

Tutorial ICT

Bersama **FIRKHAN**

firkhan.uthm@yahoo.com



Teknologi data raya

Baru-baru ini, seorang ahli politik tempatan menerusi satu pertubuhan bukan kerajaan (NGO) menampilkan hasil dengan kajian politiknya. Hasil kajiannya mendapati banyak negeri di Semenanjung Malaysia akan tumbang kepada pihak pembangkang dan satu parti politik lama di negara ini tidak menang apa-apa kerusi pada pilihan raya umum akan datang.

Hasil kajiannya menyamai apa yang beliau jangkakan sebelum ini dengan mengaku menggunakan teknologi data raya atau data besar atau *big data* dalam kajian tersebut. Namun begitu, beliau tidak nyatakan jenis teknologi, kaedah atau pendekatan data raya yang digunakan dengan memberitahu jumlah data yang diperolehi adalah dalam sekitar 500,000.

Kaedah dapatan data pula adalah menerusi temu bual di telefon dengan soalan-soalan yang telah ditetapkan. Jadi, jenis data yang diperolehi adalah jenis berstruktur dan separa struktur.

Adakah ini dikatakan teknologi data raya?

Teknologi data raya tidak lagi terhad kepada sistem pangkalan data atau sistem maklumat yang mengoperasikannya. Data raya juga merangkumi data jenis tidak berstruktur seperti *metadata*, data-data yang diperolehi daripada log fail seperti log web serta dari sumber laman web sosial, blog, peranti elektronik seperti peranti Identifikasi Frekuensi Radio (RFID), Wi-Fi dan sebagainya.

Saiz keseluruhan data ini sangat besar sehingga

ini sangat besar sehingga boleh mencapai kepada unit Terabait (TB), Pentabait (PB), Exabait (EB) atau Zettabait (ZB). Disebabkan saiz data ini terlalu besar sehinggakan kadangkala agak sukar untuk mejangkakan saiznya, ia dikenali sebagai data raya.

Data raya ini bukan sahaja boleh diklasifikasikan ciri-cirinya dari segi saiz atau jumlahnya (*volume*) tetapi berdasarkan kepada kadar kelajuan aliran data (*velocity*) dan pelbagai jenis data atau *variety*.

Setakat jumlah data sebanyak 500,000 boleh dipertikai penggunaan teknologi data raya dalam kajian tersebut.

Kebanyakan teknologi data raya bertransaksi melalui talian Internet sama ada berwayar atau tanpa wayar daripada pengguna akhir atau pelanggan ke organisasi. Oleh itu, organisasi perlu pastikan kemudahan infrastruktur untuk aliran data bagi data raya ini adalah terbaik.

Sumber teknologi data raya berpunca dari pelbagai jenis seperti web, peranti elektronik, telefon pintar dan sebagainya. Data raya juga terdiri daripada pelbagai jenis dan aspek integrasi antaranya perlu diambil perhatian. Biasanya antara format yang terbaik untuk mengintegrasikan data raya ini adalah format XML atau menggunakan aplikasi NoSQL, Hadoop dan lain-lain.

Sebab itu, dapatan daripada teknologi data raya ini memerlukan kaedah analitik data yang sangat rumit dengan menggunakan pendekatan kepintaran buatan (AI) dan pembelajaran mesin.

Janganlah sewenang-wenangnya menggunakan terma teknologi untuk kepentingan peribadi masing-masing bagi mendapat perhatian khalayak ramai tanpa berpijak di bumi yang nyata. Jadi, pendekatan kajian yang boleh dipertikai membolehkan hasil kajian juga boleh dipersoal.