UNIVERSITI TEKNOLOGI MARA

EXTRACTING DATA FROM A LARGE DATASET OF SONG TITLES AND REVIEWS TOWARDS MALAY SONGS

NUR HAJAR AMIRAH BINTI ABD AZIZ@ABD MANAF
NUR ATIKAH BT WAN MOHAMAD KAMIL
NURUL AFIQAH BT RAMLI

BACHELOR OF SCIENCE (HONS.) STATISTICS
FACULTY OF COMPUTER AND MATHEMATICAL SCIENCES

JULY 2019

ABSTRACT

Music is the written signs representing instrumental sound that produce beauty of form, harmony and expression of emotion. It is the act or art of singing burst into song. This study focused on analyzing Malay's song titles and reviews. The data was generated from the JOOX music player and Youtube channel. The objectives of this study are to identify the similar words used for song titles and reviews and to determine the relationship between words used for song titles and reviews. There are 250 songs have been collected in various genre with 609 reviews. The data is in a text form and being analyze by extracting word by word for each song's title and reviews. The most common word in Malay song titles and reviews are determine by using tibble function in R programming and presented in cloud form. The relationship among words in song's reviews are measured using Pairwise correlation coefficient. The Malay songs titles are also presented via visualization of a network of bigrams. The analysis of results found that the most common word used for Malay songs titles is *"cinta"*. While the most common word used for Malay songs reviews is song. The relationship among the words in song reviews are statistically significant with positive relationship between sad and you, sad and all, songs and singer, deep and you, deep and sad, and deep and all. *'cinta'* is a common center of node, which is often followed by the song title.

ACKNOWLEDGEMENTS

IN THE NAME OF ALLAH, THE MOST GRACIOUS, THE MOST MERCIFUL

Firstly, we are grateful to Allah S.W.T for giving us the strength to complete this project successfully. Without His permission, we could never finish this project within the appropriate time as we had faces many hurdles and challenges during the implementation process of this research project.

First of all, we wanted to emphasize million thanks Madam Nurul Nisa' bt Khairol Azmi our supervisor for this project who supervised us tolerantly and patiently by providing valuable guidance, advice, support as well as motivation from the very beginning until the completion of this project. She has clearly explained to us the procedure we should consider when doing the project to avoid us from making any mistake during implementing this project. Honestly, all guidance and support that she has given really helps us to overcome our mistakes and errors in the project.

Besides, special thanks also to our parents and family who provided much support, both financial and emotional. Thanks to all the group members for the encouragement and cooperation in term of money, time and effort during conduct this research project.

Last but not least, we wanted to record our appreciation and thanks to one and all who has directly or indirectly participated in the process of implementing our project. We would like to apologise if we have cause any disaffection to any party throughout this project. The best indeed comes from Allah S.W.T and the worst actually comes from our side.
Thank you.

NUR HAJAR AMIRAH BINTI ABD AZIZ @ ABD MANAF
NUR ATIKAH BT WAN MOHAMAD KAMIL
NURUL AFIQAH BT RAMLI

TABLE OF CONTENTS

TOPIC		PAGE
ABSTRACT		i
ACKNOWLEDGEMENT		ii
TABLE OF CONTENTS		iii
LIST OF TABLES		iv
LIST OF FIGURES		v
LIST OF ABBREVIATION		vi
LIST OF APPENDICES		vii
CHAPTER 1:	INTRODUCTION	
	1.1 Introduction	7
	1.2 Problem Statement	9
	1.3 Research Questions	10
	1.4 Research Objectives	10
	1.5 Scope of Study	10
	1.6 Significance of Study	10
CHAPTER 2:	LITERATURE REVIEW	
	2.1 Introduction	11
	2.2 Common word used in song titles and music reviews	12
	2.3 Relationship between word used and song reviews	13
	2.4 Conclusion	13
CHAPTER 3:	METHODOLOGY	
	3.1 Introduction	14
	3.2 Sources of Data	14
	3.3 Method of Analysis	15
	3.4 Conclusion	19
CHAPTER 4:	RESULTS AND DISCUSSION	
	4.1 Introduction	20
	4.2 Organize song according titles and reviews	20
	4.3 Analyzing words used in song titles and reviews	21
	4.4 Relationship between words in song reviews	25
	4.5 Visualizing a network of bigrams in song titles	29

CHAPTER 5: CONCLUSION AND RECOMMENDATION

5.1 Conclusion 30
5.2 Recommendation 32

REFERENCES **33**

APPENDIX A:SINGS DATASET AND R CODING **36**

APPENDIX B: RESEARCH SCHEDULE **41**