CHAPTER 1

INTRODUCTION

1.1 Introduction

In this age of information, web user is able to access vast amount of data through the advancement of web technology. Users use the web's available resources, and participate in providing comments, which resulting in the generation of additional data. Most of the text information available in the web such as social media is unstructured data. Even thought those review, post and comment are unstructured data but they contain massive amount of information that could provide valuable insight. Especially for marketing business analysis and future improvement guideline for an organization. Hence, organising, analysing, and exploration of web user feedback and idea in an efficient manner is important and crucial to provide an effective decision-making. However, text mining and sentiment Analysis (SA) on web text using natural language processing (NLP) techniques is required (Ferdous, Syed, Bin, & Uddin, 2024).

Bidirectional Encoder representations from Transformer (BERT) is one of the most well-known deep learning models for natural language processing (Gardazi et al., 2025). It is a pre-trained model based on transformer architecture. During pre-trained, BERT's mask language modelling is trained to enable it predict masked word based on surrounding token content. Hence, BERT can learn and understand the contextual information and relationship between token of word causing it effective for language processing task. There is a research show that BERT is applicable in mining nuance information of the language which is beneficial for sentiment analysis by capturing the content of word (Sornalakshmi et al., 2024). Sentiment analysis is typically classifying text based on sentiment expression of positive, negative and neutral. Hence, BERT is suitable to be implement in mining contextual information from a social media text. Research by Bhola et.al. (Bhola et al., 2022). conducted

research on text mining utilizing BERT for sentiment analysis. Sentiment analysis for monitoring customer feedback on social media to improve service and product.

A study analysing on 36000 Facebook posts, identified that there is cluster of discussion on topic related to economic, politic and legal aspect of electric vehicle (Debnath, Ronita Bardhan, Reiner, & Miller, 2021). There another study on cross-multiple social media platforms of Facebook, X and Instagram found out that electric vehicle had high user engagement and towards positive sentiment (Hafize et.al., 2024). This research also exploring on how social media would influence electric vehicle adoption by the consumer. Research by Zhao et.al. (Zhao et.al.,2024) found out that media would influence the market diffusion of electric vehicle. Thereby, this research project aims to assist in providing nuanced information and evaluation by using structured sentiment analysis in gaining real time user's feedback information on strength, limitation and future prospective of the electric vehicle. This would help fostering the market growth of EV industry and ensure wider market acceptance for consumer in making wiser decision.

1.2 Problem Background

The transport sector contributes significant to the greenhouse gas (GHG) emission. Statistical result from climate change scientific research shows that transportation is the largest contribution to the United State GHG emissions in 2022 with a total percentage of 28% ("Climate Change 2022: Mitigation of Climate Change," 2022). While a statistical report by International Energy Agency (IEA) also shows that carbon dioxide (CO_2) emission from fossil fuel transportation is about 22% (Malaysia - Countries & Regions - IEA, 2025). Thereby many countries aim for carbon neutrality and intend to achieve net-zero for carbon dioxide emission by the year of 2050 while ensuring affordable and stable renewable energy (Fam & Fam, 2024). Hence, this crisis had triggered a worldwide action in boosting the usage of electric vehicle and minimize the burning of fossil fuel energy vehicles. As the electric vehicle has lesser carbon emission reduction (Littlejohn & Stef Proost, 2022). Many efforts such as improve the infrastructure and tax intensive policy had been implemented for the adaption of electric vehicle (Wibowo & Dovi Septiari, 2023).

Nevertheless, Malaysia also follows closely the global climate issue with commitment and collaboration of Malaysia regarding the global climate issue is also important as one of the major oil and gas producers. Based on the Paris agreement, Malaysia targeted to reduce GHG emission by 35% for the gross domestic product (GDP) by 2030 (Fernandez et. al., 2024). Thereby, Malaysia had a strategies plan in energy transition effort from non-renewable energy to renewable energy resource. (Majekodunmi et al., 2023). One of the strategies plans is targeting 15% of the total industry volume (TIV) by year of 2030 and 80% by 2050 (Siew et. al., 2024). A transition strategy guideline by the Natural Resources and Environmental Sustainability Ministry (NRES) shows that the government intend to transition 50% of government vehicle to electric vehicle ("Portal Rasmi Kementerian Sumber Asli Dan Kelestarian Alam," 2025). Besides, several initiatives had been implemented by the Malaysia government in full exemption of import and excise duties to encourage manufactured of electric vehicle and tax relief for the electric vehicle owner ("Tax Reliefs | Lembaga Hasil Dalam Negeri Malaysia," 2025).

With the effort from the Malaysia government, the Electric Vehicle in Malaysia automotive market is experiencing significant growth, with yearly sales increasing substantially. As the statistic from JPJ Malaysia shows that at the first quarter of year 2025 there were 6827 electric vehicles registered as compared to the first 3 months of 2024 with a record of 4689 electric vehicles registered, which marks an increase of 45.6% (Government of Malaysia, 2025). However, according to Ernst & Young's (EY) fifth annual EY Global Mobility Consumer Index on year 2024 shows that only 25% of Malaysian consumers intent on buying an electric vehicle (EV). Nearly 40 % of Malaysians would intend to buy an internal combustion engine, and 20% would purchase a hybrid car (EY, 2024). Despite implementation of several incentive and more choices of electric vehicle brands, most of the automotive buyer in Malaysia still remain elusive in purchasing the electric vehicle.

The research by Higueras-Castillo et al. (Higueras-Castillo et.al., 2019) proposed that sentiment of customer in electric vehicle would influence the purchase intention. Moreover, research by Kutabish et. Al. (Kutabish, Soares, & Casais, 2023) indicated that consumer nowadays influence by and rely on online reviews to assist

their decision before purchasing. Online reviewer platforms such as YouTube, social media, automotive news and community. Recently, the electric vehicle revolution of Malaysia automotive market had led to surge of online discussion. Online platform such as social media, forums or video review had become key venue for consumer to share experience, discussing and debating about their concern (Ruan & Qin Lv, 2023). Hence, to leverage those reviews on online platform, research on sentiment analysis of electric vehicle has gained significant traction in recent years as text analysis is important for business and organisation to have a insight and understand customer perception. There various recent studies on text mining regarding electric vehicle related topic mostly on conducted on global, with limited focus on Malaysia's unique market dynamics.

There a study implemented deep learning model such as ERNIE combined with deep CNNs to improve sentiment classification accuracy (Wang et.al., 2023). Even though it achieves a high accuracy, there persisting gap lies in the limited sentiment diversity which model only classify sentient into basic categories: positive, neutral and negative. As noted by Wang et. Al. such simplistic classification cannot capture the full spectrum of aspect-based sentiment analysis in public discourse.

There advance natural language processing model, which is bidirectional encoder representations from transformer (BERT) which had significant contribution for sentiment analysis and text mining for text-related data as compared to traditional method. Research by Guven et.al. (Guven, 2021) shows that BERT outperformed Logistic Regression in classifying sentiment analysis of tweet dataset with accuracy of 98.75%. Besides, there is another research shows that BERT model also outperformed model such as Naïve Bayse and Support Vector Machine (SVM) with a accuracy of 95% in analysing customer feedback (Rahman & Maryani, 2024). In addition, there research by Gaurav et.al. (Gaurav, Gupta, & Chui, 2024) proposed that BERT is capable and outperformed traditional model in capturing contextual information and nuanced sentiment classification.

There is research leveraging Large Language Model which is BERT-based model in electric vehicle have demonstrating promising result. However, there is

limited data diversity in this research as it solely relies on small dataset from single platform. This narrow data source may result in potential bias. As proposed by Sharma et. Al. (Sharma, Din, & Ogunleye, 2024) future research can be done on wider dataset to include user generated content from other platform is crucial for more unbiased sentiment analysis. Research done by Wu et.al. (Wu et.al., 2023) also proposed to expand the research study on multiple social media platform and other foreign or domestic platform as the sentiment of public would be different in different countries.

1.3   Problem Statement

Lack of localized research hinder on electric vehicle discourse in Malaysia. There is previous research conducted on other countries and global such as China electric vehicle (Liang, Li, & Chen, 2023) and global YouTube comment on Tesla Motor and Lucid Motor (Sharma, Din, & Ogunleye, 2024) but lack of text mining research done on topic regarding electric vehicle in Malaysia. The second problem statement is the lack of diverse data sources will result in potential bias of the analysis result and incomplete insights into EV-related discussions. Although there is research applied for sentiment analysis, however existing approach rely on single source dataset which might lead to bias data insight (Sharma, Din, & Ogunleye, 2024). The third problem statement is simplistic sentiment classification fails to capture the full spectrum of aspect-based sentiment analysis in public discourse about EVs. There is some of the current research on sentiment classification for electric vehicle using traditional model is simplified, cause failing in capturing nuance consumer discourse information. As proposed by Jena et.al (Jena, 2020) state that there is neglection of public sentiment analysis on topic related to new electric vehicle. Hence, there gap remains in current sentiment analysis of EV domain

1.4   Research Question

The research question is specified as below:

(a)   What is the sentiment analysis of the electric vehicle discourse in Malaysia?

(b)   How can BERT outperform traditional models and which BERT-based model is better in text mining and sentiment analysis?

(c)   How does sentiment and public prospective will vary across platforms?

1.5   Research Objectives

This research aims to mining public discourse on topic related to electric vehicle with the following objective:

(a)   To identify, preprocess and explore electric vehicle related text data from multiple online social media and implementing preprocessing procedure to clean data and discover underlying patterns.

(b)   To Implement and compare pre-trained Bert-based model in determining structured sentiment analysis.

(c)   To analyse the sentiment insight of Malaysian about electric vehicle.

1.6   Research Scope

The following scope are included in this research:

(a)   This study will focus exclusively on analysis of discourse within Malaysian social media websites.

(b)   The web crawler will be use to crawl on text data published between year 2021 until year 2025.

(c)   The analysis only coverage English-language electric vehicle related textual data.

(d)   BERT-based model will be implemented for aspect-based analysis, which analysis not just solely on sentiment classification.

1.7   Significant of the study

The significant of this research is that provide an insight of public concerns and adoption barriers surrounding electric vehicle from a user perspective. This would ensure the government and organization can have a valuable understanding on customer opinion that would ensure better solution and incentive provided in order to attract and boost confident of the consumer. Besides, implementation of Bert-based model would provide more nuance analysis as compared to traditional sentiment analysis. In addition, web scraping of the text dataset would provide a valuable real time research study on electric vehicle related discourse. Can be also the released of preprocessed version of electric vehicle social media dataset.

1.8   Thesis Structure

This research project report consists of a total of six chapters and each chapter entails a systematic way for solving the research question posed on title "Sentiment Analysis of Electric Vehicle Discourse Using BERT-Based Language Model". Hence, thereby this subsection outlines the scope of those chapter.

Chapter 1 is the introduction to the proposed project. This chapter includes an overview of the chapter, introduction, problem background, problem statement, research question, project objectives, project scope and organization of the chapter.

Chapter 2 is the literature review. This chapter provides a systematic review of current work through published articles. The published literature articles are obtained from Scopus, Web of Science, Science direct and other sources. This chapter will review all the existing research articles such as datasets, ML algorithms, performance metrics evaluation, challenges, and opportunities.

Chapter 3 is about methodology. Each of the model phases, methods, algorithms used, and requirements for this project are described and presented in this

chapter. This chapter provides a definite, comprehensive and structural understanding of the project.

Chapter 4 is the implementation of experiments. This chapter presents the implementation of the experiments of this project. Implementation of BERT-based model for text mining is conducted.

Chapter 5 is the result & discussion. This chapter showcase the results of the experiments conducted. Suitable evaluation metric is used to show the accuracy and reliability of the results. The discussion of the experiment results is presented and visualized in this chapter. Besides, analysis of the result is also discussed.

Chapter 6 is the conclusion. This chapter concludes the achievements on the project objectives and future work based on the experiment's implementation and the results.

1.9   Summary

This chapter discuss an overview of electric vehicle and the BERT method used for extracting meaningful and valuable insight from the electric vehicle related text. Besides, provide insight on the problem background of the electric vehicle discourse analysis, in order to identifies the research gap and provides significant mapping to the research question, problem statement and research objective. Research scope is also state in this chapter to clarify the specific boundary of the study. A comprehensive overview for the following chapter in this research is also outlined to have a understanding on each chapter specific scope.

REFERENCES

Bhola, A., Senthil Athithan, Singh, S., Mittal, S., Sharma, Y. K., & Jagjit Singh Dhatterwal. (2022). Hybrid Framework for Sentiment Analysis Using ConvBiLSTM and BERT. *2022 2nd International Conference on Technological Advancements in Computational Sciences (ICTACS)*, *14.2*, 309–314. https://doi.org/10.1109/ictacs56270.2022.9987774

Debnath, R., Ronita Bardhan, Reiner, D. M., & Miller, J. R. (2021). Political, economic, social, technological, legal and environmental dimensions of electric vehicle adoption in the United States: A social-media interaction analysis. *Renewable and Sustainable Energy Reviews*, *152*, 111707–111707. https://doi.org/10.1016/j.rser.2021.111707

EY. (2024, October 15). Malaysia's EV market slows amid consumer worries over charging docks and high maintenance costs. Retrieved May 8, 2025, from Ey.com website: https://www.ey.com/en_my/newsroom/2024/10/malaysias-ev-market-slows-amid-consumer-worries-over-charging-docks-and-high-maintenance-costs

Fam, A., & Fam, S. (2024). Review of the US 2050 long term strategy to reach net zero carbon emissions. Energy Reports, 12, 845–860. https://doi.org/10.1016/j.egyr.2024.06.031

Ferdous, S. M., Syed, Bin, S., & Uddin, M. (2024). Sentiment Analysis in the Transformative Era of Machine Learning: A Comprehensive Review. *Statistics Optimization & Information Computing*, *13*(1), 331–346. https://doi.org/10.19139/soic-2310-5070-2113

Fernandez, M. I., Go, Y. I., Wong, D., & Wolf-Gerrit Früh. (2024). Malaysia's energy transition and readiness towards attaining net zero: review of the potential, constraints, and enablers. *Renewable Energy Focus*, *51*, 100640–100640. https://doi.org/10.1016/j.ref.2024.100640

Gardazi, N. M., Daud, A., Malik, M. K., Bukhari, A., Tariq Alsahfi, & Bader Alshemaimri. (2025). BERT applications in natural language processing: a review. *Artificial Intelligence Review*, *58*(6). https://doi.org/10.1007/s10462-025-11162-5

Gaurav, A., Gupta, B. B., & Chui, K. T. (2024). BERT Based Model for Robust Mental Health Analysis in Clinical Informatics. *2024 21st International Joint Conference on Computer Science and Software Engineering (JCSSE)*, 153–160. https://doi.org/10.1109/jcsse61278.2024.10613729

Government of Malaysia. (2025). Car Popularity Explorer | data.gov.my. Retrieved May 8, 2025, from Data.gov.my website: https://data.gov.my/dashboard/car-popularity

Guven, Z. A. (2021). Comparison of BERT Models and Machine Learning Methods for Sentiment Analysis on Turkish Tweets. *2021 6th International Conference on Computer Science and Engineering (UBMK)*, 98–101. https://doi.org/10.1109/ubmk52708.2021.9559014

Hafize Nurgul Durmus Senyapar. (2024). Electric Vehicles in the Digital Discourse: A Sentiment Analysis of Social Media Engagement for Turkey. *SAGE Open*, *14*(4). https://doi.org/10.1177/21582440241295945

Higueras-Castillo, E., Molinillo, S., Coca-Stefaniak, J. A., & Liébana-Cabanillas, F. (2019). Perceived Value and Customer Adoption of Electric and Hybrid Vehicles. *Sustainability*, *11*(18), 4956. https://doi.org/10.3390/su11184956

IPCC, 2022: Climate Change 2022: Mitigation of Climate Change. Contribution of Working Group III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [P.R. Shukla, J. Skea, R. Slade, A. Al Khourdajie, R. van Diemen, D. McCollum, M. Pathak, S. Some, P. Vyas, R. Fradera, M. Belkacemi, A. Hasija, G. Lisboa, S. Luz, J. Malley, (eds.)]. Cambridge University Press, Cambridge, UK and New York, NY, USA. doi: 10.1017/9781009157926

Liang, D., Li, F., & Chen, X. (2023). Failure mode and effect analysis by exploiting text mining and multi-view group consensus for the defect detection of electric vehicles in social media data. *Annals of Operations Research*, *340*(1), 289–324. https://doi.org/10.1007/s10479-023-05649-z

Littlejohn, C., & Stef Proost. (2022). What role for electric vehicles in the decarbonization of the car transport sector in Europe? *Economics of Transportation*, *32*, 100283–100283. https://doi.org/10.1016/j.ecotra.2022.100283

Majekodunmi, T. B., Mohd Shahidan Shaari, Nor Fadzilah Zainal, Harun, N. H., Ridzuan, A. R., Noorazeela Zainol Abidin, & Nur. (2023). Gas Consumption

as a Key for Low Carbon State and its Impact on Economic Growth in Malaysia: ARDL Approach. *International Journal of Energy Economics and Policy*, *13*(3), 469–477. https://doi.org/10.32479/ijeep.14134

Malaysia - Countries & Regions - IEA. (2025). Malaysia - Countries & Regions - IEA. Retrieved May 8, 2025, from IEA website: https://www.iea.org/countries/malaysia/emissions

Portal Rasmi Kementerian Sumber Asli dan Kelestarian Alam. (2025). Retrieved May 8, 2025, from Nres.gov.my website: https://www.nres.gov.my/ms-my/_layouts/15/osssearchresults.aspx#k=50%25electric%20vehicle#l=1086

Ruan, T., & Qin Lv. (2023). Public perception of electric vehicles on Reddit and Twitter: A cross-platform analysis. *Transportation Research Interdisciplinary Perspectives*, *21*, 100872–100872. https://doi.org/10.1016/j.trip.2023.100872

Rahman, B., & Maryani. (2024). Optimizing Customer Satisfaction Through Sentiment Analysis: A BERT-Based Machine Learning Approach to Extract Insights. *IEEE Access*, *12*, 151476–151489. https://doi.org/10.1109/access.2024.3478835

Wibowo, A. S., & Dovi Septiari. (2023). How Does the Public React to the Electric Vehicle Tax Incentive Policy? A Sentiment Analysis. *Journal of Tax Reform*, *9*(3), 413–429. https://doi.org/10.15826/jtr.2023.9.3.150

Wu, Z., He, Q., Li, J., Bi, G., & Antwi-Afari, M. F. (2023). Public attitudes and sentiments towards new energy vehicles in China: A text mining approach. *Renewable and Sustainable Energy Reviews*, *178*, 113242–113242. https://doi.org/10.1016/j.rser.2023.113242

Saleh Mohammed Kutabish, Ana Maria Soares, & Casais, B. (2023). The Influence of Online Ratings and Reviews in Consumer Buying Behavior: a Systematic Literature Review. *Lecture Notes in Business Information Processing*, *485*, 113–136. https://doi.org/10.1007/978-3-031-42788-6_8

Siew, T., Hui, N., & Rebecca, Y. (2024). *A Study of the Emerging Electric Vehicle (EV) Supply Chain in Malaysia*. Retrieved from https://www.iseas.edu.sg/wp-content/uploads/2024/04/ISEAS_Perspective_2024_33.pdf

Sornalakshmi, R. R., Ramu, M., Raghuveer, K., Vuyyuru, V. A., Venkateswarlu, D., & Balakumar, A. (2024). Harmful News Detection using BERT model through sentimental analysis. *2024 International Conference on Intelligent Systems*

and *Advanced Applications (ICISAA)*, 1–5. https://doi.org/10.1109/icisaa62385.2024.10828690

Sharma, H., Din, F. U., & Ogunleye, B. (2024). Electric Vehicle Sentiment Analysis Using Large Language Models. *Analytics*, *3*(4), 425–438. https://doi.org/10.3390/analytics3040023

Tax Relief | Lembaga Hasil Dalam Negeri Malaysia. (2025). Retrieved May 8, 2025, from Lembaga Hasil Dalam Negeri Malaysia website: https://www.hasil.gov.my/individu/kitaran-cukai-individu/lapor-pendapatan/pelepasan-cukai/