NEWSLETTER

Department of Statistics Malaysia

STATS BDA
BIG DATA ANALYTICS

www.dosm.gov.my

@StatsMalaysia

DOSM/BMP/2.2019/Series 4

WHAT IS BIG DATA ?

There another 35 V's...

Validity · Vastness · Varifocal · Victual · Vogue · Valor · Verdict · Venue · Vexed · Voodoo · Vet · Veil · Visibility · Varnish · Volatility · Varmint · Vane · Vault · Vivify · Voice · Vagueness · Veer · Viability · Virtuosity · Vantage · Voyage · Vanilla · Versed · Vibrant · Vulpine · Viscosity · Vaticination · Version Control · Vocabulary · Viral

How much data are we talking about ?

Data is exploding
Today, every minute we are generating the same amount of data that was created from the beginning of time until the year 2000.

Over **100 hours of video** are uploaded to YouTube every minute and it would take around 15 years to watch every video uploaded by users in one day

Google alone processes on average over 40 thousand search queries per second, making it over 3.5 billion in a single day.

Every minute over 200 million emails send, generate almost 2 million Facebook likes, send over 250 thousand Tweets, and upload over 200,000 photos to Facebook.

Over 90% of all the data in the world was created in the past 18 months.

PUBLIC MATURITY ASSESSMENT ON OFFICIAL STATISTICS (PMAOS)

REAL TIME NEWS ON OFFICIAL STATISTICS (RTOS)

PMAOS and RTOS created by DOSM are used to obtain the sentiments of social media such as Facebook comments and tweets. It determines the degree of "happiness" of the community statistics.

Sentiment Analysis

MAIN PAGE

INFLUENCER

DASHBOARD

MEDIA SHARE

DASHBOARD STATISTICS

CLUSTERING

TOP PERSON

SENTIMENT

ONTOLOGY

➤ Calculated based on comparison of positive/neutral/negative tone exposed on the media.
➤ Measure the headlines of articles as well as statements exposed in each article.
➤ The best measurement on the heat of an issue. The more negative tone is exposed, the hotter the issue from public perspective.

SOCIAL MEDIA ANALYSIS
- Statistics of general trend of media account activities
- Top active account contributing comments intensity analysis (drilldown until hour level)
- Fans growth

DEPENDENCY
- Timeline exposure dependency
- Influencer dependency
- Media share dependency

☐ Work based on accuracy whereby the system is trying to detect the sentiment tone of specific sentence automatically.
☐ Perform according to machine learning algorithm on the keywords
☐ Not 100% accurate as it cannot understand the context of the sentence
☐ Different with human being who are able to use reasoning ability

TOPIC REPORT
- Social media account
- Monitoring report
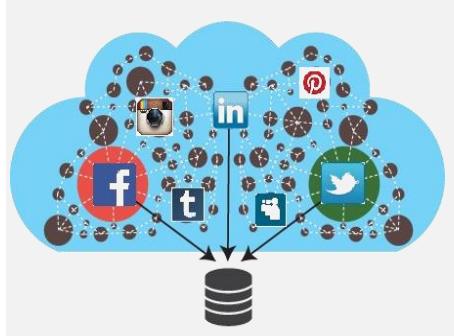
✓ Based on the limitation of research, sentiment analysis is not advised to be used as measurement to find out who is against or who is supporting the issue.
✓ In the media industry, sentiment analysis is performed to measure the heat of the issues as part of overall media intelligence measurements

SOURCE: https://www.kdnuggets.com/2017/04/42-vs-big-data-data-science.html

#StatsMalaysia   #mystatsday   #myretailcensus2019   #hies2019   #isiwsc2019   #mycensus2020   #LeaveNoOneBehind