

UNIVERSITI TEKNOLOGI MARA

**ANALYZING STUDENT ACADEMIC
PERFORMANCE ACROSS STREAMS
AT SMK SERI PAGI USING
K-MEANS CLUSTERING**

NUR ZAHIRAH BINTI ZAINUDDIN

**BACHELOR OF INFORMATION SYSTEMS (Hons.)
BUSINESS COMPUTING**

JULY 2025

ACKNOWLEDGEMENT

First and foremost, I would like to thank Allah, the Most Gracious and Most Merciful, for giving me the sustenance, perseverance, and intention to complete this work. It would not have been possible without His never-ending blessings and guidance. Other than that, special thanks to my passionate supervisor, Madam Siti 'Aisyah binti Sa'dan who helped me throughout the project with her support, motivation, encouragement and guidance. This work has in large part benefited from her encouragement and expertise.

I would like to also express my deep gratitude to my inspiring CSP650 lecturer, Miss Nor Hasnul Azirah binti Abdul Hamid whose lectures, encouragement and guidance were indispensable input to my understanding and growth. And I would like to thank my parents for the support that enable me to continue studying academically, and their patience put me through all this while. My perseverance and success is founded on their belief in me. Finally, I would like to thank my peers, whose friendship, encouragement, and willingness to work together have made this process so much more rewarding and memorable.

ABSTRACT

The academic performance of students is a critical concern for schools, yet many institutions such as SMK Seri Pagi continue to rely heavily on manual spreadsheet analysis, which is time-consuming and lacks visual clarity. This project aims to improve the analysis of student academic performance across streams by implementing a descriptive data analytics solution using K-Means clustering and interactive visualization. Following the CRISP-DM methodology, data preprocessing was conducted using Microsoft Excel and Google Spreadsheet, while RapidMiner was employed for K-Means clustering analysis. The clustering model utilized selected attributes including Gender, Stream, marks for Bahasa Melayu, Bahasa Inggeris, Matematik, Sejarah and the Average Grade Point (GP). The optimal number of clusters was determined using the Davies-Bouldin Index. The final Power BI dashboard features multiple interactive pages presenting overall statistics, clustering insights, and subject-wise analysis. Expert evaluation rated the dashboard positively for clarity and usefulness. The dashboard significantly improved the school's ability to visualize and interpret student performance patterns, facilitating more efficient and informed decision-making by educators. Challenges related to data preprocessing and integration were addressed through iterative refinement and the use of complementary tools, ensuring data quality and reliability. This project offers a data-driven approach that supports teachers in interpreting student performance patterns and making informed decisions. Future work may involve integrating real-time data updates, expanding clustering attributes, and enhancing dashboard interactivity for broader educational applications.

TABLE OF CONTENTS

CONTENT	PAGE
SUPERVISOR'S APPROVAL	ii
STUDENT DECLARATION	iii
ACKNOWLEDGEMENT	iv
ABSTRACT	v
TABLE OF CONTENTS	vi
LIST OF TABLES	ix
LIST OF FIGURES	x
LIST OF ABBREVIATIONS	xii
CHAPTER 1 INTRODUCTION	1
1.1 Background of Study	1
1.2 Current Business Process	3
1.3 Problem Statement	5
1.4 Objectives	6
1.5 Scope	6
1.6 Significance	7
1.7 Project Framework	8
1.8 Gantt Chart	11
1.9 Conclusion	12
CHAPTER 2 LITERATURE REVIEW	13
2.1 Introduction	13
2.2 Business Intelligence	13
2.2.1 Business Analytics	15
2.2.2 Data Mining	15
2.2.3 Educational Data Mining (EDM)	17
2.3 Descriptive Analysis	18
2.3.1 Measures of Central Tendency	19
2.3.2 Measure of Dispersion and Variability	19
2.3.3 Descriptive Analysis in Education	20
2.4 Clustering	21
2.4.1 Clustering Algorithm	22

2.4.2	K-Means Clustering	23
2.4.3	Performance Evaluation	24
2.4.4	Clustering in Education	25
2.5	CRISP-DM	26
2.5.1	Business Understanding	27
2.5.2	Data Understanding	28
2.5.3	Data Preparation	29
2.5.4	Modeling	29
2.5.5	Evaluation	30
2.5.6	Deployment	30
2.5.7	CRISP-DM in Education	31
2.6	Similar Existing Research	33
2.7	Implementations of Literature Review	35
2.8	Conclusion	36
CHAPTER 3 PROJECT METHODOLOGY		37
3.1	Introduction	37
3.2	Project Methodology	37
3.3	Business Understanding	41
3.4	Data Understanding	43
3.4.1	Data Collection	43
3.4.2	Data Description	46
3.4.3	Data Exploration	49
3.4.4	Data Quality Verification	50
3.5	Data Preparation	52
3.5.1	Data Selection	53
3.5.2	Data Cleaning	54
3.5.3	Data Construction	57
3.5.4	Data Integration	60
3.5.5	Data Formatting	61
3.6	Modeling	62
3.6.1	Descriptive Analysis	63
3.6.2	Clustering	64
3.7	Evaluation	67