

AI & UNCONVENTIONAL DATA SOURCES EMPOWERED DOSM DATA ANALYTICS

Faiza Rusrianti Tajul Arus, Fatin Ezzati Mohd Aris, Nur Fathin Mohd Khalil Big Data Analytics Core Team, Department of Statistics, Malaysia

TATISTIK & SCIENTIFIC POSTER 2021 21 SEPTEMBER 2021 | SELASA









PROBLEM STATEMENTS





PROBLEM STATEMENTS

- **1.** How can Artificial Intelligence (AI) be used in everyday life?
- 2. How AI and unconventional data enable organisations to leverage their data accurately?
- 3. How DOSM leverages AI in compilation of statistics / big data analytics?





RESEARCH OBJECTIVES





RESEARCH OBJECTIVES



6



INTRODUCTION





OECD definition

"An AI system is a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations, or decisions influencing real or virtual environments. It does so by using machine and/or human-based inputs to: i) perceive real and/or virtual environments; ii) abstract such perceptions into models through analysis in an automated manner (e.g. with machine learning, or manually); and iii) use model inference to formulate options for information or action. AI systems are designed to operate with varying levels of autonomy." (OECD, 2019a)

Eurostat definition

"Artificial intelligence refers to systems that use technologies such as: text mining, computer vision, speech recognition, natural language generation, machine learning, deep learning to gather and/or use data to predict, recommend or decide, with varying levels of autonomy, the best action to achieve specific goals. ."

English Oxford Living Dictionary definition

STATISTIK & SCIENTIFIC POSTER

"The theory and development of computer systems able to perform tasks normally requiring human intelligence, such as visual perception, speech recognition, decision-making, and translation between languages."

AI DEFINITION





UNSD definition

"Experimental data, data under development and particular useful indicators (i.e.business cycle turning points)."

Source: Eurostat; United Nations Statistics Division (UNSD), "Handbook on Rapid Estimates", Publications Office of the European Union, Luxembourg, 2017

World Bank definition

"Valuable information such as social media, discussion forums, mailing lists, mobile phone data, websites and news outlets. These channels can complement traditional sources to help authorities monitor results and feedback in real-time and improve performance and outcomes"

Source: the 2017 World Government Summit on Big Data for Government in collaboration with the World Bank

STATISTIK & SCIENTIFIC POSTER

UNCONVENTIONAL DATA DEFINITION





Automation process

- automating analytics tasks;
- analyse data autonomously; and
- useful as long as it is accessible.

Wider scope and increased efficiency

- expand the scope of analytics where the semi-structured and unstructured data is becoming analysable; and
- advance of AI algorithms and machine learning.

Real-time data

- Solutions to conventional data problems;
- Data quality improvement by combination of conventional and unconventional data sources; and
- Main contributions of AI to the analytics capabilities

STATISTIK & SCIENTIFIC POSTER

Privacy

-

- the demand for anonymized data; and
- the data protection legislation.

• Manual to AI techniques.





STATISTIK & SCIENTIFIC POSTER 2021



EVERYDAY EXAMPLES OF AI & ML

Everyday examples of Artificial Intelligence and Machine Learning:

- <u>Self-driving cars</u>
- Google's AI-Powered Predictions: <u>location data from smartphones</u>
- <u>Uber's use of machine learning</u> to identify the pickup location of passengers
- Commercial Flights Use an AI Autopilot
- E-mail Spam Filters Gmail successfully filters 99.9% of spam
- Plagiarism Checkers
- Fraud Prevention uses neural networks to predict fraudulent transactions
- Credit Decisions for a loan or credit card <u>machine learning could be used to reduce a</u> <u>bank's losses</u>
- Pinterest automatically identify objects in images
- Instagram uses machine learning to identify the contextual meaning of emoji
- Snapchat using machine learning to track movements in video
- Online Shopping Recommendations <u>Amazon uses artificial neural networks</u> to generate these product recommendations
- Voice-to-Text: Google <u>uses artificial neural networks to power voice search</u>





Automatic Speech Natural Langue Recognition Processing

Text to Speech





LITERATURE REVIEW





AI IN EVERYDAY LIFE

How can Artificial Intelligence (AI) be used in everyday life?

D.W. HALL AND **JAIN AND GANDHI MIRIALYS ET AL. JEROME** (2021)(2018)(2017)1. Al can reduce the burden of 1. Al have been applied to 1. Al has the potential to searching large datasets different areas related to the enhance the retailers and especially in legal sector. transportation environment. consumer engagement 2. In healthcare, data from 2. Al can control the vehicle 2. Al assist human in smartphones and fitness trackers system, traffic flow prediction computerized methods is analyzed using ML techniques in urban area and road safety such as GPS and Waze, to improve management of and accident prediction. Face recognition by chronic conditions. The result of this study reveal uploading pictures on 3. Early identification of potential that the combination of social media by identifies pandemics and tracking incidence different AI techniques can the person and tags them of the disease to help prevent and manage and analyze the 3. Al influenced in contain its spread. massive amount of data medical fields 4. Fraud detection and anti-money generated in transportation. laundering. **STATISTIK & SCIENTIFIC POSTER**

14



TO OVERCOME LACK OF KNOWLEDGE WORKERS

How can Artificial Intelligence (AI) be used in everyday life?



STATISTIK & SCIENTIFIC POSTER



POTENTIAL IMPACTS ON SOCIETY

STATISTIK & SCIENTIFIC POSTER

1. Application of AI allows preventing damage to human health from agricultural chemicals during reinforcement agriculture.

ELAHI ET AL. (2019)





THE USE OF UNCONVENTIONAL DATA

BERCHIALLA ET AL. CAO Q ET AL. Multitudes of Content Content (2012)(2016) 1. Unconventional data 1. Forecast production for sources like newspaper article existing and new wells by has the advantage to make use geological maps, early alerting on Injury production history, Surveillance System Architecture of pressure data and 2. Use Machine Learning Artificial Neural Network operational constraints through text analysis 2. Use Machine Learning 3. The data provide details method – Artificial Neural that are not available from traditional public health Netwrok datasets Output Inpu

Hidden

STATISTIK & SCIENTIFIC POSTER



POTENTIAL FOR DECISION MAKING

STATISTIK & SCIENTIFIC POS

LAKSHMI AND DUBEY ET AL. **CORBETT (2020)** (2020)1. Big data analytics (BDA) powered by artificial intelligence to extract more meaningful information with which to organisations can improve their decision-making skills 2. BDA powered by AI enables interpret and combine complex information derived from various sources to reduce uncertainties

1. Al is primarily being applied to increase productivity and efficiency and secondarily address labor shortages and environmental sustainability concerns 2. Al in industrial transformation, increase profitability and productivity

THE ESSENTIAL PREPARATION FOR THE AI IMPLEMENTATION IN NATIONAL STATISTICS OFFICE





PREPARATION FOR AI IMPLEMENTATION IN NSO

Educating people about AI and Digital Technology

- Basic AI concepts and AI programming;
- Data and digital literacy;
- Ethics of AI; dan
- Online safety.



STATISTIK & SCIENTIFIC POSTER

Business Priority and Strategy

- How can AI help for business delivery; and
- What problems need to solve.

Ethical and Legal Issues

- How to avoid invading people's privacy; and
- Are there any legal implications of using AI.



Technology Issues

- The right technology in place; and
- The right system to put in place.

Skills and capacity

- Skills gaps; and
- Talent and train existing staff.



Change management issues

- Change process management; and
- Whom will be impacted by this AI implementation.



No.	NSO	Project	Remarks
1.	Statistics Canada	Industry and Occupation Coding	https://github.com/UNECE/CodingandClas sification Statcan
2.	Statistics Belgium	Sentiment Coding of Flemish Tweets	https://github.com/jmaslankowski/WP7- Population-Life-Satisfaction
3.	Statistics Poland	eCOICOP classification	https://github.com/statisticspoland/ecoic op_classification
4.	National Center for Statistics and Information, Oman	Utilizing mobile positioning data for official statistics	Webinar Presentation
5.	Badan Pusat Statistik (BPS), Indonesia	Commuting statistics using mobile positioning data in Indonesia	ESCAP Stats Brief Issues Big Data for Population and Social Statistics

KOLOKUM STATISTIK & SCIENTIFIC POSTER 2021



AI IMPLEMENTATION IN OTHER NSO

NSO Project Remarks No. National Bureau of Using web scraping price data for the price index of e-commerce 6. 1. Statistics of China Crop survey by farmland: using satellite and aerial remote sensing to 2. Readiness help estimate agricultural statistics of National Comparison of data of interbank transactions with retail sales: credit 3. **Statistical** card data for use in verifying retail sales Systems in Application of big data for highway and waterway transport statistics 4. Asia and Online price changes of means of production 5. the Pacific Big data enterprise statistical indicator 6. for 7. **Statistics Korea** 1. Online price index Leveraging Daily migration of population: using mobile call detail record data for 2. Big Data to daily migration data Monitor the 8. Singapore Department of Integrated environment system: using environmental sensing systems 1. SDGs Statistics (DOS) and data analytics for real-time environment information Population estimates: using administrative data from many sources for population estimates

STATISTIK & SCIENTIFIC POSTER



DOSM'S AI INITIATIVE





BIG DATA ANALYTICS BY NSO

National Statistical Office are increasingly use data from the internet as a source for their statistics CUMPANY

> SCHEMAS, CHARTS, TABLES GRAPHS

1-1-1

Modernization of data collection tools

Adoption of web scraping techniques Automated processes implemented using a web crawler

> Data is gathered and copied from the web

• A collection of automated and semi automated techniques

Data retrieval and analysis

Discovering; Previously unknown patterns in data; and

Relationship that can be used for prediction of user-relevant quantiles





Big data implementation in Malaysia Trade and Price domain





INITIATIVE 1: INTERNAL PORTAL FOR TRADE BY ENTERPRISE CHARACTERISTICS (TEC)

The initiative is to produce an insights of trade statistics. The integration able to identify the enterprises that are engaged in international markets as well as to describe their characteristics.

To initiate TEC database, DOSM has to attribute trade flows to enterprises with different characteristics by merging data on international trade from Royal Malaysian Custom Department with statistical business register information on enterprises at the individual enterprises level.





INITIATIVE 2: INTERNAL PORTAL FOR PRICE INTELLIGENCE (PI)

STATISTIK & SCIENTIFIC POSTER



Modernization of data collection tools for improving quality of Consumer Price Index (CPI). The modernization of data collection mainly consist of the adoption of web scraping techniques to scrape price data from related website for CPI compilation.



PUBLIC MATURITY ASSESSMENT ON OFFICIAL STATISTICS

INITIATIVE 3: PUBLIC MATURITY ASSESSMENT ON OFFICIAL STATISTICS

The analysis and assessment of the degree of "happiness" of Malaysia community with regards to official statistics published by DOSM. The data is obtained from online social media.

STATISTIK & SCIENTIFIC POSTER





REAL TIME BUSINESS STATUS

INITIATIVE 4: REAL TIME BUSINESS STATUS

STATISTIK & SCIENTIFIC POSTER





An intelligent patching system is to link the current status and financial report of enterprises from Companies Commission of Malaysia website (e-Search & e-Info System) with MSBR in NEWSS at real time.





REAL TIME NEWS ON OFFICIAL STATISTICS

INITIATIVE 5: REAL TIME NEWS ON OFFICIAL STATISTICS

STATISTIK & SCIENTIFIC POSTER

A daily comprehensive up-to-date news on official statistics which is aggregated from sources worldwide. The news are selected ranked based on certain characteristics of news such as topic, freshness, location, relevance etc.







UNCONVENTIONAL DATA SOURCES IN DOSM









Social Media

- Public Maturity Assessment On Official Statistics (PMAOS)
- Informal Sector and Industry Micro (SIIM)

Satellite Imagery by Earth Observation

- DOSM Smart Listing
- DOSM Smart Map
- Malaysia Statistical Address Register
- Collaboration project DOSM, MySA and Malaysian Rubber Board

STATISTIK & SCIENTIFIC POSTER

Web Crawling

Newspaper Articles

- Price Intelligence (PI)
- Informal Sector and Industry Micro (SIIM)

 Real Time News On Official Statistics (RTOS)

31

ARTIFICIAL INTELLIGENCE BRANCHES IN DOSM





ARTIFICIAL INTELLIGENCE PROJECT IN DOSM

DOSM AI PROJECTS: CURRENT AND FUTURE





MOVING FORWARD





LEVERAGING STATSBDA PLATFORM





Berchialla, P., Scarinzi, C., Snidero, S., Rahim, Y., & Gregori, D. (2012). Information extraction approaches to unconventional data sources for "Injury Surveillance System": the case of newspapers clippings. Journal of medical systems, 36(2), 475-481.

Boztas, S. (2017). Automated holidays: how AI is affecting the travel industry, The Guardian, Sustainable Business, <u>www.theguardian.com/sustainable-business/2017/feb/17/holidays-travel-automatedlastminute-expedia-skyscanner</u>

Cao, Q., Banerjee, R., Gupta, S., Li, J., Zhou, W., & Jeyachandra, B. (2016, June). Data driven production forecasting using machine learning. In SPE Argentina Exploration and Production of unconventional resources symposium. OnePetro.

Damioli, G., Van Roy, V., & Vertesy, D. (2021). The impact of artificial intelligence on labor productivity. *Eurasian Business Review*, 11(1), 1-25.

Davenport, T. H. (2018). From analytics to artificial intelligence. *Journal of Business Analytics*, 1(2), 73-80.

Dubey, R., Gunasekaran, A., Childe, S. J., Bryde, D. J., Giannakis, M., Foropon, C., ... & Hazen, B. T. (2020). Big data analytics and artificial intelligence pathway to operational performance under the effects of entrepreneurial orientation and environmental dynamism: A study of manufacturing organisations. *International Journal of Production Economics*, 226, 107599.

Elahi, E., Weijun, C., Zhang, H. and Nazeer, M. (2019), "Agricultural intensification and damages to human health in relation to agrochemicals: application of artificial intelligence", Land Use Policy, Vol. 83, pp. 461-474.

Farrow, E. (2019), "To augment human capacity—artificial intelligence evolution through causal layered analysis", Futures, Vol. 108, pp. 61-71.

TATISTIK & SCIENTIFIC POSTER



Fountaine, T., McCarthy, B., & Saleh, T. (2019). Building the AI-powered organization. *Harvard Business Review*, 97(4), 62-73.

Furman, J., & Seamans, R. (2019). AI and the Economy. Innovation policy and the economy, 19(1), 161-191.

Hall, W. and Pesenti, J. (2017). Growing the Artificial Intelligence Industry in the UK. [online] Available at: https://assets.publishing.service.gov.uk/ government/uploads/system/uploads/attachment_data/file/652097/ Growing_the_artificial_intelligence_industry_in_the_UK.pdf

Jain, S., & Gandhi, A. V. (2021). Impact of artificial intelligence on impulse buying behaviour of Indian shoppers in fashion retail outlets. *International Journal of Innovation Science*.

Lakshmi, V., & Corbett, J. (2020, January). How artificial intelligence improves agricultural productivity and sustainability: A global thematic analysis. In *Proceedings of the 53rd Hawaii International Conference on System Sciences*.

Majumdar, D. & Chattopadhyay, H.K. (2020). "Artificial intelligence and its impact on the society", International Journal of Law, Vol. 6 No. 5, pp. 306-310.

Mazilescu, V. (2019). Tourism dan travel can effectively benefit from technologies associated with Industry 4.0. In *International Conference on Risk in Contemporary Economy.*

Mirialys, M., Julio, A.S., Piedad, G. and Francisco, J. M. (2018, April). "On the use of Artificial Intelligence techniques in Intelligent Transportation Systems". In 2018 IEEE Wireless Communications and Networking Conference Workshop (WCNCW)

Pappalardo, L., Smoreda, Z., Pedreschi, D. and Gianotti, F. (2015). Using Big Data to study the link between human mobility and socio-economic development. Paper presented at the IEEE International Conference on Big Data, Santa Clara, CA.

Popkova, E.G., Litvinova, T., Mitina, M.A. and French, J. (2018b), "Social advertising: a Russian perspective", *Espacios*, Vol. 39 No. 1, pp. 17-26.

Sanchez, J.A. (2018). The use of machine learning in official statistics.



THANK YOU





SECARA ATAS TALIAN

M