**Project 1**

**STQD6114 – Unstructured Data Analytics**

**P152419 – Hazim Fitri Bin Ahmad Faudzi**

**Part 2 – Task 1**

**INTRODUCTION**

This study analyses a collection of news articles from technology fields. These articles are extracted from various new site such as The Edge, News Staits Times, Malay Mail, and many more. It covers topics such as recent update from leading tech companies, financial performance reviews, AI breakthroughs, and new global business trends. The goal of this topic-modelling analysis using Latent Dirichlet Allocation (LDA) is to clearly uncover the main themes across these varied articles and highlight the key conversations driving these industries.

**DISCUSSION ON FINDINGS**

**2.1 TOPIC 1: TECHNOLOGY AND BUSINESS INNOVATIONS**

The frequent use of words like “new”, “technology”, “Tesla” and “China” suggest that a suitable topic for these news is technology and business innovations. These words indicates that the main focus is on the latest developments in the technology sector and major changes in the automotive industry. The repeated mention of "Tesla" points to strong attention on the company’s growth and innovations, while the word "China" reflects the country’s increasing role and influence in technology.

**2.2 TOPIC 2: FINANCIAL PERFORMANCE AND MARKET TRENDS**

The dominant terms associated with this topic include "billion," "data," "according," "Chinese," and "June." These keywords collectively point to a thematic concentration on financial reporting, prevailing economic conditions, and market analyses. The recurring presence of "billion" underscores the discussion of substantial financial movements, while "data" and "according" suggest a reliance on financial metrics, market predictions, and company-specific reports. The inclusion of "Chinese" further indicates a focus on financial and market trends within the Chinese economy.

**2.3 TOPIC 3: ARTIFICIAL INTELLIGENCE AND SCALABILITY**

This topic is demonstrably AI-centric, with prominent terms such as "OpenAI," "scale," "models," and numerical indicators like "one" and "billion." The co-occurrence of "OpenAI" strongly suggests a focus on leading AI technology companies and their contributions to the field. Furthermore, the terms "scale" and "models" indicate discussions pertaining to advancements in AI model development, the scalability of technological applications, and associated investments within the artificial intelligence ecosystem.

**2.4 COMPARISON OF TERMS ACROSS TOPIC**

**A graph with different colored bars

AI-generated content may be incorrect.**

Figure 1 Top Terms per Topic

**A graph with red and blue bars

AI-generated content may be incorrect.**

Figure 2 Topic 1 VS Topic 3

**A graph of a bar graph

AI-generated content may be incorrect.**

**CONCLUSION**

The application of topic modelling to the news dataset successfully yielded coherent and clearly distinguishable thematic clusters. This analysis highlights a clear thematic separation among technology innovations, financial market dynamics, and developments within artificial intelligence. While the current analysis provides valuable insights into the predominant discourse trends captured by these news articles, it is posited that further preprocessing, specifically the exclusion of overly general terms, could potentially enhance the precision and interpretability of the identified topics. Nonetheless, the present findings contribute to a better understanding of the key themes permeating contemporary news coverage in the specified sectors.