Project 1

STQD6114 – Unstructured Data Analytics

P152419 – Hazim Fitri Bin Ahmad Faudzi

Part 1 – Task 1

---

1. **Pick one example of raw data. Explain how this raw data turns into wisdom using DIKW pyramid.**

DIKW stands for Data, Information, Knowledge and Wisdom. Data is a raw fact and one example of a raw facts from an insurance underwriter perspective is the claim records of a customer. This includes policyholders age, claim amount, date of claim, type of claim, policy type, health condition and many more.

Next is the information. This is where data is being structured and aggregated to give it more context and meaning. This step helps answer the 'who', 'what', 'where', and 'when' question. By cleaning and summarising the data, the data will show what is real root cause of a certain problem. For example, actuaries might find that a person with more than 20 speeding ticket history tends is two times more likely to files a claim.

Knowledge is next step where pattern recognition from domain knowledge meets with statistics knowledge. By using statistical methods to solve real-world business problems, in this case to assess the risk carried by someone and predict future claims based on historical trends to give an insured person a suitable premium.

Lastly, wisdom where smart data-driven decision making happens. Insurer or insurance company make informed decisions such as adjusting premium for drivers with accidental history, just recently got their driving license, or has a lot of speeding ticket history. Wisdom transforms insights (knowledge) into actionable strategies that balance out risk and profitability.

2. **Based on your opinion, explain why unstructured data gains its popularity nowadays?**

Nowadays, unstructured data is everywhere from social media post, emails, images, videos, and sensor data. It can be said that from all the data in this world, 80% of it is unstructured while remaining 20% is structured data and recently unstructured data has gained its popularity due to several reasons.

The first reason is rise of social media and streaming platforms such as Facebook, TikTok, YouTube and many more. Every type of unstructured data such as text, image, audio, and video is being generated by user each time they do a posting in the social media account. These platforms generate massive amounts of data daily, from tens of terabytes up to petabytes of data.

The next reason is the continuous and rapid advancement of Artificial Intelligence and Big Data Technology. With this, meaningful insights can be extracted from unstructured data such as images, audio, and text by using advanced machine learning and natural language processing techniques.

Unstructured data also gained its popularity not only on academic, but it has helped a lot of industry with its business and personalization needs. For example, sentiment analysis can be used to help companies in analysing their brand perceptions from social media or reviewer. Recommendation systems also help companies like Facebook and Netflix to suggest content based on each customer's viewing habits.

The technology of cloud storage and cloud computing has also become a reason for unstructured data for its increasing popularity. Cloud provider companies such as Amazon, Google and Microsoft has been able to provide lower storage cost and scalable cloud computing make storing and processing large amounts of unstructured data become more affordable and efficient.

In conclusion, unstructured data has become popular and the skills to analysing the data has become important to company to make smarter decisions. Unstructured data will only

become more popular in the future with its flexibility, rich insight potential and technological advancements.