

Project 1

STQD6114 – Unstructured Data Analytics

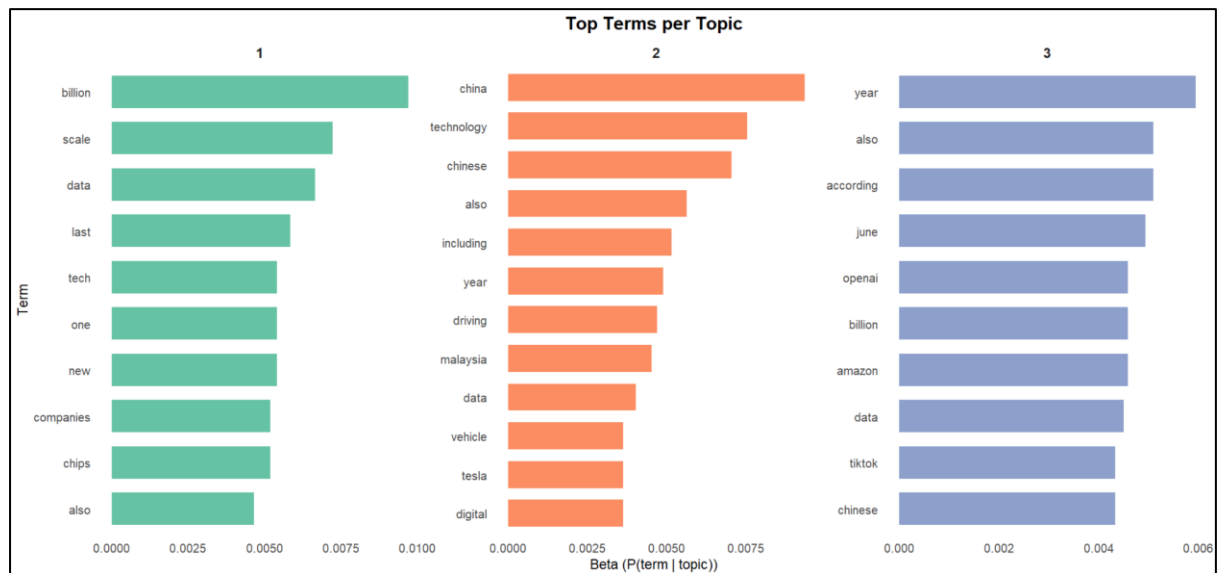
P152419 – Hazim Fitri Bin Ahmad Faudzi

Part 2 – Task 1

INTRODUCTION

Technology has been selected as the theme for this topic modelling analysis using Latent Dirichlet Allocation (LDA). This analysis will focus on the latest trends and advancements changing technology around the world. The news has been compiled from various news websites such as The Edge, The Malay Mail, New Strait Times, etc. and covers topic including artificial intelligence advancements, cybersecurity concerns, telecommunication infrastructure, growth within semiconductor industry, strategic move from leaders of technology companies, innovations in consumer electronics, and the technology regulation adapting. Artificial Intelligence (AI).

DISCUSSION ON FINDINGS



2.1 TOPIC 1: TECH INDUSTRY GROWTH & SEMICONDUCTOR INVESTMENT

Terms like billion, scale, data, last, tech, one, new, companies, chip, appears to be the top 10 terms from topic 1 indicates that this topic focuses on business of technology such as large scale industry development, big tech companies, and hardware. Terms like “billion” and “companies” suggest news about massive investment, market valuation, or revenues in the tech sector. The presence of “chips” together with “billions” shows that focus on semiconductor industry and a massive amount of funding or subsidies for chip manufacturing. Words like “scale”, “data”, “tech”, and “new” further proved that discussion of improvement in technology or data-driven ventures.

2.2 TOPIC 2: CHINESE TECH EXPANSION AND ELECTRIC VEHICLES

From the figure, we can see that the top terms are China, technology, Chinese, also, including, year, driving, Malaysia, data, vehicle, tesla, and digital. These terms clearly shows that this topic is focus on a sector and geospatial, with main concern on China’s influence in technology and the rise of Electric Vehicle (EV). The word “China” and “Chinese” dominating this topic shows that many article are discussing about China advancement in the tech industry. Next, the term “Malaysia” as the top 10 terms shows that there is some regional context. For example, Malaysia and China are working together to boost emerging industries. Lastly, the terms such as “vehicle”, “driving”, and “Tesla” indicate that the topic covers EV. Several news highlights how Chinese are aggressive within the EV industry and has been expanding to southeast Asia especially Malaysia.

2.3 TOPIC 3: ARTIFICIAL INTELLIGENCE BREAKTHROUGH & BIG TECH TRENDS

This topic appears to represent current tech trends and big-name companies, with a strong emphasis on Artificial Intelligence (AI) and major platforms. The presence of “Open AI” is a clear point to mark this topic related to AI. The term “billion” here as in topic 1 likely related to the investment make to the development of AI. Besides, we also see the terms like “Amazon” and “TikTok”. This shows that this topic also covers industry leaders and big tech companies.

2.4 COMPARISON OF TERMS ACROSS TOPIC

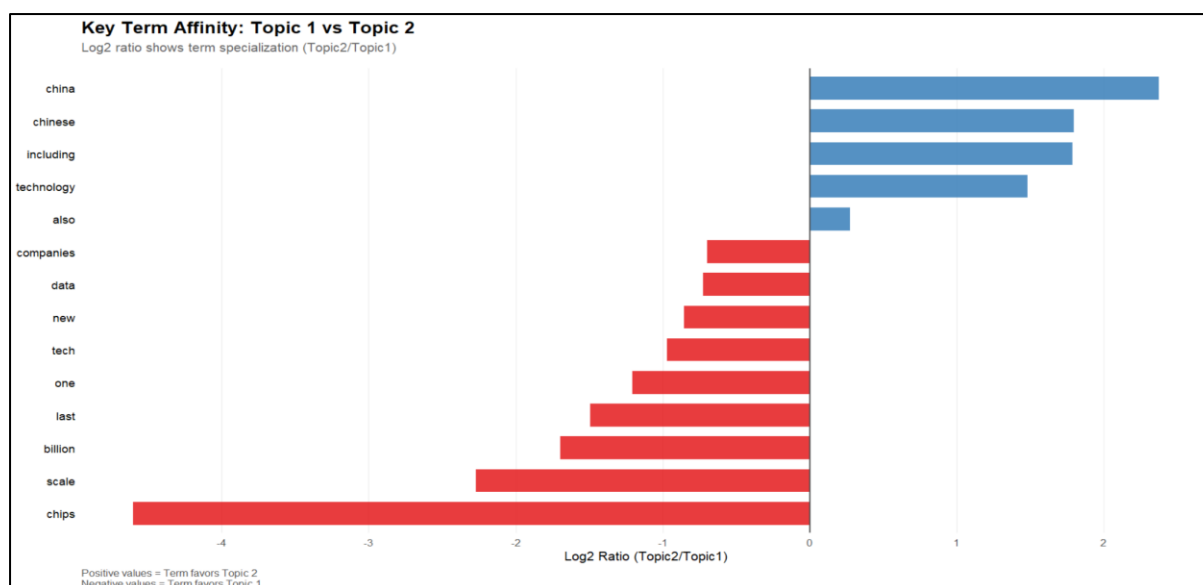


Figure 2 Topic 1 VS Topic 2

The spread in beta values topic 1 and topic 2 shows a clear, defining vocabulary for each. Terms like “chips”, “scale”, and “billion” have very large negative ratios, this means that topic 1 has much higher probability to these terms compared to topic 2. On the other hand, words like “China” and “Chinese” show positive log ratios. This is a strong indicators that both topics are well separated.

CONCLUSION

The application of topic modelling to the news dataset successfully t and clearly distinguish thematic clusters. This analysis highlights a clear thematic separation among EV, semiconductor and AI advancements. This analysis gives useful insights into the main trends in these news articles. However, removing very common words could help make the topics clearer and easier to understand while showing the key themes in today’s news coverage within technology industry.