



UNIVERSITI KEBANGSAAN MALAYSIA

The National University of Malaysia

PROJECT 1

STQD6114

UNSTRUCTURED DATA ANALYTICS

DR. NOR HAMIZAH BINTI MISWAN

HAZIM FITRI BIN AHMAD FAUDZI

P152419

**1. Two popular examples of big data technology are Netflix and credit-card.
Explain how big data is used in these examples.**

Netflix: Netflix using big data technology to learn about their customers' preferences and behaviours in watching and selecting movies. This can help customers to find their favourite movies without searching for it and Netflix will be able to retain customers.

Credit card: Credit card provider uses big data technology to detect fraud. For example, if a card was used at a certain place, then it was used again in a short time but at a very far location from the previous use which is impossible to cover by human in that short period of time. It will suspect of fraud.

2. For EVERY application of big data found in different fields (e.g., banking, telecommunication, healthcare, media, etc.) mentioned in the video, describe two real-life examples of each field.

Banking:

- Credit scoring: use customer data such as bank balance, default rate to determine the creditworthiness of a customer.
- fraud detection: detect unusual or impossible activity such as sudden large withdrawal or card used in two very far places in a very short time.

Telecommunication:

- network optimization: predict congestion using call data record to improve coverage
- customer churn prediction: use customer usage behaviour data to predict if customer is likely to change telco provider.

Healthcare:

- disease identification: using deep learning to identify disease from result of X-ray, CT scan, MRI
- precision medicine: optimize trial and error on drug test to identify potential medicine and optimize existing medicine

Media:

- content personalization: identify user habit and preferences to recommend similar content to user
- target marketing: display only related marketing advertisement to user based on their preferences

Advertising:

- enabling targeted advertising: discover what influences customer decision and display only related content to gain the highest engagement.
- optimize marketing: discover which marketing strategies that works the best and spend money accordingly to avoid wasting money on ineffective marketing.

Manufacturing:

- Predicting maintenance schedule using predictive analytics to predict when a machine should be maintained to make sure machine always in a good condition.
- quality control: use deep learning and AI to inspect products for abnormalities, faults, and defects.

Transportation:

- GPS: using GPS and weather data can optimize route to reduce delay in delivery and optimize fuel.
- Predictive maintenance: use predictive analytics on data from sensors and operational logs to predict equipment failures to minimize downtime.

Retail:

- price optimization: use predictive analytics on price history to determine to best price to put on each season and on promotion.
- enhance customer service: use customer service data and develop chatbot to help solve customer's problems faster and always available 24/7.

3. Construct the table to differentiate between structured, unstructured and semi-structured. Give at least 3 points, including sources of the data.

	Structured	Unstructured	Semi-Structured
Storage	Database	No clear format	Not in database
Sources	Machine (sensors, financial systems, weblogs, etc.) and human (personal details, behaviour)	Machine (satellite image, scientific data, radar data) and human (social media, website content, mobile data)	Machine (IoT devices, log files, NoSQL database) and human (social media image & caption, emails)
Examples	Medical devices, GPS, stock market	Facebook post, blackhole image, youtube video	API response, webpages, open-ended survey