UNIVERSITI TEKNOLOGI MARA

TEXT ANALYSIS OF 2019 AUDITOR GENERAL'S REPORT

AISYAH HAMIZAH BINTI AZMI
FARESHA FARHANA BINTI RAHMAN
NOOR DZULAIQA IZAFITRIAH BINTI MOHD ALIAS

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

JULY 2021

ABSTRACT

Text data analysis has become an essential tool in extracting information from enormous amount of online documents. One of the documents that can be analyzed is the Malaysian Auditor General's report. This research was inspired to assist the National Audit Department collect valuable details from the report and to visualise it into a simplest form to monitor. The first objective of this research is to explore the word pattern of Auditor General's Report for 2019. The method used to achieve this is by using the collocation analysis. It is found that the collocation of *telaga tiub* has the highest association strength, measured by lambda which has been standardized. Since *telaga tiub* has the highest probability that exactly follow each other, this research also investigate the words that relate to *telaga* by employing cluster analysis, which is the second objective. The method of clustering used is the Ward's Minimum Variance. There are two clusters of words formed. The first cluster can be classified as authorities that are responsible for the *Telaga Tiub* project which are Kementerian Pendidikan Malaysia and Jabatan Mineral dan Geologi. The second cluster represents the agencies that can get benefits from the *Telaga Tiub* project. As for the third objective, this research also focus on determining the words that are significantly related to specific terms such as *penyelewengan*, *pembaziran*, *gagal*, *kecuaian* and *ketirisan*, using the multiple Fisher's Exact test. The term *penyelewengan* is found to be highly significant with the words *wujud* and *pengawal*. The words *hpkk*, *pengawal*, *memandang*, *diharapkan* and *mengelakkan* are found to be highly significant with term *pembaziran*. As for the term *gagal*, it is found that the term is highly significant with the words *bayaran*, *deposit*, *membayar*, *syarat* and *guaman*. Whereas, the words *skop*, *kolam*, *spesifikasi* and *uwet* are found to be highly significant with the term *kecuaian*. The last term *ketirisan* is found to be highly significant with the words *lkim*, *hasil*, *sewa* and *dikutip*.

ACKNOWLEDGEMENT

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

First and foremost, we are extremely grateful to our supervisor Dr. Nurul Nisa' Khairol Azmi for her invaluable advice, unwavering support, and patience which enabled us to successfully complete the project. Her dynamism, vision, genuineness, and motivation have left an indelible impression on us. She has taught us how to conduct research and present our findings in the most clear and concise manner possible. It was a great privilege and honour to work and study under her guidance.

Apart from our advisor, we would like to thank Madam Zaitul Anna Melisa binti Md Yasin, our lecturer for the Final Year Project subject, for her encouragement, support, and advice during the process of writing and finishing this study. We would also like to express our gratitude for her time as she was always available for consultation whenever needed.

Last but not least and the ones to never be forgotten are our parents. We would like to dedicate our appreciation to our parents for their endless support in making sure we are able to complete this study healthy physically, mentally and emotionally. We are greatly indebted for their love and encouragement throughout this study.

TABLE OF CONTENTS