

**MAKLUMAT PALSU COVID-19: ANALISIS CORAK  
TOPIK TEKS LAMAN WEB *SEBENARNYA.MY*  
MENGUNAKAN WEKA**

**SITI FATIMAH BINTI HUZAIDI**

**UNIVERSITI SAINS MALAYSIA**

**2023**

**MAKLUMAT PALSU COVID-19: ANALISIS CORAK  
TOPIK TEKS LAMAN WEB *SEBENARNYA.MY*  
MENGGUNAKAN WEKA**

oleh

**SITI FATIMAH BINTI HUZAIDI**

**Tesis yang diserahkan untuk  
memenuhi keperluan bagi  
Ijazah Sarjana Seni**

**April 2023**

## PENGHARGAAN

Syukur ke hadrat Ilahi kerana dengan limpah dan kurniaNya, kajian ini berjaya disiapkan sepenuhnya. Setinggi-tinggi penghargaan dan ucapan terima kasih ditujukan khas kepada Dr Nik Norma Nik Hasan, selaku penyelia yang banyak membantu dan memberikan tunjuk ajar, perhatian semangat serta nasihat sewaktu kajian ini dilaksanakan. Begitu juga kepada semua kakitangan Pusat Pengajian Komunikasi (terutamanya Noor Azlina, Mohd Asyraf Hazmi, Muhammad Hafizi dan Mohamed Haris) serta *Institute of Postgraduate Studies* (IPS) yang memberi banyak bantuan secara teknikal yang berharga. Jutaan kalungan kasih kepada keluarga tercinta (Huzaidi Husain, Siti Dora Daud dan Muhammad Danial Huzaidi) yang tidak jemu memberikan nasihat, semangat dan sokongan fizikal serta mental sepanjang tempoh kajian disediakan. Saya juga ingin merakamkan jutaan terima kasih buat suami, Aiman Fayyadh Mohamad Ziyad yang sentiasa menemani dan memberi pandangan sewaktu diperlukan. Terima kasih sahabat-sahabat seperjuangan yang sama-sama mengikuti program pascasiswazah kerana sering berkongsi maklumat semasa terutamanya berkenaan program pascasiswazah agar saya tidak ketinggalan. Rakan-rakan yang lain, terima kasih atas keprihatinan yang diberikan. Akhir sekali, terima kasih diucapkan kepada Majlis Amanah Rakyat (MARA) yang meluluskan permohonan pinjaman pelajaran sekaligus meringankan beban saya dari segi kewangan sewaktu pengajian.

## ISI KANDUNGAN

|   |             |
|---|-------------|
| <b>PENGHARGAAN.....</b>   | <b>ii</b>   |
| <b>ISI KANDUNGAN.....</b>   | <b>iii</b>  |
| <b>SENARAI JADUAL.....</b>  | <b>vi</b>   |
| <b>SENARAI RAJAH.....</b>   | <b>vii</b>  |
| <b>SENARAI SINGKATAN.....</b>                                       | <b>viii</b> |
| <b>ABSTRAK.....</b>   | <b>x</b>    |
| <b>ABSTRACT.....</b>  | <b>xi</b>   |
| <b>BAB 1 PENGENALAN.....</b>  | <b>1</b>    |
| 1.1 Pengenalan.....   | 1           |
| 1.2 Pernyataan Masalah.....   | 5           |
| 1.3 Persoalan Kajian.....   | 7           |
| 1.4 Objektif Kajian.....  | 7           |
| 1.5 Kepentingan Kajian.....   | 7           |
| 1.6 Skop Kajian.....  | 9           |
| 1.7 Konseptualisasi.....  | 9           |
| 1.7.1 Maklumat Palsu.....   | 9           |
| 1.7.2 COVID-19.....   | 10          |
| 1.7.3 Laman Web <i>sebenarnya.my</i> .....                          | 10          |
| 1.7.4 Perlombongan Teks.....  | 11          |
| 1.7.5 <i>Supervised Learning Technique</i> (Pengklasifikasian)..... | 11          |
| 1.8 Operasionalisasi.....   | 11          |
| 1.8.1 Maklumat Palsu.....   | 11          |
| 1.8.2 Laman Web <i>sebenarnya.my</i> .....                          | 12          |
| 1.8.3 Perlombongan Teks.....  | 12          |
| 1.8.4 <i>Supervised Learning Technique</i> (Pengklasifikasian)..... | 13          |
| <b>BAB 2 SOROTAN LITERATUR.....</b>                                 | <b>14</b>   |
| 2.1 Pengenalan.....   | 14          |
| 2.2 Pandemik COVID-19.....  | 14          |

|                                  |  |           |
|----------------------------------|--|-----------|
| 2.3                              | Infodemik Ketika Pandemi COVID-19.....   | 16        |
| 2.4                              | Maklumat Palsu COVID-19.....   | 18        |
| 2.5                              | Laman Web <i>sebenarnya.my</i> .....   | 20        |
| 2.6                              | Perlombongan Teks.....   | 23        |
| 2.7                              | Kerangka Teoretikal.....   | 24        |
| 2.7.1                            | Model Six-Step KDP.....  | 24        |
| <b>BAB 3 METODOLOGI.....</b>     |  | <b>29</b> |
| 3.1                              | Pengenalan.....  | 29        |
| 3.2                              | Desain Kajian.....   | 29        |
| 3.3                              | Pengumpulan Data.....  | 30        |
| 3.4                              | Unit Analisis.....   | 31        |
| 3.5                              | Analisis Data.....   | 31        |
| 3.5.1                            | Prapemprosesan Data ( <i>Data Pre-Processing</i> ).....  | 32        |
| 3.5.1(a)                         | Tokenisasi ( <i>Tokenization</i> ).....  | 33        |
| 3.5.1(b)                         | <i>Stop-Words Removal</i> .....  | 34        |
| 3.5.1(c)                         | <i>Stemming</i> .....  | 35        |
| 3.5.2                            | <i>Supervised Learning Technique</i> (Pengklasifikasian).....  | 36        |
| 3.5.2(a)                         | WEKA.....  | 36        |
| 3.5.2(b)                         | Hutan Rawak.....   | 37        |
| 3.6                              | Kesahan dan Kebolehpercayaan.....  | 37        |
| <b>BAB 4 DAPATAN KAJIAN.....</b> |  | <b>39</b> |
| 4.1                              | Pengenalan.....  | 39        |
| 4.2                              | Pra uji ciri-ciri signifikan topik COVID-19 dalam <i>sebenarnya.my</i> .....                                 | 39        |
| 4.3                              | Ciri-ciri signifikan topik COVID-19 palsu, waspada dan penjelasan dalam laman web <i>sebenarnya.my</i> ..... | 43        |
| 4.3.1                            | Waspada.....   | 43        |
| 4.3.2                            | Penjelasan.....  | 46        |
| 4.3.3                            | Palsu.....   | 49        |
| 4.4                              | Pengujian ketepatan model yang diperoleh melalui perisian WEKA.....  | 52        |

|   |  |           |
|---|--|-----------|
| 4.5   | Kesimpulan.....  | 51        |
| <b>BAB 5 PERBINCANGAN, KESIMPULAN DAN CADANGAN.....</b> |  | <b>54</b> |
| 5.1   | Pengenalan.....  | 54        |
| 5.2   | Perbincangan.....  | 54        |
|   | 5.2.1 Kepentingan masyarakat melengkapkan diri dengan kemahiran literasi ciri-ciri signifikan maklumat palsu COVID-19 di Malaysia..... | 54        |
| 5.3   | Implikasi Kajian.....  | 56        |
|   | 5.3.1 Implikasi Teoretikal.....  | 56        |
|   | 5.3.2 Implikasi Metod.....   | 57        |
|   | 5.3.3 Implikasi Praktikal.....   | 57        |
| 5.4   | Kesimpulan.....  | 58        |
| 5.5   | Limitasi Kajian.....   | 59        |
| 5.6   | Cadangan.....  | 60        |
| <b>RUJUKAN.....</b>                                     |  | <b>63</b> |

## SENARAI JADUAL

|            | <b>Halaman</b>  |
|------------|---|
| Jadual 3.1 | Jumlah sampel yang diambil dari laman web <i>sebenarnya.my</i> ..... 31 |
| Jadual 3.2 | Kaedah tokenisasi..... 34   |
| Jadual 3.3 | Proses <i>Stop-Words Removal</i> ..... 35                               |

## SENARAI RAJAH

|           | <b>Halaman</b>  |
|-----------|---|
| Rajah 2.1 | Model Six-Step KDP yang dibangunkan oleh Cios et. al. (2007).....26                           |
| Rajah 2.2 | Kerangka konseptual model Six-Step KDP kajian..... 28   |
| Rajah 3.1 | Desain kajian..... 29   |
| Rajah 4.1 | Awan perkataan ciri-ciri signifikan topik COVID-19 laman<br>web <i>sebenarnya.my</i> ..... 39 |
| Rajah 4.2 | Awan perkataan ciri-ciri signifikan topik COVID-19 waspada..... 43                            |
| Rajah 4.3 | Awan perkataan ciri-ciri signifikan topik COVID-19 penjelasan..... 46                         |
| Rajah 4.4 | Awan perkataan ciri-ciri signifikan topik COVID-19 palsu..... 49                              |



## SENARAI SINGKATAN

|          |   |
|----------|---|
| AI       | <i>Artificial Intelligent</i>                               |
| BERNAMA  | Pertubuhan Maklumat Nasional Malaysia                       |
| CoV      | <i>coronavirus</i>  |
| CRISP-DM | <i>Cross-Industry Process for Data Mining</i>               |
| HTAR     | Hospital Tengku Ampuan Rahimah                              |
| IFCN     | <i>International Fact Checking Network</i>                  |
| IR       | <i>Information Retrieval</i>                                |
| JAPEN    | Jabatan Penerangan Malaysia                                 |
| KDP      | <i>Knowledge Discovery Process</i>                          |
| KKM      | Kementerian Kesihatan Malaysia                              |
| KKMM     | Kementerian Komunikasi dan Multimedia Malaysia              |
| KWSP     | Kumpulan Wang Simpanan Kerja                                |
| MER-CoV  | <i>Middle East Respiratory syndrome-related coronavirus</i> |
| MGA      | <i>Malaysian Golf Association</i>                           |
| MKN      | Majlis Keselamatan Negara                                   |
| PDRM     | Polis Diraja Malaysia                                       |
| PKP      | Perintah Kawalan Pergerakan                                 |

|            |  |
|------------|--|
| PKPB       | Perintah Kawalan Pergerakan Bersyarat                |
| PKPD       | Perintah Kawalan Pergerakan Diperketatkan            |
| PKPP       | Perintah Kawalan Pergerakan Pemulihan                |
| PRP        | Pasukan Respons Pantas                               |
| PSA        | Pesanan Khidmat Masyarakat                           |
| Rt         | R-naught   |
| RTM        | Radio Televisyen Malaysia                            |
| SARS       | <i>severe acute respiratory syndrome coronavirus</i> |
| SARS-CoV-2 | Sindrom Pernafasan Akut Teruk 2                      |
| SKMM       | Suruhanjaya Komunikasi dan Multimedia Malaysia       |
| SOP        | prosedur operasi standard                            |
| WHO        | Pertubuhan Kesihatan Sedunia                         |

# MAKLUMAT PALSU COVID-19: ANALISIS CORAK TOPIK TEKS

## LAMAN WEB *SEBENARNYA.MY* MENGGUNAKAN WEKA

### ABSTRAK

Penyebaran maklumat palsu semakin membimbangkan terutamanya sewaktu Malaysia dilanda pandemik COVID-19. Maklumat palsu berkaitan COVID-19 mampu menimbulkan kekeliruan kepada masyarakat dalam mendepani situasi sukar ini. Tambahan pula, maklumat palsu COVID-19 memberi kesan negatif terhadap usaha dalam membendung penularan COVID-19. Justeru, kajian ini bertujuan untuk menganalisis maklumat palsu COVID-19 berdasarkan objektif iaitu untuk memahami ciri-ciri signifikan topik COVID-19 palsu, waspada dan penjelasan dalam laman web *sebenarnya.my*. Data kajian diperoleh melalui analisis kandungan kualitatif dan kuantitatif dalam laman web *sebenarnya.my* bagi mencapai objektif. Data kajian ini dianalisis menggunakan perlombongan teks dengan bantuan perisian WEKA. Hasil kajian menunjukkan kategori palsu, penjelasan dan waspada dalam laman web *sebenarnya.my* mempunyai ciri-ciri signifikan tersendiri bagi topik pengesahan COVID-19. Ciri signifikan bagi kategori palsu ialah perintah kawal gerak palsu dan tidak benar, covid tidak benar serta bantu palsu. Bagi kategori penjelasan, ciri signifikannya adalah perintah kawal gerak, prosedur operasi standard, kes, pakai gelang dan mati. Manakala kategori waspada pula, waspada aplikasi maklumat dan waspada laman sawang tawar percuma. Kajian ini memberi gambaran sebenar kepada masyarakat, kerajaan dan media bahawa mengenalpasti serta memahami ciri-ciri signifikan topik maklumat palsu COVID-19 adalah penting sewaktu pandemic tersebut melanda.

# COVID-19 MISINFORMATION: ANALYSIS OF TEXT PATTERN FROM *SEBENARNYA.MY* WEBSITE USING WEKA

## ABSTRACT

The spread of false information is increasingly worrying especially when Malaysia is hit by the COVID-19 pandemic. False information related to COVID-19 is able to cause confusion to the public in the face of this difficult situation. In addition, misinformation about COVID-19 has a negative impact on efforts to curb the transmission of COVID-19. Thus, this study aims to analyze COVID-19 misinformation based on the objective of understanding the significant features of COVID-19 topics *palsu*, *waspada* and *penjelasan* in the *sebenarnya.my* website. The study data is obtained through qualitative and quantitative content analysis in the *sebenarnya.my* website to achieve the objective. The data of this study were analyzed using text mining through WEKA software. The results of the study show that the categories of *palsu*, *penjelasan* and *waspada* on the *sebenarnya.my* website have their own significant features for the topic of COVID-19 validation. Significant features for *palsu* category are fake and untrue movement control orders, fake COVID and fake assist. For the *penjelasan* category, the significant features are movement control orders, standard operating procedures, cases, wearing bracelets and death. While the *waspada* category, wary of information application and wary of free offering site. This study provides a true picture to the public, government and media that recognizing and understanding the significant features of the COVID-19 fake information topic is important during the pandemic.

# **BAB 1**

## **PENGENALAN**

### 1.1 Pengenalan

Perkembangan teknologi yang semakin rancak ketika ini memberi impak kepada pembangunan sosial dan ekonomi di Malaysia seiring dengan perubahan globalisasi yang akan mendatangkan manfaat kepada masyarakat (Rahayu & Sensusoyati, 2020). Namun, kerancakan pertumbuhan ini juga mendatangkan kesan negatif dimana mudahnya penyebaran maklumat palsu berlaku di mana-mana. Menurut Hafidzul (2021), sebanyak 1897 kertas siasatan dibuka membabitkan penyebaran maklumat palsu di media sosial sejak Januari 2019 hingga Jun 2021. Ia menjadi satu masalah besar apabila terdapat masyarakat yang mudah menerima dan mempercayai maklumat yang diterima tanpa banyak menyoal. Berdasarkan kajian dari UCSI Poll Research Centre (2021), 62 peratus responden mengaku masih mempercayai maklumat palsu yang disebarkan dalam media sosial dan aplikasi komunikasi. Malah terdapat di antara mereka yang memilih untuk berkongsi maklumat dengan rakan-rakan tanpa merujuk dengan pihak berautoriti terlebih dahulu.

Pandemik COVID-19 yang berlaku di Malaysia meningkatkan lagi aktiviti perkongsian maklumat (Aimi et. al., 2020), hingga mencetuskan infodemik. Infodemik adalah fenomena berlakunya lambakan maklumat benar dan palsu berkenaan pandemik sehingga membuatkan masyarakat keliru antara maklumat yang sahih dan salah serta sukar untuk mencari panduan yang boleh dipercayai (Harris & Farlina, 2020). Menurut Pertubuhan Kesihatan Sedunia (WHO) (2020) dan Zaracostas (2020), infodemik mengakibatkan bilangan maklumat palsu mengenai COVID-19 melonjak naik. Sehingga November 2020, Polis Diraja Malaysia (PDRM) dan Suruhanjaya

Komunikasi dan Multimedia Malaysia (SKMM) telah membuka sebanyak 274 kertas siasatan kes maklumat palsu berhubung COVID-19 (Azman, 2020). Angka kes maklumat palsu berhubung COVID-19 kekal menunjukkan trend peningkatan menjelang tahun berikutnya. Dalam dapatan kajian oleh Zalmizy (2020) membuktikan bahawa kes maklumat palsu berkenaan COVID-19 kebanyakannya berpunca daripada penyebaran dalam platform jaringan sosial seperti Facebook, WhatsApp dan Telegram.

Pada awal pandemik, kerajaan Malaysia memberikan tumpuan penuh untuk melandaikan lengkung penularan COVID-19 dengan menggariskan prosedur operasi standard (SOP) dan menguatkuasakan Perintah Kawalan Pergerakan (PKP). Namun, tindakan itu sahaja kurang mampu mengawal penularan COVID-19 kerana lambakan maklumat palsu mengenainya mempengaruhi masyarakat mempertikaikan tindakan pihak kerajaan (Shahidah et. al., 2020). Maklumat palsu COVID-19 menambahkan lagi kegusaran dalam kalangan masyarakat sehingga mempengaruhi pemikiran dan proses membuat keputusan (Sapora, 2020). Contohnya, tindakan menyebarkan maklumat palsu COVID-19 semasa gelombang pertama menyebabkan masyarakat berasa bimbang dan mula membuat pembelian panik (Muzaffar, 2020). Ia bukan sahaja mengakibatkan keadaan menjadi huru-hara, malah meningkatkan lagi kadar kebolehjangkitan COVID-19.

Justeru, pihak kerajaan mengambil pelbagai langkah dalam membendung penularan COVID-19 dan maklumat palsu mengenainya (Rodrigues & Jian, 2020). Selain penguatkuasaan undang-undang, kerajaan menerusi beberapa agensi menyiarkan Pesanan Khidmat Masyarakat (PSA) dalam memberi kesedaran pada masyarakat berkenaan maklumat palsu COVID-19 (Fazreen, 2020). Kerajaan juga menggunakan laman web *sebenarnya.my* yang berperanan sebagai pusat sehenti untuk

memberi verifikasi maklumat palsu COVID-19 melalui 3 kategori iaitu palsu, waspada dan penjelasan (Olivia et. al., 2020). Setiap kategori dalam *sebenarnya.my* mempunyai ciri dan fungsi yang tertentu. Bagi kategori palsu, kandungan pengesahannya mempunyai kata kunci seperti palsu, tidak benar, kurang tepat, tidak tepat dan menafikan. Kategori penjelasan, ia mempunyai kandungan pengesahan biasanya terdapat kata kunci seperti penjelasan, menerangkan dan memberitahu. Manakala waspada mempunyai kata kunci kandungan pengesahan seperti waspada, hati-hati dan semak semula. Dari segi kefungsiannya, kategori palsu adalah sebuah pengesahan bahawa maklumat yang tular adalah tidak benar. Bagi kategori penjelasan, ia berperanan sebagai maklumat benar yang memberi penerangan mengenai sesuatu hal yang disalah erti atau tidak difahami masyarakat. Kategori waspada pula berperanan sebagai maklumat untuk masyarakat pertimbangannya kerana ia boleh jadi benar atau palsu dalam masa yang sama. Oleh itu, setiap kategori penting untuk dikaji kerana ia dapat memberi pemahaman kepada masyarakat bahawa pengkategorian pada laman web *sebenarnya.my* mempunyai fungsinya tertentu dalam mengesahkan laporan maklumat palsu COVID-19 yang diterima. Selain itu, corak topik dalam setiap kategori berfungsi dan perlu dikenalpasti kerana ia mampu membantu masyarakat untuk sentiasa peka dengan topik COVID-19 yang sering dimanipulasikan dan disalah erti serta perlu diselidiki sebelum mempercayainya.

Banyak kajian tentang COVID-19 di Malaysia telah dijalankan terutamanya mengenai persepsi masyarakat mengenai PKP, perkhidmatan kesihatan awam, tindakan kerajaan dalam mitigasi pandemik secara survei (Arina et. al., 2020; Khayriyyah & Wan, 2020; Elengoe, 2020; Mohamad et. al., 2020; Merdeka Center, 2020; Center Force Group, 2020) dan analisis sentimen masyarakat dalam Twitter (Nik Norma & Suratnoaji, 2020). Namun, kajian mengenai maklumat palsu COVID-19

masih belum diterokai sepenuhnya. Antara kajian berkaitan ialah *Fake News on Social Media During COVID-19 Crisis* (Mohd Hanif et. al., 2020), *Misinformation And Disinformation During Covid19: The Effects and The Relevant Laws in Malaysia* (Hanis, 2020) dan *The COVID-19 Infodemic in Malaysia* (Harris & Farlina, 2020) secara temubual dan analisis isi kandungan. Justeru, kajian ini penting dikaji untuk merapatkan jurang kajian dengan meneroka kategori palsu, waspada dan penjelasan pada laman web *sebenarnya.my* dengan menggunakan kaedah perlombongan teks.

Definisi berita palsu masih diperdebatkan hingga kini (Ireton & Posetti, 2020). Dalam konteks akademik, definisi berita palsu mempertikaikan kredibiliti bidang kewartawanan. Contohnya, istilah berita palsu digunakan pada maklumat tidak benar dalam format berita sahaja (Allcott & Gentzkow, 2017; Conroy et. al., 2015). Southwell et. al. (2017) dan Jack (2017) memperjelaskan istilah berita palsu dan hubungkaitnya dengan *misinformation* dan *disinformation* kerana ia berkenaan informasi yang salah tetapi membawa kepada konteks dan makna berbeza. Pemilihan istilah bersesuaian untuk menerangkan manipulasi media membawa kepada andaian mengenai kaedah penyebaran, individu penyebar dan penerimanya. Terminologi *misinformation* dan *disinformation* lebih sesuai digunakan dan menggantikan berita palsu kerana sifat perkataan berita adalah informasi yang benar secara faktual (Ireton & Posetti, 2020; Wardle & Derakhshan, 2018). Bagi kajian ini, penyelidik menggunakan istilah maklumat palsu (*misinformation*) dan bukannya *disinformation* kerana Kuskridho (2019) mendefinisikannya sebagai informasi salah dan pembuat serta penyebarannya mengetahui bahawa ia adalah maklumat yang salah. Laman web *sebenarnya.my* mentakrifkan maklumat palsu sebagai informasi yang tidak dipastikan kesahihannya sehingga boleh mendatangkan perselisihan faham serta syak wasangka.



## 1.2 Pernyataan Masalah

Dari sudut etimologi, penggunaan maklumat palsu mula direkodkan pada tahun 1547 dengan definisi sebagai tindakan memberi maklumat yang salah atau keadaan apabila sesuatu informasi dimaklumkan secara tidak betul tanpa tujuan khusus (Simpson & Weiner, 1991; Tumber & Waisbord, 2021). Manakala, SANTOS-D'AMORIN & de Oliveira Miranda (2021) menegaskan maklumat palsu mula ditemui dalam Kamus Rusia pada tahun 1949 yang bermaksud tindakan untuk mengelirukan masyarakat melalui penggunaan serta penyebaran informasi yang salah. Maklumat palsu juga diterjemahkan sebagai informasi yang dimanipulasi sehingga ia kelihatan benar dan diterima oleh masyarakat (MacDougall, 1958; Fathiah & Zuliza, 2018). Terma COVID-19 bukti menunjukkan pula ia mula digunakan pada akhir Disember 2019. Ia adalah Ia berlaku apabila sistem kesihatan China melaporkan sekumpulan pesakit menghidapi radang paru-paru atas sebab yang tidak diketahui. Kemudian, ia dikaitkan pula dengan pasar basah di Wuhan, wilayah Hubei, China (Anumudu et. al., 2020; Mackenzie & Smith, 2020).

Antara satu masalah besar yang dihadapi sewaktu pandemik COVID-19 adalah penyebaran maklumat palsu berkenaannya. Maklumat palsu COVID-19 mampu memberi impak terhadap segenap aspek kehidupan (Hou et. al., 2020). Penularannya mengakibatkan masyarakat ragu dengan usaha kerajaan dalam mengawal selia pandemic (Hao & Basu, 2020; Wan, 2020). Masyarakat juga akan menjadi panik sekaligus membuat keputusan terburu-buru (Nicomedes & Avila, 2020), tidak patuh arahan kerajaan (Kadam & Atre, 2020). Sungguhpun maklumat palsu COVID-19 meninggalkan kesan negatif, angkanya didapati terus meningkat saban hari (Apuke & Bahiyah, 2020). Setakat awal November 2020 sahaja, PDRM dan SKMM sudah membuka 274 kertas siasatan berkenaan penyebaran maklumat palsu berkaitan wabak

tersebut. Sehingga September 2021, jumlah kumulatif kertas siasatan berhubung penyebaran maklumat palsu berkaitan COVID-19 yang dibuka sejak Mac 2020 adalah 307. Su et. al. (2020) menerangkan bahawa media sosial serta aplikasi Whatsapp menjadi punca kepada perkongsian maklumat palsu COVID-19 berlaku dengan begitu pantas dalam kalangan masyarakat.

Ekoran itu, kerajaan mula menggunakan *sebenarnya.my*, sebuah laman web yang ditubuhkan SKMM pada 14 Mac 2017 untuk memberi verifikasi informasi serta membolehkan masyarakat membuat aduan (Md Azlanshah & Mohd Nazri, 2018) tentang maklumat palsu berkenaan wabak tersebut. Pengesahan tersebut dikumpulkan dibawah tab COVID-19 yang mula diperkenalkan semenjak pandemik. Namun, bilangan maklumat palsu COVID-19 didapati tidak menunjukkan tanda penurunan, sebaliknya terus meningkat setiap hari. Ini kerana masyarakat mempunyai sikap lebih mempercayai maklumat palsu yang tular dan kurang inisiatif atau tidak mengetahui bagaimana cara untuk menyemak kesahihannya. Pertubuhan Maklumat Nasional Malaysia (BERNAMA) (1 Februari 2018) dan USCI Poll Research Centre (2021) melaporkan bahawa masyarakat masih kurang faham fungsi laman web *sebenarnya.my* ditubuhkan oleh kerajaan dalam mengawal selia isu maklumat palsu. Ditambah pula dengan tab baharu seperti COVID-19 dalam *sebenarnya.my*, masyarakat diandaikan masih belum cakna dan mahir untuk membuat semakan maklumat berkenaannya.

Kajian lepas mengenai maklumat palsu COVID-19 kebanyakannya menghujahkan perihal yang berkaitan dengan bidang perubatan dan sains sosial (Noor et. al., 2020). Antara kajian berkaitan adalah berkenaan kesan, punca dan skop maklumat palsu COVID-19 yang tular di media sosial. Namun, corak kandungan maklumat palsu yang tersebar di Malaysia perlu diterokai lagi (Sarah, 2019; Tee &

Raja, 2019). Shiang & Edmund (2021) menegaskan kajian semasa yang memfokuskan kandungan maklumat palsu COVID-19 di Malaysia masih lagi kurang. Masyarakat lebih mudah untuk dimanipulasi sekiranya tahap pemahaman kandungan maklumat palsu COVID-19 seperti topik yang sering menjadi sasaran untuk mengelah masih lagi cetek dan kabur. Malah, ia juga merencatkan strategi pihak kerajaan, agensi media dan organisasi kesihatan awam dalam membendung penularan maklumat palsu COVID-19.

### 1.3 Persoalan Kajian

1.3.1 Bagaimanakah ciri-ciri signifikan topik COVID-19 palsu, waspada dan penjelasan dalam laman web *sebenarnya.my*?

### 1.4 Objektif Kajian

1.4.1 Untuk memahami ciri-ciri signifikan topik COVID-19 palsu, waspada dan penjelasan dalam laman web *sebenarnya.my*

### 1.5 Kepentingan Kajian

Kajian ini penting untuk dijalankan kerana COVID-19 adalah pandemik yang tidak pernah wujud sebelum ini dan sangat merbahaya. Banyak spekulasi timbul akibat kejahilan masyarakat mengenai pandemik ini dan lambakan maklumat tersebut boleh menggugat usaha kerajaan membendung penyebarannya. Oleh itu, kajian ini meninjau bagaimana laman web *sebenarnya.my* membantu orang ramai untuk mengenalpasti kesahihan berita yang mereka terima. Kajian lepas mengenai pandemik dan maklumat palsu mengenai COVID-19 banyak membincangkan isu-isu maklumat palsu di media sosial (Mohd Hanif et.al., 2020; Shahi et. al., 2021), kesan penyebarannya (Hanis, 2020 & Shahidah et. al., 2020), pengurusan maklumat palsu (Kasirye, 2021; Schuldt, 2021) dan punca penyebaran (Apuke & Bahiyah, 2020;

Balakrishnan et. al., 2021). Justeru, kajian mengenai maklumat palsu COVID-19 perlu dijalankan agar ia mampu memberi sumbangan dari segi akademik dengan mengisi jurang penyelidikan. Selain itu, hasil dapatan kajian ini dapat menjadi sumber rujukan kepada penyelidik lain untuk melakukan kajian berkaitan maklumat palsu sewaktu pandemik melanda.

Manakala secara teoritis, kajian ini dapat memberikan sumbangan pemikiran dan perkembangan ilmu kepada bidang silang disiplin. Dalam hal ini, kajian menerapkan model dari satu bidang yang memerlukan pengetahuan informatik seperti sains komputer dalam bidang sains sosial khususnya komunikasi. Kajian ini secara tidak langsung akan menentusahkan kefungisian dan menyumbang kepada perkembangan model sedia ada iaitu model Six-Step KDP yang digunakan untuk memberi penerangan secara teknikal mengenai prosedur perlombongan teks meskipun tanpa kewujudan teori tambahan.

Dapatan kajian juga dapat membantu pihak kerajaan yang meliputi beberapa agensi seperti Jabatan Penerangan Malaysia (JAPEN), Suruhanjaya Komunikasi dan Multimedia Malaysia (SKMM) dan Radio Televisyen Malaysia (RTM) dalam menambahbaik strategi mengawal selia masalah maklumat palsu terutamanya sewaktu pandemik. Ia juga boleh dijadikan penanda aras kepada keberkesanan laman web *sebenarnya.my* dalam mengekang penularan maklumat palsu di Malaysia. Hasil kajian ini juga mampu membantu pihak kerajaan dalam merencanakan pelbagai strategi baru dalam menyelesaikan masalah berhubung maklumat palsu terutamanya sewaktu wabak penyakit melanda.

## 1.6 Skop Kajian

Bagi kajian ini, penyelidik menetapkan bahawa hanya *sebenarnya.my* yang akan dikaji, meskipun terdapat beberapa lagi laman web disediakan untuk semakan maklumat palsu COVID-19 seperti *mycheck.my* milik BERNAMA dan *faqcheck.org* oleh aktivis media. Laman web *sebenarnya.my* dipilih kerana ia merupakan suatu platform semakan fakta dan kesahihan maklumat pertama ditubuhkan oleh kerajaan Malaysia (Hussain, 2018). Meskipun *sebenarnya.my* mempunyai media sosial yang tersendiri seperti Facebook dan Twitter, penyelidik memilih untuk mengkaji laman web. Ini kerana, penentusahan maklumatpalsu pada laman web *sebenarnya.my* masih lagi menjadi lampiran (*attachment*) menerusi pautan dalam hantaran (*post*) bagi kedua-dua platform media sosial itu. Kajian ini juga tertumpu kepada pengesahan pada tab COVID-19 bagi kategori palsu, waspada dan penjelasan. Setiap kategori yang terdapat pada laman web *sebenarnya.my* mempunyai fungsi yang tertentu dalam memberi verifikasi. Justeru, penyelidik ingin mengenalpasti corak maklumat palsu bagi setiap kategori itu. Kajian ini memfokuskan pada pengesahan bagi kategori palsu, waspada dan penjelasan pada tab COVID-19 dalam laman web *sebenarnya.my* selama setahun iaitu bermula Mac 2020 hingga Februari 2021. Penyelidik memilih Mac 2020 kerana Malaysia dilaporkan mula menjalani PKP bermula minggu ketiga pada bulan tersebut.

## 1.7 Konseptualisasi

### 1.7.1 Maklumat palsu

Penyebaran maklumat palsu didapati semakin berleluasa di platform media sosial seperti Twitter serta Facebook dengan tujuan mewujudkan kebencian dan provokasi dalam kalangan masyarakat (Farris et. al., 2019). Allcott & Gentzkow (2017) menegaskan maklumat palsu adalah artikel yang terbukti tidak benar dan

mampu menyesatkan pembaca. Demartini et. al. (2020) menegaskan bahawa maklumat palsu boleh terdiri daripada bentuk format seperti teks, iklan, gambar dan video berdurasi pendek. Wardle & Derakhshan (2018) menegaskan maklumat palsu terdiri daripada tujuh kategori iaitu parodi, kandungan mengelirukan, kandungan penipuan, kandungan dibuat-buat, perkaitan dimanipulasikan, salah konteks dan perkaitan palsu.

### 1.7.2 COVID-19

Menurut laporan WHO, COVID-19 merupakan suatu penyakit berjangkit yang baru sahaja ditemui dalam strain coronavirus dan penyakit ini boleh tersebar dengan mudah melalui titisan cair yang keluar dari badan manusia. Wan et. al. (2020) menerangkan bahawa COVID-19 adalah sejenis virus yang boleh mengakibatkan jangkitan pada saluran pernafasan. Secara dasarnya, virus ini pernah digelar sebagai ‘2019 Novel Coronavirus’, ‘Wuhan Coronavirus’ dan ‘2019-nCoV’. Namun pada 13 Februari 2020, WHO telah menggazet nama virus ini kepada COVID-19 supaya ia mudah dirujuk terutamanya bagi tujuan penyelidikan dan penyelenggaraan kes.

### 1.7.3 Laman web *sebenarnya.my*

Laman web *sebenarnya.my* merupakan laman web yang dibangunkan oleh Suruhanjaya Komunikasi dan Multimedia (SKMM) dengan beberapa agensi kerajaan pada 14 Mac 2017 bagi membolehkan masyarakat menyemak status kesahihan maklumat dengan cepat dan mudah secara atas talian (Swee et. al., 2020). Laman web ini hanya menumpukan kepada isu-isu tempatan yang memberi kesan terhadap kepentingan awam atau nasional (Syed, 2018). Laman web *sebenarnya.my* boleh mendapatkan mengesahkan sesuatu maklumat yang tersebar di media sosial sejak

tahun 2001 (Sulaiman et. al., 2019). Kesahihan maklumat disahkan oleh saluran rasmi atau agensi kerajaan berdasarkan peranan, bidang kuasa dan kepakaran agensi.

#### 1.7.4 Perlombongan teks

Menurut Dan et. al. (2020), perlombongan teks merupakan kaedah menganalisis teks dengan tujuan untuk mengekstrak pengetahuan dan hubungan dari dokumen yang berbentuk tekstual. Selain itu, perlombongan teks juga sebagai suatu kaedah mengumpul dan mencari corak bagi sejumlah set data yang besar serta tertentu. Saire (2020) dalam kajiannya turut menyifatkan perlombongan teks sebagai kaedah yang dapat mendekati dan memahami beberapa bidang seperti kesihatan awam, sosial, dan banyak lagi. Bagi menjalankan kaedah, alat atau algoritma analisis diperlukan.

#### 1.7.5 *Supervised Learning Technique* (pengklasifikasian)

*Supervised Learning Technique* atau pembelajaran terselia sesuai untuk tugas klasifikasi dan regresi (Siti Hajar & Hasmawati, 2021) kerana ia adalah kaedah yang mementingkan konsep sasaran dalam set data, di mana ia dijadikan sebagai panduan untuk mengetahui hubungan antara atribut yang terlibat dengan konsep sasaran. *Supervised Learning Technique* sering digunakan untuk menganalisis data yang telah dilabel atau diberi keterangan kumpulannya (Jajat & Tacbir, 2008). Antara algoritma di bawah *supervised learning technique* adalah Hutan Rawak, Naive Bayes, Pohon Keputusan dan Regrasi Logistik.

### 1.8 Operasionalisasi

#### 1.8.1 Maklumat Palsu

Pengesahan maklumat palsu yang terdapat dalam tab COVID-19 di laman web *sebenarnya.my* telah dibahagikan kepada tiga kategori yang memainkan peranan

masing-masing dalam memberitahu status kesahihan sesuatu perkara. Antara tiga kategori yang telah dibahagi dan dilabel oleh laman web *sebenarnya.my* ialah palsu, waspada dan penjelasan. Bagi kajian ini, penyelidik mengambil pengesahan dengan label palsu, waspada dan penjelasan yang terdapat pada tab COVID-19 di laman web *sebenarnya.my*. Kemudian, pengesahan tersebut dibahagikan mengikut kategori masing-masing agar ia dapat dikaji satu per satu dengan mudah oleh penyelidik. Ini secara tidak langsung membolehkan penyelidik mengenalpasti persamaan dan perbezaan topik bagi setiap kategori.

### 1.8.2 Laman Web *sebenarnya.my*

Secara amnya, laman web *sebenarnya.my* mempunyai beberapa tab di bahagian atasnya yang terdiri daripada UTAMA, NASIONAL, SOSIAL, INFO, LOGO, COVID-19, SALUR KEPADA KAMI dan TENTANG KAMI. Bagi kajian ini, penyelidik hanya memilih tab COVID-19 laman web *sebenarnya.my* untuk dikaji. Pada tab COVID-19, terdapat sebanyak 439 pengesahan yang dimuatnaik oleh laman web *sebenarnya.my* dari Mac 2020 hingga Februari 2021. Selain itu, bagi mencapai objektif kajian, penyelidik hanya memfokuskan kepada teks yang terdapat pada pengesahan tersebut. Penyelidik tidak menganalisa sebarang bentuk gambar, infografik mahupun video berdurasi pendek pada pengesahan dalam laman web *sebenarnya.my*.

### 1.8.3 Perlombongan Teks

Pengukuran dibuat dengan mengumpul pengesahan COVID-19 dalam kategori palsu, waspada dan penjelasan pada laman web *sebenarnya.my* untuk menjalankan analisis perlombongan teks. Pengesahan tersebut perlu melalui prapemprosesan data secara manual yang terdiri daripada tokenisasi, *stop-word removal* dan *stemming*



dengan menggunakan perisian Microsoft Excel. Kemudian, set data melalui kaedah Hutan Rawak yang tergolong dalam *supervised learning technique* untuk mencapai objektif kajian.

#### 1.8.4 *Supervised learning technique* (pengklasifikasian)

Dalam kajian ini, *supervised learning technique* digunakan untuk mengklasifikasikan setiap set data pengesahan di laman web *sebenarnya.my* mengikut ciri-ciri signifikan topik COVID-19. Teknik pengklasifikasian ini dibuat dengan menggunakan algoritma Hutan Rawak menerusi perisian untuk perlombongan teks, WEKA. Bagi *supervised learning technique*, setiap set data pengesahan dibahagikan kepada beberapa buah kumpulan dengan mengikut ciri-ciri dominan padanya yang boleh dikesan oleh algoritma Hutan Rawak. Setelah itu, penyelidik dapat mengenalpasti ciri-ciri signifikan topik bagi maklumat palsu yang dilabel ‘palsu’, ‘waspada serta ‘penjelasan’ yang dimuatnaik oleh laman web *sebenarnya.my*. Ia juga sekaligus membolehkan penyelidik menganalisis persamaan dan perbezaan untuk setiap satu kategori.

## BAB 2

### SOROTAN LITERATUR

#### 2.1 Pengenalan

Bab ini membincangkan mengenai infodemik, maklumatpalsu dan perlombongan teks. Pertama, kajian ini menyetengahkan isu berkenaan pandemik COVID-19 dan infodemik yang berlaku sewaktu krisis. Seterusnya, bab ini akan membincangkan tentang maklumat palsu yang melibatkan isu-isu COVID-19 dan peranan sistem *fact checker* dalam membendung masalah ini. Pengaplikasian kaedah analisis perlombongan teks juga diuraikan. Akhir sekali, bab ini juga akan membincangkan Model Six-Steps KDP yang akan menjadi panduan kepada kajian ini.

#### 2.2 Pandemik COVID-19

Kementerian Kesihatan Malaysia (KKM) (2020) melaporkan coronavirus adalah sejenis virus yang mampu menyebabkan jangkitan pada sistem pernafasan. Kebiasaannya, coronavirus hadir dengan spektrum gejala ringan seperti selesema, batuk dan demam. Malah ia juga boleh mengakibatkan simptom yang lebih serius seperti kegagalan pada beberapa organ, radang paru-paru (*pneumonia*) dan kejutan septik sehingga mampu mengakibatkan kematian. Sebelum ini, Kementerian Kesihatan Malaysia (KKM) telah mengenalpasti dua *strain* coronavirus berbeza iaitu *severe acute respiratory syndrome coronavirus* (SARS) dan *Middle East Respiratory syndrome-related coronavirus* (MER-CoV). Namun pada penghujung tahun 2019, satu lagi coronavirus (CoV) atau lebih dikenali sebagai COVID-19 telah dikenalpasti.

Virus ini mula ditemui di negara China dengan mengakibatkan pesakit yang menghidapinya mengalami kesukaran untuk bernafas. Malah, terdapat beberapa kes

dimana pesakit COVID-19 turut mengalami masalah pernafasan yang teruk sehingga memerlukan rawatan khas dengan menggunakan alat bantuan seperti *ventilator*. COVID-19 didapati boleh menjadi lebih serius sekiranya ia dihidapi oleh pesakit dengan masalah kesihatan lain seperti penyakit paru-paru kronik. Selain itu, warga emas, kanak-kanak dan bayi adalah antara golongan berisiko tinggi untuk dijangkiti COVID-19 dengan memperoleh simptom yang jauh lebih serius (Siti Masayu & Fatimah, 2020).

Sejak awal tahun 2020, WHO mengistiharkan COVID-19 sebagai pandemik kerana penularannya secara global tidak dapat dikawal (Nor Faziana, 2020). COVID-19 menjadi sangat serius kerana ia merebak melalui populasi manusia dan merentasi kawasan yang luas sehingga melepasi sempadan benua tanpa tanda-tanda kebolehan kawalan. Ernest et. al. (2020) mendefinisikan pandemik sebagai penyakit berjangkit yang tiba-tiba menimpa sebilangan besar penduduk dunia. Pandemik bukanlah frasa yang boleh dipandang remeh dan digunakan dengan sewenang-wenangnya, kerana ia merujuk kepada penularan penyakit melalui manusia secara global. Justeru, pengklasifikasian sesuatu keadaan sebagai pandemik perlulah dilakukan oleh pihak berautoriti seperti WHO dengan teliti, agar ia tidak memberi kesan terhadap beberapa aspek seperti pembangunan sosial, ekonomi, kesihatan dan pendidikan.

Malaysia telah melaporkan kes pertama berkenaan COVID-19 pada 25 Januari dengan melibatkan tiga warga China yang memasuki negara ini melalui Johor dari Singapura (Rozita, 2020). Hakikatnya, Malaysia telah pun berdepan dengan dua gelombang jangkitan COVID-19 yang masing-masingnya mempunyai kadar jangkitan berbeza. Sewaktu gelombang pertama (25 Januari hingga 15 Februari 2020), kadar jangkitan COVID-19 masih dalam keadaan yang terkawal dan hanya melibatkan kes-kes import. Manakala, pada gelombang kedua (bermula 27 Februari 2020), kadar

jangkitan didapati semakin meningkat secara mendadak ekoran daripada kluster tempatan yang mulai muncul.

Setelah melihat kepada keadaan yang semakin meruncingkan, kerajaan Malaysia telah mengambil tindakan dengan melaksanakan Perintah Kawalan Pergerakan (PKP) selama dua minggu bermula 18 Mac 2020. Malaysia merupakan antara negara terawal yang melaksanakan PKP di rantau Asia Tenggara bagi mengekang penularan COVID-19 (Wan Faizah et. al., 2020). PKP digerakkan dan dipantau oleh beberapa agensi kerajaan di bawah Majlis Keselamatan Negara (MKN) melalui Akta Pencegahan dan Pengawal Penyakit Berjangkit 1988 dan Akta Polis 1967. Sehingga kini, Malaysia telah melaksanakan beberapa fasa PKP dengan tempoh tertentu yang mengikut pada kadar kes positif COVID-19 di negara ini.

### 2.3 Infodemik Ketika Pandemik COVID-19

Menurut laporan keadaan (*situation report*) yang dikeluarkan oleh WHO (15 April 2020) infodemik COVID-19 boleh didefinisikan sebagai ledakan maklumat dimana yang benar dan palsu telah pun bercampur baur sewaktu pandemik itu melanda. Selain itu, WHO (2020) menyifatkan infodemik ini seolah-olah sama sahaja seperti pandemik COVID-19. Ini kerana infodemik COVID-19 telah tersebar dalam kalangan masyarakat melalui teknologi sebagai medium. Manakala, pandemik COVID-19 pula disebarkan melalui sejenis virus yang dikaitkan dengan Sindrom Pernafasan Akut Teruk 2 (SARS-CoV-2). Justeru, WHO menggesa orang ramai untuk sama-sama memerangi isu infodemik dan pandemik COVID-19 ini secara serentak (KKM, 2020).

Berbeza pula dengan kajian oleh Tangcharoensathien et. al. (2020) yang menerangkan bahawa isu infodemik COVID-19 masih mempunyai titik perbezaan

dengan pandemiknya. Hal ini adalah kerana infodemik COVID-19 didapati lebih sukar untuk dihentikan, berbanding wabak penyakit tersebut. Penyebaran lebih mudah melalui media sosial seperti Facebook, Instagram, Twitter yang menghubungkan manusia sama ada dengan jarak dekat mahupun jauh. Manakala, pandemik COVID-19, masih lagi sukar untuk tersebar selagi manusia saling berjauhan. Sekiranya bilangan maklumat COVID-19 yang tersebar berjaya disusutkan, ia mampu mengurangkan keparahan wabak penyakit itu (Shukla, 2020).

Menurut Rahkman (2020), infodemik terjadi apabila masyarakat cenderung untuk mendapatkan maklumat berkaitan COVID-19 dengan kadar yang cepat dan mudah. Sewaktu kejadian seperti pandemik COVID-19 berlaku, masyarakat dilihat sentiasa menginginkan penjelasan mengenai segala perkara berkaitan penyakit berjangkit tersebut. Hakikatnya, masyarakat terlalu taksib untuk mendapatkan maklumat kerana ia membolehkan mereka mengawal dan meramal situasi semasa. Malah dapat dikatakan juga bahawa masyarakat akan mula berasa takut (*paranoid*) sekiranya mereka tidak mampu untuk memperoleh maklumat dengan kuantiti yang banyak berkenaan sesuatu kejadian seperti COVID-19.

Infodemik telah menimbulkan rasa kurang selesa dalam kalangan masyarakat sewaktu pandemik COVID-19 (Rudy, 2020). Masyarakat mengalami kecelaruan dalam menentukan maklumat yang diperolehi sama ada benar atau palsu. Bahkan, kecelaruan dalam menentukan maklumat benar dan palsu boleh mencetuskan rasa panik dalam kalangan masyarakat sepertimana yang pernah berlaku sebelum ini. Oleh itu, pelbagai usaha wajarlah dilakukan bagi memastikan infodemik kekal terkawal terutamanya sewaktu pandemik. Hakikatnya, isu infodemik yang tidak dikawal dengan baik pastinya akan mendatangkan kesan negatif yang bukan sahaja terhadap masyarakat semata, malah juga pada negara itu sendiri.

## 2.4 Maklumat Palsu COVID-19

Semenjak awal pandemik COVID-19 lagi, penyebaran maklumat palsu berlaku secara meluas dan pantas di media sosial seperti Facebook, Twitter serta Instagram melalui kaedah *shares*, *likes* atau *retweets* (Apuke & Bahiyah, 2020; Barua et. al., 2020). Menerusi kaedah ini, ia memberi kesan pada algoritma media sosial tersebut dengan mempamerkan hantaran yang mempunyai bilangan *share*, *like* atau *retweet* tertinggi terlebih dahulu. Di Malaysia, Harris & Farlina (2020) pula menyatakan bahawa WhatsApp merupakan medium utama yang digunakan dalam menyebarkan maklumat palsu berkaitan COVID-19. Hakikatnya, penyebaran maklumat palsu COVID-19 di Malaysia akan terus berlaku selagi mana masih terdapat sesetengah pihak yang menyalahgunakan jaringan sosial yang sedia ada.

Lamos et. al. (2020) menyifatkan maklumat palsu COVID-19 boleh mengakibatkan ancaman kepada masyarakat dengan mencetuskan rasa bimbang. Ia mengakibatkan masyarakat tidak membuat pertimbangan sebelum melakukan tindakan. Contohnya, masyarakat mula berasa bimbang dan akhirnya membuat pembelian panik setelah mempercayai maklumat palsu (Nurul Hidayah, 2020). Melalui pembelian panik, ia meningkatkan kadar kebolehjangkitan kerana penjarakan sosial tidak diamalkan. Maklumat palsu COVID-19 turut memberi impak kepada tahap kesihatan masyarakat (Noor Rahmawati et. al., 2020; Rodriguez et. al, 2020). Misalnya, masyarakat terpedaya dengan maklumat palsu dimana air kencing lembu dan bahan peluntur mampu mencegah penyakit berjangkit tersebut. Pelbagai jenis nasihat dan tuntutan bagi mencegah COVID-19 yang mengelirukan turut dikeluarkan (Naeem et. al., 2020)

Kerajaan telah melaksanakan pelbagai usaha dalam membanteras maklumat palsu. Antaranya, dengan mewujudkan Akta Antiberita Tidak Benar 2018 pada bulan April 2018 (BERNAMA, 2018). Namun, pada Ogos 2018 ia dimansuhkan kerana akta sebelumnya sudah cukup memadai (Mohd Iskandar, 2019). Kini, pihak kerajaan menetapkan bahawa pesalah akan berdepan dengan tindakan undang-undang di bawah Seksyen 211 dan 233 Akta Komunikasi dan Multimedia 1998 (SKMM, 2020). Jika disabitkan kesalahan, pesalah boleh disaman tidak melebihi RM50, 000, dipenjarakan tidak melebihi setahun atau kedua-duanya. Perdana Menteri sedang mempertimbangkan untuk mewujudkan kembali Akta Antiberita Tidak Benar 2018 (Rahimah, 2020) kerana bilangan maklumat palsu dikesan mula meningkat pada tahun 2020 ekoran daripada pandemik COVID-19.

Maklumat palsu COVID-19 turut berlaku di Sepanyol, Itali, Perancis dan Jerman (Murray, 2020). Maklumat palsu yang tersebar di negara tersebut adalah berkenaan peranti elektronik dengan ciri 5G menjadi punca kepada kewujudan penyakit COVID-19. Malah, terdapat maklumat palsu yang menyifatkan ciri 5G tersebut membuatkan simptom COVID-19 bertambah teruk dan sukar untuk dirawat. WHO menegaskan maklumat palsu mengenai penyakit COVID-19 yang berpunca dari 5G adalah tidak benar kerana virus tidak boleh tersebar melalui gelombang radio atau peranti elektronik (Foo, 2020). Tambahan pula, terdapat beberapa buah negara yang tidak mempunyai peranti elektronik 5G turut berdepan dengan pandemik itu.

Harris & Farlina (2020) melaporkan maklumat palsu COVID-19 di Malaysia kebanyakannya adalah mengenai tindakan pihak berautoriti dalam mengurangkan kadar kebolehjangkitan virus dan punca penyebaran COVID-19. Antara maklumat palsu yang tular ialah mengenai penangguhan rawatan atau pembedahan bagi pesakit kanser bagi mengelakkan berlakunya peningkatan pada kadar kebolehjangkitan

COVID-19 di hospital. Selain itu, maklumat palsu mengenai punca penyebaran COVID-19 yang terdiri daripada golongan seperti tabligh, pelancong dari negara China, migran dan pelarian (Wan, 2020). Setelah menerima maklumat palsu begitu, masyarakat mula menyalahkan kumpulan tertentu sebagai penyebab berlakunya penyakit berjangkit berdasarkan andaian yang tidak berasas.

## 2.5 Laman Web *sebenarnya.my*

Laman web *sebenarnya.my* adalah strategi kerajaan menerusi SKMM dalam mengawal selia gejala penularan maklumat palsu di Malaysia (Emma & Nur Hazwani, 2019). Ia ditubuhkan pada 14 Mac 2017 sebagai pusat sehati yang membolehkan masyarakat menyemak kesahihan maklumat (Olivia et. al., 2020). Laman web *sebenarnya.my* membuat verifikasi dengan memuatnaik tangkap skrin (*screen shot*) maklumat palsu yang tular beserta satu artikel pendek. Selain itu, ia membolehkan masyarakat melaporkan maklumat palsu yang diterima melalui media sosial, perkhidmatan mesej segera, blog, laman web dan lain-lain. Pada tahun pertama penubuhan, laman web ini hanya mempunyai satu kategori iaitu palsu yang berfungsi sebagai pemakluman maklumat tidak sah (BERNAMA, 1 Februari 2018).

Pada tahun 2018, laman web ini menambah skop kategorinya daripada hanya palsu kepada waspada dan penjelasan. Menurut BERNAMA (1 Februari 2018), penjelasan berfungsi sebagai penerangan yang dikeluarkan oleh agensi berkaitan mengenai sesuatu isu. Manakala, waspada adalah peringatan kepada masyarakat agar berjaga-jaga terhadap segala jenis bentuk maklumat palsu. Semenjak itu, laman web *sebenarnya.my* mempunyai tiga kategori verifikasi mengikut pengkelasan warna seperti palsu (merah), waspada (kuning) dan penjelasan (biru). Dalam masa sama, SKMM melancarkan aplikasi mudah alih *sebenarnya.my* yang boleh dimuatnaik oleh



masyarakat menerusi peranti pintar. Kemudian, ia memperluaskan lagi jangkauannya dengan menggunakan saluran media sosial Twitter dan Facebook (The Sun Daily, 20 Mac 2018). Sehingga Februari 2021, laman web ini mencatatkan jumlah pengikut seramai 9242 (Twitter) dan 16787 (Facebook).

Sewaktu Malaysia berdepan dengan pandemik pada tahun 2020, SKMM memperkenalkan tab COVID-19 dalam laman web *sebenarnya.my* bagi memudahkan masyarakat untuk menyemak kesahihan maklumat mengenai penyakit berjangkit tersebut. Menurut Arumugam (2020), SKMM melantik Pasukan Respons Pantas (PRP) untuk memantau maklumat COVID-19 yang mencurigakan di media sosial. Kemudian, ia dinilai dan dimuatnaik di laman web *sebenarnya.my* sebagai verifikasi. Sehingga pertengahan bulan Mei 2020, sebanyak 256 maklumat palsu COVID-19 berjaya dibongkar dan disebar kepada masyarakat oleh laman web *sebenarnya.my* (Aimi et. al., 2020). Pada tahun yang sama, laman web *sebenarnya.my* menubuhkan satu saluran di Telegram bagi membolehkan masyarakat menyemak maklumatpalsu dengan lebih mudah (Angelin, 2020).

*Snopes.com* yang dibangunkan oleh Snopes Media Group mempunyai titik persamaan dari segi fungsinya dengan laman web *sebenarnya.my* di Malaysia (Samantha, 2017). Dalam konteks ini, *Snopes.com* turut membolehkan masyarakat menyemak kesahihan bagi sesuatu maklumat yang tular di media sosial. Menurut Botnevik (2020), laman web ini menyemak kesahihan maklumat secara manual dengan melakukan analisis kontekstual. Selain itu, masyarakat boleh membuat pemakluman mengenai sumber maklumatpalsu berdasarkan kata kunci dan capaian pautan yang berkaitan. Sewaktu pandemik COVID-19 melanda dunia, laman web *Snopes.com* turut menyediakan verifikasi maklumat berkaitan penyakit berjangkit tersebut mengikut kategori dan topik tertentu (Shahi & Nandini, 2020).

Pada Mac 2020, BERNAMA menubuhkan *mycheck.my* sebagai laman web semakan fakta secara bebas untuk membanteras maklumat palsu di Malaysia. Ia beroperasi sepenuhnya pada Ogos 2020 dengan mematuhi piawaian *International Fact Checking Network* (IFCN). Laman web ini hanya menyemak maklumat yang memberi impak kepada kehidupan masyarakat di Malaysia. Proses semakan bermula daripada pemantauan sumber yang berpotensi menjadi tempat sebaran maklumat palsu. Setelah itu, semakan maklumat dibuat melalui proses perbincangan dan kerja berkumpulan. Selain daripada menerbitkan artikel pengesahan, *mycheck.my* turut menyediakan ringkasan maklumat dalam dua kotak iaitu 'Benarkah' dan 'Hasil' bagi memudahkan pembaca untuk memahami apa yang disampaikan. Kemudian, diikuti dengan 'Meter Kebenaran' berdasarkan empat kriteria penarafan iaitu benar, bercampur, tidak pasti dan tidak benar) (MyCheck Malaysia, 2020).

Laman web *faqcheck.org* merupakan salah satu inisiatif pengesahan fakta selain *sebenarnya.my* dan *mycheck.my*. Ia ditubuhkan oleh aktivis media di Malaysia sebagai perkhidmatan awam yang bebas dengan menggunakan IFCN sebagai panduan untuk mengatasi penularan maklumatpalsu. Laman web ini mengutamakan maklumat yang merangkumi kesihatan, keselamatan awam, keharmonian masyarakat dan hak asasi manusia untuk disemak terlebih dahulu. Bagi mengelakkan penyebaran khabar angin dan menerbitkan laporan bertanggungjawab, *faqcheck.org* hanya menyemak maklumat yang mencapai *tipping point*. Maklumat yang disemak oleh laman web ini diberi penilaian (palsu, benar, campuran, tiada bukti, di luar konteks, pengaitan palsu atau satira) di bahagian kesimpulan artikel yang dimuatnaik. Kemudian, ia diikuti dengan teg semakan sama ada disinformasi, misinformasi atau malinformasi (Faqcheck Lab, 2020).

## 2.6 Perlombongan Teks (*Text Mining*)

Perlombongan teks adalah suatu kaedah yang digabungkan dalam bidang penyelidikan seperti linguistik *Information Retrieval* (IR) dan pengiraan data (Said et al., 2017). Dalam konteks ini, perlombongan teks tidak terhad dan hanya boleh digunakan untuk penyelidikan dalam bidang kejuruteraan, sains komputer dan matematik. Malah, ia boleh digunakan untuk bidang sastera atau sains sosial seperti bahasa dan komunikasi. Namun begitu, setiap analisis perlombongan teks yang dijalankan adalah berbeza dari segi prosesnya kerana ia mengikut kepada bidang penyelidikan tersebut. Contohnya analisis perlombongan teks bagi bidang sains komputer, ia dilihat lebih kompleks dan praktikal.

Kenyataan di atas turut disokong oleh Dan et. al. (2020) yang menjelaskan perlombongan teks adalah teknik analisis data dengan tujuan untuk mengekstrak pengetahuan dan hubungan dari dokumen yang dipilih sebagai sampel. Kaedah perlombongan teks turut dianggap sebagai alat pintar dan amat sesuai digunakan oleh penyelidik untuk menyelesaikan masalah seperti peramalan, pengkelasan, pengelompokan dan sebagainya. Bagi menyelesaikan masalah tersebut, beberapa teknik perlombongan data yang boleh digunakan. Antaranya, pengelompokan hieraki, pepohon keputusan, teknik rangkaian neural dan peraturan kesatuan (Mahadi et. al., 2006).

Teks merupakan medium terpenting yang digunakan dalam perkongsian data. Secara amnya, teks telah terbahagi kepada dua jenis iaitu berstruktur dan tidak berstruktur (Sharma & Srivastana, 2016). Perlombongan teks merupakan kaedah yang boleh menganalisis teks atau data yang tidak berstruktur. Malah, perlombongan teks juga mampu menganalisis data yang tidak dilabelkan oleh penyelidik. Dalam konteks

ini, data yang dilabel boleh menjalani analisis perlombongan teks dengan melalui *supervised learning technique*. *Supervised learning technique* merupakan kaedah yang terdapat pada perlombongan teks dengan tujuan mengelompokkan data megikut ciri-ciri tertentu.

Perlombongan teks turut berperanan sebagai kaedah yang mampu mengkaji dan memahami fenomena sosial (Ignatow & Mihalcea, 2017). Dalam hal ini, kaedah perlombongan teks dapat merungkai dan menemukan perhubungan antara data, sekaligus membantu penyelidik meramal keadaan pada masa hadapan berdasarkan dapatan dari kajian-kajian terdahulu. Misalnya, kaedah perlombongan teks telah digunakan dalam bidang kesihatan awam serta pengurusan krisis untuk memahami dan mengesan penyebaran bencana alam dan wabak yang berlaku (Ampofo et. al., 2013).

## 2.7 Kerangka Teoretikal

### 2.7.1 Model Six-Step KDP

*Six-Step Knowledge Discovery Process* (KDP) ialah model yang dibangunkan oleh Krzysztof J. Cios, seorang pengerusi di Jabatan Sains Komputer, Virginia Commonwealth University. Model Six-Step KDP dibangunkan pada tahun 2010 untuk mencari pengetahuan baru mengenai sesuatu domain. Ia dikembangkan dari model terdahulu iaitu *Cross-Industry Process for Data Mining* (CRISP-DM). CRISP-DM ditubuhkan oleh Integral Solution Ltd., NCR, DaimlerChrysler dan OHRA pada tahun 1996 untuk memberi panduan kepada sektor industri mengenai perlombongan data. Ini bermakna, CRISP-DM adalah satu standar penganalisan dalam bidang perindustrian sebagai strategi pemecahan masalah. Selain itu, CRISP-DM telah mendokumentasikan langkah secara teliti bagi perlombongan data dengan menggunakan bahasa yang mudah difahami (Cios et. al., 2007).