectively. Furthermore, two respondents

come from the Institute of Chinese Studies (ICS) and two from the Faculty of Medicine and Health Sciences (FHMS), accounting for 2.0% and 1.0% of total respondents, respectively.

4.2.2 Central Tendencies Measurement of Variables

Variables	Frequency	Mean	Standard Deviation
Home Environment	101	3.612	0.566
Student Interest	101	2.785	0.683
Study Habits	101	3.017	0.586
Academic Performance	101	3.272	0.412

Table 4.4: Descriptive Statistic on Variables

According to Table 4.5, there is a descriptive statistic of personal factors such as home environment, student interest, study habits, and lastly academic performance. A collection of responses that provide an overview of the sample's data is referred to as this statistic. The highest value, 3.612, was achieved by the mean of Home Environment in the preceding table, followed by Academic Performance and Study Habits which are 3.272 and 3.017 respectively. On the other hand, Study Interest has the lowest mean value among these four variables which is 2.785.

In statistics, the standard deviation is a measure of how much a random variable is projected to differ from its mean. Student Interest had the highest standard deviation value of 0.683, indicating that the data are dispersed apart from large values and that the collected data are distinct from one another. In addition, the standard deviation of Study Habits and Home Environment is 0.586 and 0.566 respectively. In the meantime, Academic Performance has the lowest standard deviation (0.412), indicating that the information gathered about Academic Performance is comparable to one another and extremely near to the same value.

4.2.3 Pearson's Correlation

Variable	Home Environment	Student Interest	Study Habits
Home Environment	1	0.192	0.089
Student Interest		1	0.127
Study Habits			1

Table 4.5 Pearson's Correlation

As demonstrated in the above table, multicollinearity does not exist among any of the independent variables, because the values all less than 0.30, which are negligible correlation.

4.3 Data Preprocessing

In the process of handling missing values and preparing the dataset for analysis, a structured and meticulous approach is adopted, ensuring accuracy and integrity in the data preprocessing stage. The initial step involves a thorough examination for missing values within the dataset. 7 of the questionnaires that were gathered were removed since 4 were missing values and SPSS determined 3 value that they were outliers. As a result, only 101 of the surveys could be carried out.

After the identification of missing values, attention is directed towards rectifying column names. This step is vital to maintain consistency and avoid potential errors in data processing. Any leading spaces in the column names are meticulously corrected, ensuring uniformity and accuracy in naming conventions across the dataset. The integrity of the data is further bolstered by eliminating duplicate records.

4.4 Modelling

4.4.1 Multiple Linear Regressions

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1.346	3	0.449	2.780	0.045^{b}
	Residual	15.650	97	0.161		
	Total	16.996	100			

a. Predictor: (Constant), Home Environment, Student Interest, Study Habits

b. Dependent variable: GPA

Table 4.6: ANOVA for Multiple Regression Analysis

The F value is 2.780, and the corresponding significance level is <0.045, which is below the alpha value of 0.05, as shown in Table 4.7. Therefore, the overall regression model effectively explains the variability in GPA using the following three predictors: home environment, student interest, and study habits.

		Unstandardized Coefficients		Standardized Coefficients		
Model		B	Std.Error	Beta	t	Sig.
1	(Constant)	3.468	0.333	-	10.420	<0.001
	(HE)	-0.035	0.072	-0.048	-0.478	0.633
	(SI)	0.132	0.060	0.219	2.192	0.031
	(SH)	-0.145	0.069	-0.207	-2.100	0.038

Table 4.7: Coefficients of Multiple Regression Analysis

Output from Table 4.8, the equation is formed as below:

GPA = 3.468 + 0.132(Student Interest) + -0.145 (Study Habits)

4.4.2 Hypothesis

Hypothesis 1

H_0: The home environment does not affect the academic performance among UTAR students.

H_1: The home environment affects the academic performance among UTAR students.

Conclusion: Accept the null hypothesis.

The obtained HE value of 0.633 surpasses the predetermined significance level of 0.05. Consequently, the evidence presented in this sample is adequate to substantiate the null hypothesis, H_0.

Hypothesis 2

H_0: The student's interest does not affect the academic performance among UTAR students.

H_1: The student's interest affects the academic performance among UTAR students.

The SI obtained as a significant value of 0.031 falls below the predetermined significance level of 0.05. Moreover, the inference that academic performance among UTAR students is influenced by student interest is supported by the evidence presented in this sample data, which is adequate to reject the null hypothesis, H_0.

Hypothesis 3

H_0: The study habits do not affect the academic performance among UTAR students.

H_1: The study habits affect the academic performance among UTAR students.

SI was determined to have a significant value of 0.038, which is below the predetermined significance level of 0.05. Furthermore, the evidence presented in this sample is adequate to refute the null hypothesis, H_0, and support the conclusion that academic achievement is indeed influenced by study habits among students enrolled at UTAR.

4.4.3 Conclusions

The Pearson Correlation was employed in this chapter to analyse the association among the subsequent independent variables: student interest, home environment, and study practices. Besides, the MLR also run to indicate which personal factors are significantly associated with GPA, which are student interest and study habits affect academic performance. The predictive of the model and model development in Power Bi for visualisation will be continued in the Chapter 5.

CHAPTER 5

Development

5.1 Data Preprocessing and Modelling

To construct a predictive model in Power BI using the survey that collected, which includes independent factors such as Home Environment (HE), Student Interest (SI), and Study Habits (SH), and a dependent variable GPA, by follow these instructions:

1. **Prepare the Dataset**
   - To examine the dataset, access the Excel file.
   - Ensure that each row corresponds to the data of a student, and that the independent variables (HE, SI, SH) and the dependent variable (GPA) are organized in columns.
   - Perform data cleansing to verify the absence of any missing or erroneous records.

2. **Load the Dataset into Power BI**
   - To import the Excel file, open Power BI and select 'Get Data' > 'Excel'.
   - Choose the document Excel that contains the data and import it into Power BI.

3. **Create a Regression Model**
   - A Multiple Linear Regression analysis is conducted on a dataset by fill in Python script to model the relationship between a dependent variable (GPA) and three independent variables (HE, SH, SI).
   - Add the python script into the Power BI:

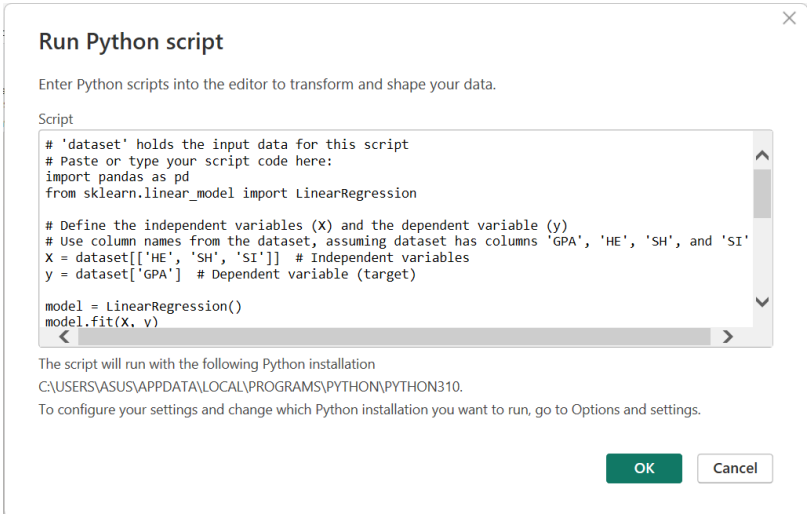


Figure 5.1 Python Script into Power BI

```
# 'dataset' holds the input data for this script
# Paste or type your script code here:
import pandas as pd
from sklearn.linear_model import LinearRegression

# Define the independent variables (X) and the dependent variable (y)
# Use column names from the dataset, assuming dataset has columns 'GPA', 'HE', 'SH',
and 'SI'
X = dataset[['HE', 'SH', 'SI']]  # Independent variables
y = dataset['GPA']  # Dependent variable (target)

model = LinearRegression()
model.fit(X, y)

# Make predictions
y_predictions = model.predict(X)

# Get model coefficients
coef = model.coef_
intercept = model.intercept_

# Create a DataFrame to store coefficients and intercept
coefficients = pd.DataFrame({
  'component': ['HE', 'SH', 'SI', 'intercept'],
  'value': list(coef) + [intercept]
})
```

Figure 5.2 Python Script in Word File to Show the Code

- The script assumes that the dataset contains columns HE, SH, SI (independent variables) and GPA (dependent variable).
- X is established for storing the independent variables (HE, SH, SI), while y is designated to store the goal variable (GPA).
- Scikit-learn initialises a linear regression model.
- Using the independent variables (HE, SH, SI), the model.fit(X, y)function trains the regression model by identifying the best-fitting line (or hyperplane) that minimises the error between actual and predicted GPA values.
- Following the training of the model, the function model.predict(X) is employed to generate predictions on the GPA using the given values of HE, SH, and SI.
- The regression model calculates a coefficient for each independent variable (HE, SH, SI). These coefficients indicate how much GPA is expected to change with a one-unit change in each independent variable, holding others constant.

- An intercept is also derived. This is the anticipated grade point average (GPA) when all the independent coefficients are equal to zero.
- A DataFrame named coefficients is instantiated by the script to contain the coefficients of the independent variables and the intercept. Every row in the table represents either an independent variable or the intercept, and the corresponding coefficient or intercept value is recorded in the value column.

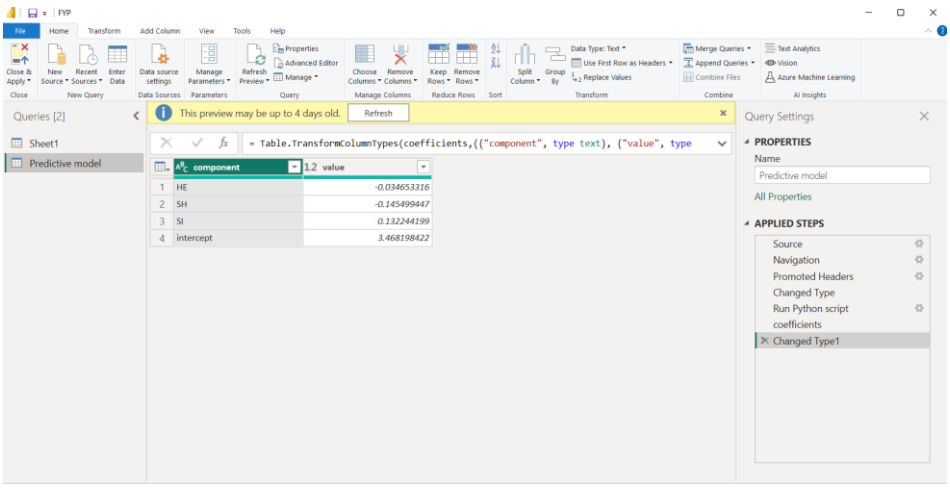


Figure 5.3 Coefficient of Independent Variables and Intercept

5.2 Predictive Models

Individual input of independent variable factors (Home Environment, Student Interest, Study Habits) in Power BI for GPA prediction can be accomplished through the use of parameters and What-If methodology. Below is a detailed, sequential manual on how to go with setting things up:

1. **Create What-If Parameters**

Power BI enables the creation of What-If parameters, which users may modify to evaluate the impact of various values of independent factors on the dependent variable (GPA). Procedure for generating What-If parameters for each independent variable:

- To use the What-If Parameter option in Power BI Desktop, navigate to the Modelling tab and select New Parameter.
- Specify a parameter for each independent variable (Home Environment, Student Interest, Study Habits):

1. Specify a parameter for each independent variable (Home Environment, Student Interest, Study Habits).
2. Designate the data type as to decimal values, based on the desired range of values.
3. Specify the suitable range according to the scale of the variable (e.g., for Likert scale, designate the minimum as 1 and the maximum as 5).
4. Place the increment (e.g., 1 for a Likert scale).
5. This procedure should be repeated for the remaining two variables, namely Student Interest (SI) and Study Habits (SH).

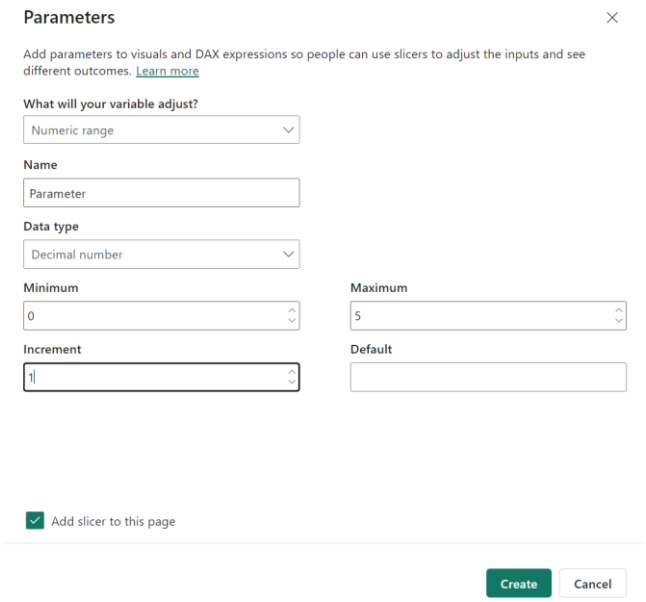


Figure 5.4 Create the Parameter for Independent Variables

2. **Add Slicers for Parameters**

- Upon creation, the parameters are displayed as newly generated fields within the Fields Pane.
- Then move each parameter (HE Parameter, SI Parameter, SH Parameter) into the designated report canvas.
- Specify the visual representation as Slicer. This functionality enables users to enter or adjust the values for each parameter.

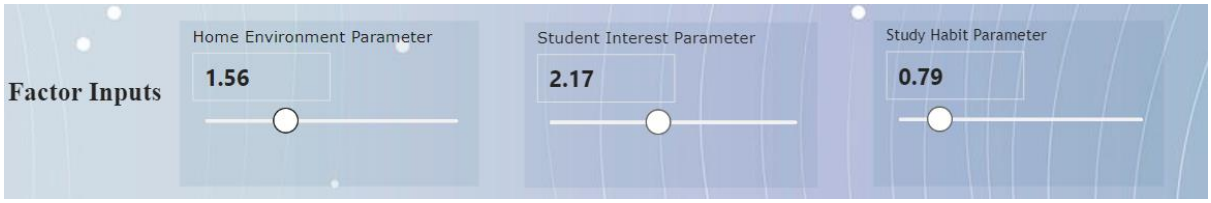


Figure 5.5 Parameter for Independent Variables

3. Create a Calculated Measure for Prediction

Given the set of parameters, generate a computed metric to forecast GPA using the user's inputs.

- Go to the Modeling tab and click on New Measure.
- Specify the formula for predicting GPA based on the regression model equation obtained from the investigation. For Example:

  GPA = Intercept + [HE] × Coefficient + [SI] × Coefficient + [SH] × Coefficient

```
1 Predictive GPA = [Home Environment Parameter Value]*[HE_coef] + [Student Interest Parameter Value]*[SI_coef] +
  [Study Habit Parameter Value]*[SH_coef] + [Intercept]
```

Figure 5.6 Formula for Predictive GPA

4. **Visualize the Prediction**

- The predictive GPA can be displayed using either a Card graphic or a Table.
- Allocate the recently generated metric (Predictive GPA) to the visual representation.
- Now, as users modify the sliders for Higher Education (HE), Student Intelligence (SI), and Student Health (SH), their predicted GPA will be automatically updated according to the inputted values.

Figure 5.7 Predictive GPA Models

5.3 Dashboard and Visualization

5.3.1 Dashboard of Predictive GPA Models

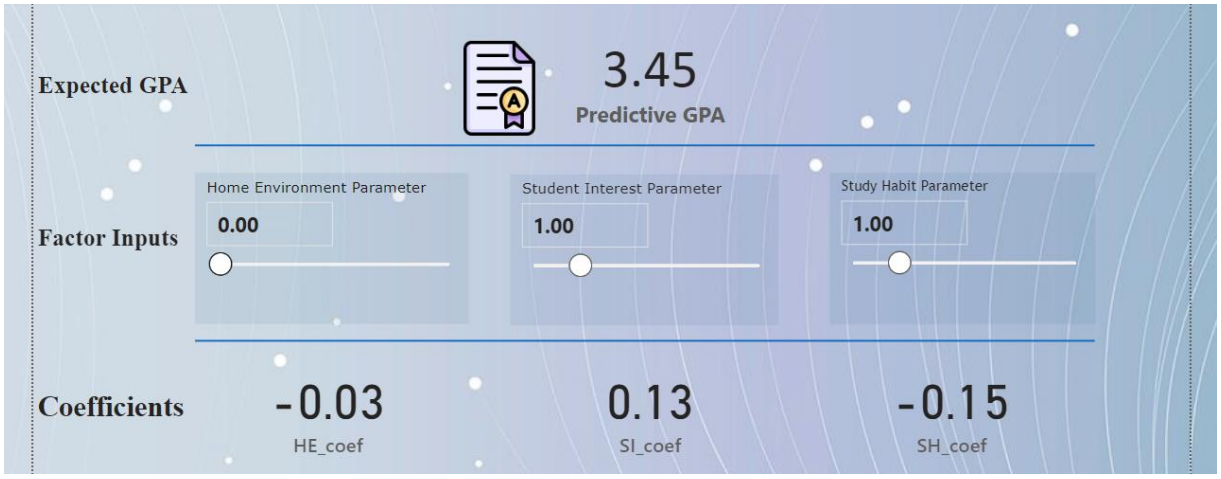


Figure 5.8 Dashboard of Predictive GPA Models

5.3.2 Rule Based of Predictive GPA Models

1. Home Environment:

Initially:

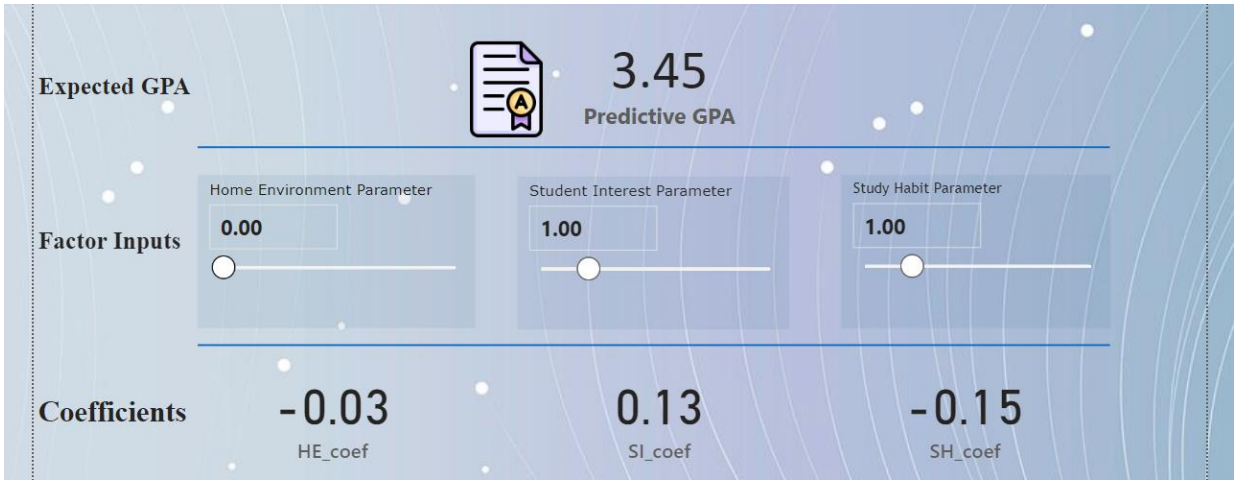


Figure 5.9 Predictive GPA Before Home Environment Factors Change

After:

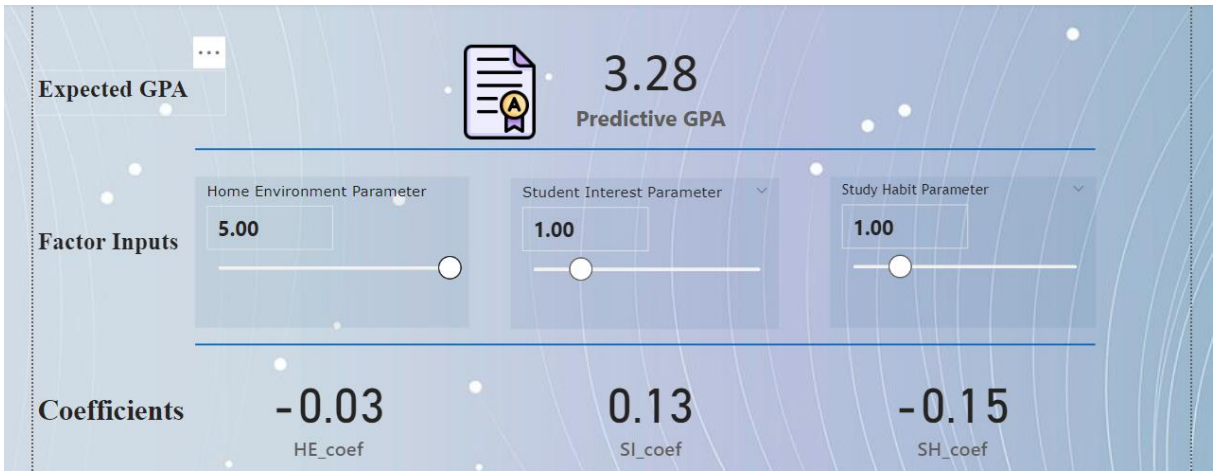


Figure 5.10 Predictive GPA After Home Environment Factors Change

When the Home Environment (HE) parameter is set to 0 (SI and SH constants), the predicted GPA is 3.45 before modification. Setting the HE parameter to 5 results in a reduction in the predictive GPA to 3.28. As the Home Environment factor increases, there is a very small decrease in the predicted GPA. These findings indicate an inverse correlation between the Home Environment (HE) factor and GPA. Comparing the expected GPA for HE = 0 and HE = 5 reveals the following gap: Difference=3.45−3.28=0.17, therefore, a 5-point rise in HE (from 0 to 5) results in a 0.17-point decrease in the expected GPA. These findings suggest that variations in the Home Environment are not significantly affecting the GPA. A p-value of 0.633 is relatively large and much above the conventional significance level, often established at 0.05. Hence, the correlation between Home Environment (HE) and GPA lacks statistical significance. Put simply, the research lacks compelling evidence that variations in the HE variable can significantly or reliably impact GPA. Despite the observed negative relationship between HE and GPA, the high p-value suggests that this relationship could be due to chance rather than a true causal effect. Thus, the existing data does not provide sufficient evidence to support the conclusion that the Home Environment has a statistically significant influence on GPA. In future work the HE will be removed because when non-significant variables are included, they might introduce noise and diminish the accuracy and simplicity of the model. Given that the p-value above the conventional threshold of 0.05, the evidence indicates that HE does not have a substantial impact on predicting GPA.

2. Student Interest:

Initially:

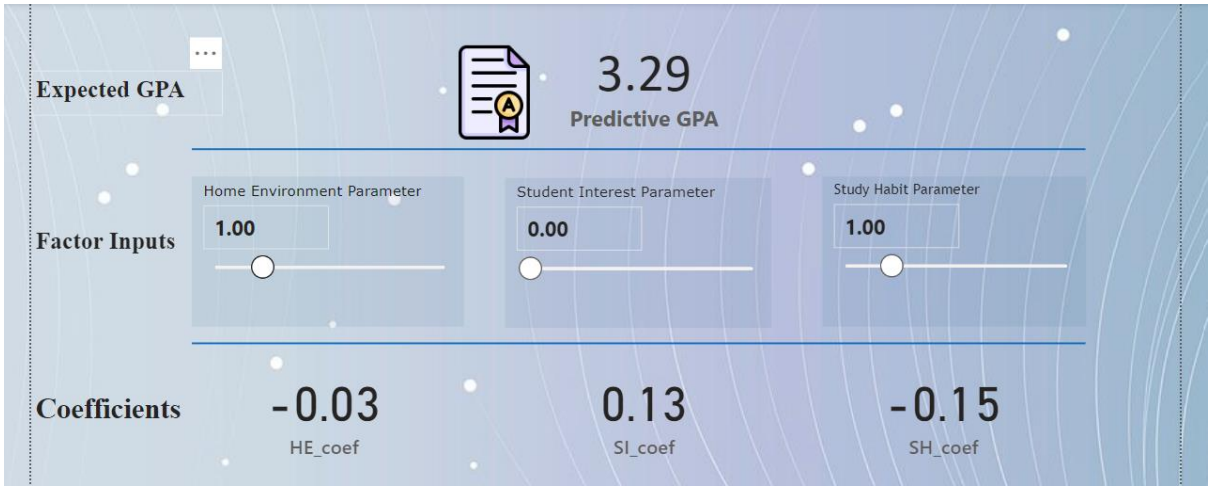


Figure 5.11 Predictive GPA Before Student Interest Factors Change

After:

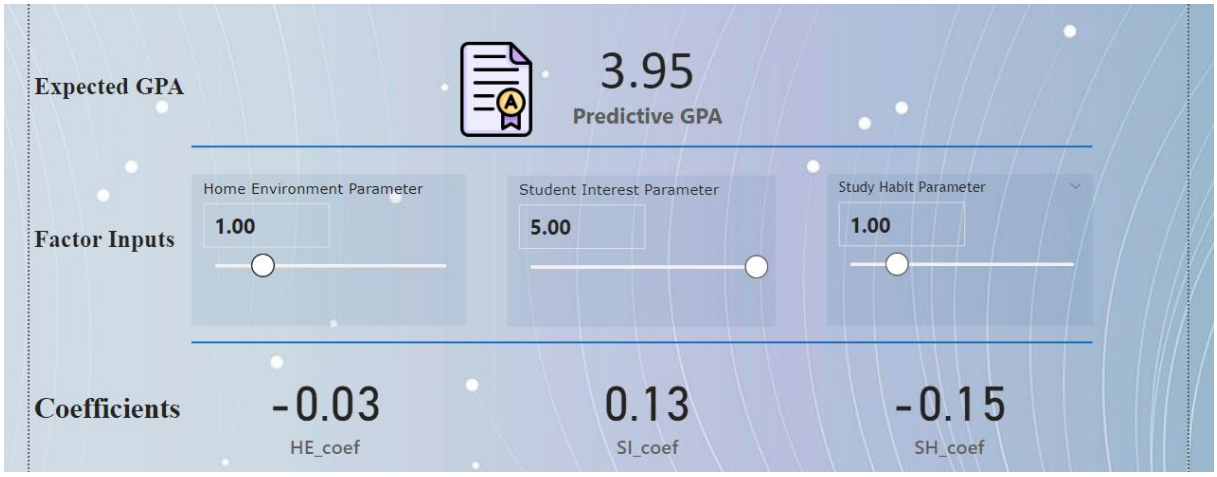


Figure 5.12 Predictive GPA After Student Interest Factors Change

When the Student Interest (SI) parameter is set to 0 (HE and SH constants), the predicted GPA is 3.29 before modification. Setting the SI parameter to 5 results in an increase in the predictive GPA to 3.95. This finding suggests a direct correlation between Student Interest and GPA, whereby a greater degree of interest exhibited by the student is linked to a higher predicted GPA. Comparing the expected GPA for SI = 0 and SI = 5 reveals the following gap: Difference=3.95−3.29=0.66, therefore, a 5-point rise in SI (from 0 to 5) results in a 0.66-point increase in the expected GPA. These findings suggest that variations in the Student Interest are positive significantly affecting the GPA. The obtained p-value of 0.031 falls below the conventional significance level of 0.05, so suggesting that there is a statistically significant

association between SI and GPA. Consequently, there is strong evidence indicating that Student Interest has a significant and dependable impact on prediction of GPA. The observed increase in GPA from 3.29 to 3.95 with the sequential increase in the SI parameter from 0 to 5 indicates a substantial predictive relationship between Student Interest and GPA.

3. Study Habit:

Initially:

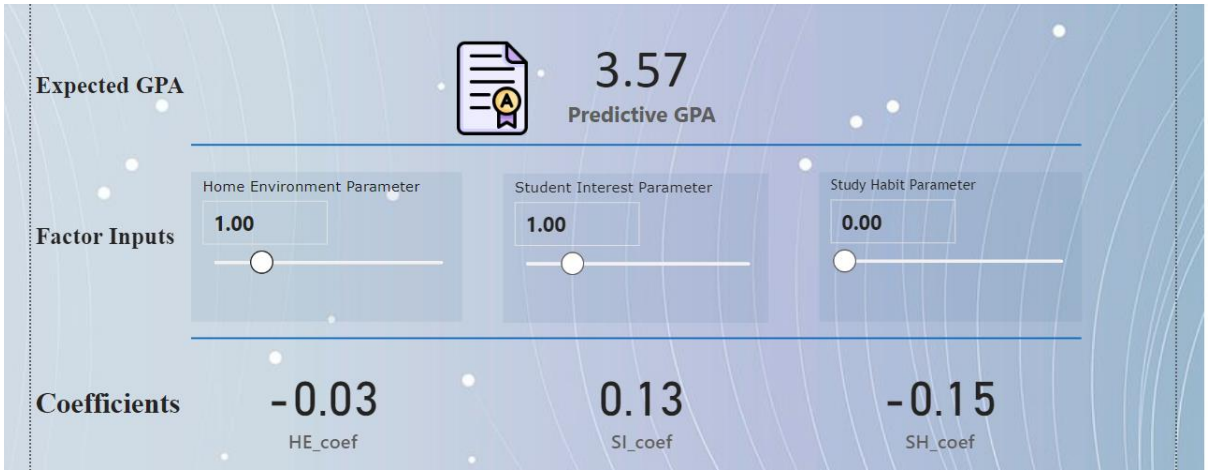


Figure 5.13 Predictive GPA Before Study Habit Factors Change

After:

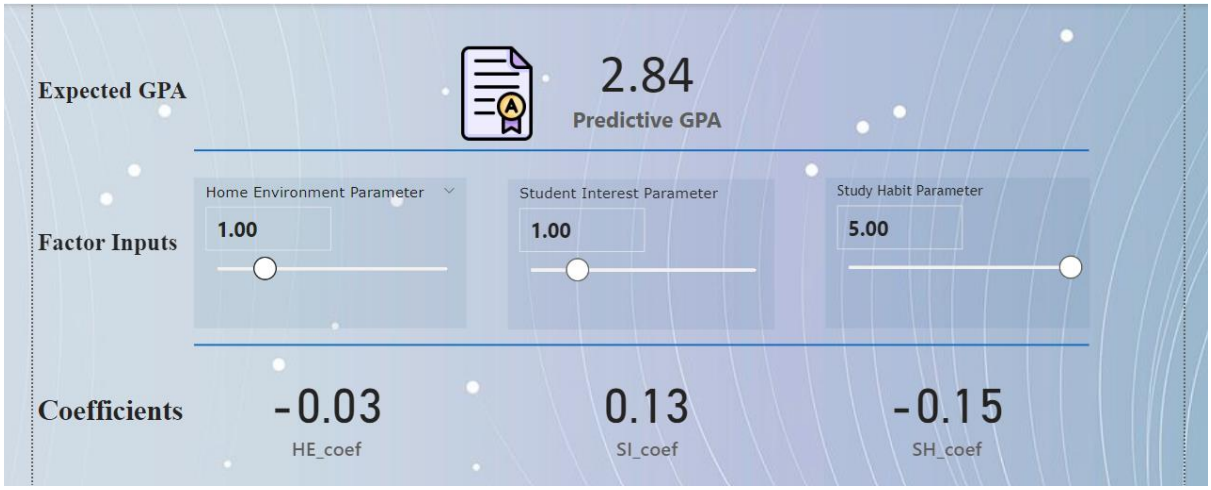


Figure 5.14 Predictive GPA After Study Habit Factors Change

When the Study Habit (SH) parameter is set to 0 (HE and SI constants), the predicted GPA is 3.57 before modification. Setting the SI parameter to 5 results in a reduction in the predictive GPA to 2.84. These results suggest a negative correlation between Study Habits and GPA, where a higher Study Habit score is associated with a lower predicted GPA. Comparing the

expected GPA for SH = 0 and SH = 5 reveals the following gap: Difference=3.57−2.84=0.73, therefore, a 5-point rise in SI (from 0 to 5) results in a 0.73-point decrease in the expected GPA. The obtained p-value of 0.038 falls below the conventional significance threshold of 0.05, indicating that there is a statistically significant association between Study Habits and GPA. These findings indicate that the adverse effect of Study Habits on GPA is not a result of random variation, and there is empirical evidence that supports the existence of this correlation in the dataset. The observed decreasing trend in GPA from 3.57 to 2.84 with increasing SH parameter from 0 to 5 indicates a negative correlation between higher Study Habits and lower GPA in this specific dataset. Although this finding is statistically significant, it is contradictory as we typically predict a positive correlation between good study habits and academic achievement. High Study Habits scores among students may indicate either excessive studying or the use of ineffective study methods, resulting in fatigue or poor academic performance. Data or measurement challenges may arise in the survey regarding the measurement of study habits, or the very notion of "study habits" may be encompassing unforeseen factors such as stress or imbalance.

5.3.3 Other Visualization

1. **Comparison between Actual and Predicted GPA**

- By graphically representing both the real GPA (derived from the dataset) and the predicted GPA (generated by the model), individuals can evaluate the performance of the model.
- Identification of possible gaps or biases in the model is facilitated by detecting substantial deviations between the forecasts and the actual data.
- A scatter plot is a graphical representation of the relationship between the real GPA as determined by the dataset and the predicted GPA as determined by the model. Achieving optimal alignment of the points along a diagonal line suggests that the forecasts are near the true values.

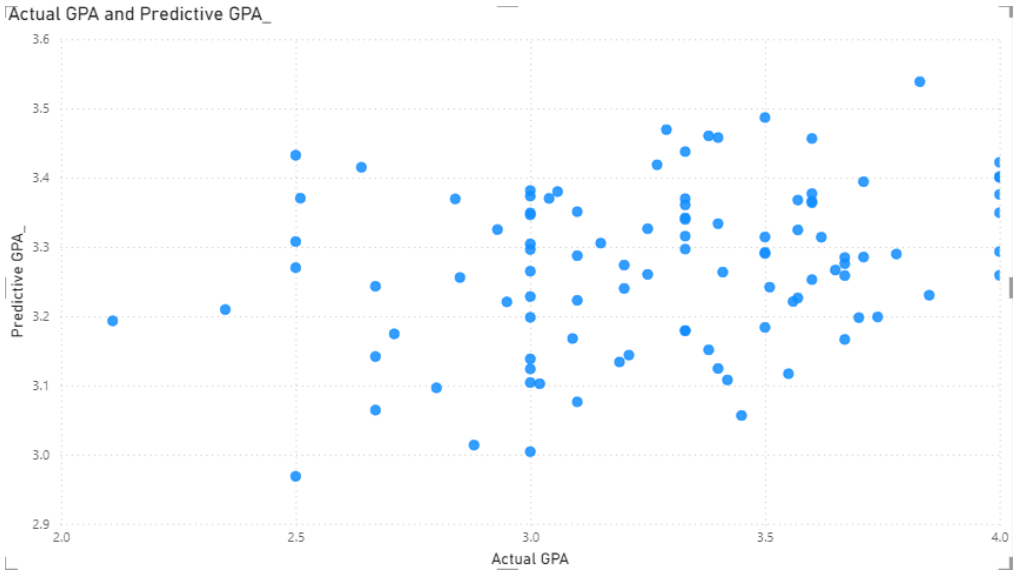


Figure 5.15 Predictive GPA vs Actual GPA

Within an ideal prediction model, among expects the points to closely align along a diagonal line where the Actual GPA is equivalent to the Predicted GPA. The dispersion of the data points in the scatter plot indicates that the model exhibits a certain degree of variability between the actual and predicted values. There exist regions in which the predictive model either underestimates or overestimates GPA, resulting in data points deviating from the optimal diagonal line. Statistical analysis of the data suggests that the predictive model does not accurately represent the actual variability in GPA. The model's accuracy decreases as the dispersion broadens. The plot in this case displays points that are situated at a significant distance from the diagonal in certain regions, suggesting a predictive accuracy that is moderate to low. The majority of actual GPA values are observed to be between 2.5 and 4.0, while the majority of predicted GPA values are between 3.0 and 3.6. In spite of the presence of actual GPAs, the predictive model appears to be less sensitive to lower GPA values (below 3.0), as there are few predictions in that range. This could suggest a bias in the model, as it consistently over-predicts or under-predicts specific ranges. In future work, reevaluate the independent variables (Home Environment, Student Interest, Study Habits) and potentially incorporate additional predictive factors if available to further refine the model.

2. **Comparison between Residual and Predicted GPA**

- The residual is the difference between the actual value and the predicted value.
- The model's error or inaccuracy is demonstrated by residuals. The model is making unbiased predictions when the residuals are modest and randomly distributed around zero. This is the ideal situation.
- Benefits of Visualizing Residuals:
    1. **Model Accuracy & Bias Detection**:
        - The model is functioning effectively when residuals are distributed randomly around zero.
        - This suggests that the model is biassed or fails to capture certain critical relationships in the data if the residuals exhibit a systematic pattern (e.g., residuals increase or decrease as predictive GPA increases or decreases).
    2. **Detecting non-linearity:**
        - The presence of patterns in the residuals (e.g., U-shape or inverted U-shape) may suggest that the relationship between the independent variables and the dependent variable (GPA) is not linear. In such instances, a linear model may not be the most appropriate fit, and a more intricate model could be investigated.

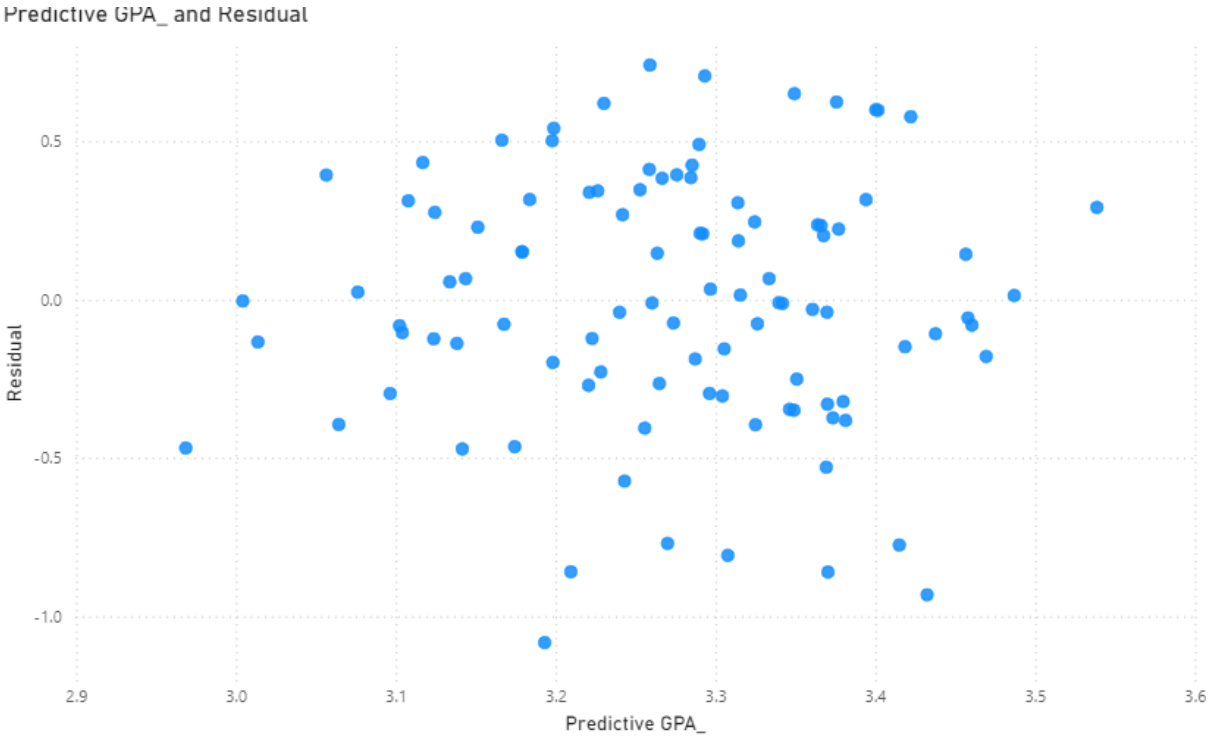


Figure 5.16 Residual vs Predictive GPA

The residuals are somewhat erratically distributed around the zero line, which is a positive indicator. This implies that the model is not systematically biassed, meaning that it is not consistently underpredicting or overpredicting. Nevertheless, the model has a moderate level of prediction error, as there is still some dispersal. The residuals do not exhibit a robust pattern or trend. For instance, there is no distinct funnel shape (which would suggest that prediction errors escalate with higher estimated GPA). This lack of pattern means that the model's errors are reasonably consistent across different predicted GPA levels. Certain data points deviate more severely from the zero line, particularly in the lower region (with residuals approximately -1). These examples illustrate situations when the model's predictions deviate significantly from the true values. In future work, undertaking a more thorough investigation of these outliers may be beneficial. Resolving the reasons behind the model's difficulty in predicting these points can enhance the model's accuracy.

Chapter 6

CONCLUSION AND FUTURE WORK

6.1 Conclusions

The first objective achieved by determine which factors have the impact on a student's academic performance, and the result is Student Interest and Study Habits. I conduct the literature review and it is undertaken in accordance with my chosen subject. Out of the 50 research papers in collection, 30 are relevant to my topic, but only 15 of them are appropriate for referencing. Most of the academic researchers examine personal variables such as study habits, home surroundings, and student interests, which have been recognised as determinants of academic success. This finding was derived from a thorough examination of existing literature. Furthermore, I became aware that the researcher was about to use GPA or CGPA as numerical measures for assessing academic achievement.

The second objective which is to develop the predictive model based on the identified personal factors by utilizing machine learning methods. After identifying those factors and parameters, I make the hypothesis and start to research the questionnaire from researcher for conduct a survey form and delivery it to hundreds of UTAR students. I also research the way or tools to analysis the questionnaire which are using SPSS and multiple linear regression. By using the Data Science Lifecycle, we analysis the data by understand business understanding, data understanding, data preprocessing, modeling, evaluation and development. I also study the Pearson correlation, Descriptive and Inferential Analysis. Lastly, we use MLR for modelling and conclude the equation GPA = 3.468 + 0.132(Student Interest) + -0.145 (Study Habits).
After discovering significant factors and coefficient, to create the predictive models by run Python script with Scikit-learn library and predict student performance (GPA) in POWER BI. By giving the formula, the predictive model to predict student performance was successful create.

The third objective to create an interactive dashboard that explains the model's predictions to educators in an efficient manner. is achieved by using Power BI. The Power BI dashboard serves as a comprehensive tool for educators and managers to easily visualise and interpret student performance in a glance of view.

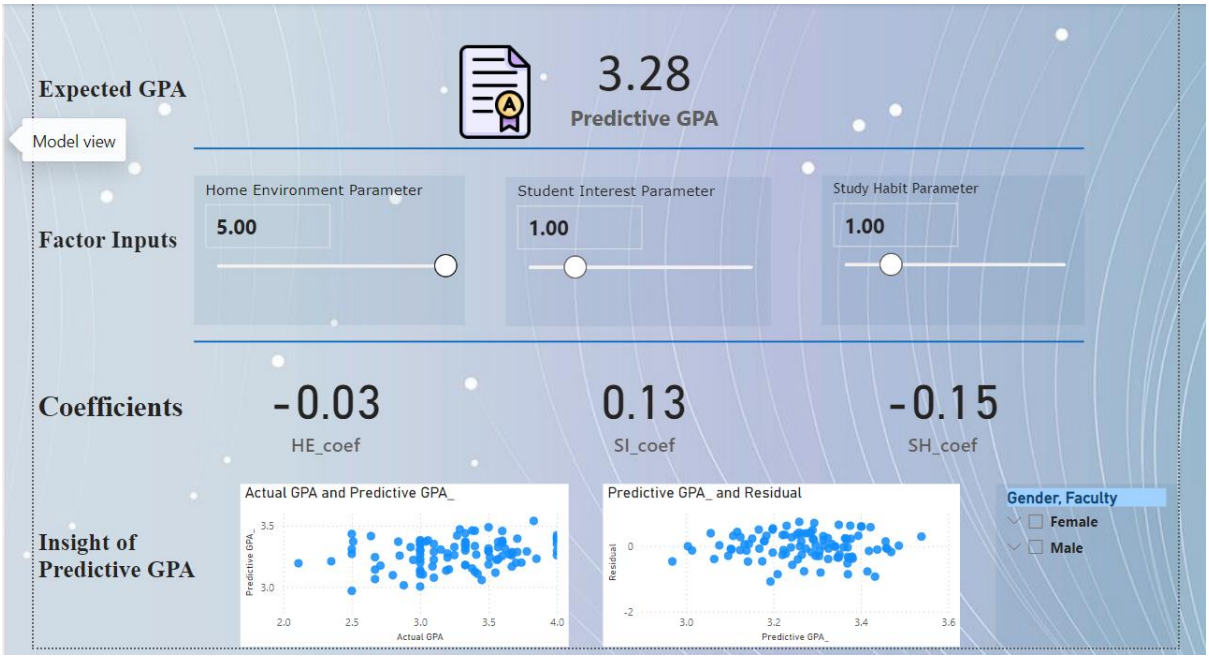


Figure 6.1 Overview of Dashboard

6.2 Future Works

1. Collect more data attributes

The characteristics utilised in this project are largely restricted to the demographic variables and solely include of combined GPA and CGPA. Personal Factors such as gender and health details may influence student performance, including academic results and timely graduation. Augmenting the data attribute can enhance the analytical outcome and reliability in forecasting student achievement.

2. Increase the Size of the Dataset

Larger datasets provide more examples for the model to learn from, reducing overfitting and improving the generalizability of the model to unseen data. With more data points, the model becomes better at recognizing patterns and trends, which can lead to higher accuracy and stability.

3. Improve Data Quality

Data of low quality, such as missing values, errors, or inconsistencies, might deceive the model and diminish its accuracy. Cleaner and more reliable data will give the model more trustworthy information to base predictions on, leading to more accurate results.

4. Broader area of analysis

This project involves the development of a dashboard that specifically analyses student performance in terms of academic achievement, and current study status. Nevertheless, there is potential for more investigation into other domains such as the attainment of program learning outcomes and course learning outcomes, which can be effectively examined by visualising students' coursework and assessment results.

REFERENCES

[1] M. Dev, "Factors Affecting the Academic Achievement: A Study of Elementary School Students of NCR Delhi, India.," Journal of Education and Practice, vol. 7, no. 4, pp. 70–74, 2016, Available: https://eric.ed.gov/?id=EJ1092343

[2] A. M. Adeyemi and S. B. Adeyemi, "Personal factors as predictors of students academic achievement in colleges of education in South Western Nigeria," ir.unilag.edu.ng, 2014, Accessed: Dec. 11, 2023. [Online]. Available: https://ir.unilag.edu.ng/items/625ec4a8-c2f4-4b96-a2f5-8cc76b2eff74

[3] N. Arora and N. Singh, "Factors Affecting the Academic Performance of College Students," Journal of Educational Technology, vol. 14, no. 1, pp. 47–53, 2017, Accessed: Mar. 20, 2022. [Online]. Available: https://eric.ed.gov/?id=EJ1268938

[4] S. K. Dr. M. A. Agarwal Dr. Nimmi, "Defining and Measuring Academic Performance of HEI Students-A Critical Review," turcomat.org, Apr. 2021, doi: 10.17762/turcomat.v12i6.6952.

[5] S. J. H. Yang, O. H. T. Lu, A. Y. Q. Huang, J. Huang, H. Ogata, and A. J. Q. Lin, "Predicting students' academic performance using multiple linear regression and principal component analysis," Journal of Information Processing, vol. 26, no. 0, pp. 170–176, Jan. 2018, doi: 10.2197/ipsjjip.26.170.

[6] "Primary and secondary data: Concepts, concerns, errors, and issues - ProQuest." https://search.proquest.com/openview/68bfe6580297494936f0126ed5d3d017/1?pq-origsite=gscholar&cbl=35147

[7] S. Stratton, "Population Research: Convenience sampling strategies," Prehospital and Disaster Medicine, vol. 36, no. 4, pp. 373–374, Jul. 2021, doi: 10.1017/s1049023x21000649.

[8] Obeta, A. (2014). Home environmental factors affecting students' academic performance in Abia State, Nigeria. [Online]. Available: https://llufb.llu.lv/conference/REEP/2014/Latvia-Univ-Agricult-REEP-2014proceedings-141-149.pdf

[9] R. A. Kusurkar, O. T. Cate, C. M. P. Vos, P. Westers, and G. Croiset, "How motivation affects academic performance: a structural equation modelling analysis," Advances in Health Sciences Education, vol. 18, no. 1, pp. 57–69, Feb. 2012, doi: 10.1007/s10459-012-9354-3.

[10] Tus, J. (2020, October). The Influence of Study Attitudes and Study Habits on the Academic Performance of the Students. [Online]. Available: https://www.researchgate.net/profile/Jhoselle-Tus/publication/344659816_The_Influence_of_Study_Attitudes_and_Study_Habits_on_the_Academic_Performance_of_the_Students/links/5f877620299bf1b53e28d4f5/The-Influence-of-Study-Attitudes-and-Study-Habits-on-the-Academic-Performance-of-the-Students.pdf

[11] N. Shrestha, "Detecting Multicollinearity in Regression Analysis," American Journal of Applied Mathematics and Statistics, vol. 8, no. 2, pp. 39–42, Jun. 2020, doi: https://doi.org/10.12691/ajams-8-2-1.

APPENDIX

Questionnaire Sample

1. Poor Parents/students relationship in the home.
A. Strongly disagree
B. Disagree
C. Neutral
D.Agree
E. Strongly Agree

2. Socio-economic status of the student's family has influence on the students' academic performance.
A. Strongly disagree
B. Disagree
C. Neutral
D.Agree
E. Strongly Agree

3. Adequate Provision of the student's educational needs, teaching, and supervision of the students work at home by parents can enhance the student's academic performance.
A. Strongly disagree
B. Disagree
C. Neutral
D.Agree
E. Strongly Agree

4. Educated and high income earner parents can provide the educational needs of the student.
A. Strongly disagree
B. Disagree
C. Neutral
D.Agree
E. Strongly Agree

5. Family size and Parental attitudes (interest) towards the student's academic work/activities enhances student's academic performance.
A. Strongly disagree
B. Disagree
C. Neutral
D.Agree
E. Strongly Agree

6. The academic level of the student's parents and positive attitude towards education can enhances the student's academics performances.
A. Strongly disagree
B. Disagree
C. Neutral
D.Agree

E. Strongly Agree
1. I find that at times studying gives me a feeling of deep personal satisfaction
A. Never or only rarely true of me
B. Sometimes true of me
C. True of me about half the time
D. Frequently true of me
E. Always or almost always true of me

2. I feel that virtually any topic can be highly interesting once I get into it.
A. Never or only rarely true of me
B. Sometimes true of me
C. True of me about half the time
D. Frequently true of me
E. Always or almost always true of me

3. I find most new topics interesting and often spend extra time trying to obtain more information about them.
A. Never or only rarely true of me
B. Sometimes true of me
C. True of me about half the time
D. Frequently true of me
E. Always or almost always true of me

4. I find that studying academic topics can at times be as exciting as a good novel or movie.
A. Never or only rarely true of me
B. Sometimes true of me
C. True of me about half the time
D. Frequently true of me
E. Always or almost always true of me

5. I work hard at my studies because I find the material interesting.
A. Never or only rarely true of me
B. Sometimes true of me
C. True of me about half the time
D. Frequently true of me
E. Always or almost always true of me

6. I spend a lot of my free time finding out more about interesting topics which have been discussed in different classes.
A. Never or only rarely true of me
B. Sometimes true of me
C. True of me about half the time
D. Frequently true of me
E. Always or almost always true of me

7. I make a point of looking at most of the suggested readings that go with the lectures. A. Never or only rarely true of me
B. Sometimes true of me
C. True of me about half the time

D. Frequently true of me
E. Always or almost always true of me

1. When my assigned homework is extra-long or unusually hard, I either quit or study only the easier parts of the lesson.
A. Strongly disagree
B. Disagree
C. Neutral
D.Agree
E. Strongly Agree

2. If I have to be absent from class, I make up missed lessons without being reminded by the teacher.
A. Strongly disagree
B. Disagree
C. Neutral
D.Agree
E. Strongly Agree

3. I do not bother to correct errors on the papers of my teachers have graded and returned to me.
A. Strongly disagree
B. Disagree
C. Neutral
D.Agree
E. Strongly Agree

4. It takes a long time for me to get warmed up to the job of studying
A. Strongly disagree
B. Disagree
C. Neutral
D.Agree
E. Strongly Agree

5. I am unable to study well because I get restless, moody, or have the blues.
A. Strongly disagree
B. Disagree
C. Neutral
D.Agree
E. Strongly Agree

6. I put off doing written assignments until the last minute
A. Strongly disagree
B. Disagree
C. Neutral
D.Agree
E. Strongly Agree

7. I do poorly on tests because I find it hard to think clearly and plan my study within a short period of time.
A. Strongly disagree
B. Disagree
C. Neutral
D.Agree
E. Strongly Agree

FINAL YEAR PROJECT WEEKLY REPORT

(Project II)

Trimester, Year: Y3S2	**Study week no.: 4**
Student Name & ID: Tan Chung Hao 2005504	
Supervisor: Ts Dr Shakiroh binti Khamis	
Project Title: Student Performance Prediction based on Personal Factors by Using Dashboard	

1. WORK DONE

-

2. WORK TO BE DONE

Summarize the Literature Review into the table form.

3. PROBLEMS ENCOUNTERED

Challenges to summarize as a few research papers no defined clearly.

4. SELF EVALUATION OF THE PROGRESS

Need able to summarise that have more information research paper.

Supervisor's signature

Student's signature

FINAL YEAR PROJECT WEEKLY REPORT

(Project II)

Trimester, Year: Y3S2	**Study week no.: 6**
Student Name & ID: Tan Chung Hao 2005504	
Supervisor: Ts Dr Shakiroh binti Khamis	
Project Title: Student Performance Prediction based on Personal Factors by Using Dashboard	

1. WORK DONE

Able to summarize literature review in paper.

2. WORK TO BE DONE

Create the predictive models.

3. PROBLEMS ENCOUNTERED

No familiar with python script, therefore, need time to research for create the predictive models.

4. SELF EVALUATION OF THE PROGRESS

Should to effort to more familiar to python.

Supervisor's signature

Student's signature

FINAL YEAR PROJECT WEEKLY REPORT

(Project II)

Trimester, Year: Y3S2	**Study week no.: 8**
Student Name & ID: Tan Chung Hao 2005504	
Supervisor: Ts Dr Shakiroh binti Khamis	
Project Title: Student Performance Prediction based on Personal Factors by Using Dashboard	

1. WORK DONE

Able to familiar python language.

2. WORK TO BE DONE

Create Dashboard using Power BI.

3. PROBLEMS ENCOUNTERED

No familiar with Power BI, need times to get use to it.

4. SELF EVALUATION OF THE PROGRESS

Should to effort to for more familiar with Power BI.

Supervisor's signature

Student's signature

FINAL YEAR PROJECT WEEKLY REPORT

(Project II)

Trimester, Year: Y3S2	**Study week no.: 10**
Student Name & ID: Tan Chung Hao 2005504	
Supervisor: Ts Dr Shakiroh binti Khamis	
Project Title: Student Performance Prediction based on Personal Factors by Using Dashboard	

1. WORK DONE

Familiar and visualize dashboard using Power BI.

2. WORK TO BE DONE

To visualize the data.

3. PROBLEMS ENCOUNTERED

Need some mindset to understand which is important to visualize which is no important to supervise.

4. SELF EVALUATION OF THE PROGRESS

Should to effort to know how to visualize.

Supervisor's signature

Student's signature

FINAL YEAR PROJECT WEEKLY REPORT

(Project II)

Trimester, Year: Y3S2	**Study week no.: 12**
Student Name & ID: Tan Chung Hao 2005504	
Supervisor: Ts Dr Shakiroh binti Khamis	
Project Title: Student Performance Prediction based on Personal Factors by Using Dashboard	

1. WORK DONE

Familiar to visualize.

2. WORK TO BE DONE

Conclusion and Future Work.

3. PROBLEMS ENCOUNTERED

Models doesn't seem performing well with limited dataset.

4. SELF EVALUATION OF THE PROGRESS

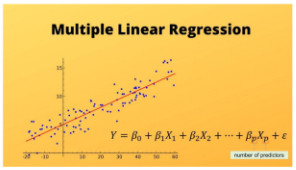
Should speed up the progress.

Supervisor's signature

Student's signature

POSTER

Faculty of Information and Communication Technology (FICT)



Student Performance Prediction based on personal factors by using dashboard

1 Introduction

Understanding and predicting student performance is crucial for effective educational strategies. Various personal factors can influence academic success, making it imperative to develop accurate predictive models. In this study, we delve into the realm of student performance prediction leveraging personal factors through an innovative dashboard approach.

2 Objective

To predict student performance based on personal factors using machine learning and visulise with Power BI.

3 Methodology

Using Data Science Life Cycle. Throughout the process, numerous steps are performed, including business understanding, exploratory data analysis, data preparation, modeling, model evaluation, and development.

4 Proposed method

Our proposed method involves the creation of a dynamic dashboard system that integrates diverse personal factors such as socio-economic background, student interest, and study habits. Leveraging machine learning algorithms such as multi linear regression and data visualization techniques such as Power BI, this dashboard will show the predictive models provide educators with actionable insights into student performance trends and potential areas of improvement.

5 Why Proposed Method is Better than Existing:

Unlike traditional methods that often rely on static data analysis and limited variables, our dashboard approach offers real-time updates and the predictive models. Additionally, the interactive nature of the dashboard enables educators to quickly identify patterns and intervene proactively, leading to more targeted support and improved outcomes.

6 Conclusion

In conclusion, our study highlights the importance of considering personal factors in student performance prediction and underscores the value of dashboard systems in this endeavor. By harnessing the power of data analytics and visualization, educators can gain deeper insights into student behavior and tailor interventions to foster academic success. Through ongoing refinement and integration with educational practices, our proposed method holds the potential to positively impact student outcomes and promote a more inclusive learning environment.

Related literature
References can take up a lot of space, so cite only the key references used in the study.

PLAGIARISM CHECK RESULT

2005504_fyp_turnitin.docx

ORIGINALITY REPORT

12%	8%	6%	5%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	eprints.utar.edu.my Internet Source	2%
2	Abisola Moradeyo Adeyemi,, Semiu Babatunde Adeyemi,. "Personal factors as predictors of students academic achievement in colleges of education in South Western Nigeria", Educational Research and Reviews, 2014 Publication	1%
3	Submitted to University of Nottingham Student Paper	1%
4	Submitted to Management Development Institute Of Singapore Student Paper	1%
5	www.coursehero.com Internet Source	<1%
6	core.ac.uk Internet Source	<1%
7	Seema Rawat, V. Ajantha Devi, Praveen Kumar. "Advancement of Data Processing	<1%

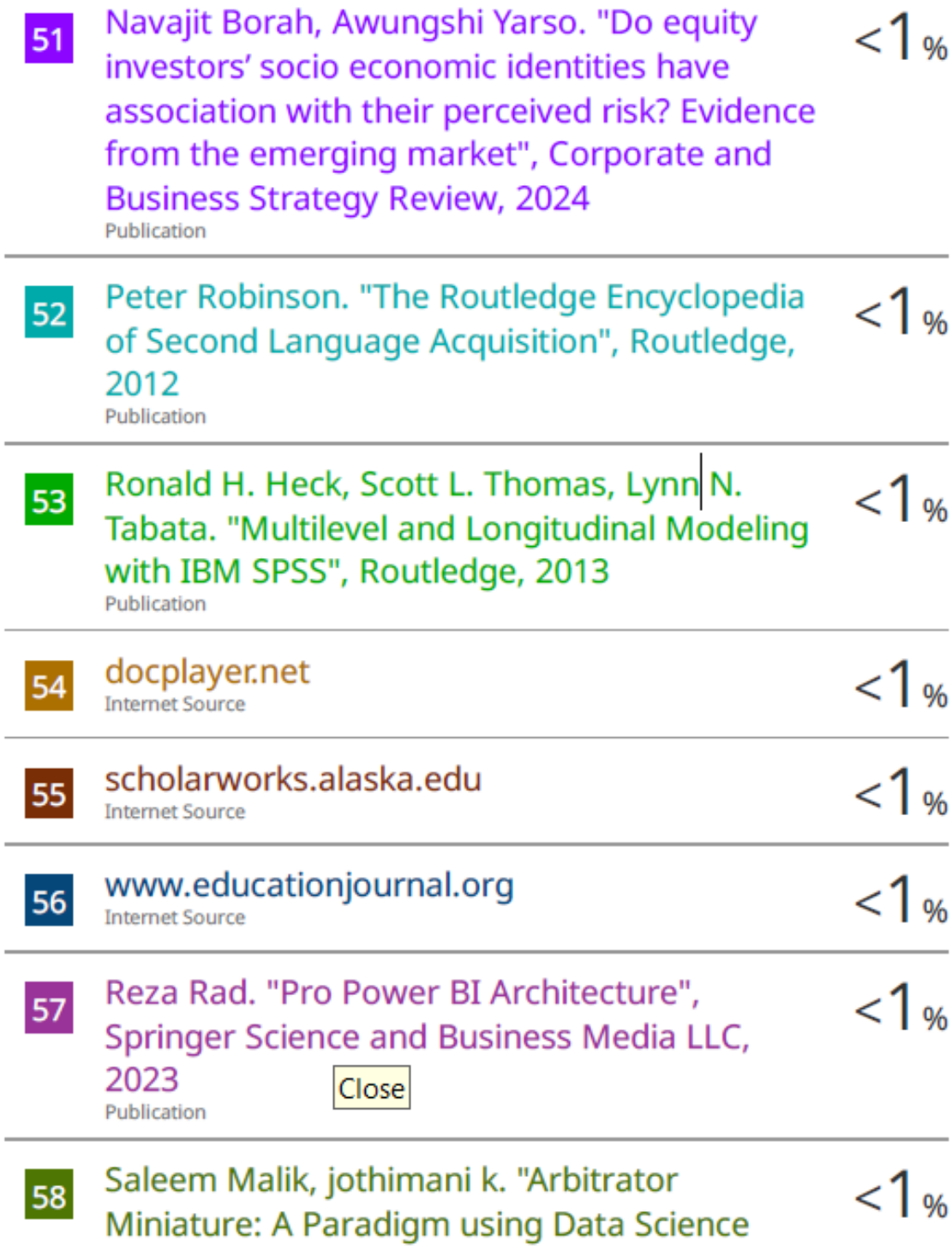
	Methods for Artificial and Computing Intelligence", River Publishers, 2024 Publication	
8	Ahmed Mohammed Abdelalim, Ruqaya Al-Sabah, Mohamed Salem, Salah Omar Said, Mohamed Tantawy, Mohamed Ramadan Ezz Al-Regal. "Variations and Claims in International Construction Projects in the MENA Region from the Last Decade", Buildings, 2024 Publication	<1 %
9	Submitted to The Hong Kong Polytechnic University Student Paper	<1 %
10	www.slideshare.net Internet Source	<1 %
11	link.springer.com Internet Source	<1 %
12	vital.seals.ac.za:8080 Internet Source	<1 %
13	Alan R. Jones. "Best Fit Lines and Curves, and Some Mathe-Magical Transformations", Routledge, 2018 Publication	<1 %
14	Submitted to Submitted on 1685861417976 Student Paper	<1 %

15	etd.uum.edu.my Internet Source	<1 %
16	Submitted to RMIT University Student Paper	<1 %
17	academicjournals.org Internet Source	<1 %
18	www.hindawi.com Internet Source	<1 %
19	Submitted to Adtalem Global Education, Inc. Student Paper	<1 %
20	Submitted to University College Technology Sarawak Student Paper	<1 %
21	Submitted to Western Mindanao State University Student Paper	<1 %
22	Submitted to HTM (Haridus- ja Teadusministeerium) Student Paper	<1 %
23	Submitted to Valdosta State University Student Paper	<1 %
24	Submitted to Middle East College of Information Technology Student Paper	<1 %
25	Submitted to UC, San Diego	

	Student Paper	<1%
26	archive.org Internet Source	<1%
27	Submitted to American University of Central Asia Student Paper	<1%
28	Submitted to Higher Education Commission Pakistan Student Paper	<1%
29	Sandip J. Gami, Balaji Dhamodharan, Pushan Kumar Dutta, Vaibhav Gupta, Pawan Whig. "chapter 23 Data Science for Personalized Nutrition Harnessing Big Data for Tailored Dietary Recommendations", IGI Global, 2024 Publication	<1%
30	Submitted to Troy University Student Paper	<1%
31	personalpages.manchester.ac.uk Internet Source	<1%
32	Submitted to Middlesex University Student Paper	<1%
33	Submitted to Pennsylvania State System of Higher Education Student Paper	<1%

34	Submitted to Swiss School of Business and Management - SSBM Student Paper	<1 %
35	Submitted to University Tunku Abdul Rahman Student Paper	<1 %
36	Submitted to University of Kent at Canterbury Student Paper	<1 %
37	Submitted to University of the West Indies Student Paper	<1 %
38	Submitted to University of Greenwich Student Paper	<1 %
39	qspace.library.queensu.ca Internet Source	<1 %
40	rosap.ntl.bts.gov Internet Source	<1 %
41	es.scribd.com Internet Source	<1 %
42	www.ncbi.nlm.nih.gov Internet Source	<1 %
43	www.researchgate.net Internet Source	<1 %
44	www.yumpu.com Internet Source	<1 %

45	Alireza Kokabi, Mina Fayazi. "Boron and nitrogen doping engineering of biphenylene electronic and optical properties using deep learning", Materials Today Communications, 2024 Publication	<1 %
46	Satria Bangsawan, MS Mahrinasari, Ernie Hendrawaty, Rindu Rika Gamayuni et al. "The Future Opportunities and Challenges of Business in Digital Era 4.0", Routledge, 2020 Publication	<1 %
47	elearning.ovgu.de Internet Source	<1 %
48	www.arcjournals.org Internet Source	<1 %
49	www.docstoc.com Internet Source	<1 %
50	Junianto, Dhoriva Urwatul Wutsqa. "Comparison of Effectiveness between Contextual Teaching and Learning (CTL) and Problem-Based Learning (PBL) Approach on the Interest of Junior High School Students", Proceedings of the 2019 International Conference on Mathematics, Science and Technology Teaching and Learning - ICMSTTL 2019, 2019 Publication	<1 %



51	Navajit Borah, Awungshi Yarso. "Do equity investors' socio economic identities have association with their perceived risk? Evidence from the emerging market", Corporate and Business Strategy Review, 2024 Publication	<1 %
52	Peter Robinson. "The Routledge Encyclopedia of Second Language Acquisition", Routledge, 2012 Publication	<1 %
53	Ronald H. Heck, Scott L. Thomas, Lynn N. Tabata. "Multilevel and Longitudinal Modeling with IBM SPSS", Routledge, 2013 Publication	<1 %
54	docplayer.net Internet Source	<1 %
55	scholarworks.alaska.edu Internet Source	<1 %
56	www.educationjournal.org Internet Source	<1 %
57	Reza Rad. "Pro Power BI Architecture", Springer Science and Business Media LLC, 2023 Publication	<1 %
58	Saleem Malik, jothimani k. "Arbitrator Miniature: A Paradigm using Data Science	<1 %

Close

Methods to Predict Academic Performance",
Research Square Platform LLC, 2022
Publication

Exclude quotes	Off	Exclude matches	Off
Exclude bibliography	Off		

Universiti Tunku Abdul Rahman			
Form Title : Supervisor's Comments on Originality Report Generated by Turnitin for Submission of Final Year Project Report (for Undergraduate Programmes)			
Form Number: FM-IAD-005	Rev No.: 0	Effective Date: 01/10/2013	Page No.: 1of 1

FACULTY OF INFORMATION AND COMMUNICATION TECHNOLOGY

Full Name(s) of Candidate(s)	TAN CHUNG HAO
ID Number(s)	20ACB05504
Programme / Course	FICT /IB
Title of Final Year Project	Student Performance Prediction based on personal factors by using dashboard

Similarity	**Supervisor's Comments (Compulsory if parameters of originality exceeds the limits approved by UTAR)**
Overall similarity index: 12 % **Similarity by source** Internet Sources: 8 % Publications: 6 % Student Papers: 5 %	
Number of individual sources listed of more than 3% similarity: 0	
Parameters of originality required and limits approved by UTAR are as Follows: **(i) Overall similarity index is 20% and below, and** **(ii) Matching of individual sources listed must be less than 3% each, and** **(iii) Matching texts in continuous block must not exceed 8 words** ***Note: Parameters (i) – (ii) shall exclude quotes, bibliography and text matches which are less than 8 words.***	

Note Supervisor/Candidate(s) is/are required to provide softcopy of full set of the originality report to Faculty/Institute

Based on the above results, I hereby declare that I am satisfied with the originality of the Final Year Project Report submitted by my student(s) as named above.

Signature of Supervisor

Name: DR SHAKIROH BINTI KHAMIS

Date: 13/9/2024

Signature of Co-Supervisor

Name: _____

Date: _____

UNIVERSITI TUNKU ABDUL RAHMAN

FACULTY OF INFORMATION & COMMUNICATION TECHNOLOGY (KAMPAR CAMPUS)

CHECKLIST FOR FYP2 THESIS SUBMISSION

Student Id	20ACB05504
Student Name	TAN CHUNG HAO
Supervisor Name	TS DR SHAKIROH BINTI KHAMIS

TICK (√)	DOCUMENT ITEMS Your report must include all the items below. Put a tick on the left column after you have checked your report with respect to the corresponding item.
√	Title Page
√	Signed Report Status Declaration Form
√	Signed FYP Thesis Submission Form
√	Signed form of the Declaration of Originality
√	Acknowledgement
√	Abstract
√	Table of Contents
√	List of Figures (if applicable)
√	List of Tables (if applicable)
	List of Symbols (if applicable)
√	List of Abbreviations (if applicable)
√	Chapters / Content
√	Bibliography (or References)
√	All references in bibliography are cited in the thesis, especially in the chapter of literature review
√	Appendices (if applicable)
√	Weekly Log
√	Poster
√	Signed Turnitin Report (Plagiarism Check Result - Form Number: FM-IAD-005)
√	I agree 5 marks will be deducted due to incorrect format, declare wrongly the ticked of these items, and/or any dispute happening for these items in this report.

*Include this form (checklist) in the thesis (Bind together as the last page)

I, the author, have checked and confirmed all the items listed in the table are included in my report.

_____

(Signature of Student)

Date: 12/9/2024