

# ANALISIS DESKRIPTIF DAN PERLOMBONGAN PETUA SEKUTUAN UNTUK COVID-19 SEMASA VARIAN DELTA

Nazifa Adriana Binti Nazarrudin<sup>1</sup>

Zalinda Othman

<sup>1</sup>*Fakulti Teknologi & Sains Maklumat, Universiti Kebangsaan Malaysia, 43600 UKM Bangi, Selangor Darul Ehsan, Malaysia*

## Abstrak

Penyakit yang berasal dari Wuhan, China dikenali sebagai COVID-19 seperti novel coronavirus-Severe Acute Respiratory Syndrome-Corona Virus2 (SARS-CoV2). Penyakit ini boleh menyebabkan jangkitan pada sistem saluran pernafasan melalui penularan. Dengan peningkatan bilangan pesakit yang menghidap COVID-19, jelas sekali penyelesaian pantas untuk membantu organisasi kesihatan mengatasi masalah ini diperlukan. Pendekatan pembelajaran mesin digunakan untuk membangunkan model mesin dengan ciri-ciri data pesakit. Dalam kajian ini, tumpuan diberikan kepada daerah Hulu Langat di Malaysia di mana data diperolehi daripada Pusat Kesihatan Daerah (PKD) Hulu Langat sebanyak 161,556 rekod namun hanya data pada tahun 2021 sahaja digunakan. Kajian ini menggunakan kaedah analisis deskriptif dan perlombongan petua sekutuan bagi mencari hubungan antara komorbiditi, gejala dan hasil (pulih/meninggal dunia). Serta mengetahui sama ada terdapat sebarang corak antara status (kuarantin/ward) pesakit dengan pemulihan dan kematian. Hasil kajian ini mendapati hasil analisis bagi atribut umur, status, mukim dan jantina. Umur yang terendah adalah 0 tahun, menunjukkan jangkitan pada usia tidak sampai satu tahun (bayi dalam usia harian atau bulan terlibat). Kategori dewasa merupakan kategori yang paling tinggi dijangkiti Covid-19. Status tertinggi adalah kuarantin dan mukim tertinggi adalah Kajang. Atribut jantina mempunyai hubungan dengan atribut mukim dan jenis penyebaran. Menunjukkan perkaitan antara dua atribut. Perlombongan petua sekutuan berjaya menjana peraturan bagi algoritma Apriori dan algoritma FP-Growth dengan nilai parameter sokongan, keyakinan dan lif. Peraturan daripada FP-Growth dipilih kerana ianya lebih banyak menjana peraturan dan lebih menjimatkan masa.

Satu papan pemuka dihasilkan dengan tiga fungsi iaitu dua daripada objektif dan satu fungsi tambahan untuk melihat bilangan pemulihan atau kematian bagi setiap bulan pada tahun 2021.

**Kata kunci: COVID-19, PKD, Apriori, FP-Growth**

## Pengenalan

Penyakit berjangkit sedang menular di seluruh dunia dengan impak besar yang memberi kesan kepada kesihatan, ekonomi dan kehidupan sosial. Dunia dikejutkan dengan berita tentang penyakit yang merebak dengan pantas yang boleh menyebabkan kematian pada tahun 2019. Wuhan, China adalah asal usul COVID-19 yang disebabkan oleh virus yang dinamakan sebagai Coronavirus-Severe Acute Respiratory Syndrome-Corona Virus2 (SARS-CoV2). Wabak itu adalah dari pasar tempatan, 'Pasar Makanan Laut Huanan' dengan sekurang-kurangnya 41 orang dilaporkan menghidap penyakit itu. The World Health Organization (WHO) mengisytiharkan wabak COVID-19 pada 30 Januari 2020, sebagai kecemasan kesihatan awam yang membimbangkan peringkat antarabangsa. Penyakit ini merebak melalui penularan seperti jangkitan pernafasan iaitu daripada titisan air yang berlainan saiz di mana seseorang boleh berjangkit apabila bersentuhan rapat (dalam jarak 1 meter) yang mempunyai gejala COVID-19. Batuk, bersin, menyanyi, bercakap atau terdedah kepada titisan air daripada pesakit boleh menjejaskan orang lain. Orang ramai juga boleh dijangkiti dengan menyentuh mata, mulut atau hidung mereka sendiri tanpa mencuci tangan selepas menyentuh permukaan yang tercemar. Gejala yang paling biasa dilaporkan ialah demam, batuk kering, selsema dan keletihan. Berdasarkan COVIDNOW di Malaysia yang merupakan laman web rasmi kerajaan Malaysia yang memainkan peranan dalam perkongsian data dan pandangan tentang COVID-19 yang telah dan sedang berlaku di negara ini, terdapat kira-kira 86.1% rakyat Malaysia telah mengambil dos pertama, 84.2% daripada dos kedua, 49.7% penggalak pertama dan 1.6% daripada penggalak

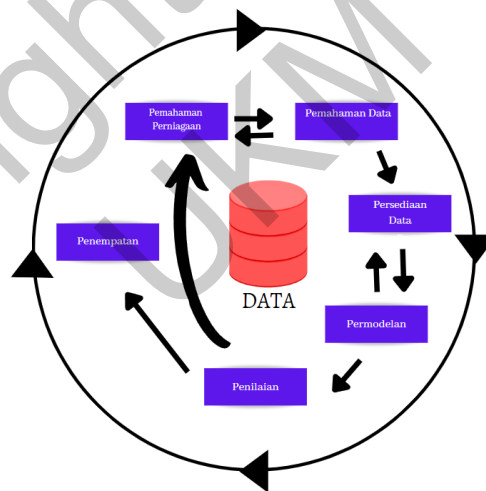
kedua. Sehingga 5 Oktober 2022, 0.1% berada di ICU, 95.5% dalam kuarantin di rumah dan 4.2% berada di hospital. Ini menunjukkan bahawa walaupun dengan kewujudan vaksin, COVID-19 juga boleh dijangkiti dan perlindungan diri yang tinggi diperlukan terutamanya bagi orang tua dan juga individu yang mempunyai penyakit.

Pendekatan yang pantas dan cekap diperlukan untuk membantu organisasi kesihatan dalam mengatasi masalah agar dapat membantu pesakit yang menghidap COVID-19. Teknik analisis deskriptif digunakan dalam kajian ini untuk mengenal pasti hubungan agar dapat mencapai objektif. Analisis deskriptif merupakan proses dimana data baharu dan data yang lepas diambil bagi tujuan mengenal pasti trend dan hubungan antara data-data tersebut. Selain itu, perlombongan petua sekutuan turut digunakan dalam kajian ini dimana ia digunakan bagi membentangkan hubungan antara dua perkara. Hulu Langat masih menjadi lima daerah teratas yang mempunyai jumlah kematian tertinggi akibat COVID-19 sejak bermulanya penyebaran sehingga kini. Dengan adanya masalah begini, penyelesaian diperlukan agar dapat membantu untuk mengetahui corak hubungan yang memainkan peranan dalam COVID-19 dimana adakah komorbidity, gejala dan hasil mempunyai kaitan dengan jangkitan COVID-19. Perlombongan data digunakan untuk mencari data masa lalu. Data akan dianalisis, dan trend boleh didapati sebagai keputusan pada akhir kajian. Perlombongan petua sekutuan boleh digunakan. Corak status dengan pemulihan atau kematian harus disiasat. Oleh itu, set data yang merangkumi ciri butiran untuk menentukan hasil diperlukan.

Terdapat dua objektif bagi kajian ini iaitu untuk mengetahui antara komorbidity, gejala dan keputusan (pulih/meninggal dunia). Seterusnya, untuk menyiasat jika terdapat sebarang corak antara status (kuarantin/ward) dengan pemulihan atau kematian. Bagi skop penyelidikan pula, tiada data terkumpul daripada orang ramai secara tinjauan kerana data tersebut diterima daripada Pejabat Kesihatan Daerah Hulu Langat, Selangor. Tarikh data yang digunakan adalah dari 1 Januari 2021 hingga 31 Disember 2021 yang mengandungi semua data pesakit COVID-

19 dari daerah Hulu Langat. Terdapat banyak kekangan daripada kajian ini. Yang pertama ialah lokasi kajian. Sumber data adalah dari daerah Hulu Langat. Jadi, hanya data penduduk di Hulu Langat yang akan dikaji. Kedua, data diterima daripada Pejabat Kesihatan Daerah Hulu Langat di Selangor, tiada data yang diperolehi daripada sumber berbeza. Akhir sekali, data adalah mengenai pesakit yang telah didiagnosis dengan COVID-19 sahaja. Sekiranya pesakit tidak mempunyai penyakit COVID-19 maka data pesakit tidak disertakan dalam kajian.

Metodologi atau kaedah yang digunakan untuk mencapai objektif kajian adalah CRISP-DM iaitu 'Cross Industry Standard Process for Data Mining' yang merupakan satu kaedah yang terdiri daripada 6 fasa seperti rajah 1. Kaedah ini sesuai digunakan kerana melibatkan teknik perlombongan data yang dilaksanakan seperti kajian ini. CRISP-DM merupakan garis panduan sepanjang perjalanan proses. Ianya selalu digunakan semasa melakukan proses yang melibatkan sains data.



Rajah 1 : 6 Fasa dalam CRISP-DM

Terdapat perbezaan antara kajian lepas dengan kajian analisis deskriptif untuk COVID-19 semasa varian delta. Kajian lepas yang pertama adalah kajian lepas yang menggunakan kaedah SVM bertajuk 'Machine learning and comorbidity network analysis for hospitalized

patients with COVID-19 in a city in Southern Brazil’, ia merupakan kajian daripada Passarelli-Araujo et al. 2022. Kajian lepas ini juga menggunakan model Hutan Rawak. Bagi kajian lepas daripada Almalki et al. 2022 bertajuk ‘Regression Analysis for COVID-19 Infections and Deaths Based on Food Access and Health Issues’, ia menggunakan model Mesin Vektor Sokongan dan juga Hutan Rawak. Kajian lepas ketiga adalah kajian lepas daripada Ong et al. 2022 bertajuk ‘Utilization of Random Forest Classifier and Artificial Neural Network for Predicting Factors Influencing the Perceived Usability of COVID-19 Contact Tracing “MorChana” in Thailand’, model yang digunakan adalah Hutan Rawak. Seterusnya, kajian lepas daripada Bhatnagar et al. 2020 bertajuk ‘Descriptive analysis of COVID-19 patients in the context of India’, menggunakan model Pohon Keputusan. Bagi kajian lepas yang kelima iaitu kajian Khan et al. 2022 bertajuk ‘Deep Learning and Ensemble Learning Mechanism for Delta-Type COVID-19 Detection’, ia menggunakan gambaran X-ray sebagai input kajian untuk mengesan kehadiran COVID-19. Dua kajian terakhir adalah kajian Tandan et al. 2021 bertajuk ‘Discovering symptom patterns of COVID-19 patients using association rule mining’ dan Arreras et al. 2019 bertajuk ‘Association Rule Mining Tourist-Attractive Destinations for the Sustainable Development of a Large Tourism Area in Hokkaido Using Wi-Fi Tracking Data’ yang merupakan kajian lepas yang menggunakan kaedah perlombongan petua sekutuan. Daripada ketujuh-tujuh kajian lepas ini, kajian yang hampir sama dengan kajian analisis deskriptif dan perlombongan petua sekutuan untuk COVID-19 semasa varian delta adalah kajian ke-empat daripada Bhatnagar et al. 2020 dan kajian ke-enam oleh Tandan et al. 2021. Bagi kajian Bhatnagar et al. 2020, objektif yang dinyatakan adalah juga hampir sama dengan kajian ini namun berbeza dari segi kawasan dan kaedah.

## Metodologi Kajian

Keperluan perisian yang digunakan seperti Google Colab, Tableau dan Sublime Text. Windows 10 dan Google Chrome turut digunakan bagi pencarian maklumat sehingga ke persiapan model. Penggunaan perisian Sublime Text yang menggabungkan bahasa pengaturcaraan python bersama dengan reka bentuk web HTML menjadi sasaran terakhir dimana ianya akan digunakan di akhir kajian bagi penggabungan kajian dengan perisian Tablau. The Cross Industry Standard Process for Data Mining (CRISP-DM) terdiri daripada 6 fasa iaitu pemahaman perniagaan, pemahaman data, persediaan data, permodelan, penilaian dan penempatan. Pelaksanaan implementasi atau pelan pelaksanaan dilaksanakan mengikut kitaran Crisp-DM. Pelaksanaan implementasi di lakukan bagi meneruskan pembinaan atau mendedahkan perjalanan pengekodan bagi melengkapkan projek. Bagi projek ini, pengekodan telah bermula daripada fasa pra-pemprosesan dimana dengan kod dan perpustakaan yang sesuai pelaksanaan projek berjaya dilakukan. Dalam pra-pemprosesan terdapat beberapa langkah pembersihan yang dilaksanakan seperti fasa pemilihan data, fasa pembersihan, fasa penyatuan data dan fasa transformasi.

Fasa pemahaman perniagaan adalah fasa pertama dalam CRISP-DM. Sefahaman perlu dicapai agar dapat melaksanakan objektif dengan sempurna. Projek atau kajian yang dijalankan ini merupakan kajian bersama dengan Pusat Kesihatan Daerah (PKD) Hulu Langat di mana tujuan utamanya adalah berkaitan dengan penyakit berjangkit yang berbahaya dikenali sebagai COVID-19. Seterusnya, fasa pemahaman data merupakan proses dimana penambahan daripada pemahaman perniagaan. Ia lebih tertumpu kepada mengenalpasti, mengumpul dan menganalisa set data bagi menjalankan kajian yang ingin dicapai. Data yang didapati bagi kajian ini adalah sebanyak 161,556 rekod data mentah namun pemilihan data yang dijalankan hanya melibatkan data selama satu tahun sahaja dibandingkan dengan data mentah diterima melibatkan tiga tahun, hasil akhir didapati sebanyak 104,923 sahaja digunakan. Fasa ketiga adalah fasa persediaan data melibatkan empat

peringkat iaitu pemilihan, pembersihan, mengintegrasikan dan transformasi data. Bagi proses pemilihan, hanya data yang terlibat sahaja akan digunakan. Bagi kajian ini untuk mencari hubungan berkaitan komorbiditi, kategori dan hasil diagnosis serta corak antara tahap kategori yang lebih tinggi dengan pemulihan atau kematian, Terdapat kira-kira 55 atribut yang diperolehi bagi tujuan kajian ini. Namun hanya lapan atribut yang dipilih untuk digunakan. Pilihan kelapan-lapan atribut ini berdasarkan kepentingan kajian ini dan hanya atribut ini yang akan memberi keputusan yang diinginkan.

Seterusnya, fasa pembersihan dimana pengendalian nilai hilang, pengecualian nilai hilang bagi atribut 'id\_pesakit', pengendalian nilai tidak konsisten dan penukaran kepada huruf kecil. Fasa integrasi data adalah fasa di mana penggabungan fail yang diterima daripada pihak PKD sebanyak empat fail. Digabungkan menjadi hanya satu fail. Fail 1 mempunyai data daripada 20 Oktober 2020 sehingga 26 April 2021, fail 2 daripada 27 April 2021 sehingga 5 Mei 2021 dan fail 3 mempunyai data daripada 6 Mei 2021 sehingga 7 Mac 2022. Tiga fail pertama ini adalah merupakan fail data pesakit covid-19 di Hulu Langat dari 2020 sehingga 2022, namun fail ini telah disatukan dengan penggunaan aplikasi Excel tanpa melibatkan pengkodan dan hanya mengambil pesakit dari tahun 2021 sahaja. Manakala fail ke-empat ialah fail data pesakit covid-19 yang meninggal dunia. Sekiranya terdapat sebarang persamaan daripada atribut 'id\_pesakit' di dalam kedua-dua fail, kemudian ianya akan di labelkan kepada data meninggal di atribut baru yang di namakan sebagai 'Keputusan'. Jika tiada persamaan maka ianya di labelkan sebagai pulih. Terakhir adalah transformasi data. Format data amat penting dalam menjamin kualiti data yang baik. Kesemua atribut di pilih ditukarkan kepada nilai bernombor.

Fasa ke-empat dalam CRISP-DM adalah pemodelan. Terdapat dua jenis algoritma digunakan bagi permodelan yang terdiri dalam perlombongan petua sekutuan iaitu algoritma apriori dan algoritma FP-Growth. Algoritma apriori mencari kekerapan set item di dalam set data untuk perlombongan petua sekutuan. Algoritma apriori boleh digunakan dalam python. Ia merupakan algoritma pembelajaran mesin yang digunakan untuk memahami hubungan hierarki antara pelbagai

objek yang terlibat. Seterusnya adalah algoritma FP-Growth. Ia juga boleh menggunakan bahasa pengaturcaraan Python. Seterusnya fasa penilaian di mana tiga jenis ukuran di ambik iaitu sokongan, keyakinan dan lif. Ketiga-tiga penilaian ini akan digunakan bagi menilai permodelan perlombongan petua sekutuan yang dicipta. Setelah model dinilai, ia akan ditentukan semula sama ada dimulakan semula pada fasa pertama iaitu pemahaman perniagaan atau diteruskan ke peringkat akhir iaitu penempatan projek. Sekiranya kajian di mulakan semula ke fasa pertama, bermaksud terdapat beberapa aspek yang perlu diperbetulkan semula. Kajian ini akan diakhiri dengan penempatan projek. Projek ini akan dimasukkan ke papan pemuka yang akan digunakan bagi memvisualisasi hasil kajian ini. Pihak PKD akan menggunakan kajian analisis deskriptif dan perlombongan petua sekutuan untuk COVID-19 semasa varian delta ini di dalam perisian tersebut. Namun kajian ini akan terus dipantau dan sekiranya data baharu diterima, penambahan akan dilakukan dan corak kajian akan mengalami perubahan mengikut data semasa. Akhir sekali, sebuah laporan akhir akan dilengkapkan beserta dengan keputusan hasil dapatan.

### **Keputusan dan Perbincangan**

Implementasi merupakan sebuah langkah yang dilaksanakan apabila perlaksanaan proses-proses sebelumnya telah selamat atau lengkap dijalankan seperti mengeksplorasi data, pembersihan data dan permodelan sejajar dengan kitaran CRISP-DM. Ini merupakan fasa implementasi projek dalam bentuk dokumen yang menerangkan secara terperinci mengenai proses keputusan yang terlibat seperti permodelan dan antara muka serta pendokuman tentang peraturan-peraturan yang ditetapkan atau ditentukan dari proses sebelumnya. Pelan pelaksanaan ialah manual langkah-langkah yang perlu dilakukan untuk mencapai matlamat tersebut. Matlamat akan dicapai jika pelan strategik menentukan teknik yang akan digunakan (Asana, T., 2022). Dengan menggariskan prosedur tepat yang mesti diambil, pelan pelaksanaan memudahkan untuk melaksanakan pelan, idea, model, reka bentuk, spesifikasi, standard,



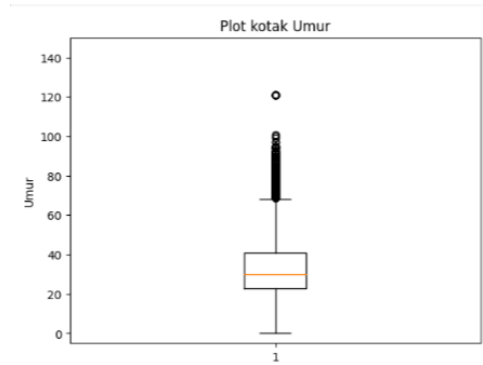
algoritma atau dasar. Maka, pelan pelaksanaan ialah senarai langkah yang mesti diikuti untuk pencapaian matlamat dengan cara yang terperinci (team,E.,2022).

Bagi pengujian di fasa pra-pemprosesan data daripada data mentah yang terdiri daripada 161,556 hanya 104,923 yang digunakan merujuk kepada jadual 1.

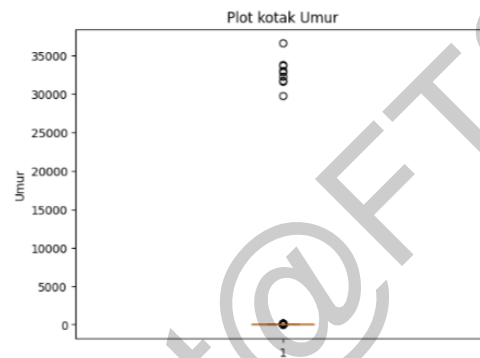
Jadual 1: Penerangan mengenai data setelah fasa pra-pemprosesan

Atribut/ Pembolehubah	Data	Maksud data
Umur	Muda, dewasa, tua	Kategori setiap pesakit
Komorbidity	Komorbidity:0, Komorbidity:1	Komorbidity:0=Tiada komorbidity Komorbidity:1=Ada komorbidity
Jantina_num	Jantina:0, Jantina:1	Jantina:0 = Perempuan Jantina:1 = Lelaki
Mukim_num	Mukim:0, Mukim:1, Mukim:2, Mukim:3, Mukim:4, Mukim:5	Mukim:0 = Kajang Mukim:1 = Semenyih Mukim:2 = Cheras Mukim:3 = Ampang Mukim:4 = Beranang Mukim:5 = Hulu langat
Gejala_num	Gejala:0, Gejala:1	Gejala:0 = Tiada gejala Gejala:1 = Ada gejala
Jenis_penyebaran_num	Jenis_penyebaran:0, Jenis_penyebaran:1	Jenis_penyebaran:0=Lokal Jenis_penyebaran:1=Import
Status_num	Status:0, Status:1	Status:0 = Kuarantin Status:1 = Warded
Keputusan_num	Keputusan:0, Keputusan:1	Keputusan:0 = Pulih Keputusan:1=Meninggal dunia

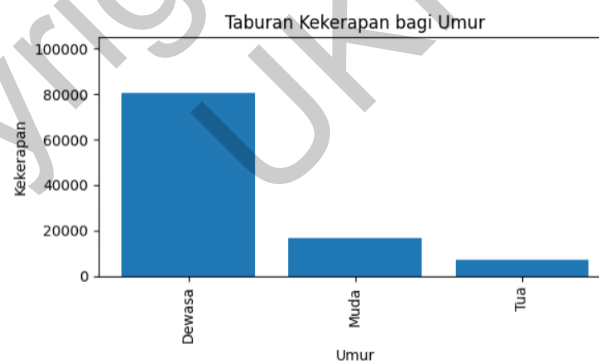
Rajah 2 merupakan visual plot kotak bagi analisis umur yang dihadkan kepada 100. Manakala, Rajah 3 mewakili data yang sama tetapi tanpa had umur. Membenarkan semua umur pesakit dimasukkan. Rajah 3 memaparkan sejumlah besar outlier tanpa mengandungi kuartil kerana set data mempunyai nilai yang banyak apabila sekatan atau had umur tidak digunakan manakala rajah 4 menunjukkan taburan kekerapan bagi umur dengan nilai 'Dewasa' tertinggi.



Rajah 2 : Hasil output analisis umur terhad sehingga 105



Rajah 3 : Hasil dan output plot kotak tanpa had



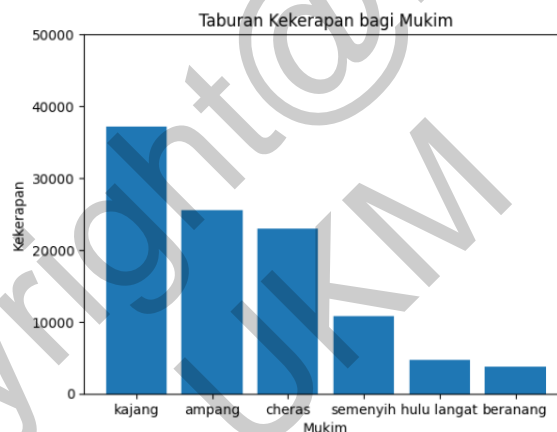
Rajah 4 : Taburan kekerapan bagi umur

Berdasarkan analisis, minimum umur bagi pesakit Covid-19 adalah '0' kerana bayi yang tidak mencecah umur setahun di masukkan sebagai kosong tahun manakala maksimum umur adalah 100 tahun. Jadual 2 merupakan nilai plot kotak dengan penggunaan had. Outlier dalam set data ini mempunyai umur antara 79 hingga 86 dan seterusnya. Ini adalah titik data yang menyimpang dengan ketara daripada majoriti data dan boleh menunjukkan nilai yang melampau.

Jadual 2: Nilai plot kotak

Parameter	Nilai penggunaan had 100
Minimum	0
Misai rendah ( <i>Lower Whisker</i> )	0.0
Misai tinggi ( <i>Upper Whisker</i> )	68.0
Terhad atas ( <i>Upper limit</i> )	68
Terhad bawah ( <i>Lower limit</i> )	0
Median	30
Maksimum	100
Outliers	[79 83 73 ... 77 71 86]

Berdasarkan rajah 5, analisis mukim tertinggi adalah kajang yang mencatatkan jumlah tertinggi diikuti oleh ampang, cheras, semenyih, hulu langat dan beranang.



Rajah 5 : Hasil output kod analisis mukim

Analisis jantina digunakan bagi mengkaji hubungan antara jantina dan jenis penyebaran serta jantina dan mukim. Penggunaan khi-kuasa dua dan nilai-p memerlukan hipotesis nol. Dengan merumuskan hipotesis nol, keperluan menyediakan ujian statistik untuk menentukan sama ada terdapat cukup bukti dalam data untuk menolak hipotesis nol yang memihak kepada hipotesis alternatif. Nilai-p, yang diperolehi daripada ujian khi kuasa dua, akan membantu dalam membuat penentuan ini. Hipotesis nol bagi hubungan adalah seperti berikut:

- i) Tiada hubungan antara jantina dan jenis penyebaran
- ii) Tiada hubungan antara mukim dan jantina

Jadual 3 mempunyai nilai berkaitan hubungan antara jenis penyebaran dan jantina. Menunjukkan jumlah statistik khi kuasa dua dan nilai-p yang berjaya dijana. Berdasarkan keputusan, statistik khi kuasa dua ialah 36.05 dan nilai-p ialah  $1.9200283862484957e-09$ . Nilai-p mewakili kebarangkalian memperoleh perbezaan yang diperhatikan atau yang lebih ketara jika hipotesis nol adalah benar. Dalam kes ini, hipotesis nol ialah dua pembolehubah kategori, "Jenis\_penyebaran" dan "Jantina". Oleh kerana nilai-p adalah kurang daripada 0.05 iaitu nilai bagi tahap alfa biasa yang digunakan untuk menentukan hubungan. Berdasarkan hipotesis nol, ianya boleh ditolak dan kesimpulannya bahawa terdapat perkaitan yang signifikan secara statistik antara kedua-dua pembolehubah. Kesimpulannya, berdasarkan nilai khi kuasa dua dan nilai p yang diperolehi, menyatakan bahawa terdapat hubungan yang signifikan antara jantina dan jenis penyebaran, dan hipotesis nol tiada hubungan boleh ditolak.

Jadual 3 juga mempunyai nilai hubungan antara mukim dan jantina. Jumlah statistik khi kuasa dua dan nilai-p yang berjaya dijana. Berdasarkan hasil output ujian khi kuasa dua, nilai statistik khi kuasa dua ialah 593.97 dan nilai p adalah sangat kecil ( $4.0608401935983715e-126$ ), yang menunjukkan terdapat hubungan yang signifikan antara kategori mukim dan jantina dalam set data. Oleh itu, ianya boleh menolak hipotesis nol iaitu tiada hubungan antara kedua-dua pembolehubah dan membuat kesimpulan bahawa terdapat hubungan yang signifikan antara mukim dan Jantina.

Jadual 3 : Rumusan analisis jantina

Hipotesis Nol	Khi Kuasa Dua	Nilai-p	Terima atau tolak null hipotesis
Tiada hubungan antara jantina dan jenis penyebaran	36.05	$1.92e-09$ (atau $0.00000000192$ )	Tolak
Tiada hubungan antara mukim dan jantina	593.97	$4.0608401935983715e-126$	Tolak

Penggunaan algoritma Apriori telah berjaya menjana sebanyak 100. Namun, dalam 100 peraturan tersebut, tidak ada yang mempunyai nilai keputusan kematian. Maka, kaedah persampelan berlebihan digunakan. Kaedah persampelan berlebihan (*oversampling*) turut digunakan bagi mengatasi risiko ketidakseimbangan set data. Terdapat sebanyak 3,399 peraturan dijana.

Namun, apabila menggunakan kod persampelan berlebihan dalam algoritma apriori, terdapat jumlah perbezaan yang berlebihan sekiranya di bandingkan dengan tanpa penggunaan persampelan berlebihan iaitu perbezaan sebanyak 3,299. Persampelan berlebihan mempengaruhi metrik evaluasi seperti sokongan, keyakinan dan lif. Dalam ujian ini, dapat dilihat bahawa algoritma apriori tanpa persampelan berlebihan menghasilkan nilai yang lebih tinggi untuk metrik tersebut. Ini menunjukkan bahawa menggunakan data asli tanpa modifikasi dapat menghasilkan asosiasi yang lebih kuat dan signifikan. Selain itu, kaedah persampelan berlebihan tidak dapat mengatasi isu ketidakseimbangan di mana keputusan meninggal dunia turut tidak dijana seperti algoritma tanpa penggunaan persampelan berlebihan.

Manakala penggunaan algoritma FP-Growth telah berjaya menjana sebanyak 11,861 peraturan. Berdasarkan penjaan kedua-dua algoritma dimana apriori telah menjana 100 peraturan manakala frequent pattern growth menjana sebanyak 11,861. Daripada peraturan algoritma apriori yang terdiri daripada 100 peraturan, hanya 47 yang melibatkan keputusan bahawa pesakit pulih daripada covid-19 dan sifar bagi kematian. Manakala untuk 11,861 peraturan yang dijana oleh algoritma apriori, terdapat 4,855 peraturan berkaitan dengan pesakit pulih dan sifar bagi kematian. Jadual 4 menunjukkan nilai sokongan, keyakinan dan lif daripada algoritma apriori manakala jadual 5 menunjukkan nilai sokongan, keyakinan dan lif daripada algoritma FP-growth.

Jadual 4 : Nilai sokongan, keyakinan dan lif bagi algoritma apriori

Anteseden	Akibat	Sokongan	Keyakinan	Lif
Gejala:0	Keputusan:0	0.7838	0.9979	0.9998
Komorbid:0	Keputusan:0	0.9252	0.9981	1.0000
Status:1	Keputusan:0, Jenis_penyebaran:0	0.8604	0.995	1.0000
Gejala:0,Komorbid:0, Jenis_penyebaran:0	Keputusan:0	0.7473	0.9979	0.9998
Status:1,Gejala:0, Komorbid:0	Keputusan:0	0.6743	0.9978	0.9996
Jantina:1,Komorbid:0, Jenis_penyebaran:0	Keputusan:0	0.5731	0.9981	0.9999
Status:1,Jantina:1, Komorbid:0	Keputusan:0	0.5014	0.9980	0.9999
Status:1,Komorbid:0, Jenis_penyebaran:0	Keputusan:0	0.8024	0.9981	0.9999
Status:1,Gejala:0, Komorbid:0	Keputusan:0, Jenis_penyebaran:0	0.6722	0.9947	1.0002
Status:1, Gejala:0	Keputusan:0, Komorbid:0, Jenis_penyebaran:0	0.67224	0.9555	1.0365

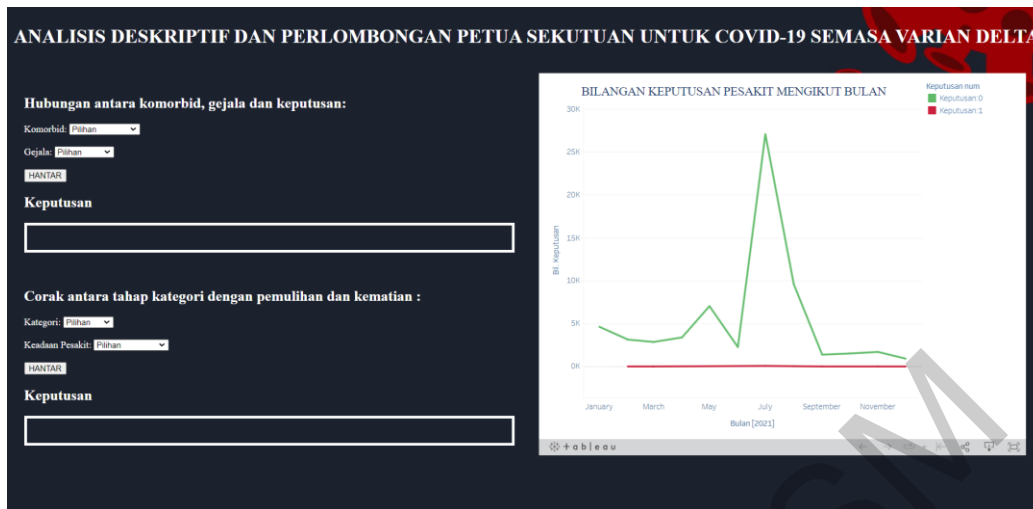
Berdasarkan jadual 4 di atas, daripada peraturan ke-5 menunjukkan anteseden data iaitu Status:1, Gejala:0 dan Komorbid:0 manakala akibatnya adalah Keputusan:0. Daripada peraturan ini, nilai sokongan adalah 0.6743, keyakinan 0.9978 dan lif 0.9996. Dalam erti lain, sokongan merupakan kekerapan item set berlaku dalam set data. Manakala keyakinan mengukur kekerapan akibat (Keputusan:0) muncul dalam transaksi yang mengandungi anteseden (Status:1, Gejala:0, Komorbid:0). Ia dikira sebagai nisbah sokongan keseluruhan peraturan iaitu kedua-dua anteseden dan akibat kepada sokongan anteseden. Dalam kes ini, peraturan mempunyai keyakinan kira-kira 99.78%, yang bermaksud bahawa hampir semua urus niaga yang mengandungi anteseden juga mengandungi akibat (Keputusan:0). Keyakinan bermaksud nilai keyakinan (99.78%) dalam peraturan yang dijana menunjukkan bahawa apabila anteseden (Status:1, Gejala:0, Komorbid:0) diperhatikan dalam urus niaga, terdapat kemungkinan yang sangat tinggi bahawa akibat (Keputusan:0) akan turut hadir. Nilai lif (0.9996) menghampiri 1

menunjukkan bahawa hampir tiada korelasi antara anteseden dan akibat, atau mereka berkorelasi positif secara lemah. Ini bermakna bahawa kehadiran (Status:1, Gejala:0, Komorbid:0) tidak memberi kesan yang ketara kepada akibat (Keputusan:0), kerana tingkatan adalah sangat hampir dengan 1.

Jadual 5 : Nilai sokongan, keyakinan dan lif bagi algoritma FP-Growth

Anteseden	Akibat	Sokongan	Keyakinan	Lif
Komorbid:0	Keputusan:0	0.9269	0.9981	0.9252
Status:1	Keputusan:0	0.8651	0.9981	0.8634
Gejala:0	Keputusan:0	0.7855	0.9981	0.7838
Gejala:1	Keputusan:0	0.0109	0.9981	0.0109
Komorbid:1	Keputusan:0	0.0731	0.9981	0.0729
Status:0, Mukim:1	Keputusan:0	0.0189	0.9981	0.0189
Komorbid:0, Gejala:0	Keputusan:0	0.7514	0.9981	0.7498
Komorbid:1, Gejala:0	Keputusan:0	0.0341	0.9981	0.0340
Status:1	Jenis_penyebaran:0, Keputusan:0	0.8651	0.9946	0.8605
Status:1	Keputusan:0	0.8651	0.9252	0.8053

Berdasarkan implementasi projek, dapat disimpulkan bahawa iainya berjaya dilaksanakan seperti jangkaan. Namun perbandingan antara algoritma apriori dan algoritma frequent pattern growth menunjukkan bahawa algoritma frequent pattern growth merupakan algoritma yang paling terbaik yang boleh di gunakan bagi meneruskan projek ini. Berdasarkan nilai keputusan antara tiga ukuran iaitu sokongan, keyakinan dan lif serta penjana peraturan yang lebih banyak berbanding algoritma apriori. Algoritma frequent pattern growth juga merupakan sebuah algoritma yang lebih baik daripada algoritma dari pelbagai segi seperti penjimatan masa. Kesimpulannya, peraturan yang dijana, digunakan di bahagian seterusnya iaitu pelaksanaan papan pemuka projek seperti rajah 6.



Rajah 6 : Papan pemuka utama

Bagi objektif pertama iaitu di bahagian hubungan antara komorbid, gejala dan keputusan.

Peraturan yang dirujuk adalah:

- i) Jika komorbid : Tiada dan gejala : Tiada. Maka, keputusan : Pulih
- ii) Jika komorbid : Ada dan gejala : Tiada. Maka, keputusan : Pulih

Sekiranya tiada di dalam peraturan di atas, ianya dianggap tiada hubungan.

Rajah 7 menunjukkan contoh bagi peraturan pertama di mana apabila pengguna membuat pilihan dan menghantar butang 'HANTAR', keputusan akan dipaparkan di bahagian bawah.



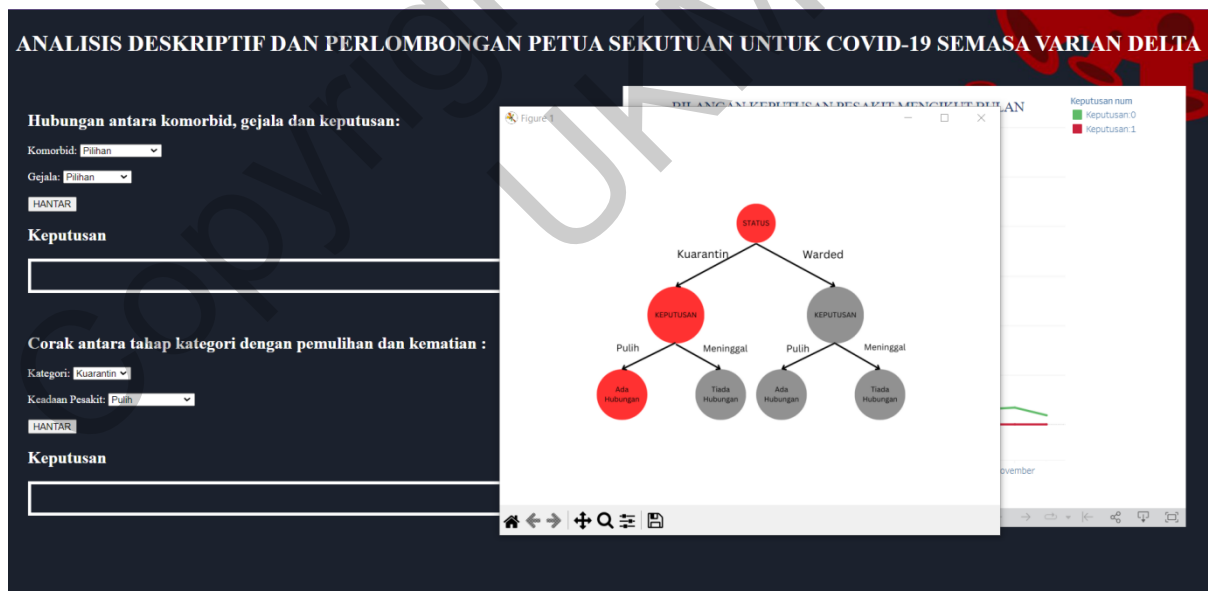
Rajah 7 : Contoh fungsi pertama



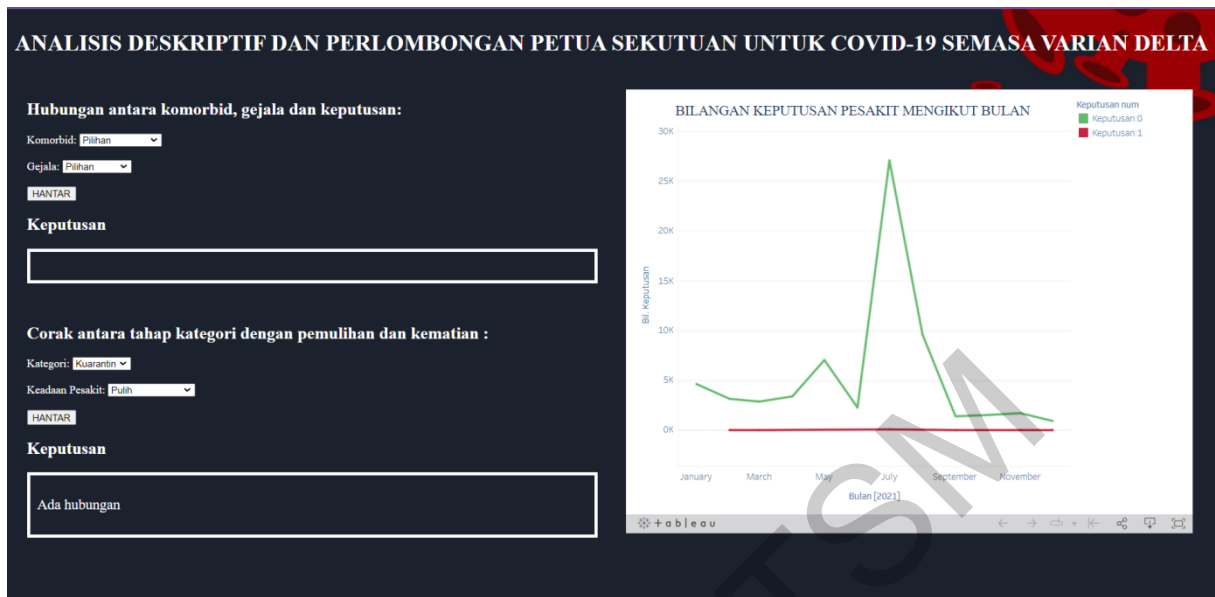
Bagi objektif kedua iaitu corak antara tahap kategori dengan pemulihan dan kematian. Peraturan yang dirujuk adalah:

- i) Jika status : Kuarantin dan keadaan pesakit : Pulih. Maka, ada hubungan.
- ii) Jika status : Kuarantin dan keadaan pesakit : Meninggal dunia. Maka, tiada hubungan.
- iii) Jika status : Warded dan keadaan pesakit : Pulih. Maka, ada hubungan.
- iv) Jika status : Warded dan keadaan pesakit : Meninggal dunia. Maka, tiada hubungan.

Rajah 8 menunjukkan contoh bagi peraturan pertama di mana apabila pengguna membuat pilihan dan menghantar butang 'HANTAR', graf akan dipaparkan menunjukkan aliran hubungan antara status dan keputusan. Rajah 9 adalah apabila tanda pangkah ('X') pada terminal graf ditekan. Keputusan akan dipaparkan.



Rajah 8 : Contoh fungsi kedua (1)



Rajah 9 : Contoh fungsi kedua (2)

Rajah 10 menunjukkan fungsi graf bilangan keputusan pesakit mengikut bulan di bahagian kanan papan pemuka. Apabila pemanah tetikus melalui garis pada graf, beberapa maklumat akan dipaparkan iaitu keputusan data, bulan data dan bilangan data tersebut seperti rajah 10.



Rajah 10 : Fungsi graf

## Kesimpulan

Daripada mulanya bab pertama iaitu bab pengenalan sehingga bab terakhir iaitu bab kesimpulan. Dapat disimpulkan terdapat banyak maklumat yang telah dilengkakan di dalam laporan ini bagi tujuan pengenalan projek, data pembinaan dan penganalisaan data sehingga semua objektif yang di usulkan berjaya dilaksanakan. Projek ini dilaksanakan bagi memenuhi keperluan atau objektif iaitu mencari hubungan antara komorbid, gejala dan keputusan serta corak antara status (kuarantin/ward) dengan pemulihan dan kematian. Kedua-dua objektif ini dicapai dan penggunaan analisis deskriptif dalam menganalisa data dan penggunaan algoritma dalam perlombongan petua sekutuan juga telah berjaya dicapai dalam menjana peraturan yang dikehendaki. Projek ini juga berjaya dilaksanakan dengan penambahan fungsi di dalam papan pemuka yang menunjukkan trend bagi setiap bulan dan keputusan akhir pesakit sama ada pulih atau meninggal dunia.

Projek ini berjaya dilaksanakan dan papan pemuka telah direka dengan fungsi yang ditetapkan. Penggunaan papan pemuka yang mesra pengguna dan berinformasi memudahkan lagi pengguna memahami topik berkaitan COVID-19 yang melanda daerah Hulu Langat pada tahun 2021. Penggunaan algoritma dalam penjanaan peraturan membantu dalam penghasilan akhir setiap fungsi dan melengkapi objektif kajian. Analisis deskriptif yang telah dilaksanakan juga berjaya menunjukkan hubungan yang diinginkan seperti analisis umur yang berjaya mengeluarkan hasil output berkaitan nilai maksimum dan minimum serta penimbunan outliers yang dapat dikesan menggunakan plot kotak. Analisis mukim juga berjaya menunjukkan mukim yang tertinggi dan terendah di antara enam mukim yang terdapat di Hulu Langat. Selain itu, analisis status juga dapat menghasilkan taburan kekerapan berkaitan status pesakit yang tertinggi dan terendah iaitu status bagi kuarantin dan warded. Akhir sekali, analisis jantung juga telah berjaya dilaksanakan di mana dengan analisis ini, dapat diketahui sama ada atribut atau pembolehubah Jantina mempunyai hubungan atau tidak dengan atribut Mukim dan

Jenis\_penyebaran. Kedua-dua hipotesis nol berjaya ditolak dan menyimpulkan bahawa kedua-dua atribut tersebut mempunyai hubungan dengan atribut Jantina.

Kekuatan projek ini adalah di mana ianya terdiri daripada penggunaan dua pemodelan iaitu daripada perlombongan petua sekutuan dikenali sebagai algoritma apriori dan algoritma FP-Growth. Dengan penggunaan kedua-dua model ini, ianya telah dibandingkan dan peraturan yang terbaik telah dijana dan dipilih bagi penciptaan projek ini. Ini merupakan sebuah kajian atau projek di mana dua jenis algoritma daripada sebuah model iaitu Perlombongan Petua Sekutuan di laksanakan. Ianya dikaji dan dibandingkan agar peraturan terbaik di ambil untuk ditempatkan di dalam papan pemuka. Selain itu, kekuatan projek ini juga adalah di mana data COVID-19 di ambil daripada pihak PKD dalam pelaksanaan penganalisaan data iaitu analisis deskriptif berjaya dilaksanakan. Tiga fungsi daripada projek ini yang diletakkan di dalam papan pemuka merupakan fungsi yang berguna dan dapat memberi kemudahan pada pihak pengguna iaitu Pusat Kesihatan Daerah Hulu Langat dalam konteks COVID-19 pada tahun 2021.

Berdasarkan pengalaman dalam menjayakan projek ini yang dikenali sebagai 'Analisis deskriptif dan Perlombongan Petua Sekutuan untuk COVID-19 semasa varian delta', ianya memberi banyak pengalaman yang bermakna namun terdapat kekangan dimana data yang diberikan tidak seimbang. Terdapat terlalu banyak bias atau berat sebelah yang diperolehi. Jadi hasil keputusan daripada kajian ini mempunyai hanya satu hasil iaitu 'Pulih' apabila dijana daripada kedua-dua algoritma iaitu algoritma apriori dan algoritma FP-Growth. Ianya tidak dapat diseimbangkan menggunakan kaedah persampelan berlebihan kerana keputusannya masih sama iaitu hanya keputusan 'Pulih' sahaja dijana. Selain itu, nilai-nilai parameter daripada kaedah persampelan berlebihan iaitu sokongan, keyakinan dan lif adalah sangat rendah berbanding algoritma tanpa penggunaan persampelan berlebihan.

Pada masa hadapan, projek ini dapat digunakan dengan lebih baik seperti penggunaan sistem pakar dalam menjana peraturan daripada kedua-dua algoritma dalam perlombongan petua

sekutuan. Dalam penggunaan sistem yang lebih efisien, projek mampu mencari nilai peraturan yang terbaik tidak kira jumlah peraturan yang dijana. Projek ini hanya memerincikan kepada sepuluh peraturan, namun bagi masa akan datang, setiap peraturan yang dijana, boleh digunakan bagi penghasilan peraturan yang terbaik. Setiap peraturan yang dijana boleh digunakan semuanya agar penghasilan atau penemuan dalam mencari atribut menjadi lebih banyak dan mudah.

Kesimpulan keseluruhan bagi pembangunan projek ini dapat dirumuskan, setiap perancangan yang di beri telah berjaya dilaksanakan dan setiap objektif berjaya dicapai. Di akhir pelaksanaan projek ini, setiap kekuatan dan kekangan projek serta penambahbaikan di masa hadapan di senaraikan agar setiap penemuan yang di jumpai boleh dibaiki pada masa akan datang. Sebagai contoh, sekiranya terdapat sebuah kekuatan dalam penghasilan projek, maka ianya seharusnya dikekalkan manakala sekiranya terdapat kekangan yang ditemui, maka kekangan itu seharusnya di atasi agar tidak mendatangkan akibat yang buruk pada masa akan datang atau kekangan harus di atasi agar projek itu mudah untuk di percayai. Penambahbaikan di ambil kira agar projek yang lebih baik dapat dilaksanakan.

## Penghargaan

Syukur Alhamdulillah dan setinggi-tinggi kesyukuran dipanjatkan kehadiran ilahi kerana dengan izin kurnianNya, saya mampu menyiapkan laporan yang bertajuk 'Analisis deskriptif dan perlombongan petua sekutuan untuk COVID-19 semasa varian delta' ini.

Setinggi-tinggi penghargaan saya rakamkan kepada penyelia saya iaitu Prof. Madya Dr. Zalinda Binti Othman diatas segala budi bicara, tunjuk ajar dan bantuan yang diberikan kepada saya sepanjang pelaksanaan kajian ini. Segala jasa, ilmu dan tunjuk ajar beliau akan saya kenang sepanjang hayat.

Saya juga mengambil kesempatan ini untuk memberikan penghargaan dan ucapan ribuan terima kasih kepada ibubapa, keluarga dan rakan-rakan atas sokongan yang diberikan. Sokongan tersebut menjadi sumber semangat dan kekuatan saya dalam melaksanakan projek tahun akhir ini.

Akhir sekali, terima kasih kepada semua pihak yang terlibat secara langsung atau tidak langsung dan saya ingin memohon ribuan kemaafan sekiranya terdapat kesilapan sebelum, semasa dan setelah kajian ini berlangsung.

Sekian, terima kasih.

**RUJUKAN**

- Almalki, A., Gokaraju, B., Acquaah, Y., & Turlapaty, A. 2022. Regression analysis for COVID-19 infections and deaths based on food access and health issues. *Healthcare, 10*(2), 324.
- Arreeras, T., Arimura, M., Asada, T., & Arreeras, S. (2019). Association Rule Mining Tourist-Atractive Destinations for the Sustainable Development of a Large Tourism Area in Hokkaido Using Wi-Fi Tracking Data. *Sustainability, 11*(14), 3967.
- Asana, T. 2022. Understanding the iterative process, with examples. Asana.
- Bhatnagar, V., Poonia, R. C., Nagar, P., Kumar, S., Singh, V., Raja, L., & Dass, P. 2020. Descriptive analysis of COVID-19 patients in the context of India. *Journal of Interdisciplinary Mathematics, 24*(3), 489–504.
- Khan, H. U., Khan, S., & Nazir, S. 2022. A novel deep learning and ensemble learning mechanism for delta-type COVID-19 detection. *Frontiers in Public Health, 10*.
- Ong, A. K., Prasetyo, Y. T., Yudianto, N., Nadlifatin, R., Persada, S. F., Robas, K. P., Chuenyindee, T., & Buaphiban, T. 2022. Utilization of random forest classifier and artificial neural network for predicting factors influencing the perceived usability of COVID-19 contact tracing “Morchana” in Thailand. *International Journal of Environmental Research and Public Health, 19*(13), 7979.
- Passarelli-Araujo, H., Urbano, M. R., & Pescim, R. R. 2022. Machine learning and comorbidity network analysis for hospitalized patients with COVID-19 in a city in southern Brazil. *Smart Health, 26*, 100323.
- Tandan, M., Acharya, Y., Pokharel, S., & Timilsina, M. 2021. Discovering symptom patterns of COVID-19 patients using association rule mining. *Computers in biology and medicine, 131*, 104241.
- team, E. 2022. Implementation plan: What is it & how to create it? (Steps & Process). Bit Blog.

Nazifa Adriana Binti Nazarrudin (A180241)  
Prof. Madya Dr. Zalinda Binti Othman  
Fakulti Teknologi & Sains Maklumat,  
Universiti Kebangsaan Malaysia

Copyright@FTSM  
UKM