

MODEL MULTIVARIATE BI-LSTM UNTUK RAMALAN KES COVID-19 MENGUNAKAN INTEGRASI DATA METEOROLOGI DI MALAYSIA

HAREENRAJ CHANTHIRASEKARAN
DR AZURALIZA ABU BAKAR

^{1,2}*Fakulti Teknologi & Sains Maklumat, Universiti Kebangsaan Malaysia, 43600 UKM Bangi,,
Selangor Darul Ehsan, Malaysia*

Abstrak

Memahami cara menangani wabak ini sebagai sebahagian daripada kehidupan seharian adalah penting memandangkan penyebaran kes COVID-19 yang meluas dan kesan buruknya terhadap pelbagai aspek kehidupan kita. Selepas pandemik COVID-19 yang telah berlarutan selama hampir empat tahun, telah menjadi jelas bahawa pelbagai faktor, termasuk parameter meteorologi, mempengaruhi kadar penyebaran penyakit tersebut serta kemungkinan kematian. Ini penting, kerana kajian telah menunjukkan bahawa pada masa lalu penyakit virus seperti influenza dan sindrom pernafasan akut parah (SARS) banyak dipengaruhi oleh meteorologi. Data besar sedia ada kes covid memerlukan teknologi untuk dimodelkan dan dikendalikan. Pembelajaran mendalam ialah kaedah yang popular dan berkuasa untuk mempelajari sejumlah besar data multivariate. Dalam kajian ini, teknik rangkaian neural berdasarkan bidirectional long short-term memory akan digunakan untuk mencipta model pembelajaran mendalam (BiLSTM). Perbandingan dibuat antara regresi vektor sokongan (SVR) dan rangkaian neural unidirectional long short-term memory (LSTM). Melalui penggunaan metrik penilaian seperti MAPE, RMSE, MAE dan R2-Score, prestasi model ramalan akan dinilai. Model ini sesuai digunakan dalam ramalan kes aktif COVID-19 bagi setiap negeri, mengikut skor penilaian model tertinggi. Model BiLSTM, dijangka yang boleh meramalkan kedua-dua aliran dinamik jangka panjang dan jangka pendek kes COVID-19 dengan berkesan, mempunyai

ketepatan ramalan yang paling tinggi berbanding dengan model LSTM, SVR. Hasil daripada projek ini adalah untuk membina sebuah papan muka yang dapat menunjukkan kajian korelasi antara faktor berbeza yang dikaitkan dengan COVID-19 untuk mengurus dan mengawal wabak semasa serta lebih bersedia sekiranya berlaku wabak seperti tersebut pada masa hadapan.

Kata kunci: COVID-19, ML, BI-LSTM, LSTM, SARS

Pengenalan

Projek ini bertujuan untuk mengembangkan model ramalan menggunakan data meteorologi untuk meramalkan kasus COVID-19 aktif harian di setiap negeri di Malaysia. Latar belakang projek ini dipicu oleh impak luas yang ditimbulkan oleh pandemi COVID-19 yang telah mengganggu kehidupan global. Penyebaran virus ini dipengaruhi oleh berbagai faktor, termasuk kondisi cuaca, yang mungkin memiliki keterkaitan yang signifikan dengan penyebaran virus. Oleh karena itu, pemahaman yang lebih mendalam tentang hubungan antara data meteorologi dan penularan COVID-19 dapat membantu dalam mengambil langkah-langkah pencegahan yang lebih efektif dan pengelolaan pandemi.

Projek ini relevan dan penting untuk dikaji kerana menggunakan data meteorologi sebagai faktor penting dalam meramalkan penyebaran COVID-19. Keterlibatan faktor cuaca dalam model ramalan dapat memberikan pandangan yang lebih holistik dan komprehensif tentang hubungan antara kondisi cuaca dan penularan penyakit. Hasil penelitian ini dapat memberikan wawasan baru tentang peranan data meteorologi dalam mempengaruhi dinamika penyebaran virus dan membantu pihak yang berkenaan dalam mengidentifikasi area yang hotspot dan mengambil tindakan yang sesuai untuk mengurangi penularan COVID-19.

Kajian ini juga relevan mengingat pentingnya mengintegrasikan data dan pendekatan yang pelbagai untuk menghadapi pandemik. Dengan menggunakan teknik pembelajaran mesin seperti LSTM dan BI-LSTM, penelitian ini menyediakan kerangka kerja yang dapat diterapkan untuk

pemodelan ramalan pandemi di masa depan. Hal ini dapat membantu dalam membangun sistem peringatan dini yang lebih akurat dan respons yang lebih efektif terhadap perubahan kondisi pandemi. Dengan demikian, projek ini memiliki potensi untuk memberikan kontribusi yang signifikan dalam upaya global untuk mengatasi pandemi COVID-19 dan memberikan manfaat nyata bagi masyarakat dan pihak berkepentingan.

Objektif utama projek ini adalah untuk membangunkan model ramalan yang berkesan bagi meramalkan kes COVID-19 aktif harian di setiap negeri di Malaysia dengan menggunakan data meteorologi sebagai faktor prediksi. Kajian ini bertujuan untuk mencapai beberapa matlamat seperti berikut:

1. Meningkatkan ketepatan ramalan: Dengan menggabungkan data meteorologi dalam model ramalan, diharapkan hasil ramalan kes COVID-19 menjadi lebih tepat dan boleh dipercayai. Penggunaan teknik pembelajaran mesin seperti LSTM dan BI-LSTM akan membantu model mengenal pasti corak dan trend yang relevan dari data sebelumnya, meningkatkan ketepatan ramalan.

2. Memahami hubungan antara data meteorologi dan penyebaran COVID-19: Melalui kajian ini, diharapkan dapat mengenal pasti dan memahami hubungan yang lebih mendalam antara faktor-faktor cuaca dengan penularan COVID-19. Penemuan ini akan memberikan wawasan baru tentang bagaimana keadaan cuaca dapat mempengaruhi tingkat penyebaran virus di pelbagai wilayah.

3. Meningkatkan kapasiti tanggapan darurat dan perancangan pandemik: Dengan mempunyai model ramalan yang berkesan, pihak berkuasa dan authoriti kesihatan dapat merancang langkah-langkah pencegahan yang lebih baik untuk mengatasi penyebaran COVID-19. Ramalan yang tepat dapat memperkuat kapasiti tindak balas darurat dan membantu memperuntukkan sumber daya dengan lebih efisien dalam menangani pandemik.

Secara keseluruhan, tujuan utama kajian ini adalah untuk menghasilkan model ramalan yang memberi manfaat yang nyata dalam penanganan pandemik COVID-19, serta memberi sumbangan yang positif dalam ilmu pengetahuan dan pemahaman tentang hubungan antara data meteorologi dan

penyebaran penyakit. Projek ini berfokus pada meramalkan kes COVID-19 aktif harian di setiap negeri di Malaysia dengan menggunakan data meteorologi sebagai pemboleh ubah dalam model prediksi.

Skop kajian ini terhad kepada analisis dan perbandingan model LSTM dan BI-LSTM dalam meramalkan penyebaran virus berdasarkan data cuaca. Isu-isu kajian termasuk keupayaan kedua-dua model dalam menangkap hubungan antara faktor meteorologi dan penyebaran COVID-19, serta pengaruh atribut meteorologi harian terhadap prestasi ramalan. Batasan projek ini ialah menggunakan data harian kes COVID-19 dan data meteorologi yang tersedia dan hanya memfokuskan pada negeri-negeri di Malaysia sebagai kawasan kajian.

Projek ini perlu dilakukan kerana penyebaran COVID-19 adalah isu kesihatan global yang mendesak dan memerlukan pendekatan yang holistik untuk mengurusnya. Dengan mengintegrasikan data meteorologi dalam model ramalan, projek ini dapat memberikan pandangan yang lebih komprehensif tentang hubungan antara cuaca dan penyebaran virus, serta meningkatkan ketepatan ramalan kes aktif harian COVID-19 di setiap negeri di Malaysia. Hasil kajian ini akan memberi manfaat kepada bidang ilmu epidemiologi dan kesihatan awam dengan meningkatkan pemahaman tentang faktor-faktor yang mempengaruhi penyebaran virus dan membantu dalam merancang strategi pencegahan dan pengendalian yang lebih efektif. Selain itu, hasil kajian ini juga dapat memberikan sumbangan kepada industri teknologi informasi dan aplikasi berbasis data, yang dapat memanfaatkan model ramalan yang canggih untuk memberikan informasi penting tentang situasi COVID-19 kepada masyarakat secara tepat dan cepat.

Dalam projek ini, metodologi yang digunakan ialah CRISP-DM yang merangkumi langkah-langkah untuk membangunkan model ramalan menggunakan teknik pembelajaran mesin berdasarkan data meteorologi dari beberapa atribut seperti suhu, kelembapan, dan tekanan udara, serta kes aktif harian COVID-19. Pertama, data harian kes COVID-19 dan data meteorologi dikumpulkan dan diproses untuk memastikan kualiti dan kesahihan. Selanjutnya, data ini digunakan untuk melatih dan

menguji model LSTM (Memori Jangka Pendek Panjang) dan BI-LSTM (LSTM Dwi Arah) untuk meramalkan kes aktif harian COVID-19 di setiap negeri di Malaysia. Proses latihan dan penilaian dilakukan dengan menggunakan metrik prestasi seperti MAE (Mean Absolute Error), RMSE (Root Mean Squared Error), dan R2-Score untuk menilai ketepatan dan keberkesanan model. Langkah-langkah ini diulangi dan disesuaikan untuk mencari model terbaik yang memberikan hasil ramalan paling tepat untuk setiap negeri. Metodologi ini memastikan penggunaan data meteorologi dan algoritma yang sesuai untuk mencapai objektif kajian dengan efektif dan memberikan hasil yang bermakna dalam meramalkan penyebaran COVID-19 dengan mengambil kira faktor-faktor cuaca yang relevan.

Berikut adalah ringkasan mengenai struktur keseluruhan laporan teknikal untuk projek Ramalan Kes COVID-19 menggunakan Integrasi Data Meteorologi Di Malaysia:

- **Halaman Sampul:**

Halaman ini akan mencakup judul laporan, nama penulis, institusi, tarikh, dan maklumat penting lain yang berkaitan dengan laporan.

- **Abstrak:**

Abstrak adalah ringkasan singkat projek dan laporan, termasuk tujuan, metodologi, hasil, dan kesimpulan utama. Ia memberi pembaca gambaran keseluruhan tentang kandungan laporan.

- **Pendahuluan:**

Bahagian ini akan menerangkan latar belakang projek, tujuan projek, maksud laporan, dan gambaran keseluruhan tentang apa yang akan dibincangkan dalam laporan.

- **Metodologi Kajian:**

Metododologi Kajian dalam sebuah laporan teknik adalah bahagian yang menjelaskan tentang kaedah dan pendekatan yang digunakan dalam menjalankan kajian. Ia juga menerangkan model proses pembangunan khusus yang digunakan serta jelaskan mengapa model proses berkenaan dipilih.

• **Keputusan dan Perbincangan:**

Bahagian ini akan menganalisis hasil ujian dan prestasi sistem dengan merujuk kepada objektif projek. Penemuan dan perbincangan tentang kelebihan dan kelemahan sistem juga akan dikongsikan.

• **Kesimpulan:**

Kesimpulan adalah bahagian penting dalam laporan teknik kerana ia memberi gambaran terhadap hasil dan maklumat yang diperolehi dari kajian. Ia juga adalah bahagian di mana penulis menyimpulkan keseluruhan kajian dan memberi ringkasan tentang hasil kajian serta implikasinya.

• **Penghargaan:**

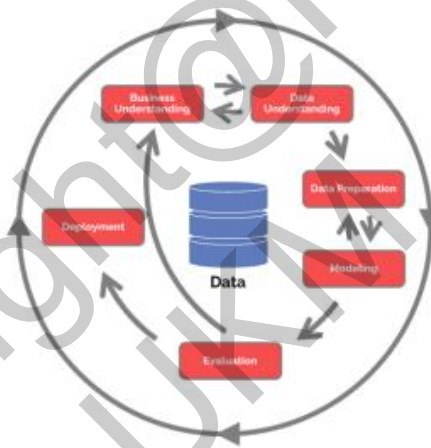
Penghargaan dalam sebuah laporan teknik adalah bahagian di mana penulis menyatakan terima kasih kepada individu, kumpulan, atau pihak yang telah memberikan sokongan, bantuan, atau sumbangan dalam menjalankan kajian.

• **Rujukan:**

Senarai rujukan atau sumber yang telah digunakan dalam projek ini akan dicantumkan di sini. Seterusnya, laporan akan diatur dalam urutan ini untuk memberikan paparan yang sistematik dan mudah difahami bagi pembaca.

Metodologi Kajian

Projek ini menggunakan model proses pembangunan CRISP-DM (The Cross Industry Standard Process for Data Mining) yang terdiri dari enam fasa, iaitu Pemahaman Bisnes, Pemahaman Data, Penyediaan Data, Permodelan, Penilaian, dan Penggunaan. Fasa-fasa ini memastikan pemahaman matlamat bisnes, pemilihan data berkualiti, penggabungan data kes COVID-19 dan data meteorologi, pemodelan menggunakan model BI-LSTM dan LSTM, penilaian prestasi model, dan akhirnya, penerbitan hasil ramalan untuk kegunaan awam. Model CRISP-DM ini sesuai dengan objektif projek dalam meramalkan penyebaran COVID-19 berdasarkan data meteorologi dengan pendekatan yang terstruktur dan efektif.



Rajah 1.0: Fasa-fasa Metodologi CRISP-DM

Projek ini menggunakan kaedah pengumpulan data dengan mengambil data harian kes COVID-19 dan data meteorologi dari sumber dipercayai seperti Repositori github Kementerian Malaysia dan laman web yang relevan. Penggunaan kaedah ini justifikasi kerana data harian kes COVID-19 memungkinkan analisis perubahan dan perkembangan infeksi, sementara data cuaca harian memainkan peranan penting dalam mempengaruhi penularan virus. Selain itu, sumber data yang dipercayai memberikan keyakinan terhadap keandalan dan kualitas data yang digunakan dalam projek ini.

Kaedah Analisis Data yang digunakan dalam projek pembangunan model ramalan kes COVID-19 melalui data meteorologi adalah sebagai berikut:

1. Penerokaan Data (Data Exploration):

Pada fasa pemahaman data, dilakukan penerokaan data untuk memahami struktur dan karakteristik set data. Ini melibatkan pemeriksaan visual data menggunakan grafik, carta, dan statistik deskriptif untuk mendapatkan gambaran yang jelas tentang distribusi atribut dan pola hubungan antara atribut. Penerokaan data membantu mengidentifikasi outlier, kesalahan data, dan trend dalam set data.

2. Pembersihan Data (Data Cleaning):

Setelah penerokaan data, dilakukan pembersihan data untuk menangani data yang hilang, duplikat, atau tidak konsisten. Jika terdapat data yang hilang, langkah-langkah pengisian nilai atau imputasi dilakukan untuk menggantikan nilai yang hilang. Data duplikat juga dihapuskan untuk memastikan integritas set data.

3. Penggabungan Data (Data Integration):

Data dari sumber yang berbeza, seperti data kes COVID-19 dan data meteorologi, perlu digabungkan menjadi satu set data yang utuh. Penggabungan data ini memastikan semua atribut yang relevan tersedia dalam satu dataset untuk analisis dan model.

4. Visualisasi Data (Data Visualization):

Data dianalisis melalui visualisasi data menggunakan grafik, carta, dan plot lainnya. Visualisasi data membantu dalam memahami pola, korelasi, dan distribusi data secara lebih intuitif dan mudah difahami. Visualisasi membantu mengidentifikasi hubungan antara atribut, mengidentifikasi outlier, dan memahami pola perubahan seiring waktu.

5. Korelasi Data (Data Correlation):

Dilakukan analisis korelasi untuk memahami hubungan antara atribut dalam dataset. Korelasi digunakan untuk mengukur tingkat keterkaitan antara dua atribut. Korelasi positif

menunjukkan bahwa dua atribut bergerak ke arah yang sama, sedangkan korelasi negatif menunjukkan bahwa dua atribut bergerak ke arah yang berlawanan. Informasi ini membantu dalam pemilihan fitur penting untuk pembangunan model prediksi.

6. Pemilihan Fitur Penting (Feature Selection):

Berdasarkan analisis korelasi dan pengaruh atribut terhadap prediksi, dilakukan pemilihan fitur penting. Fitur-fitur yang memiliki dampak signifikan dalam pembangunan model dipilih untuk digunakan dalam model prediksi. Pemilihan fitur penting membantu menyederhanakan model dan meningkatkan efisiensi proses prediksi.

Kaedah pendekatan analisis yang digunakan dalam proyek ini adalah analisis eksploratori data (data exploration) yang melibatkan penerokaan data, pembersihan data, penggabungan data, dan visualisasi data. Selain itu, dilakukan analisis korelasi untuk memahami hubungan antara atribut dan pemilihan fitur penting untuk menentukan atribut yang memiliki dampak signifikan dalam pembangunan model. Semua langkah analisis ini membantu dalam memahami dataset, mengidentifikasi pola dan hubungan antara atribut, dan mempersiapkan data yang berkualitas untuk pembangunan model prediksi COVID-19.

Pengukuran dan Alat Ukur yang digunakan untuk menilai keberkesanan hasil proyek ramalan COVID-19 berdasarkan data meteorologi.

1. Ketepatan Ramalan:

Ketepatan ramalan adalah salah satu ukuran penting untuk menilai keberkesanan proyek ini. Ia mengukur sejauh mana model ramalan dapat memprediksi jumlah kasus COVID-19 dengan tepat berdasarkan data meteorologi. Untuk mengukur ketepatan ini, kita membandingkan jumlah kasus yang diprediksi oleh model dengan jumlah kasus sebenar. Semakin tinggi ketepatan ramalan, semakin baik hasil proyek ini dalam meramalkan kes COVID-19.

2. Purata Ralat Mutlak (MAE):

Purata Ralat Mutlak (MAE) adalah satu metrik yang mengukur purata perbezaan antara nilai yang diramalkan oleh model dengan nilai sebenar jumlah kes COVID-19. MAE memberikan gambaran tentang sejauh mana model ini mempunyai ralat yang kecil dalam ramalannya. MAE yang lebih rendah menunjukkan keberkesanan yang lebih baik dalam meramalkan kes COVID-19.

3. Purata Ralat Mutlak (MAE):

Punca Kuasa Dua Ralat Mutlak (RMSE) adalah satu lagi ukuran statistik yang mengukur purata perbezaan kuasa dua antara nilai yang diramalkan oleh model dengan nilai sebenar jumlah kes COVID-19. RMSE memberikan tumpuan kepada ralat yang lebih besar dan memberikan gambaran tentang kesesuaian model dengan data. RMSE yang lebih rendah menunjukkan keberkesanan yang lebih baik dalam meramalkan kes COVID-19.

4. Purata Peratusan Ralat Mutlak (MAPE):

Purata Peratusan Ralat Mutlak (MAPE) mengukur purata perbezaan peratusan mutlak antara nilai yang diramalkan oleh model dengan nilai sebenar jumlah kes COVID-19. MAPE memberikan pemahaman tentang ralat relatif dalam bentuk peratusan. MAPE yang lebih rendah menunjukkan keberkesanan yang lebih baik dalam meramalkan kes COVID-19.

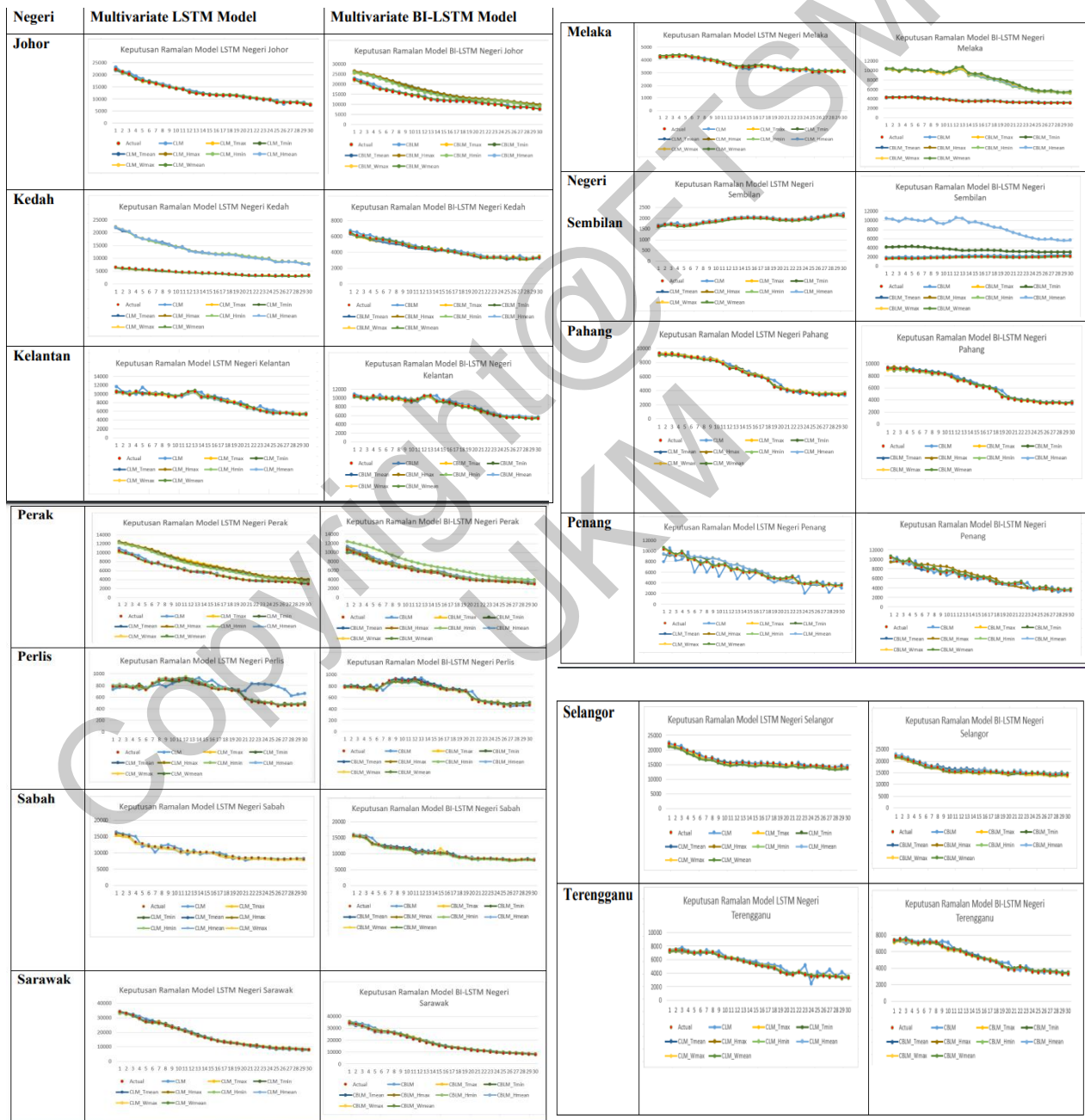
5. Skor R-squared:

Skor R-squared atau R-squared mengukur sejauh mana variasi dalam data jumlah kes COVID-19 dapat dijelaskan oleh model. Skor R-squared yang lebih tinggi menunjukkan keberkesanan yang lebih baik dalam menjelaskan variasi dalam data dan kemampuan model untuk meramalkan jumlah kes COVID-19 dengan tepat.

Pengukuran dan alat ukur di atas akan membantu menilai keberkesanan hasil projek ramalan COVID-19 berdasarkan data meteorologi dan memastikan model ini memberikan hasil yang tepat dan berguna bagi pengambilan keputusan dan pengurusan penyebaran COVID-19.

Keputusan dan Perbincangan

Berikutan adalah hasil kajian berdasarkan eksperimen yang telah dilakukan untuk model ramalan LSTM dan BI-LSTM dalam meramalkan kes COVID-19 aktif di setiap negeri di Malaysia. Rajah 1.2, telah paparkan trend model ramalan bagi semua negeri di Malaysia berdasarkan model LSTM dan BI-LSTM.



Rajah 1.2: Graf keputusan ramalan model LSTM dan BI-LSTM multivariate bagi setiap negeri

Bahagian tersebut menunjukkan hasil keputusan penilaian metrik bagi setiap model ramalan mengikut negeri-negeri di Malaysia. Rajah 5.4 di bawah menunjukkan keputusan penilaian metrik yang dipaparkan dalam bentuk graf bar.



Rajah 1.3, telah menunjukkan graf perbandingan metrik antara model bagi setiap negeri.

Hasil kajian menunjukkan bahawa model LSTM (Memori Jangka Pendek Panjang) adalah lebih sesuai dan memberikan prestasi yang lebih baik dalam meramalkan penyebaran COVID-19 aktif harian di Malaysia berbanding dengan model BI-LSTM (LSTM Dwi Arah). Penggunaan model LSTM univariat dan multivariat dengan mempertimbangkan data meteorologi, seperti atribut angin minimum harian (CLM_Tmin) dan angin maksimum harian (CLM_Wmax), telah menunjukkan prestasi yang lebih baik dalam menangkap corak dan hubungan dalam penyebaran COVID-19.

Walaupun model LSTM univariat dan multivariat menunjukkan keunggulan dalam meramalkan kes COVID-19 di kebanyakan negeri di Malaysia, model BI-LSTM tetap relevan dalam beberapa senario tertentu dan memberikan hasil yang baik di beberapa negeri. Oleh itu, dalam menghadapi situasi COVID-19 yang kompleks dan berubah-ubah, penting untuk melakukan pemantauan dan penyesuaian berkala terhadap model ramalan untuk memastikan ketepatan dan keberkesananannya.

Hasil kajian ini memberikan panduan yang berharga kepada para pengambil keputusan dan pegawai kesihatan awam dalam merancang langkah-langkah pencegahan dan pengendalian yang tepat untuk menghadapi penyebaran COVID-19 di setiap negeri di Malaysia. Dengan mempertimbangkan faktor meteorologi dalam model ramalan, keupayaan model LSTM untuk meramalkan kes aktif harian COVID-19 telah meningkat secara signifikan.

Oleh itu, dapat disimpulkan bahawa model LSTM adalah pilihan yang lebih sesuai dalam meramalkan penyebaran COVID-19 di Malaysia, terutamanya apabila mempertimbangkan faktor meteorologi. Namun, model BI-LSTM tetap relevan dalam beberapa skenario tertentu. Keputusan ini memberikan panduan yang berharga bagi pihak berkepentingan dalam mengambil tindakan pencegahan dan pengendalian yang tepat untuk menghadapi situasi COVID-19.

Berdasarkan data di atas, keputusan kajian menunjukkan bahawa model LSTM (Memori Jangka Pendek Panjang) adalah lebih sesuai dan memberikan prestasi yang lebih baik dalam meramalkan penyebaran COVID-19 aktif harian di Malaysia berbanding dengan model BI-LSTM (LSTM Dwi Arah). Model LSTM univariat dan multivariat dengan mempertimbangkan data meteorologi, seperti

atribut angin minimum harian (CLM_Tmin) dan angin maksimum harian (CLM_Wmax), menunjukkan prestasi yang lebih baik dalam menangkap corak dan hubungan dalam penyebaran COVID-19.

Analisis makna di sebalik keputusan ini mengindikasikan bahawa faktor meteorologi, khususnya angin minimum dan maksimum harian, memainkan peranan penting dalam meramalkan penyebaran COVID-19. Penggunaan model LSTM yang mampu mengatasi ketidakpastian dan menangkap kebergantungan jangka panjang dalam data siri masa telah menghasilkan ramalan yang lebih tepat dan akurat. Model LSTM mempunyai keupayaan untuk memproses data berurutan dan memahami trend jangka panjang, yang sangat relevan dalam meramalkan pandemi seperti COVID-19 yang berlangsung dalam jangka waktu yang panjang.

Hubungkaitan keputusan ini dengan tujuan kajian yang telah ditetapkan adalah sangat signifikan. Tujuan utama kajian ini adalah untuk memilih model ramalan yang paling sesuai dan berkesan dalam meramalkan kes COVID-19 aktif harian di setiap negeri di Malaysia. Dengan menggunakan model LSTM yang mempertimbangkan data meteorologi, kajian ini telah mencapai tujuannya dengan kesan yang positif. Model LSTM univariat dan multivariat telah membuktikan keunggulannya dalam meramalkan dengan tepat penyebaran COVID-19, dan penggunaan atribut meteorologi telah meningkatkan ketepatan ramalan.

Selain itu, hasil kajian ini memberikan wawasan yang berharga kepada para pengambil keputusan dan pegawai kesihatan awam dalam merancang strategi pencegahan dan pengendalian yang lebih efektif. Dengan mempertimbangkan faktor cuaca dalam meramalkan penyebaran COVID-19, tindakan pencegahan dapat diambil dengan lebih tepat sasaran dan efisien. Model LSTM yang berdasarkan data meteorologi juga dapat membantu meramalkan situasi pandemi pada masa depan dan menyediakan panduan bagi pengambilan keputusan yang lebih baik dalam menangani penyebaran wabak.

Dalam kesimpulannya, analisis keputusan menunjukkan bahawa penggunaan model LSTM yang mempertimbangkan data meteorologi adalah pendekatan yang berkesan dalam meramalkan penyebaran COVID-19 di Malaysia. Penggunaan atribut cuaca telah meningkatkan ketepatan ramalan dan memberikan panduan yang lebih baik dalam menangani situasi pandemi. Dengan demikian, kajian ini memberikan sumbangan penting dalam usaha menangani wabak COVID-19 dan memberikan manfaat kepada masyarakat luas.

Hasil kajian yang telah dilakukan dalam laporan ini dapat dibandingkan dengan kajian-kajian terdahulu yang telah dilakukan oleh penyelidik lain dalam bidang yang sama, iaitu meramalkan penyebaran COVID-19 dengan mempertimbangkan faktor meteorologi. Dalam perbandingan ini, beberapa perbezaan dan persamaan dapat dikenal pasti.

Beberapa persamaan antara kajian ini dengan kajian-kajian terdahulu adalah penggunaan model pembelajaran mesin, seperti LSTM dan BI-LSTM, dalam meramalkan penyebaran COVID-19. Ini menunjukkan bahawa kedua-dua kajian mengakui keupayaan model pembelajaran mesin untuk meramalkan dengan berkesan dalam konteks pandemik. Selain itu, kedua-dua kajian juga menekankan pentingnya mempertimbangkan faktor meteorologi, seperti data cuaca, sebagai pemboleh ubah dalam proses ramalan. Ini menunjukkan bahawa faktor cuaca memainkan peranan yang relevan dalam mempengaruhi penyebaran COVID-19 dan perlu diambil kira dalam model ramalan.

Walaupun terdapat persamaan, terdapat beberapa perbezaan antara kajian ini dan kajian-kajian terdahulu. Salah satu perbezaan utama adalah dalam jenis model yang digunakan. Kajian ini menggunakan model LSTM univariat dan multivariat, sementara kajian lain mungkin menggunakan model yang berbeza atau variasi lain daripada LSTM. Ini menunjukkan bahawa tidak ada model yang tunggal dan standard untuk meramalkan pandemi, dan pemilihan model bergantung pada keunikan data dan objektif kajian.

Selain itu, perbezaan lain mungkin terletak dalam atribut meteorologi yang dipilih untuk dimasukkan dalam model. Setiap kajian mungkin memilih atribut cuaca yang berbeza-beza berdasarkan ketersediaan data dan penilaian penyelidik terhadap kepentingan atribut tersebut dalam meramalkan COVID-19. Ini menjelaskan mengapa laporan ini memilih untuk memasukkan atribut angin minimum harian (CLM_Tmin) dan angin maksimum harian (CLM_Wmax) dalam model, yang mungkin berbeza dengan kajian-kajian lain.

Perbezaan dalam jenis model dan atribut cuaca yang dipilih mungkin disebabkan oleh kepelbagaian data dan tujuan kajian yang berbeza. Setiap negara atau kawasan mungkin menghadapi corak penyebaran COVID-19 yang unik dan dipengaruhi oleh faktor cuaca yang berbeza. Oleh itu, penyelidik perlu mengadaptasi model dan memilih atribut cuaca yang paling relevan berdasarkan data yang ada dan tujuan kajian mereka.

Dalam kesimpulannya, hasil kajian dalam laporan ini menunjukkan persamaan dengan kajian-kajian terdahulu dalam penggunaan model pembelajaran mesin dan pentingnya mempertimbangkan faktor cuaca. Walau bagaimanapun, perbezaan dalam jenis model dan atribut cuaca yang dipilih adalah normal dalam penyelidikan saintifik yang melibatkan kepelbagaian data dan corak penyebaran yang berbeza. Kajian ini memberikan pandangan yang unik dan relevan dalam konteks Malaysia dan dapat memberikan sumbangan penting dalam memahami dan meramalkan penyebaran COVID-19 dengan lebih efektif.

Hasil kajian menunjukkan bahawa penggunaan model LSTM dan BI-LSTM dengan data cuaca adalah kaedah yang berkesan untuk meramalkan kes COVID-19 di Malaysia. Kedua-dua model berfungsi dengan baik dan dapat menangkap corak penyebaran dengan tepat. Model LSTM univariat lebih tepat dalam beberapa kes, manakala model BI-LSTM multivariat lebih baik di negeri tertentu. Penggunaan ciri cuaca tertentu juga memberi kesan pada ketepatan ramalan. Kajian ini memberi sumbangan penting dengan menilai kesan cuaca terhadap COVID-19 di setiap negeri. Keputusan ini

dapat membantu pihak berkuasa dalam pengurusan dan tindakan pencegahan yang tepat. Walau bagaimanapun, hasil ini perlu dikemaskini dengan data terkini secara berkala.

Implikasi hasil kajian ini adalah bahawa penggunaan model LSTM dan BI-LSTM dengan data cuaca memberikan sumbangan berharga dalam meramalkan kes COVID-19 di setiap negeri di Malaysia. Keputusan ini memiliki potensi untuk memberikan manfaat besar kepada bidang ilmu epidemiologi dan pengurusan penyakit berjangkit. Dengan memahami hubungan antara faktor cuaca dan penyebaran COVID-19, langkah-langkah pencegahan dan penanggulangan dapat dilaksanakan secara lebih tepat dan berkesan. Model ramalan ini juga dapat membantu pihak berkuasa dalam membuat keputusan termaklum dan merancang langkah-langkah penanganan wabak di masa depan.

Di samping itu, hasil kajian ini memberi panduan kepada industri pengurusan krisis dan kesihatan awam dalam meramal dan merancang strategi penanganan COVID-19. Data cuaca sebagai pemboleh ubah tambahan memberi informasi penting dalam meramalkan penyebaran virus dan membolehkan tindakan awal untuk mengurangkan kesan wabak.

Namun, terdapat beberapa kekurangan dalam kajian ini. Pertama, data cuaca mungkin tidak lengkap dan tidak selalu tepat. Oleh itu, pemantauan dan pengumpulan data yang tepat adalah penting untuk meningkatkan ketepatan model ramalan. Kedua, model-model yang digunakan mungkin perlu diperbaiki dan ditingkatkan lagi untuk meningkatkan keakuratan hasil ramalan.

Cadangan kajian masa hadapan yang boleh dikaji oleh penyelidik lain adalah menyelidik hubungan lebih mendalam antara faktor cuaca dan penyebaran COVID-19. Selain itu, kajian lanjut boleh memperluas analisis ke kawasan lain di seluruh negara atau di negara lain untuk memahami variasi kesan cuaca pada penyebaran virus. Penggunaan data cuaca yang lebih terperinci dan tepat juga dapat diterokai untuk meningkatkan kualiti model ramalan. Selain itu, model-model lain yang menggunakan teknik pembelajaran mesin yang berbeza boleh diuji untuk melihat prestasi mereka dalam meramalkan kes COVID-19 dengan menggunakan data cuaca.

Kesimpulan

Kajian ini telah berjaya menghasilkan model ramalan LSTM dan BI-LSTM menggunakan data cuaca untuk meramalkan kes COVID-19 aktif harian di negeri-negeri Malaysia dengan tepat. Hasil kajian menunjukkan bahawa kedua-dua model ini memberikan prestasi yang baik berdasarkan metrik seperti MAE, RMSE, MAPE, dan R2-Score. Penggunaan data cuaca sebagai pemboleh ubah tambahan memberikan pandangan yang lebih komprehensif tentang hubungan antara kondisi cuaca dan penyebaran COVID-19.

Objektif kajian untuk memilih model ramalan yang sesuai untuk meramalkan kes COVID-19 aktif harian di setiap negeri di Malaysia telah berjaya dicapai. Kedua-dua model LSTM dan BI-LSTM telah menunjukkan ketepatan dan keberkesanan dalam meramalkan penyebaran COVID-19 berdasarkan data cuaca.

Hasil kajian ini mempunyai implikasi penting dalam bidang ilmu epidemiologi dan pengurusan wabak. Pengetahuan tentang hubungan antara cuaca dan penyebaran COVID-19 dapat membantu dalam merancang strategi pencegahan dan penanggulangan wabak. Model ramalan yang berkesan ini juga dapat digunakan oleh pihak berkuasa untuk membuat keputusan yang berasaskan data dan merancang langkah-langkah penanganan yang lebih efektif.

Walaupun kajian ini menghadapi kekurangan dalam data cuaca yang mungkin tidak lengkap dan tepat, namun ia membuka peluang untuk penyelidikan lanjutan. Cadangan masa hadapan adalah untuk menyelidik hubungan lebih mendalam antara cuaca dan penyebaran COVID-19 serta meningkatkan lagi prestasi model-model ramalan ini. Penggunaan data cuaca yang lebih terperinci dan tepat harus dieksplorasi untuk meningkatkan ketepatan dan kebolehpercayaan hasil ramalan.

Secara keseluruhan, kajian ini memberikan sumbangan berharga kepada pengetahuan sedia ada dalam pengurusan krisis dan kesihatan awam. Dengan hasil yang memuaskan dari model LSTM dan BI-LSTM, kajian ini membuka jalan untuk lebih banyak penemuan dalam memahami hubungan antara cuaca dan penyebaran penyakit berjangkit.

Penghargaan

Terlebih dahulu, saya ingin mengambil peluang ini untuk mengucapkan jutaan terima kasih kepada penyelia saya, Prof. Dr. Azuraliza binti Abu Bakar yang banyak membimbing dan memberi tunjuk ajar kepada saya sepanjang kajian ilmiah dijalankan. Jutaan terima kasih juga saya hendak ucapkan kepada para pensyarah dari FTSM yang telah menaburkan ilmu pengetahuan kepada saya sepanjang pengajian saya di Universiti Kebangsaan Malaysia (UKM). Seterusnya, saya juga ingin merakamkan ribuan terima kasih kepada ahli keluarga saya terutamanya ibu dan ayah saya, yang menjadi sumber inspirasi, memberi semangat dan sentiasa berdoa terhadap kejayaan saya dalam menyiapkan kajian ini. Tidak lupa juga kepada rakanrakan yang banyak membantu saya dalam memberikan pandangan yang bernas serta sokongan kepada saya semasa diperlukan. Akhirnya, saya ingin mengucapkan jutaan terima kasih kepada pihak-pihak yang terlibat secara langsung dan tidak langsung dalam melaksanakan projek tahun akhir. Sekian, terima kasih.

RUJUKAN

- Bhimala, K. R., Patra, G. K., Mopuri, R., & Mutheneni, S. R. (2021). Prediction of COVID-19 cases using the weather integrated deep learning approach for India. *Transboundary and Emerging Diseases*, 69(3), 1349–1363. <https://doi.org/10.1111/tbed.14102>
- Pani, S. K., Lin, N. H., & RavindraBabu, S. (2020). Association of COVID-19 pandemic with meteorological parameters over Singapore. *Science of the Total Environment*, 740, 140112. <https://doi.org/10.1016/j.scitotenv.2020.140112>
- Rashed, E. A., & Hirata, A. (2021). One-Year Lesson: Machine Learning Prediction of COVID-19 Positive Cases with Meteorological Data and Mobility Estimate in Japan. *International Journal of Environmental Research and Public Health*, 18(11), 5736. <https://doi.org/10.3390/ijerph18115736>

Science. (n.d.). AAAS. <https://www.science.org/content/article/how-will-covid-19-affect-coming-flu-season-scientists-struggle-clues>

Siami-Namini, S., Tavakoli, N., & Namin, A. S. (2019). The Performance of LSTM and BiLSTM in Forecasting Time Series. *2019 IEEE International Conference on Big Data (Big Data)*. <https://doi.org/10.1109/bigdata47090.2019.9005997>

WEI SHEN, N., & ABU BAKAR, A. (in press). MODEL PEMBELAJARAN MENDALAM UNTUK MERAMAL KES COVID-19 DI MALAYSIA MENGGUNAKAN INTEGRASI DATA CUACA. *PTA-FTSM-2022*.

Yang, M., & Wang, J. (2022). Adaptability of Financial Time Series Prediction Based on BiLSTM. *Procedia Computer Science*, 199, 18–25. <https://doi.org/10.1016/j.procs.2022.01.003>

Yuan, J., Wu, Y., Jing, W., Liu, J., Du, M., Wang, Y., & Liu, M. (2021). Association between meteorological factors and daily new cases of COVID-19 in 188 countries: A time series analysis. *Science of the Total Environment*, 780, 146538. <https://doi.org/10.1016/j.scitotenv.2021.146538>

Bhimala, K. R., Patra, G. K., Mopuri, R., & Mutheneni, S. R. (2021). Prediction of COVID-19 cases using the weather integrated deep learning approach for India. *Transboundary and Emerging Diseases*, 69(3), 1349–1363. <https://doi.org/10.1111/tbed.14102>

Chimmula, V. K. R., & Zhang, L. (2020). Time series forecasting of COVID-19 transmission in Canada using LSTM networks. *Chaos, Solitons & Fractals*, 135, 109864. <https://doi.org/10.1016/j.chaos.2020.109864>

Papers with Code - BiLSTM Explained. (n.d.). <https://paperswithcode.com/method/bilstm>

- Rashed, E. A., & Hirata, A. (2021). One-Year Lesson: Machine Learning Prediction of COVID-19 Positive Cases with Meteorological Data and Mobility Estimate in Japan. *International Journal of Environmental Research and Public Health*, 18(11), 5736. <https://doi.org/10.3390/ijerph18115736>
- Shahid, F., Zameer, A., & Muneeb, M. (2020). Predictions for COVID-19 with deep learning models of LSTM, GRU and Bi-LSTM. *Chaos, Solitons & Fractals*, 140, 110212. <https://doi.org/10.1016/j.chaos.2020.110212>
- Shahin, A. I., & Almotairi, S. (2021). A Deep Learning BiLSTM Encoding-Decoding Model for COVID-19 Pandemic Spread Forecasting. *Fractal and Fractional*, 5(4), 175. <https://doi.org/10.3390/fractalfract5040175>
- Shrivastav, L. K., & Jha, S. K. (2020). A gradient boosting machine learning approach in modeling the impact of temperature and humidity on the transmission rate of COVID-19 in India. *Applied Intelligence*, 51(5), 2727–2739. <https://doi.org/10.1007/s10489-020-01997-6>
- Siami-Namini, S., Tavakoli, N., & Namin, A. S. (2019). The Performance of LSTM and BiLSTM in Forecasting Time Series. *2019 IEEE International Conference on Big Data (Big Data)*. <https://doi.org/10.1109/bigdata47090.2019.9005997>
- WEI SHEN, N., & ABU BAKAR, A. (in press). MODEL PEMBELAJARAN MENDALAM UNTUK MERAMAL KES COVID-19 DI MALAYSIA MENGGUNAKAN INTEGRASI DATA CUACA. *PTA-FTSM-2022*.
- Hotz, N. (2023, January 8). *What is CRISP DM?* Data Science Process Alliance. <https://www.datascience-pm.com/crisp-dm-2/>
- Mungalpara, J. (2021, December 30). *What does it mean by Bidirectional LSTM?* - Analytics Vidhya. Medium. <https://medium.com/analytics-vidhya/what-does-it-mean-by-bidirectional-lstm-63d6838e34d9>
- Papers with Code - BiLSTM Explained*. (n.d.). <https://paperswithcode.com/method/bilstm>

Smart Vision Europe. (2020, June 17). *Crisp DM methodology*. <https://www.sv-europe.com/crisp-dm-methodology/>

Understanding LSTM Networks -- colah's blog. (n.d.). <http://colah.github.io/posts/2015-08-Understanding-LSTMs/>

Basaldella, M., Antolli, E., Serra, G., & Tasso, C. (2017). Bidirectional LSTM Recurrent Neural Network for Keyphrase Extraction. *Communications in Computer and Information Science*, 180–187. https://doi.org/10.1007/978-3-319-73165-0_18

Ihianle, I. K., Nwajana, A. O., Ebebuwa, S. H., Otuka, R. I., Owa, K., & Orisatoki, M. O. (2020). A Deep Learning Approach for Human Activities Recognition From Multimodal Sensing Devices. *IEEE Access*, 8, 179028–179038. <https://doi.org/10.1109/access.2020.3027979>

Hannah, J. (2021, August 6). *What Is a User Interface, and What Are the Elements That Comprise One?* CareerFoundry. <https://careerfoundry.com/en/blog/ui-design/what-is-a-user-interface/>

Scardina, J., & Horwitz, L. (2022, December 7). *Microsoft Power BI*. Content Management. <https://www.techtarget.com/searchcontentmanagement/definition/Microsoft-Power-BI>

HAREENRAJ CHANTHIRASEKARAN (A180921)
PROF.DR AZURALIZA ABU BAKAR
Fakulti Teknologi & Sains Maklumat,
Universiti Kebangsaan Malaysia

Copyright@FTSM
UKM