BAHASA MELAYU DIALECT TRANSLATOR AND DETECTOR (MALAYFY)

ADELLA JAVA DIRGANTARI

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

BORANG PENGESAHAN STATUS LAPORAN

JUDUL: BAHASA MELAYU DIALECT TRANSLATOR AND DETECTOR (MALAYFY)

SESI PENGAJIAN: 2022 / 2023

Saya: ADELLA JAVA DIRGANTARI

mengaku membenarkan tesis Projek Sarjana Muda ini disimpan di Perpustakaan Universiti Teknikal Malaysia Melaka dengan syarat-syarat kegunaan seperti berikut:

1. Tesis dan projek adalah hakmilik Universiti Teknikal Malaysia Melaka.
2. Perpustakaan Fakulti Teknologi Maklumat dan Komunikasi dibenarkan membuat salinan unituk tujuan pengajian sahaja.
3. Perpustakaan Fakulti Teknologi Maklumat dan Komunikasi dibenarkan membuat salinan tesis ini sebagai bahan pertukaran antara institusi pengajian tinggi.
4. * Sila tandakan (✓)

_____	SULIT	(Mengandungi maklumat yang berdarjah keselamatan atau kepentingan Malaysia seperti yang termaktub di dalam AKTA RAHSIA RASMI 1972)
_____	TERHAD	(Mengandungi maklumat TERHAD yang telah ditentukan oleh organisasi / badan di mana penyelidikan dijalankan)
_____✓_____	TIDAK TERHAD	

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

(TANDATANGAN PELAJAR)

Alamat tetap: JL. CANDI AGUNG 1/20 A, MALANG, JAWA TIMUR, INDONESIA, 65142

Tarikh: 13/09/2023

(TANDATANGAN PENYELIA)

Ts. DR. HALIZAH BINTI BASIRON

Nama Penyelia

Tarikh: 13/09/2023

CATATAN: * Jika tesis ini SULIT atau TERHAD, sila lampirkan surat daripada pihak

BAHASA MELAYU DIALECT TRANSLATOR AND DETECTOR (MALAYFY)

ADELLA JAVA DIRGANTARI

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

This report is submitted in partial fulfillment of the requirements for the Bachelor of Computer Science (Artificial Intelligence) with Honours.

FACULTY OF INFORMATION AND COMMUNICATION TECHNOLOGY
UNIVERSITI TEKNIKAL MALAYSIA MELAKA

2023

DECLARATION

I hereby declare that this project report entitled

BAHASA MELAYU DIALECT TRANSLATOR AND DETECTOR

(MALAYFY)

is written by me and is my own effort and that no part has been plagiarized

without citations.

STUDENT : _____ Date : 17/09/2023

ADELLA JAVA DIRGANTARI

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

I hereby declare that I have read this project report and found

this project report is sufficient in term of the scope and quality for the award of

Bachelor of Computer Science (Artificial Intelligence) with Honours.

SUPERVISOR : _____ Date : 17/09/2023

Ts. DR. HALIZAH BINTI BASIRON

DEDICATION

This final-year study is dedicated to my supervisor, Ts. Dr. Halizah binti Basiron, for her support and guidance. I also want to thank my parents and siblings for their financial and emotional support in helping me complete this project. Last but not least, I want to thank my friends for their constant encouragement and emotional support when I was a university student.

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

ACKNOWLEDGEMENTS

I would like to express my gratitude to my supervisor, Ts. Dr. Halizah binti Basiron, for her invaluable support and guidance throughout the project. She provided me with valuable feedback on my research methods, identified areas of improvement, and provided me with the necessary resources to successfully complete the project. I would also like to thank my beloved parents who have also been a constant source of support and encouragement, and I am extremely grateful for their unwavering belief in me. Additionally, I would like to extend our sincere appreciation to my seniors and friends who provided me with valuable insights and feedback that helped me to improve my project.

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

ABSTRACT

The Bahasa Melayu Dialect Translator and Detector (Malayfy) system is an innovative solution designed to overcome language barriers and facilitate effective communication between users of different dialects in the context of Bahasa Melayu. This system consists of three modules: a language translator from English to Bahasa Melayu, a Bahasa Melayu dialect translator, and a Bahasa Melayu dialect detector. The project's primary goal is to create a reliable and precise system capable of translating English text into Bahasa Melayu while also translating standard Bahasa Melayu into dialects. Additionally, the dialect detection module employs the Random Forest algorithm, renowned for its ability to handle complex classification tasks, to predict the dialect of user input sentences. The project encompasses various stages, including a literature review, data collection, system design, and implementation. The system's effectiveness is assessed based on its accuracy, efficiency, and user satisfaction. The current issues associated with the absence of dialect detection and translation in existing machine translation tools for Bahasa Melayu are anticipated to be addressed with the development of this system. Effective communication with Bahasa Melayu speakers is crucial in various fields, including tourism, education, healthcare, and business. The successful implementation of this system, driven by the power of machine learning algorithms like Random Forest, could have significant implications in these fields and others.

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

ABSTRAK

Sistem Penterjemah dan Pengesan Dialek Bahasa Melayu (Malayfy) ialah penyelesaian inovatif yang direka untuk mengatasi halangan bahasa dan memudahkan komunikasi berkesan antara pengguna dialek berbeza dalam konteks Bahasa Melayu. Sistem ini terdiri daripada tiga modul: penterjemah bahasa daripada Bahasa Inggeris ke Bahasa Melayu, penterjemah dialek Bahasa Melayu, dan pengesan dialek Bahasa Melayu. Matlamat utama projek ini adalah untuk mencipta sistem yang boleh dipercayai dan tepat yang mampu menterjemah teks Inggeris ke dalam Bahasa Melayu sambil juga menterjemah Bahasa Melayu standard ke dalam dialek. Selain itu, modul pengesanan dialek menggunakan algoritma Random Forest, yang terkenal dengan keupayaannya untuk mengendalikan tugas pengelasan yang kompleks, untuk meramalkan dialek ayat input pengguna. Projek ini merangkumi pelbagai peringkat, termasuk kajian literatur, pengumpulan data, reka bentuk sistem dan pelaksanaan. Keberkesanan sistem dinilai berdasarkan ketepatan, kecekapan dan kepuasan pengguna. Isu semasa yang berkaitan dengan ketiadaan pengesanan dialek dan terjemahan dalam alat terjemahan mesin sedia ada untuk Bahasa Melayu dijangka dapat ditangani dengan pembangunan sistem ini. Komunikasi yang berkesan dengan penutur Bahasa Melayu adalah penting dalam pelbagai bidang, termasuk pelancongan, pendidikan, penjagaan kesihatan dan perniagaan. Kejayaan pelaksanaan sistem ini, didorong oleh kuasa algoritma pembelajaran mesin seperti Random Forest, boleh mempunyai implikasi yang ketara dalam bidang ini dan lain-lain.

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

TABLE OF CONTENTS

PAGE

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

LIST OF TABLES

PAGE

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

LIST OF FIGURES

PAGE

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

UTeM
اونيۏرسيتي تيکنيکل مليسيا ملاک
UNIVERSITI TEKNIKAL MALAYSIA MELAKA

LIST OF ABBREVIATIONS

FYP	**-**	**Final Year Project**
ISDM	**-**	**Intelligent System Development Methodology**
NLP	**-**	**Natural Language Processing**
NMT	**-**	**Neural Machine Translation**
SMT	**-**	**Statistical Machine Translation**
ANN	**-**	**Artificial Neural Network**
RNN	**-**	**Recurrent Neural Network**
LSTM	**-**	**Long Short-Term Memory**
GRU	**-**	**Gated Recurrent Unit**

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

LIST OF ATTACHMENTS

PAGE

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

CHAPTER 1: INTRODUCTION

1.1 Introduction

Effective communication across language borders is now more crucial than ever thanks to globalization and the world's growing interconnectedness. Among people, communities, and nations, language plays a crucial role in communication, connection, and understanding. Language diversity can, however, provide difficulties, obstructing clear communication and resulting in misunderstandings and misinterpretations. The communication barrier that develops when persons or organizations need to interact with or convey information to people who speak or comprehend only Bahasa Melayu is the issue statement for the requirement of a Bahasa Melayu translator.

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

To bridge the language barrier and facilitate effective communication between parties, whether in business, education, healthcare, or other professions, a Bahasa Melayu translator is required. Without a Bahasa Melayu translator, crucial information could be misunderstood, misconstrued, or misread, with potentially disastrous effects. By offering a reliable and effective method for dialect detection and translation in the Bahasa Melayu language, the Bahasa Melayu Dialect Translator and Detector system (Malayfy) intends to overcome these issues. Malay, often known as Bahasa Melayu, is a widely spoken language in Southeast Asia with numerous regional varieties.

1.2 Problem Statement(s)

i. Language barriers can impede communication and lead to misunderstandings and misinterpretations.

ii. Current machine translation tools for Malay are often inaccurate and unreliable.

iii. There are challenges for businesses, organizations and individuals who need to communicate with Malay speakers in various contexts, such as tourism, education, health, and commerce.

1.3 Objective(s)

i. To gather Bahasa Melayu words and their dialect.

ii. To develop language translator from English to Bahasa Melayu and Bahasa Melayu dialect translator and detector.

iii. To evaluate Bahasa Melayu translator and Bahasa Melayu dialect translator and detector.

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

1.4 Project Scope

The project scope defines the boundaries and objectives of the Bahasa Melayu Dialect Translator and Detector system. It describes the particular system modules that will be created in detail and specifies the target users who will profit from its features. The core modules and the target user base are highlighted in this chapter's overview of the project's scope.

1.4.1 Module(s)

The table below shows the three components that will be built for the Malayfy application. Natural Language Processing (NLP) techniques will be used in the modules that translate Bahasa Melayu using the Transformers model of deep learning,

detect dialects using random forest algorithms, and translate dialects using rule-based systems.

Table 1.1: Modules in the Malayfy Application

Modules	Functions
Bahasa Melayu Translator	Able to aid user in interacting and communicating with the native in Malaysia by help user to translate the words or sentences in English to Bahasa Melayu.
Dialect Detector	For the dialect detection feature, classification machine learning classifies the sentences of Malaysian dialects such as Johor, Kedah, Kelantan, Melaka, Negeri Sembilan, Pahang, and Penang. This able to assist the user in identifying sentences in a dialect and predicting which dialect the phrases belong to.
Dialect Translator	The general bahasa melayu will be translated to the selected dialect using this dialect translator module. With a rule-based approach, if the vocabulary is present in the selected dialect dictionary, the words from the sentences will transform to that dialect.
User interface module	This module will provide a user-friendly interface for users to interact with the system, such as a website. It will include input/output text fields and language selection options.

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

1.4.2 Target User(s)

i. Businesses that operate in or trade with Malaysia and need to communicate with Bahasa Melayu-speaking customers, clients, or partners.

ii. Educational institutions that have Bahasa Melayu-speaking students, teachers, or researchers and need to communicate with them effectively.

iii. Tourists and travelers who visit Malaysia and want to communicate with locals in Bahasa Melayu.

1.5 Project Significant

The project's successful completion will significantly impact cross-cultural communication, promoting understanding and enabling global collaborations. The importance of a Bahasa Melayu translation project rests in its capacity to break down language barriers and promote interlingual communication. This is crucial in a nation like Malaysia, where a diverse populace speaks various languages in addition to Bahasa Melayu, the official language. A Bahasa Melayu translator can support increased communication and cooperation between various populations inside and outside Malaysia. A competent Bahasa Melayu translation can also enable others who may not be fluent in the language to access information, education, and opportunities, encouraging inclusivity and equality. Moreover, the project will also contribute to developing natural language processing and machine learning technologies, which will have broad applications beyond Bahasa Melayu translation.

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

1.6 Expected Outcome

The goal of the Malayfy project is to create a reliable and effective dialect translator, dialect language detector, and English to Malay translation system. The initiative aims to improve comprehension and communication of Bahasa Melayu by successfully identifying and classifying various dialects and by providing accurate and trustworthy translations. The system is anticipated to have an intuitive user interface that makes it simple to access its functions and promotes seamless communication across language boundaries. If the project is successful, people, companies, and organizations will be able to get over language obstacles and improve their Bahasa Melayu communication skills.

1.7 Conclusion

In conclusion, this project will provide a solution for native speakers when they come to Malaysia. This website will help them communicate with locals and understand the dialects used by them, as there are many different dialects. A discussion of the project's methodology and a review of relevant studies will follow in the subsequent chapter.

UNIVERSITI TEKNIKAL MALAYSIA MELAKA