CHAPTER 5

CONCLUSION AND FUTURE WORKS

5.1 Introduction

This chapter shows the overall summaries of the finding in for the sentiment analysis of electric vehicle discourse in Malaysia. The result obtain from the exploratory data analysis and sentiment classifier shows an overview result of public reception about electric vehicle related topics. In addition, it also show what possible future improvement can be done on this project for further research improvement on results finding accuracy and more in deep analysis.

5.2 Project Summary

This project is to develop a sentiment classifier model for electric vehicle sentiment analysis discourse in Malaysia. The scrap data from the Reddit consist of discourse containing Malaysian slang with a sample data of 43657. Then the raw data is undergoing data preprocessing step that require before ongoing auto sentiment labeling and model training. Then the performance of both Bert-uncased model and distilled Bert model is compared. In addition, analysis of the sentiment for the electric vehicle is also conducted.

5.3 Overall Achievement

This project had completed 5 chapter including this chapter. Chapter 1 is introduction of electric vehicle discourse, problem background, problem statement, objectives, project scope and the organization of the project. After that, chapter 2 is the literature review section which consist of recent studies and remaining research gap in electric vehicle reviews related domains and researching what type of approach or classifier that had been implement for the natural language text classification. The

next chapter is methodology, which describe every phase of the methodology part starting from data preparation to model evaluation and performance metric evaluation. Chapter 4 shows overall project flow and initial implementation of project model in detail. The initial result obtains from this project experiments also presented in this chapter and discussed.

5.4   Overall Objective Achievement

Table 5.1   Project Objective Achievement

Objective	Achievement	Percentage
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	Data collection had been done through web scraping the Reddit web and obtained a total sample of 43657. Relevant and suitable data preprocessing method had been implemented for the raw collected data.	80%
To Implement and compare pre-trained Bert-based model in determining structured sentiment analysis.	2 pretrained Bert-based model had been implemented for training and predicting the sentiment analysis of the electric vehicle discourse.	80%
To analyze the sentiment insight of Malaysian about electric vehicle	The sentiment insight of the Malaysia discourse whether towards positive, negative or neutral had been analyze.	80%

5.5   Future Works and Recommendations and Limitation

Future work that can be implemented in this project for future improvement and 100% achievement of the project is expanding the data sample by web scraping more online social media not only Reddit. Online social media such as YouTube reviews comments or web review news. This is to explore more variety and diversity of online discourse sentence pattern for the model to learn. In addition, can manually human labelling a small portion of the data sample to have a comparison whether the auto labelling and model training performed well or mostly identical to what human interpretation labelling. As the limitation of raw web scraping data lacks of pre-labelled data. Besides that, include topic modelling for in deep insight extraction. To visualized what is the most common discussion for the subtopic for electric vehicle. This would prevent limitation of just visualizing what are the most common words for each of the positive, negative and neutral sentiment.