Sains Malaysiana 53(11)(2024): 3607-3615
http://doi.org/10.17576/jsm-2024-5311-05

Al-Khawarizmi Heuristik bagi Pautan Data dalam Menganggarkan Bilangan Kemalangan Jalan Raya Tidak Terlapor

(A Heuristic Algorithm of Data Linkage in Estimating the Number of Unreported Traffic Accidents)

ZAMIRA HASANAH ZAMZURI* & NOR WAZIRAH RADZMAN SHAH

Jabatan Sains Matematik, Fakulti Sains dan Teknologi, Universiti Kebangsaan Malaysia, 43600 UKM Bangi, Selangor, Malaysia

Diserahkan: 30 April 2024/Diterima: 7 Ogos 2024

ABSTRAK

Analisis data kemalangan jalan raya adalah sangat penting bagi merancang strategi pencegahan yang optimum serta meminimumkan risiko berlakunya kemalangan. Bilangan kemalangan jalan raya yang dilaporkan sering kali menunjukkan kekerapan sifar yang tinggi, yang dipercayai berasal daripada situasi kemalangan yang tidak dilaporkan. Maka, penganggaran kemalangan tidak dilaporkan adalah amat penting bagi mengelakkan risiko terkurang anggaran dan ketidaktepatan dalam analisis kemalangan jalan raya. Salah satu cara untuk menganggarkan kemalangan tidak dilaporkan ini adalah menerusi perbandingan dua set data dan kadar entri data yang tidak dapat dipadankan menjadi kadar kemalangan tidak dilaporkan. Kajian ini menggunakan teknik pautan data berkebarangkalian bagi memautkan dua set data kemalangan jalan raya yang berasal daripada laporan polis dan rekod hospital dari Januari sehingga Mac 2011. Satu al-Khawarizmi heuristik dibangunkan berdasarkan keperluan semasa proses pautan data dijalankan. Unsur heuristik ini menitikberatkan proses pautan data secara berperingkat bagi mengenal pasti set pengecam yang tidak unik yang digunakan serta tapisan data yang bersesuaian dan rasional bagi anggaran yang ingin dicapai. Seterusnya penukaran unit bagi setiap entri data dari per individu ke per kemalangan juga diperlukan kerana matlamat akhir adalah untuk memperoleh jumlah kemalangan yang tidak dilaporkan. Pautan data yang dijalankan dalam kajian ini menggunakan pengecam bukan unik seperti jantina, umur, bangsa dan jenis kenderaan. Berdasarkan data yang dipautkan dan proses penganggaran yang dilaksanakan, dianggarkan sekitar 68% kemalangan adalah tidak dilaporkan dengan bilangan sebanyak 6366. Al-Khawarizmi heuristik yang dibangunkan ini dapat digunakan untuk pautan data kemalangan jalan raya antara laporan polis dan rekod hospital di Malaysia. Perbandingan antara kemalangan yang dilaporkan dan tidak dilaporkan dalam data hospital turut mendedahkan bahawa kebanyakan kemalangan yang tidak dilaporkan melibatkan kesalahan jenayah serius seperti penggunaan dadah dan alkohol berlebihan.

Kata kunci: Heuristik; kemalangan jalan raya; pautan data; tidak dilaporkan

ABSTRACT

Traffic accident data analysis is vital in order to plan for optimal preventive measures and minimizing the risk of accident occurrence. Oftentime, the traffic accident count data exhibit extra zeros, believed to be sourced from underreporting scenarios. Hence, the estimation of unreported accidents is needed to avoid under-estimation risk and inaccuracy of traffic accident data analysis. One of the ways in estimating the unreported accidents is through comparing two data sets and the proportion of unmatched entries is estimated to be the underreporting rate. In this study, the probabilistic data linkage techniques is used to link two traffic accident data sets sourced from police report and hospital records from Jan to Mar 2011. A heuristic algorithim is developed based on the needs found during the linkage process. The heuristic elements can be found in the staged data linkage process to establish the best set of non-unique identifiers and also on identifying suitable and rational filtered data to be used in the estimation. Then, the unit for data entry needs to be converted from per individual to per accident, since the ultimate aim of this study was to estimate the number of unreported accidents. In the performed data linkage process, the non-unique identifiers used are gender, age, race and vehicle type. Based on the linked data and estimation process performed, the estimate of unreported accidents is around 68% and the estimated number of reported accident is 6366. The developed algorithm can be used in linking traffic accident data based on police report and hospital record in Malaysia. The comparison of reported and unreported accidents in the hospital record shows that most unreported accidents are involving serious offences such as excessive drug and alcohol usage.

Keywords: Data linkage; heuristic; traffic accidents; unreported

PENDAHULUAN

Kemalangan jalan raya (KJR) yang mengakibatkan kecederaan merupakan antara penyumbang kepada kebimbangan dari segi kesihatan secara umumnya dan telah mengambil tempat dalam senarai antara pembunuh utama bagi kematian global (Ahmed et al. 2023; WHO 2023). Isu berkenaan kemalangan jalan raya sering menjadi topik perbincangan pelbagai pihak dan kebimbangan terhadap peningkatan bilangan kadar kemalangan jalan raya ini perlu diberi perhatian sewajarnya, yang mana penilaian risiko sesuatu kemalangan dapat membantu dalam strategi mencegah berlakunya kemalangan atau meminimumkan keparahan sesuatu kemalangan (Isa & Zamzuri 2022). Organisasi Kesihatan Dunia (WHO 2023) turut menyatakan 92% daripada kematian di jalan raya berlaku di negara-negara yang berpendapatan rendah dan sederhana walaupun negara-negara ini memiliki kira-kira 60% daripada jumlah kenderaan di dunia. Merujuk kepada konteks kemalangan jalan raya di Malaysia pula, terdapat peningkatan yang ketara bagi setiap tahun kajian dijalankan (tahun 2010 hingga 2019). Statistik yang direkodkan oleh Kementerian Pengangkutan Malaysia menunjukkan peningkatan hampir 37% dalam tempoh ini berdasarkan kepada 414,421 kes pada tahun 2010 dan 567,516 kes pada tahun 2019 (Laporan Statistik Pengangkutan Malaysia 2019).

Laporan polis bagi kemalangan jalan raya merupakan pangkalan data statistik rasmi bagi kebanyakan negara di dunia termasuklah Malaysia. Pelaporan kadar kemalangan jalan raya adalah berbeza bagi setiap negara tetapi hampir kesemuanya terdiri daripada beberapa kategori yang sama seperti lokasi, waktu kejadian, keadaan persekitaran, jenis jalan, umur mangsa yang terlibat, tahap kecederaan serta bilangan kenderaan yang terlibat dalam sesuatu kemalangan jalan raya (Muni et al. 2021). Walau bagaimanapun, data kemalangan jalan raya sering dilaporkan mengandungi lebihan sifar yang dipercayai disebabkan oleh kemalangan tidak terlapor (KJRTT) (Zamzuri 2021). KJRTT merujuk kepada situasi apabila kemalangan yang berlaku tidak direkodkan atau dilaporkan kepada pihak polis (Radzman Shah & Zamzuri 2023).

Kajian yang dijalankan oleh Ward, Lyons dan Thoreau (2006) bersetuju bahawa terdapat ketidaklengkapan maklumat mengenai kematian yang dilaporkan kepada polis. Perkara yang sama berlaku di sub-Sahara Afrika, kesukaran untuk memperoleh data yang tepat dan lengkap adalah terhad serta kemalangan jalan raya yang tidak terlapor merupakan perkara biasa walaupun kecederaan akibat kemalangan jalan raya adalah punca utama kematian di sana (Samuel et al. 2012). Antara faktor yang menyumbang kepada berlakunya kemalangan tidak terlapor adalah pihak yang terlibat tidak mahu memperbesar isu tersebut kerana kecederaan yang dialami adalah ringan serta penyelesaian secara peribadi antara kedua-dua pihak telah dilakukan (Radzman Shah & Zamzuri 2023). Selain itu, pengurusan sistem pelaporan yang kurang cekap dan memakan masa turut menyumbang kepada kecenderungan pemandu untuk tidak melaporkan kemalangan (Zamzuri 2021).

Namun, Maxwell et al. (2018) menegaskan bahawa adalah penting untuk merekod semua kes kemalangan jalan raya ke dalam sistem pangkalan data rasmi secara lengkap dan setepat mungkin kerana ini boleh mempengaruhi penyampaian maklumat yang tepat serta perancangan strategi yang berkaitan. Hasil daripada tinjauan soal selidik yang dijalankan oleh Nik Zamri dan Zamzuri (2019) berkenaan penganggaran kadar kemalangan jalan raya tidak terlapor di Malaysia, didapati bahawa sebanyak 46% daripada kemalangan yang berlaku tidak dilaporkan. Selain itu, Kamaluddin, Abd Rahman dan Várhelyi (2018) mendapati hasil padanan antara data kemalangan berdasarkan rekod polis dan hospital di daerah Melaka Tengah hanyalah sekadar 4.1%. Ini menegaskan bahawa terdapat kekurangan serius dalam perekodan kemalangan jalan raya, yang seterusnya memberi impak kepada ketepatan analisis berkaitan. Kenyataan ini diperkukuhkan lagi oleh kajian Shinar et al. (2018), Singh et al. (2018) dan Ytterstad, Gressnes dan Harborg (2018) yang mengesahkan bahawa kebanyakan kes kemalangan jalan raya yang tidak dilaporkan adalah kemalangan tanpa kecederaan atau hanya melibatkan kecederaan ringan.

Jika diteliti, bilangan kajian lepas berkaitan isu KJRTT di Malaysia adalah sangat terhad dengan hanya lima penerbitan berkaitan isu ini setakat tahun 2023 (Radzman Shah & Zamzuri 2023). Nilai ini jauh lebih kecil jika dibandingkan dengan penerbitan dalam isu yang sama di peringkat global atau di negara maju seperti Australia (Boufous et al. 2008; Watson, Vallmuur & Watson 2015). Maka, penyelidikan ini berhasrat untuk membantu dalam mengisi kekurangan kajian dengan meneroka analisis pautan data dan mencadangkan al-Khawarazmi heuristik dalam memadankan rekod kemalangan hospital dan polis serta menganggarkan bilangan kemalangan yang tidak dilaporkan.

BAHAN DAN KAEDAH

SUMBER DAN PEMPROSESAN DATA

Kajian ini menggunakan dua set data, iaitu rekod hospital dan rekod polis. Definisi KJRTT dalam kajian ini merujuk kepada kemalangan jalan raya yang tidak dilaporkan kepada pihak polis, memandangkan rekod polis adalah sumber utama yang digunakan dalam analisis berkaitan kemalangan jalan raya. Secara amnya, ketidaksamaan yang wujud antara rekod hospital dan polis digunakan sebagai asas dalam menganggarkan KJRTT. Jadual 1 menunjukkan perbandingan perincian bagi kedua-dua set data ini.

Memandangkan terdapat tiga blok data dalam rekod polis, ketiga-tiga blok ini digabungkan menggunakan ID Laporan untuk menghasilkan satu set data kemalangan jalan raya yang dilaporkan kepada polis. Seterusnya, kedua-dua set data ditapis dengan mengekalkan pemboleh ubah yang dikenal pasti sebagai bertindih, iaitu ada dalam kedua-dua set data hospital dan polis. Pemboleh ubah tersebut adalah umur, jantina, bangsa, tarikh dan jenis kenderaan. Bagi data hospital, ia ditapis dengan mengekalkan data dari Januari sehingga Mac 2011 kerana tempoh ini bertindih dengan rekod hospital. Secara dasarnya, jika semua kemalangan dilaporkan kepada pihak polis, maka setiap entri data hospital akan mempunyai padanannya apabila kedua-dua rekod data ini dipautkan.

Seterusnya, ketiga-tiga blok data polis dipautkan melalui ID Laporan untuk membentuk satu hamparan data yang lengkap. Dengan itu, kita memperoleh dua set hamparan data iaitu hamparan data polis dan hamparan data hospital. Dalam kedua-dua hamparan data ini terdapatnya duplikasi yang merujuk kepada dua situasi berikut: 1). Duplikasi ID Pesakit dalam data hospital - merujuk kepada keadaan di mana pesakit menerima rawatan di hospital tersebut untuk kemalangan yang sama lebih dari sekali dan 2). Duplikasi ID laporan dalam data polis - merujuk kepada bilangan individu yang terlibat dalam kemalangan yang sama.

PAUTAN DATA BERKEBARANGKALIAN/PADANAN KABUR

Pautan data merupakan teknik yang menggabungkan data daripada dua atau lebih sumber bagi entiti yang sama (Ali Omar et al. 2023). Di era Revolusi Industri 4.0 dan kepopularan data raya, analisis data tidak lagi tertumpu kepada satu-satu sumber atau set data sahaja. Dari aspek praktikaliti, sungguhpun terdapat banyak data daripada pelbagai sumber, masih ada maklumat dalam set data ini yang perlu dirahsiakan, seperti nombor kad pengenalan. Ini menimbulkan keperluan untuk teknik pautan data yang memadankan entri daripada set data berbeza berdasarkan entiti atau baris data yang sama atau hampir sama, memandangkan pengecam unik seperti nombor kad pengenalan tidak dapat digunakan.

Terdapat dua jenis utama bagi teknik pautan data iaitu secara deterministik atau berkebarangkalian (Khodabakhshian, Puolitaival & Kestle 2023). Pautan data secara deterministik menganggap dua entri data adalah sama sekiranya kedua-dua entri tersebut betul-betul sama. Kekangan bagi pautan data secara deterministik adalah apabila wujudnya ketidaksamaan antara dua entri data, seperti kesalahan ejaan, maka dua entri data tersebut tidak dapat dipautkan walaupun merujuk kepada entiti yang sama. Dalam hal ini, pendekatan berkebarangkalian adalah lebih baik dan tepat kerana ia membenarkan dua entri data yang hampir sama untuk dihubungkan. Menerusi pendekatan ini, dua entri data akan dihubungkan menggunakan beberapa pengecam bukan unik. Pengecam bukan unik merujuk kepada pemboleh ubah yang mana banyak entri data adalah sama, seperti bangsa, umur dan jantina manakala pengecam unik adalah entri atau siri nombor yang sememangnya unik kepada seseorang individu sahaja, seperti nombor kad pengenalan. Menerusi pendekatan ini, setiap pasangan data daripada dua sumber akan dihitung kebolehjadiannya bahawa kedua-dua entri data adalah sama berdasarkan pengecam atau pemboleh ubah yang dipertimbangkan. Pendekatan ini turut dikenali sebagai padanan kabur (Mack, 2014). Kajian ini menggunakan al-Khawarizmi Jaccard dalam memautkan data secara berkebarangkalian. Secara asasnya, al-Khawarizmi ini menghitung kesamaan antara dua entri data melalui nisbah bilangan karakter yang sama dengan jumlah karakter unik. Perincian bagi al-Khawarizmi ini boleh didapati daripada Mosleh et al. (2024).

JADUAL 1. Perincian maklumat bagi set data hospital dan polis

Sumber	Hospital	Polis
Tahun	2010 sehingga Mac 2011	Sepanjang 2011
Lokasi	5 hospital di Selangor	Seluruh Malaysia
Pemboleh ubah	ID Pesakit, Tarikh rawatan/ kemasukan ke hospital, jantina, umur, bangsa, masa kecederaan, tarikh kecederaan serta maklumat klinikal berkaitan rawatan yang diterima seperti oksigen, titisan IV, peratusan alkohol dan dadah dalam darah	Terdapat 3 blok data yang dihubungkan dengan rekod ID: Blok 1 – perincian kemalangan seperti lokasi, tarikh, keadaan jalan, cuaca, pengcahayaan Blok 2 – orang yang terlibat dalam kemalangan seperti umur, bangsa, jantina, jenis kenderaan Blok 3 – kecederaan bagi orang yang terlibat dalam kemalangan seperti jenis kecederaan, bahagian badan tercedera

AL-KHAWARIZMI HEURISTIK

Dalam penerangan definisi perkataan 'heuristik', acapkali ia dihubungkan dengan perkataan 'mencari' dan 'pengalaman' (Dale 2015; Shin, Rasul & Fotiadis 2021). Secara amnya, al-Khawarizmi heuristik merujuk kepada tatacara atau satu set peraturan yang dibentuk menerusi pengalaman apabila berhadapan dengan isu atau permasalahan yang ingin diselesaikan (David, Vangheluwe & Syriani 2023). Dalam kajian ini, unsur heuristik disuntik kepada teknik pautan data pada dua tahap. Pertama, dalam mempertimbangkan set pemboleh ubah yang boleh menjadi pengecam bukan unik yang sesuai. Kedua, apabila baris data diubah dari individu yang terlibat dalam kemalangan kepada bilangan kemalangan. Ini kerana matlamat akhir kajian ini adalah untuk menganggarkan bilangan kemalangan yang tidak dilaporkan, namun anggaran yang diperoleh selepas proses pautan data dilakukan adalah bilangan individu yang terlibat dalam kemalangan yang tidak dilaporkan. Maka, bagi memenuhi matlamat kajian ini, kedua-dua proses ini ditambah kepada al-Khawarizmi heuristik pautan data berkebarangkalian yang dijalankan. Pelaksanaan bagi al-Khawarizmi ini dijalankan menggunakan *Power Query*, iaitu satu perkakasan ETL (*Extact, Transform, Load*) yang dibangunkan Microsoft, menggunakan Bahasa M (Microsoft Learn 2024).

KEPUTUSAN DAN PERBINCANGAN

PERTIMBANGAN PEMBOLEH UBAH BAGI DATA YANG DI-PAUTKAN

Dalam kajian ini, beberapa set pemboleh ubah dipertimbangkan dalam memautkan rekod data hospital kepada data kemalangan yang dilaporkan kepada pihak polis. Pada peringkat pertama, data daripada kedua-dua set ini pada awalnya dipadankan menerusi pemboleh ubah 'Tarikh', yang menghasilkan padanan yang tinggi sebanyak 7654. Perlu diingat bahawa nilai yang tinggi ini diperoleh kerana hanya pemboleh ubah 'Tarikh' yang dipertimbangkan, yang memungkinkan begitu banyak pasangan data berpadanan walaupun hakikatnya ia tidak merujuk kepada kemalangan yang sama.

Oleh itu, pada peringkat kedua, lebih banyak pemboleh ubah bukan unik turut dipertimbangkan, iaitu umur, jantina dan bangsa. Didapati bilangan data yang berjaya dipautkan menurun kepada 3037, yang merupakan nilai yang lebih munasabah. Pada peringkat ketiga, data yang berpadanan daripada peringkat kedua digunakan dan proses pautan data dilaksanakan sekali lagi dengan mempertimbangkan pemboleh ubah yang sama bersama pemboleh ubah baharu, iaitu jenis kenderaan. Tetapan ini dinamakan sebagai peringkat 3(a). Satu lagi tetapan dipertimbangkan dengan menggunakan kesemua data dan data tersebut dipautkan berdasarkan set pemboleh ubah dalam peringkat 3(a). Bagi kedua-dua tetapan ini, iaitu 3(a) dan 3(b), didapati bilangan data yang berjaya dipadankan masing-masing adalah 1735 dan 2863.

Seterusnya, satu lagi tetapan dipertimbangkan, iaitu lokasi kemalangan menerusi pemboleh ubah 'Negeri'. Berdasarkan rasional bahawa terdapat kemungkinan individu yang terlibat dalam kemalangan di luar Selangor mendapatkan rawatan di Selangor, peringkat 4(a) dipertimbangkan. Data yang digunakan dalam peringkat ini adalah data yang telah berpadanan dalam peringkat (2). Ini bermaksud menggunakan data berpadanan daripada peringkat (2), data ini kemudian ditapis bagi negeri Selangor, Kuala Lumpur, Negeri Sembilan dan Perak, dan proses pautan data dijalankan sekali lagi. Bilangan data yang berjaya dipadankan bagi tetapan ini adalah yang terkecil, iaitu hanya 496. Manakala tetapan 4(b) pula menggunakan data yang berpadanan daripada peringkat (2) dan mempertimbangkan data kemalangan di seluruh Malaysia. Dengan perubahan tetapan ini, bilangan entri data yang berjaya dipautkan meningkat kepada 1255. Bagi proses penganggaran bilangan kemalangan yang tidak dilaporkan, maklumat daripada tetapan 4(b) akan digunakan. Ringkasan pemilihan pemboleh ubah ini dan proses pautan data berperingkat diberikan dalam Jadual 2.

PENGANGGARAN KADAR KEMALANGAN JALAN RAYA TIDAK TERLAPOR

Memandangkan data yang telah dihubungkan adalah bilangan orang yang terlibat dalam kemalangan jalan raya, sedangkan kajian ini cuba untuk menganggarkan bilangan kemalangan, maka proses penghapusan duplikasi perlu dilakukan bagi memastikan entri data tersebut adalah per kemalangan, bukannya per individu. Seterusnya, proses pengiraan bagi mendapatkan anggaran kadar kemalangan tidak terlapor diperincikan.

Biar A ialah data yang telah dipadankan dan dihapuskan duplikasi 'ID Pesakit'; B ialah data A yang dihapuskan duplikasi 'ID Laporan'; ialah purata bilangan individu yang terlibat dalam satu kemalangan; C ialah anggaran jumlah kemalangan berdasarkan rekod hospital (terlapor dan tidak dilaporkan); D ialah bilangan pesakit yang mendapatkan rawatan akibat kemalangan (data hospital); R ialah kadar kemalangan yang dilaporkan; H ialah bilangan kemalangan yang dilaporkan kepada polis (data polis); dan X ialah anggaran bilangan kemalangan yang tidak dilaporkan.

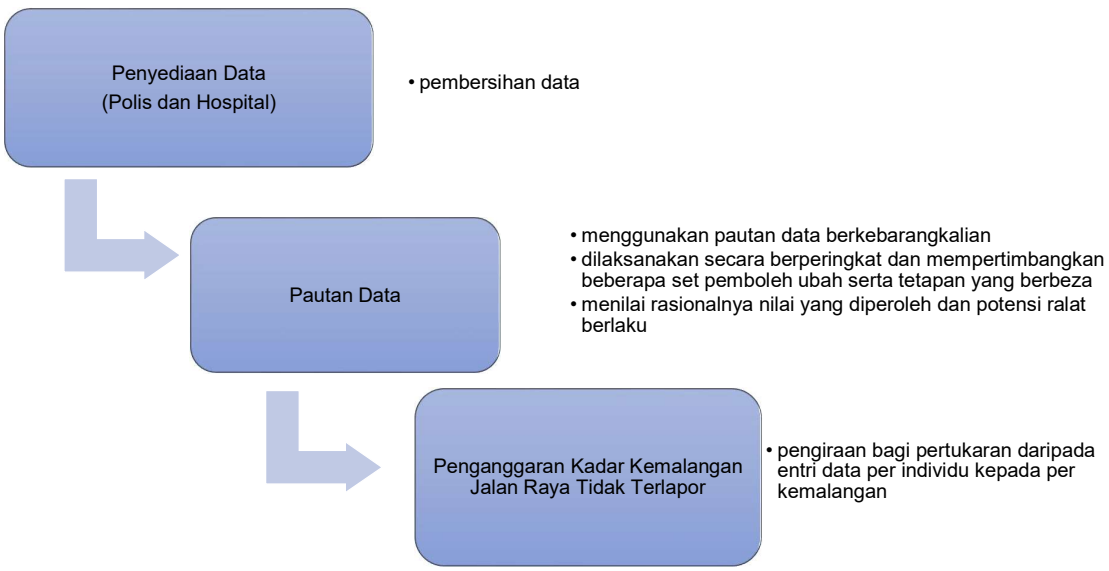
Setelah proses penghapusan duplikasi dilakukan, nilai bagi A dan B masing-masing adalah 2873 dan 1255. Ini menghasilkan , yang bermaksud bagi satu-satu kemalangan terlapor yang berlaku, sekitar dua atau tiga orang terlibat

JADUAL 2. Peringkat pautan data yang dilaksanakan

Peringkat	Data	Pemboleh ubah	Bilangan entri yang berjaya dipautkan
1	Semua	Tarikh	7654
2	Semua	Tarikh, Umur, Jantina, Bangsa	3037
3a	Data berpadanan daripada (2)	Tarikh, Umur, Jantina, Bangsa, Jenis Kenderaan	1735
3b	Semua	Tarikh, Umur, Jantina, Bangsa, Jenis Kenderaan	2863
4a	Data berpadanan daripada (2) dan ditapis bagi negeri Selangor, KL, Negeri Sembilan dan Perak	Tarikh, Umur, Jantina, Bangsa, Jenis Kenderaan	496
4b	Data berpadanan daripada (2) dan mempertimbangkan satu Malaysia	Tarikh, Umur, Jantina, Bangsa, Jenis Kenderaan	1255

JADUAL 3. Perincian nilai yang digunakan dalam penganggaran bilangan kemalangan tak terlapor

Penerangan	Nilai
bilangan kemalangan yang dilaporkan kepada polis (data polis), H	13527
bilangan pesakit yang mendapatkan rawatan akibat kemalangan (data hospital), D	9058
data yang telah dipadankan dan dihapuskan duplikasi 'ID Pesakit' , A	2873
data A yang dihapuskan duplikasi 'ID Laporan', B	1255
purata bilangan individu yang terlibat dalam satu kemalangan,	2.29
anggaran jumlah kemalangan berdasarkan rekod hospital (terlapor dan tidak dilaporkan), C	3956
kadar kemalangan yang dilaporkan , R	0.32
anggaran bilangan kemalangan yang tidak dilaporkan, X	6366



RAJAH 1. Al-Khawarizmi heuristik bagi penganggaran kadar kemalangan jalan raya tidak terlapor

dalam kemalangan tersebut. Kemudian, penganggaran jumlah kemalangan sebenar dapat dihitung dengan menggunakan data hospital sebagai sampel, iaitu *C* yang bernilai 3955.5. Seterusnya, kadar kemalangan terlapor, *R* dianggarkan pada 0.3173 ataupun 32%. Ini bermaksud kadar kemalangan tidak terlapor di Malaysia dianggarkan sekitar 68%, tekal dengan penemuan dalam kajian terdahulu (Nik Zamri, Zamzuri & Ibrahim 2018; Nik Zamri & Zamzuri 2019).

Formula bagi pengiraan yang terlibat disenaraikan seperti dalam persamaan (1) – (4) berikut.

$$\mu = \frac{A}{B} \tag{1}$$

$$C = \frac{D}{\mu} \tag{2}$$

$$R = \frac{B}{C} \tag{3}$$

$$X = \frac{H(1-R)}{R} \tag{4}$$

Bagi mendapatkan anggaran bilangan kemalangan tidak terlapor, nilai *H* iaitu sebanyak 13527 diperlukan dan menerusi persamaan (4), bilangan kemalangan tidak terlapor di Malaysia pada tahun 2011 dianggarkan sekitar 6366. Ringkasan bagi nilai penting yang terlibat dalam penganggaran ini ditunjukkan dalam Jadual 3.

Kesemua proses atau tatakerja yang terlibat dalam mendapatkan anggaran ini membawa kepada pembinaan al-Khawarizmi secara heuristik, yang diringkaskan dalam RAJAH 1. Al-Khawarizmi ini sesuai digunakan bagi pautan data kemalangan jalan raya berdasarkan rekod polis dan hospital di Malaysia.

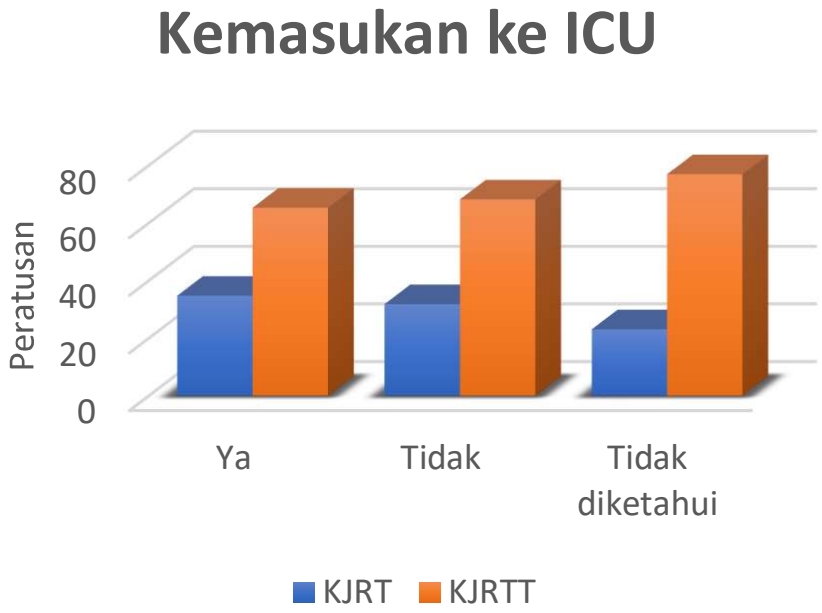
PERBANDINGAN KEMALANGAN JALAN RAYA TERLAPOR (KJRT) DAN KEMALANGAN JALAN RAYA TIDAK TERLAPOR (KJRTT) BAGI PEMBOLEH UBAH TERPILIH

Dalam bahagian ini, data hospital yang dapat dipadankan dikategorikan sebagai KJRT, manakala selebihnya dikategorikan sebagai KJRTT. Seterusnya, perbandingan dibuat untuk melihat peratusan daripada individu dalam rekod data hospital bagi kedua-dua kategori ini. Perbandingan yang dibuat adalah berdasarkan pemboleh ubah berikut iaitu kemasukan ke ICU, penggunaan alkohol dan penggunaan dadah.

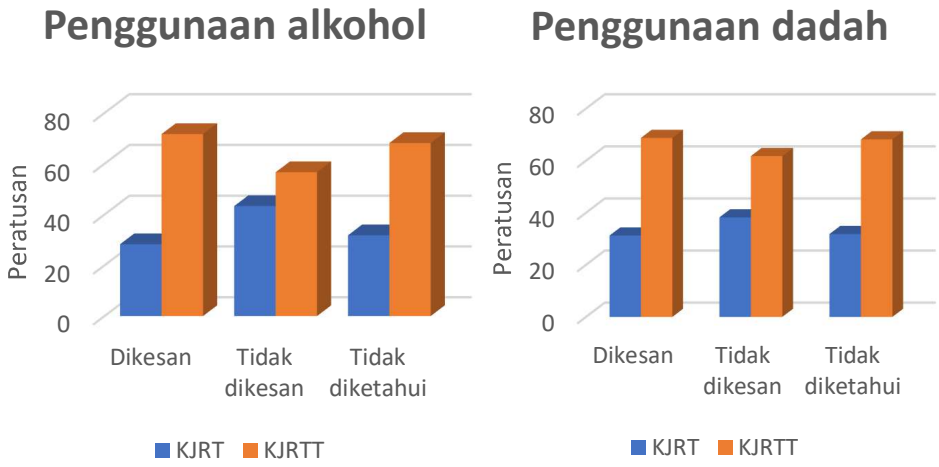
Jadual 4 menunjukkan peratusan bagi setiap kategori dalam pemboleh ubah terpilih yang dinyatakan. Didapati hanya 1.6% sahaja yang dimasukkan ke ICU dan 0.2% dengan status tidak diketahui. Bagi penggunaan dadah dan alkohol pula, kurang daripada 1% diketahui status ujian pengesanan dengan peratusan masing-masing bagi dadah (0.91%) dan alkohol (0.92%). Peratusan yang rendah ini berkemungkinan disebabkan bukan semua individu yang terlibat dalam kemalangan akan dijalankan ujian pengesanan dadah atau alkohol. Ujian ini dijalankan hanya bagi kes yang disyaki ada melibatkan penggunaan dua bahan tersebut.

JADUAL 4. Peratusan kategori bagi pemboleh ubah terpilih

Pemboleh ubah	Kategori	Peratusan (%)
Kemasukan ke ICU	Ya	1.60
	Tidak	98.11
	Tidak diketahui	0.29
Penggunaan dadah	Dikesan	0.53
	Tidak dikesan	0.38
	Tidak diketahui	99.09
Penggunaan alkohol	Dikesan	0.59
	Tidak dikesan	0.33
	Tidak diketahui	99.08



RAJAH 2. Perbandingan antara KJRT dan KJRTT bagi status kemasukan ke ICU



RAJAH 3. Perbandingan antara KJRT dan KJRTT bagi penggunaan alkohol dan dadah

Seterusnya, perbandingan antara peratusan KJRT dan KJRTT dibuat bagi setiap kategori yang dinyatakan dalam Jadual 4. Rajah 2 menunjukkan perbandingan antara KJRT dan KJRTT bagi setiap kategori dalam pemboleh ubah kemasukan ke ICU, yang mana tidak menunjukkan sebarang perbezaan ketara. Ketiga-tiga kategori secara tekal menunjukkan corak perbezaan yang sama, iaitu peratusan KJRTT adalah lebih tinggi daripada KJRT, sekitar 60% melawan 40%. Adalah tidak munasabah sekiranya kes kemalangan dengan kemasukan ke ICU mencatatkan peratusan tidak dilaporkan yang tinggi, namun ia boleh dijelaskan dengan situasi disebabkan kemalangan yang terlalu parah, mangsa terus dibawa ke hospital sebelum laporan polis bagi kemalangan tersebut dibuat.

Rajah 3 pula menunjukkan perbandingan yang sama bagi kategori dalam pemboleh ubah penggunaan alkohol dan dadah. Dapat dilihat bahawa bagi kes yang mana penggunaan alkohol dan dadah dikesan, peratusan KJRT adalah lebih rendah. Ini berkemungkinan disebabkan individu tersebut tahu akan akibat daripada kesalahan seterusnya tidak melaporkan kemalangan tersebut kepada polis.

KESIMPULAN

Kualiti data yang baik adalah amat penting sebelum analisis dapat dijalankan. Hal ini bagi memastikan ketepatan analisis dan tafsiran maklumat yang benar-benar mewakili situasi sebenar. Berdasarkan situasi kemalangan jalan raya yang kerap tidak dilaporkan, kajian ini telah membangunkan satu al-Khawarizmi heuristik yang berasaskan pautan data berkebarangkalian. Terdapat keperluan untuk menjalankan proses pautan data secara berperingkat bagi memastikan data yang dipadankan adalah hampir sama berdasarkan pengecam bukan unik yang dikenal pasti. Kemudian, anggaran bilangan kemalangan tidak terlapor dihitung dengan melibatkan proses penukaran unit data daripada per individu kepada per kemalangan. Sungguhpun set data yang digunakan adalah dicerap pada 2011, namun ia dapat menunjukkan aplikasi al-Khawarizmi heuristik ini bagi situasi yang melibatkan kemalangan jalan raya. Kajian ini hanya mampu menggunakan data pada tahun 2011 tersebut memandangkan kekangan perolehan data daripada sumber yang berbeza.

Menerusi al-Khawarizmi heuristik yang dibangunkan, dua set data kemalangan jalan raya dapat dihubungkan. Kadar ketidaksamaan antara dua set data ini dianggarkan sebagai kadar kemalangan tidak terlapor. Menerusi data yang dipautkan, ciri bagi kemalangan jalan raya tidak terlapor dapat dikenal pasti. Hal ini seterusnya memberi input kepada perancangan strategi kemalangan jalan raya pada masa akan datang, berdasarkan analisis data yang lebih jitu menggunakan data yang lebih lengkap dan berkualiti.

PENGHARGAAN

Penulis merakamkan setinggi-tinggi penghargaan kepada Kementerian Pengajian Tinggi Malaysia dan Universiti Kebangsaan Malaysia kerana menaja penyelidikan ini menerusi Skim Geran Penyelidikan Fundamental (FRGS) dengan kod FRGS/1/2021/STG06/UKM/02/1 juga kepada Institut Penyelidikan Keselamatan Jalan Raya Malaysia (MIROS) kerana menyediakan data yang digunakan dalam kajian ini. Penulis juga berterima kasih kepada penilai manuskrip ini dengan komen penambahbaikan yang konstruktif.

**Al-Khawarizmi heuristik yang digunakan untuk menjalankan analisis dan hasil yang dibentangkan dalam jadual boleh diperoleh daripada pengarang utama atas permintaan yang munasabah.

RUJUKAN

Ahmed, S.K., Mohammaed, M.G., Abdulqadir, S.O., Abd El-Kader, R.G., El-Shall, N.A., Chandran, D., Ur Rehman, M.E. & Dhama, K. 2023. Road traffic accidental injuries and deaths: A neglected global health issue. *Health Science Report* 6(5): e1240. doi: 10.1002/hsr2.1240

Ali Omar, Z., Zamzuri, Z.H., Mohd Ariff, N. & Abu Bakar, M.A. 2023. Training data selection for record linkage classification. *Symmetry* 15(5): 1060.

Boufous, S., Finch, C., Hayen, A. & Williamson, A. 2008. *Data Linkage of Hospital and Police Crash Datasets in NSW*. Technical Report. Sydney: NSW Injury Risk Management Research Centre, University of New South Wales.

Dale, S. 2015. Heuristics and biases: The science of decision making. *Business Information Review* 32(2): 93-99.

David, I., Vangheluwe, H. & Syriani, E. 2023. Model consistency as a heuristic for eventual correctness. *Journal of Computer Languages* 76: 101223.

Isa, Z. & Zamzuri, Z.H. 2022. Pengukuran risiko menggunakan Rangkaian Bayesan: Aplikasi kepada data perlanggaran kapal di Malaysia. *Sains Malaysiana* 51(7): 2305-2314

Kamaluddin, N.A., Abd Rahman, M.F. & Várhelyi, A. 2018. Matching of police and hospital road crash casualty records - a data-linkage study in Malaysia. *International Journal of Injury Control and Safety Promotion* 26(1): 52-59. doi:10.1080/17457300.2018.1476385

Kementerian Pengangkutan Malaysia. 2019. Statistik Pengangkutan Malaysia. https://www.mot.gov.my/en/Statistik%20Tahunan%20Pengangkutan/Transport%20Statistics%20Malaysia%202019.pdf (Diakses pada 1 Ogos 2024).

Khodabakhshian, A., Puolitaival, T. & Kestle, L. 2023. Deterministic and probabilistic risk management approaches in construction projects: A systematic literature review and comparative analysis. *Buildings* 13(5): 1312.

Mack, C. 2014. PS1-13: Probabilistic linkage (also known as "fuzzy matching"): The theoretical foundations of modern record linkage. *Clinical Medicine and Research* 12(1-2): 95.

Maxwell, O., Mayowa, B.A., Chinedu, I.U. & Peace, A.E. 2018. Modelling count data; A generalized linear model framework. *American Journal of Mathematics and Statistics* 8(6): 179-183.

Microsoft Learn. 2024. Power Query M Formula Language. https://learn.microsoft.com/en-us/powerquery-m/ (Diakses pada 1 Ogos 2024).

Mosleh, M.A.A., Assiri, A., Gumaei, A.H., Alkhamees, B.F. & Al-Qahtani, M. 2024. A bidirectional Arabic sign language framework using deep learning and fuzzy matching score. *Mathematics* 12(8): 1155.

Muni, K.M., Ningwa, A., Osuret, J., Zziwa, E.B., Namatovu, S., Biribawa, C., Nakafeero, M., Mutto, M., Guwatudde, D., Kyamanywa, P. & Kobusingye, O. 2021. Estimating the burden of road traffic crashes in Uganda using police and health sector data sources. *Injury Prevention* 27: 208-214.

Nik Zamri, N.S. & Zamzuri, Z.H. 2019. Estimating the proportion of non-fatality unreported traffic accidents in Malaysia. *ASM Sc. J.* 12(1): 239-245.

Nik Zamri, N.S., Zamzuri, Z.H. & Ibrahim, K. 2018. Factors influencing Malaysian drivers' tendency on underreporting. *International Journal of Engineering and Technology* 7(4): 6313-6321.

Radzman Shah, N.W. & Zamzuri, Z.H. 2023. Underreporting of road traffic accidents: A bibliometric analysis from Web of Science database. *Journal of Quality Measurement and Analysis* 19(3): 55-71.

Samuel, J.C., Sankhulani, E., Qureshi, J.S., Baloyi, P., Thupi, C., Lee, C.N., Miller, W.C., Cairns, B.A. & Charles, A.G. 2012. Under-reporting of road traffic mortality in developing countries: Application of a capture-recapture statistical model to refine mortality estimates. *PloS ONE* 7(2): e31091.

Shin, D., Rasul, A. & Fotiadis, A. 2021. Why am I seeing this? Deconstructing algorithm literacy through the lens of users. *Internet Research* 32: 1214-1234.

Shinar, D., Valero-Mora, P., van Strijp-Houtenbos, M., Haworth, N., Schramm, A., Bruyne, G.D., Cavallo, V., Chliaoutakis, J., Dias, J., Frraro, O.E., Fyhri, A., Sajatovic, A.H., Kuklane, K., Ledesma, R., Mascarell, O., Morandi, A., Muser, M., Otte, D., Papadakaki, M., Sanmartín, J., Dulf, D., Saplioglu, M. & Tzamalouka, G. 2018. Under-reporting bicycle accidents to police in the COST TU1101 international survey: Cross-country comparison and associated factors. *Accident, Analysis and Prevention* 110: 177-186.

Singh, P., Laksmi, P.V.M., Prinja, S. & Khanduja, P. 2018. Under-reporting of road traffic accidents in traffic police records - A cross-sectional study from North India. *International Journal of Community Medicine and Public Health* 5(2): 579-584.

Ward, H., Lyons, R. & Thoreau, R. 2006. *Under-reporting of Road Casualties? Phase 1*. Road Safety Research Report No. 69. London: Department for Transport.

Watson, A., Vallmuur, K. & Watson, B. 2015. How serious are they? The use of data linkage to explore different definitions of serious road crash injuries. *Proceedings of the 2015 Australasian Road Safety Conference in Gold Coast*, Australia. hlm. 1-10.

World Health Organization (WHO). 2023. Road traffic injuries. https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries (Diakses pada 1 Ogos 2024).

Ytterstad, B., Gressnes, T. & Harborg, T. 2018. PW 1663 Injury surveillance in a hospital leads to complete traffic injury data, sustainable injury prevention, and update police underreporting. *Injury Prevention* 24(2): A179.

Zamzuri, Z.H. 2021. Underreporting traffic accidents in Malaysia-A sentiment analysis. *ITM Web of Conferences* 36: 01015.

*Pengarang untuk surat-menyurat; email: zamira@ukm.edu.my