

# MODEL PEMBELAJARAN MENDALAM UNTUK MERAMAL KES COVID-19 DI MALAYSIA MENGGUNAKAN INTEGRASI DATA CUACA

NG WEI SHEN  
AZURALIZA ABU BAKAR

*Fakulti Teknologi & Sains Maklumat, Universiti Kebangsaan Malaysia*

## ABSTRAK

Kadar penularan penyakit coronavirus (COVID-19) sangat cepat sejak kes pertama yang dilaporkan pada bulan Disember 2019 di Wuhan, China. Penyakit ini telah menjangkiti lebih daripada 3 juta orang di seluruh dunia dan mengakibatkan lebih daripada 224 ribu kematian pada 1 Mei 2020 yang dilaporkan oleh *World Health Organization* (WHO). Terdapat beberapa simptom biasa COVID-19 yang telah dikenal pasti berdasarkan kes yang dikenali, termasuk demam, batuk kering, cirit-birit, keletihan dan kehilangan rasa/bau. Pada masa lalu, parameter meteorologi seperti suhu dan kelembapan adalah faktor penting dan berkesan terhadap penyakit berjangkit serius seperti influenza dan Sindrom Pernafasan Akut Parah (SARS), jadi, sangat penting untuk meneroka hubungan antara faktor-faktor meteorologi dan kes COVID-19 yang aktif. Dalam kajian ini, model pembelajaran mendalam akan dihasilkan menggunakan algoritma rangkaian neural berdasarkan *long-short term memory* (LSTM). Model ini terdiri daripada data meteorologi siri masa dan data kes aktif harian yang diperoleh daripada data sumber terbuka. Prestasi model ramalan dinilai menggunakan metrik penilaian seperti MAE, RMSE, MAPE dan Skor-R2. Skor penilaian model yang tertinggi menunjukkan bahawa model tersebut adalah sesuai untuk digunakan dalam ramalan kes aktif COVID-19 bagi negeri tersebut. Hasil projek tersebut adalah untuk membantu pihak tertentu untuk meramalkan kes aktif COVID-19 supaya dapat mencegah penyebaran virus COVID-19.

## 1 PENGENALAN

Koronavirus sindrom pernafasan akut teruk 2 (SARS CoV-2) yang melahirkan penyakit koronavirus 2019 (COVID-19) telah muncul di Wuhan, China pada awal Disember. Sejak penyakit koronavirus dijumpai, virus tersebut cepat merebak di seluruh dunia dan menjadikan penularan tempatan di banyak negara termasuk Amerika, Eropah, Afrika, dan Asia. Penyebaran pantas virus COVID-19 adalah disebabkan oleh kekurangan pemahaman yang tetap tentang etologi penyakit dan corak penghantaran semasa fasa awal wabak. Pada 30 Januari 2020, *World Health Organization* (WHO) menetapkan wabak COVID-19 sebagai kecemasan kesihatan awam yang memerlukan perhatian dan mengesahkan wabak tersebut sebagai wabak global pada 11 Mac 2020. Pandemik COVID-19 telah mengakibatkan gangguan besar kepada kehidupan manusia, sistem kesihatan awam, dan ekonomi dan akan berterusan sehingga vaksin dijumpai. Semasa gelombang awal pandemik, kebanyakan negara telah melakukan Perintah Kawalan Pergerakan (PKP), hanya perkhidmatan yang penting boleh beroperasi serta mengamalkan penjarakan sosial dan topeng muka adalah wajib untuk semua orang. Sehingga

22 Oktober 2020, lebih 40 juta jangkitan COVID-19 dan 1.1 juta kematian telah direkodkan di seluruh dunia.

Seperti negara-negara lain, kes-kes COVID-19 juga dilaporkan di negara Malaysia. Pada 4 Februari 2020, kes COVID-19 pertama telah dikesan; seorang warga emas baru-baru pulang dari Singapura dengan mengalami demam dan batuk. Disebabkan jumlah kes COVID-19 meningkat berterusan di seluruh negara, kerajaan Malaysia telah membuat keputusan untuk menggantung kemudahan VISA pengunjung luar negara mulai 18 Mac 2020. Sehubungan dengan itu, kerajaan Malaysia mengisytiharkan Perintah Kawalan Pergerakan (PKP) dari 18 Mac hingga 31 Mac 2020 untuk mengurangkan aktiviti manusia di seluruh negara. Pada 4 Mei 2020, kerajaan Malaysia telah menjalankan Perintah Kawalan Pergerakan Bersyarat (PKPB). Begitu juga, keupayaan ujian COVID-19 telah berkembang dengan pesat untuk mengesan dan mengasingkan individu yang berpotensi menyebarkan virus. Peratusan kadar positif harian adalah antara 0.4% hingga 4.0% sehingga Mei 2020, dengan negeri-negeri tertentu yang mempunyai kadar positif yang lebih tinggi, seperti negeri Selangor (10.8%), Wilayah Persekutuan Kuala Lumpur (10.8%), dan Wilayah Persekutuan Labuan (10.1%). Di Malaysia, 26 ribu kes COVID-19 dan 1130 kes kematian telah didokumentasikan sehingga Februari 2021.

Epidemiologi kebanyakan penyakit berjangkit adalah dipengaruhi oleh faktor persekitaran. Beberapa kajian mendapati bahawa keadaan iklim dan cuaca boleh memberi kesan kepada taburan penyakit berjangkit serantau dan temporal (Valsamatzis-Panagiotou, A., Penchovsky, R., 2021). Keluarga Koronavirus seperti SARS CoV-1 dan MERs CoV telah didapati bahawa virus-virus tersebut bervariasi musim dan lebih gemar suhu rendah serta kelembapan yang rendah (Pal, M., Berhanu, G., Desalegn, C., & Kandi, V., 2020). Selain itu, pada fasa awal pandemik COVID-19, penyelidik melaporkan bahawa suhu mempunyai korelasi positif dan kelembapan mempunyai korelasi negatif dengan kes di seluruh dunia (Bashir, M. F., Ma, B., Bilal, Komal, B., Bashir, M. A., Tan, D., & Bashir, M., 2020). Walau bagaimanapun, terdapat juga perkaitan linear negatif antara suhu dan kumulatif kes COVID-19 (Prata et al., 2020). Banyak kajian telah mencadangkan bahawa penyebaran COVID-19 lebih banyak berlaku dalam iklim yang sejuk dan sederhana berbanding iklim panas dan tropika, selaras dengan tingkah laku virus pernafasan selesema bermusim (Bloom-Feshbach K, Alonso WJ, Charu V, 2013).

Pembelajaran mesin dan pembelajaran mendalam (*deep learning*) ialah salah satu teknologi kecerdasan buatan (AI) yang memberikan keupayaan ramalan yang hebat dan mengatasi prestasi pemodelan statistik tradisional (Beam & Kohane, 2018; Miguel-Hurtado et al., 2016; Singal et al., 2016). Walaupun nilai ramalan memberi keputusan yang sangat baik,

algoritma ini tidak biasa digunakan dalam kajian data kesihatan awam. Tujuan penyelidikan ini adalah untuk menggunakan algoritma pembelajaran mendalam kepada set data bersepadu (epidemiologi dan data iklim) dan menggunakan rangka kerja pemodelan memori jangka pendek (LSTM) untuk menganggarkan trend COVID-19 di Malaysia. LSTM telah berjaya digunakan untuk menjangkakan wabak denggi dan *influenza* (Leonenko, V. N. , Bochenina, K. O. , & Kesarev, S. A., 2017). Tambahan pula, penyelidikan terdahulu telah mengkaji adakah kelembapan relatif dan kelembapan mutlak penting dalam penularan COVID-19, manakala kajian tentang virus *influenza* mendedahkan bahawa kelembapan khusus adalah aspek yang penting dalam penghantaran virus. Hasilnya, kajian ini adalah untuk meneliti kelembapan tertentu serta parameter iklim yang lain untuk lebih memahami penghantaran dan ramalan COVID-19 di Malaysia.

## **2 PENYATAAN MASALAH**

Penyebaran virus COVID-19 begitu pantas adalah disebabkan oleh kekurangan pemahaman yang tetap tentang etologi penyakit dan corak penghantaran semasa fasa awal wabak. Virus COVID-19 dipercayai berjangkit terus melalui titisan pernafasan daripada orang lain yang berjangkit kepada orang yang sihat dalam julat satu meter. Selain itu, orang yang sihat juga boleh dijangkiti secara tidak langsung melalui permukaan yang tercemar ("World Health Organization", 2020). Namun, penyelidikan tentang faktor-faktor yang mempengaruhi penyebaran virus COVID-19 adalah sangat kurang dan ini mungkin salah satu faktor yang menyebabkan virus COVID-19 tidak dapat ditahankan.

Terdapat ramai penyelidik telah ditanya adakan suhu tinggi boleh melambatkan penyebaran virus COVID-19 dalam kalangan orang ramai. Selain itu, terdapat juga banyak siasatan tentang kesan pembolehubah cuaca terhadap penularan COVID-19 dalam kalangan orang (A. Bernstein, 2021).

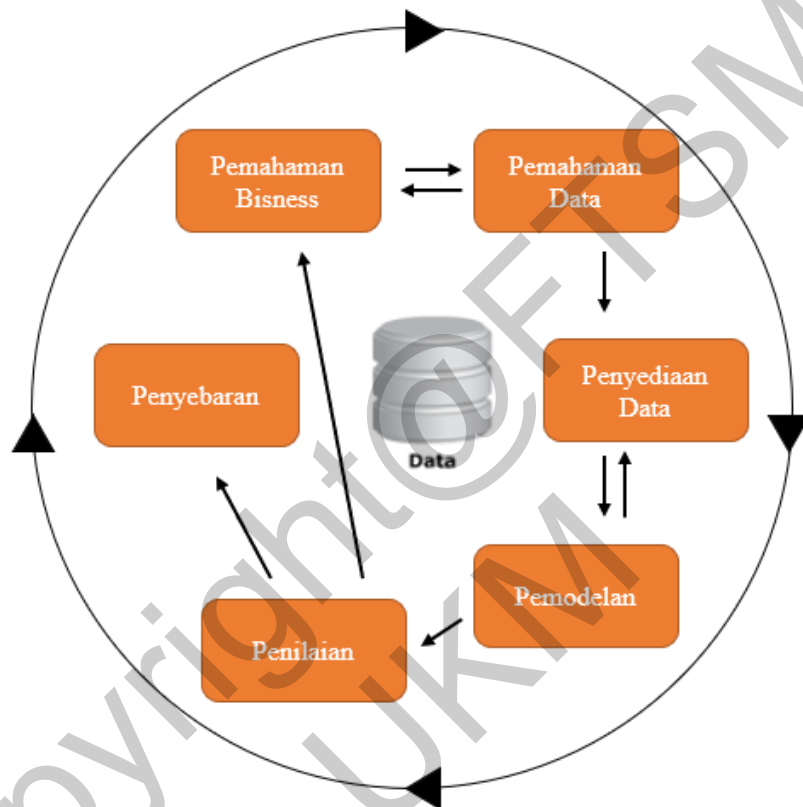
## **3 OBJEKTIF KAJIAN**

Objektif kajian ini adalah:

1. Membangun satu model pembelajaran mendalam bagi meramal kes COVID-19 di Malaysia dengan menggunakan data cuaca
2. Menilai kesan cuaca kepada corak kes COVID-19 di Malaysia

## 4 METOD KAJIAN

Metodologi yang digunakan dalam kajian ini adalah CRISP-DM (*The Cross Industry Standard Process for Data Mining*). Metodologi CRISP-DM ini merupakan model proses yang mempunyai enam fasa yang melabelkan kitaran hidup sains data secara semula jadinya (*"Data Science Process Alliance"*, 2021). Rajah 1 menunjukkan fasa-fasa metodologi CRISP-DM.



Rajah 4.1: Fasa-fasa Metodologi CRISP-DM

### 4.1 Fasa Pemahaman Bisness

Fasa pemahaman bisnes (*Business Understanding*) melibatkan proses memahami keperluan projek dari segi perniagaan dan melibatkan pengetahuan ini ke dalam definisi masalah perlombongan data dan membuat perancangan awal bagi mencapai objektif yang perlu ditentukan dengan awal dalam fasa ini supaya mendapat hala tuju yang jelas dalam menjalankan penyelidikan tersebut. Objektif bagi projek tersebut adalah untuk menghasilkan ramalan kes COVID-19 dengan menggunakan data meteorologi melalui pendekatan pembelajaran mesin. Kajian yang lepas telah menunjukkan bahawa cuaca dan iklim adalah salah satu faktor yang mempengaruhi penghantaran virus dan algoritma LSTM telah berjaya

digunakan untuk menjangkakan wabak denggi *influenza* dan SARS. Selain itu, sorotan kajian dan kajian kesusasteraan mengenai faktor iklim dan kes COVID-19 telah dilakukan dapat memudahkan proses pembangunan model bagi projek ini. Saiz data juga memainkan peranan yang penting di mana semakin besar saiz data, potensi untuk menghasilkan keputusan yang tepat semakin tinggi.

#### **4.2 Fasa Pemahaman Data**

Fasa Pemahaman Data (*Data Understanding*) merupakan fasa yang mengumpulkan data dan memahami data untuk mengenal pasti masalah-masalah yang terdapat dalam data tersebut. Fasa pemahaman data merupakan fasa yang kedua dalam meteorologi CRISP-DM yang digunakan dalam kajian tersebut. Sumber data yang digunakan untuk kajian tersebut adalah data kes COVID-19 daripada laman web repositori *github* Kementerian Malaysia, sebanyak 5161 rekod data kes aktif harian COVID-19 dalam tempoh Oktober 2020 hingga Oktober 2021 telah diekstrak untuk kajian ini. Data meteorologi yang diperolehi dari laman web *timeanddate* juga berbilang 5161 iaitu rekod cuaca harian bagi setiap negeri dari Oktober 2020 hingga Oktober 2021. Selain itu, korelasi antara atribut-atribut data akan diterokai untuk memahami hubungan antara atribut-atribut yang digunakan dengan lebih dalam. Seterusnya, data-data akan diproseskan untuk mengelakkan data yang berformat salah dan bernilai kosong untuk menghasilkan data yang berkualiti tinggi.

#### **4.3 Fasa Penyediaan Data**

Fasa penyediaan data (*Data Preparation*) merupakan fasa yang melibatkan penggabungan data serta pembersihan data. Fasa ini bertujuan untuk menghasilkan set data yang bersih dan sesuai untuk digunakan dalam pembangunan model pembelajaran mesin. Pemrosesan data kes COVID-19 dan data meteorologi akan dijalankan menggunakan *Microsoft Excel*. Fasa penyediaan data adalah sangat berguna bagi memperoleh pengetahuan melalui penyediaan dan penapisan data yang dipilih bagi kajian tersebut. Langkah-langkah pemrosesan data yang digunakan termasuklah penggabungan data, penerokaan data, pembersihan data, pendiskretan data dan pemilihan ciri-ciri-penting. Fasa ini dapat menghasilkan set data yang siap sedia digunakan untuk fasa pembangunan model.

#### **4.4 Fasa Pemodelan**

Pembangunan algoritma dalam kajian ini adalah untuk meramalkan kes COVID-19 bersama dengan data meteorologi. Algoritma pembelajaran mendalam LSTM Univariat dan LSTM

Multivariat telah dibangunkan untuk meramal kes aktif COVID-19 di setiap negeri di Malaysia. Set data bagi bulan Oktober 2021 telah dipilih sebagai set ujian dan set data yang lain digunakan sebagai set latih.

#### **4.5 Fasa Penilaian**

Fasa Penilaian menjalankan penilaian terhadap model mendalam supaya dapat mencapai objektif bisnes yang dinyatakan sebelumnya. Prestasi model akan dinilai untuk memastikan keputusan ramalan tersebut adalah memuaskan. Proses pembinaan model akan dijalankan sekali lagi sekiranya keputusan ramalan tersebut tidak memuaskan dengan mengkaji punca-punca yang berkemungkinan mengakibatkan keputusan ramalan yang tidak tepat atau mengandungi terlalu banyak ralat. Dalam kajian ini, penilaian metrik yang akan digunakan untuk menilai model ramalan kes aktif harian COVID-19 termasuklah, *Mean Absolute Error* (MAE), *Root Mean Squared Error* (RMSE), *Mean Absolute Percentage Error* (MAPE) dan Skor-R2.

#### **4.6 Fasa Penyebaran**

Fasa Penyebaran merupakan fasa terakhir proses pembangunan model ramalan kes aktif harian COVID-19. Keputusan ramalan yang dihasilkan adalah bersedia untuk disebar dan digunakan oleh orang ramai. Keputusan model ramalan bagi setiap negeri di Malaysia akan dilaporkan dan dipaparkan melalui alat visualisasi *Tableau* secara atas talian supaya pihak-pihak yang memerlukan dapat memahami dan meneroka hasil keputusan yang dikehendaki dengan cepat dan mudah. Pemantauan dan penyelenggaraan perlu dijalankan dari semasa ke semasa untuk memastikan hasil ramalan adalah betul dan tiada maklumat palsu disebar kepada orang ramai.

#### **4.7 Spesifikasi Keperluan**

Keperluan perkakasan dan perisian adalah penting kepada pengguna untuk memastikan proses pembangunan model ramalan dapat dijalankan dengan lancar. Jadual 4.1 menunjukkan spesifikasi keperluan perkakasan manakala jadual 4.2 menunjukkan spesifikasi keperluan perisian.

Jadual 4.1: Spesifikasi perkakasan

<b>Pemproses</b>	<b>Intel Core i7-9750H</b>
Pemacu keadaan pepejal	512GB
Memori Capaian Rawak (RAM)	16GB
Sistem Operasi	Windows 11
Unit Pemprosesan Grafik (GPU)	NVIDIA GeForce GTX 1050

Jadual 4.2: Spesifikasi Keperluan Perisian

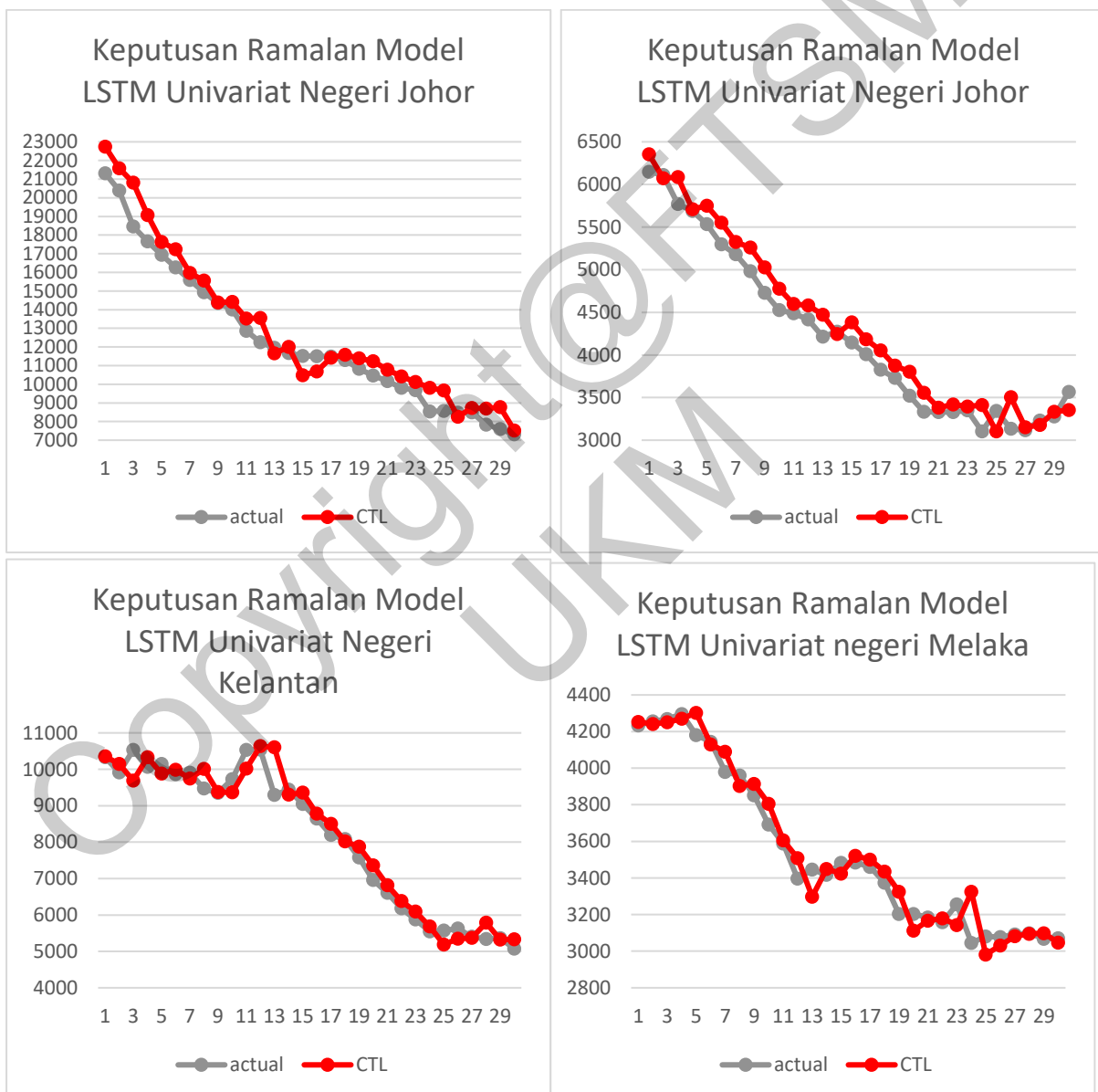
<b>Perisian</b>	<b>Perincian</b>
<i>Microsoft Word</i>	Menulis tesis projek akhir tahun
<i>Microsoft Excel</i>	Tempat simpanan maklumat dan data yang diperoleh
<i>Google Colab</i>	IDE untuk menulis algoritma <i>Python</i> bagi proses pemodelan
<i>Tableau</i>	Alat visualisasi keputusan ramalan kes COVID-19 melalui data meteorologi
<i>Windows 11</i>	Digunakan untuk membangunkan dan menyokong model ramalan

## 5 HASIL KAJIAN

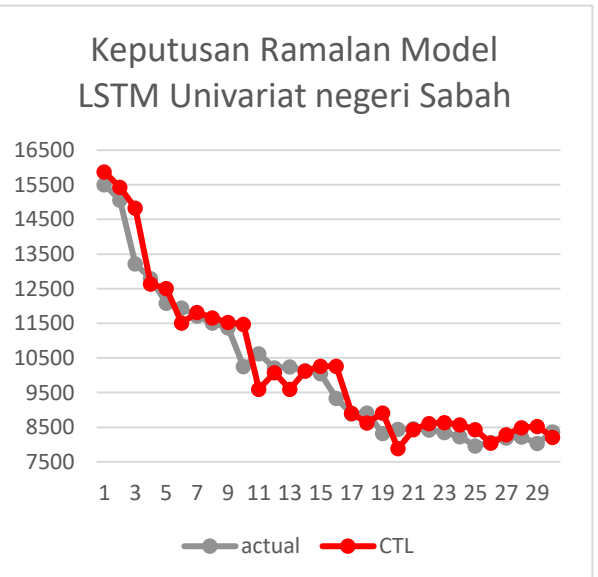
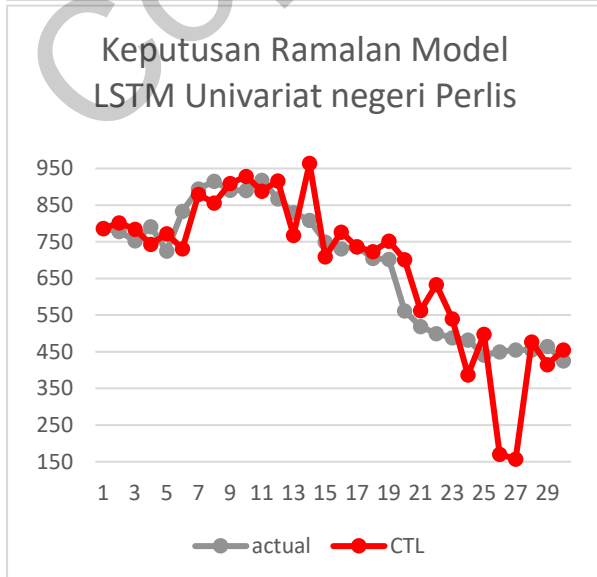
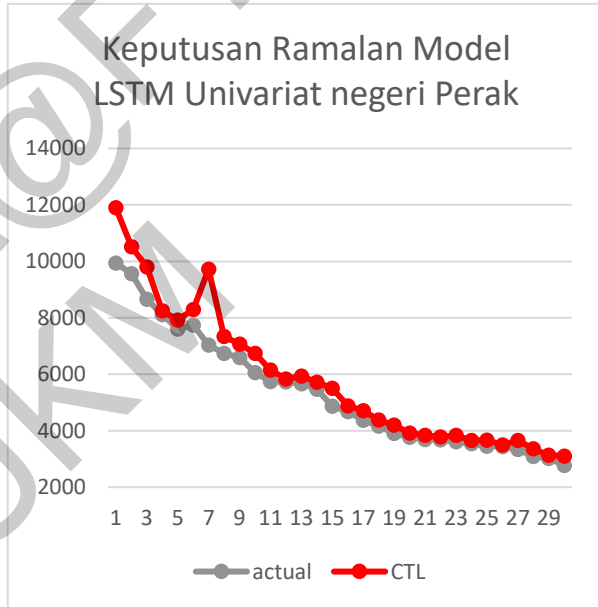
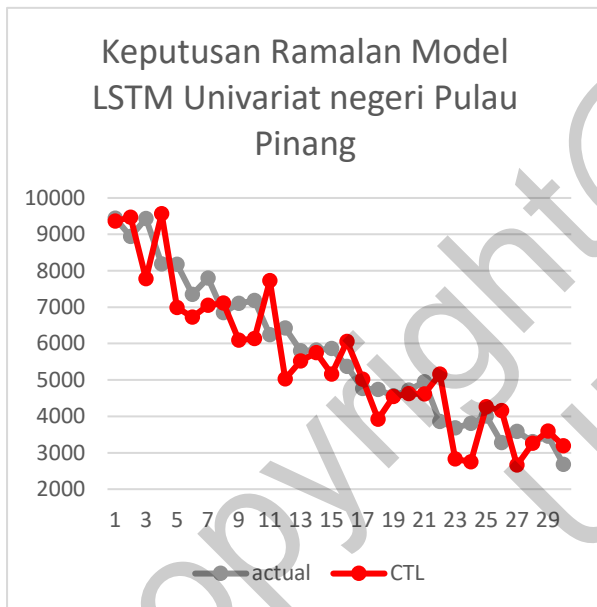
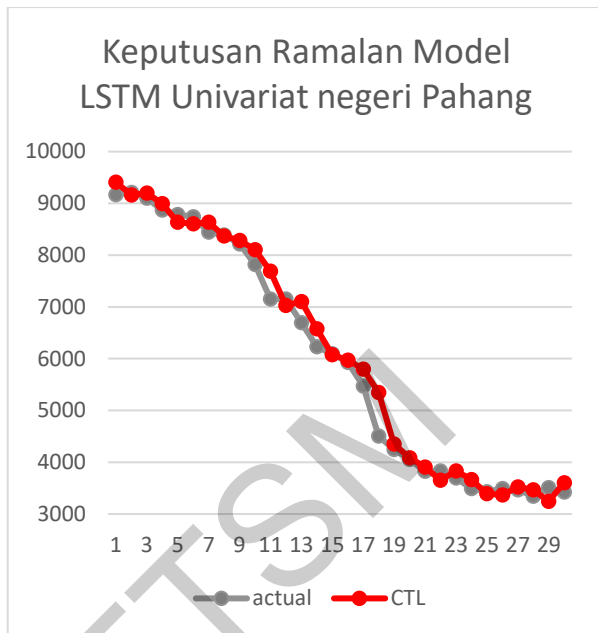
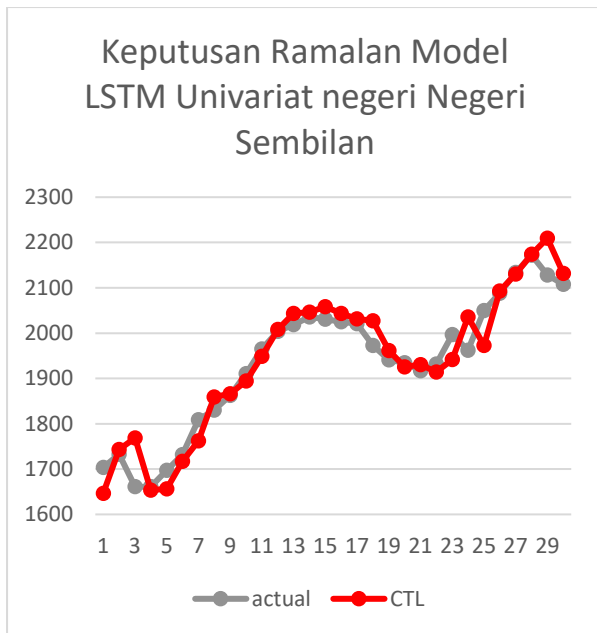
Pemodelan LSTM Univariat dan LSTM Multivariat telah dibangunkan untuk meramalkan kes aktif COVID-19 bagi setiap negeri di Malaysia. Pemodelan LSTM Multivariat telah dibangunkan untuk mendapati bahawa adakah model LSTM Multivariat dapat meningkatkan ketepatan dalam ramalan kes aktif COVID-19 harian bagi negeri masing-masing di Malaysia. Dalam kajian tersebut, set data bagi Oktober 2021 telah digunakan sebagai set ujian manakala set data yang lain digunakan sebagai set latihan.

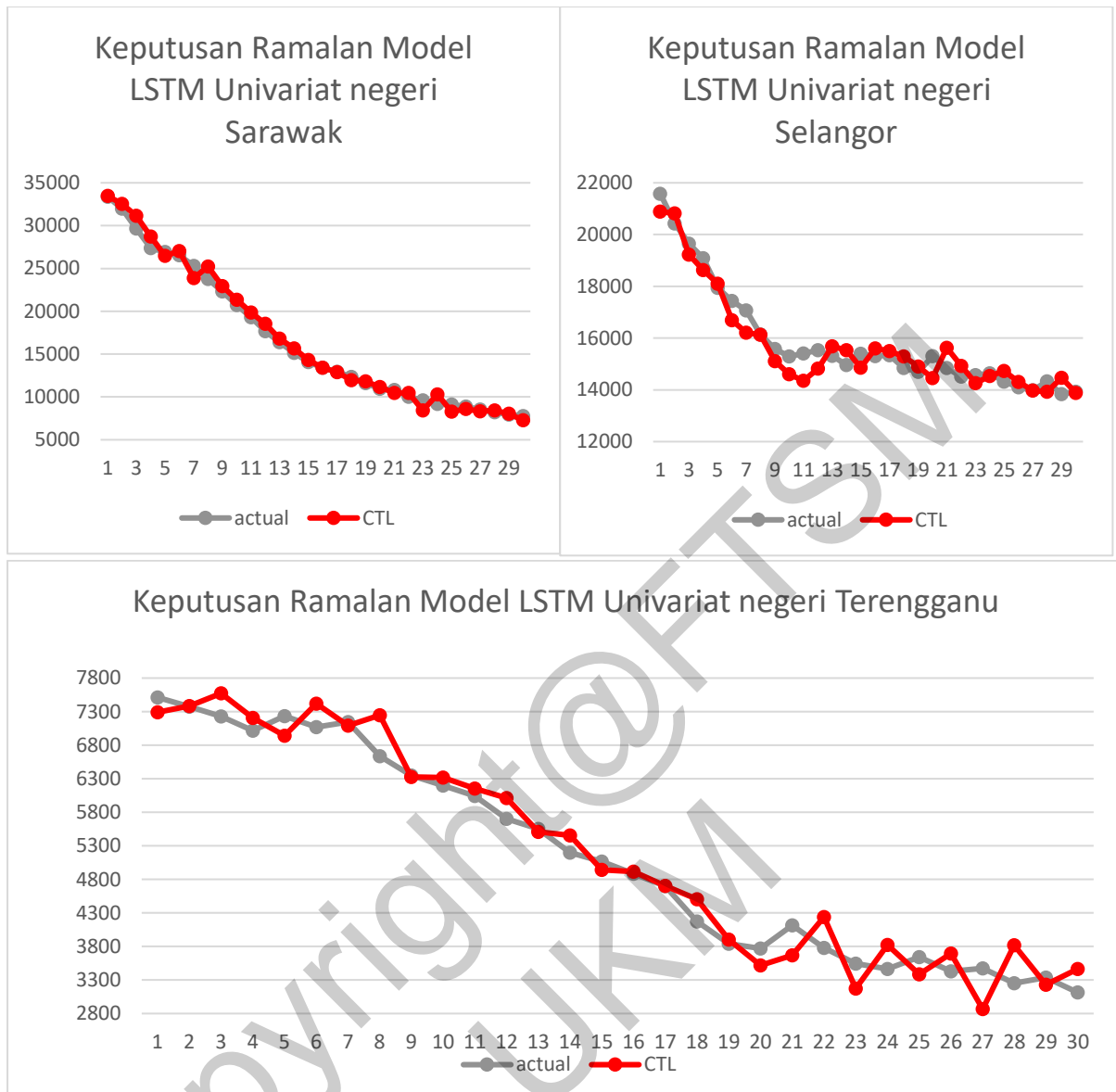
### 5.1 Keputusan Ramalan LSTM Univariat bagi setiap negeri

Pemodelan LSTM *Univariate* telah dibangunkan kepada setiap negeri untuk meramalkan kes aktif COVID-19 harian bagi setiap negeri masing-masing. Rajah 5.1 telah menunjukkan graf nilai ramalan berbanding nilai sebenar bulan Oktober 2021 bagi setiap negeri. Berdasarkan rajah-rajah yang dinyatakan boleh mendapati bahawa model LSTM Univariat dapat meramalkan kes aktif COVID-19 harian dengan baik bagi kebanyakan negeri kecuali negeri Pulau Pinang, Perlis dan Terengganu.









Rajah 5.1: Keputusan Ramalan bagi setiap negeri

## 5.2 Perbandingan penilaian model LSTM Univariat bagi setiap negeri

Dalam kajian ini, penilaian metrik yang digunakan adalah *Mean Absolute Error (MAE)*, *Root Mean Squared Error (RMSE)*, *Mean Absolute Percentage Error (MAPE)* dan Skor-R2. Jadual 5.1 telah menunjukkan metrik penilaian model LSTM Univariat yang digunakan untuk meramalkan kes aktif harian COVID-19 di setiap negeri di Malaysia. Rajah 5.2 telah menunjukkan graf perbandingan semua metrik penilaian model antara setiap negeri.

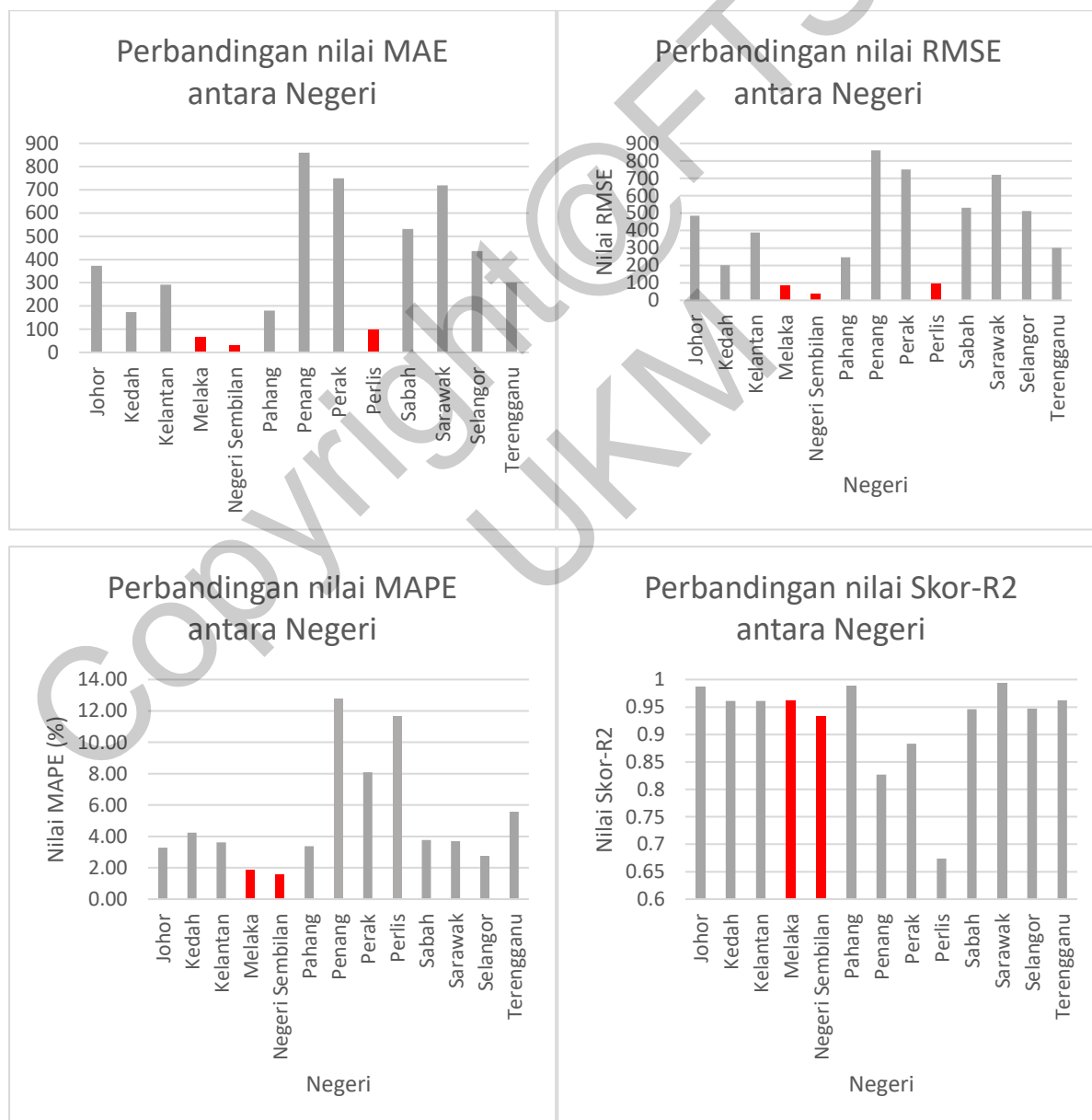
Jadual 5.1: Perbandingan metrik penilaian model LSTM Univariat antara negeri

Negeri	Penilaian Metriks			
	MAE	RMSE	MAPE	R2
Johor	373.346	484.344	3.278	0.986

Kedah	173.213	201.056	4.247	0.961
Kelantan	291.999	388.478	3.62	0.961
Melaka	63.864	85.373	1.856	0.962
Negeri Sembilan	29.44	39.695	1.56	0.932
Pahang	179.645	247.705	3.376	0.988
Penang	860.186	860.186	12.769	0.827
Perak	749.734	749.734	8.083	0.883
Perlis	95.996	95.996	11.661	0.674
Sabah	531.455	531.455	3.774	0.946
Sarawak	719.303	719.303	3.692	0.993
Selangor	436.568	512.711	2.746	0.947
Terengganu	300.994	300.994	5.578	0.962

Berdasarkan rajah 5.2 boleh mendapati bahawa nilai MAE bagi negeri Negeri Sembilan adalah terendah berikutan dengan negeri Melaka dan negeri Perlis. Nilai MAE yang lebih rendah bermaksud prestasi model bagi negeri tersebut adalah lebih tinggi. Graf perbandingan nilai MAE telah menunjukkan bahawa prestasi model LSTM Univariat untuk meramalkan kes aktif harian COVID-19 di negeri Negeri Sembilan adalah tertinggi dengan mencapai nilai MAE sebanyak 29.44 sahaja diikuti dengan negeri Melaka yang mencapai nilai 63.864 dan negeri Perlis yang mencapai nilai 95.996. Bagi metrik penilaian RMSE, nilai yang lebih rendah menunjukkan prestasi model yang lebih tinggi. Graf perbandingan nilai RMSE telah menunjukkan bahawa prestasi model LSTM Univariat bagi negeri Negeri Sembilan adalah tertinggi yang mencapai nilai 39.695 diikuti dengan negeri Melaka yang mencapai nilai 85.373 dan Perlis yang mencapai nilai 95.996 dan ini menunjukkan prestasi model bagi ketiga-tiga negeri tersebut adalah sama dengan keputusan penilaian MAE yang sebelumnya. Seperti metrik penilaian MAE, negeri Pulau Pinang juga mencapai nilai RMSE yang tertinggi iaitu sebanyak 860.186 dan ini menunjukkan bahawa model LSTM Univariat tidak dapat meramalkan kes aktif harian COVID-19 bagi negeri Pulau Pinang dengan tepat. Metrik penilaian MAPE ialah metrik ralat yang digunakan untuk mengukur prestasi model pembelajaran mesin regresi. Nilai metrik penilaian MAPE yang lebih rendah bermaksud prestasi bagi model tersebut adalah lebih tinggi. Graf perbandingan nilai MAPE antara negeri menunjukkan bahawa prestasi model bagi negeri Negeri Sembilan adalah tertinggi dengan mencapai nilai MAPE sebanyak 1.56% diikuti dengan negeri Melaka yang mencapai nilai

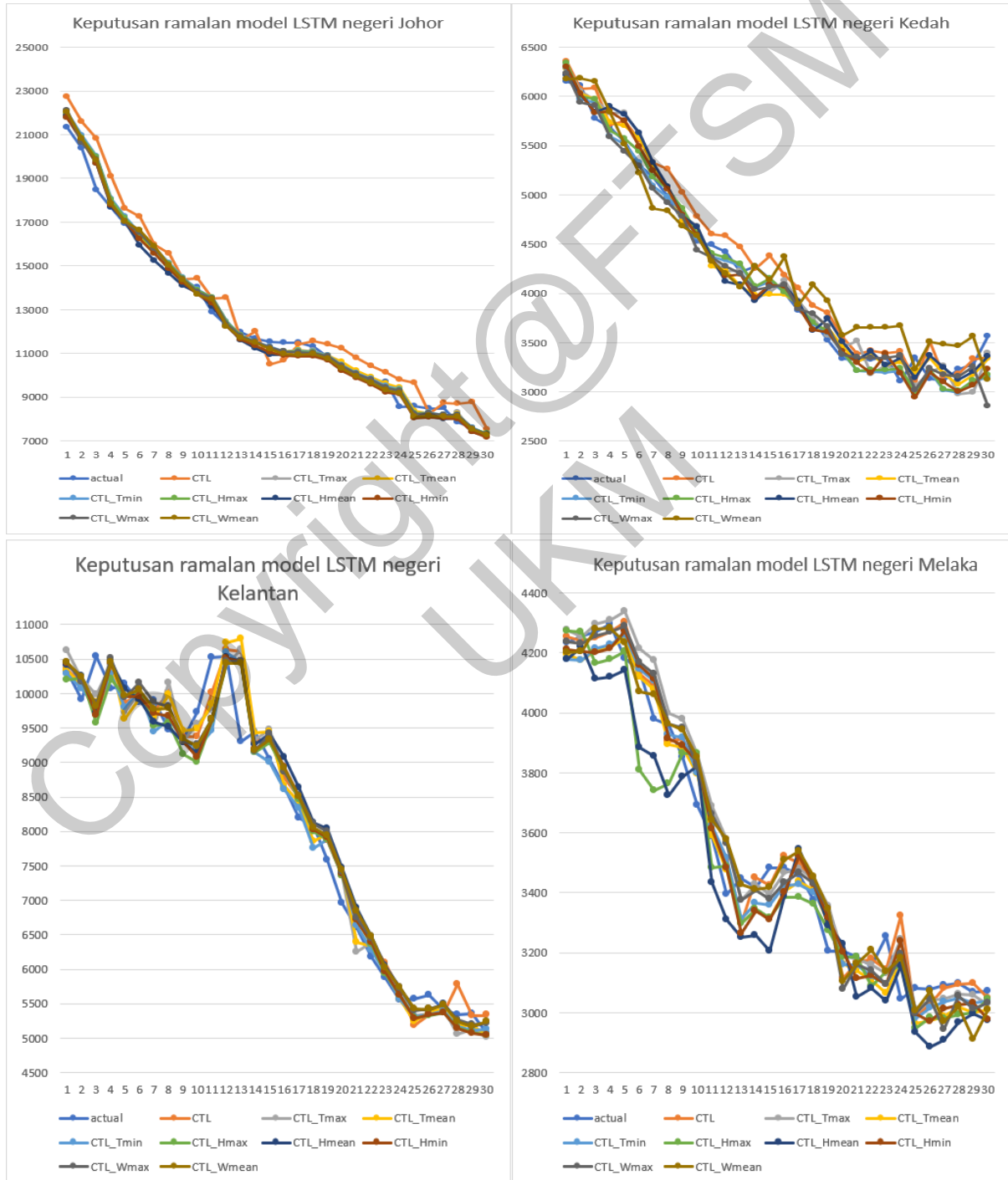
1.86%. Hanya kedua-dua negeri tersebut mencapai nilai MAPE yang rendah daripada 2% dan negeri Pulau Pinang telah mencapai nilai MAPE sebanyak 12.77%. Berdasarkan metrik-metrik penilaian MAE, RMSE, dan MAPE boleh mendapati bahawa model LSTM Univariat tidak sesuai digunakan untuk meramal kes aktif COVID-19 bagi negeri Pulau Pinang. Bagi nilai Skor-R2, nilai yang tinggi menunjukkan bahawa varians model adalah serupa dengan nilai sebenar, manakala nilai R2 yang rendah menunjukkan bahawa kedua-dua nilai tersebut tidak berkait rapat. Dalam graf perbandingan nilai Skor-R2 boleh mendapati bahawa skor R2 bagi negeri Sarawak adalah tertinggi iaitu sebanyak 0.993 diikuti dengan negeri Pahang yang mencapai nilai 0.988 dan negeri Johor yang mencapai nilai 0.986. Negeri Perlis telah mencapai nilai R2 yang terendah iaitu sebanyak 0.674.

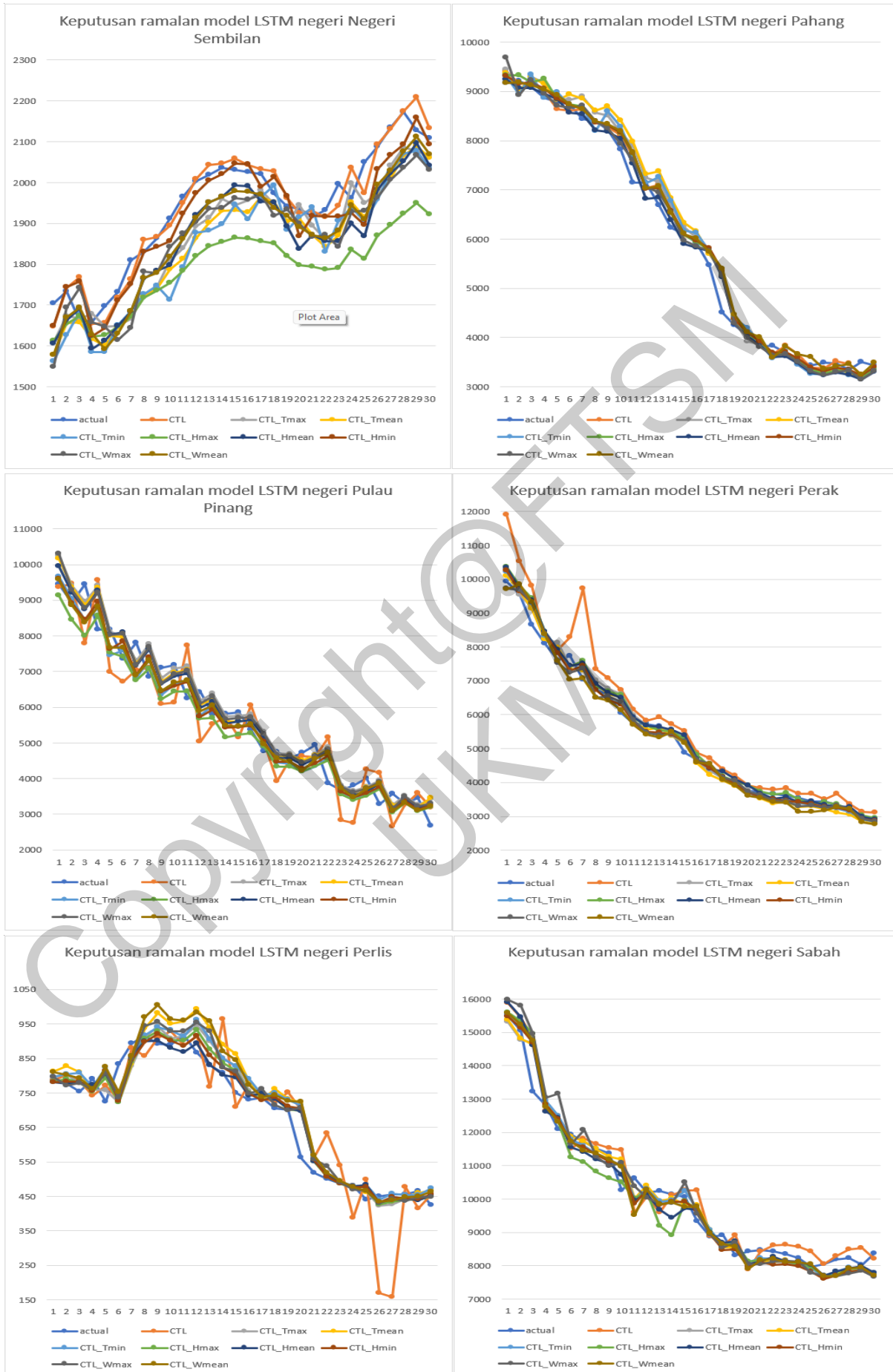


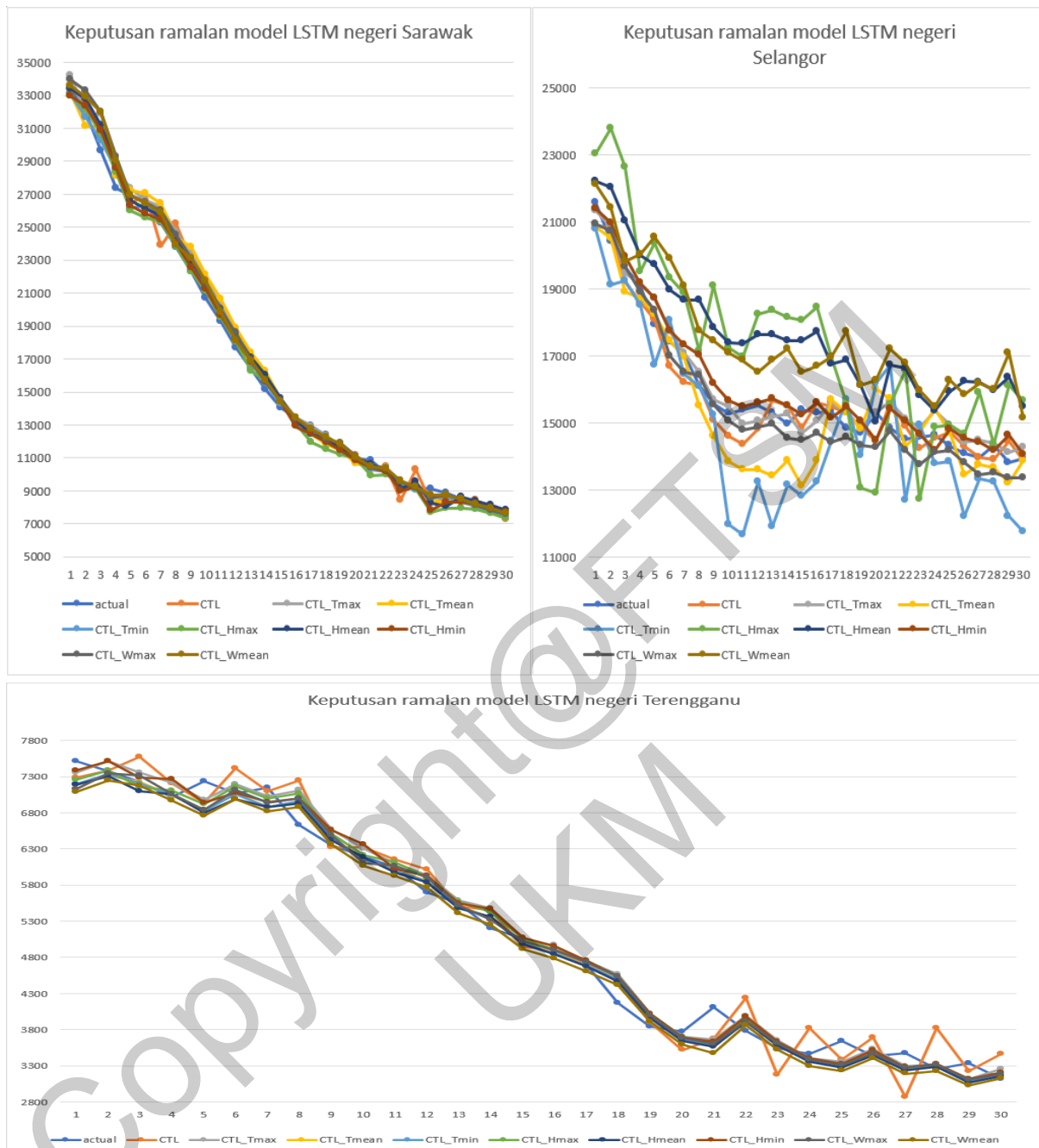
Rajah 5.2: Perbandingan metrik penilaian antara negeri

### 5.3 Keputusan Ramalan LSTM Multivariat bagi setiap negeri

Pemodelan LSTM Multivariat telah dibangunkan kepada setiap negeri untuk meningkatkan prestasi model dalam ramalan kes aktif COVID-19 harian bagi negeri masing-masing di Malaysia. Rajah 5.3 telah menunjukkan graf nilai ramalan berbanding nilai sebenar bulan Oktober 2021 bagi setiap negeri. Berdasarkan rajah-rajah di bawah, boleh mendapati bahawa prestasi model LSTM Multivariat adalah agak memuaskan dalam ramalan kes aktif COVID-19 harian.





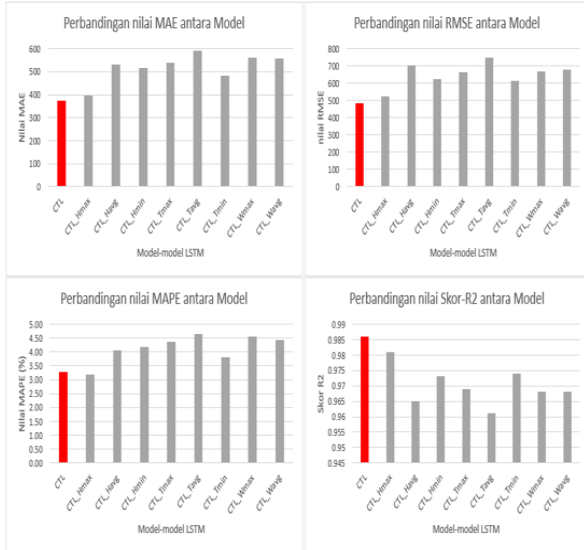


Rajah 5.3: Keputusan Ramalan bagi setiap negeri

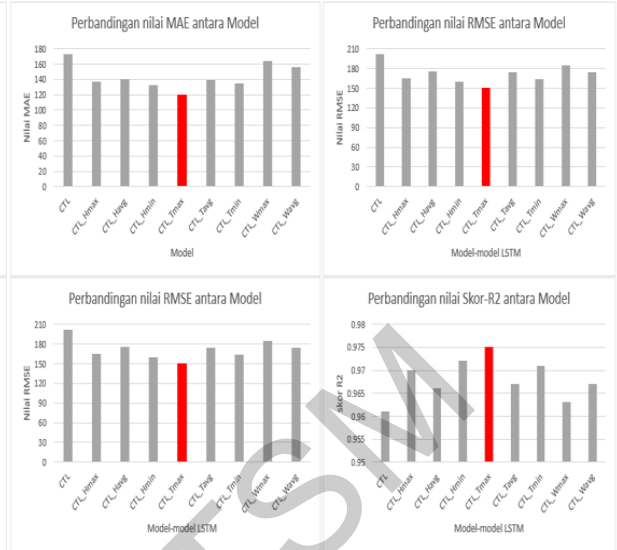
#### 5.4 Keputusan Prestasi Model Ramalan Mengikut Negeri

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

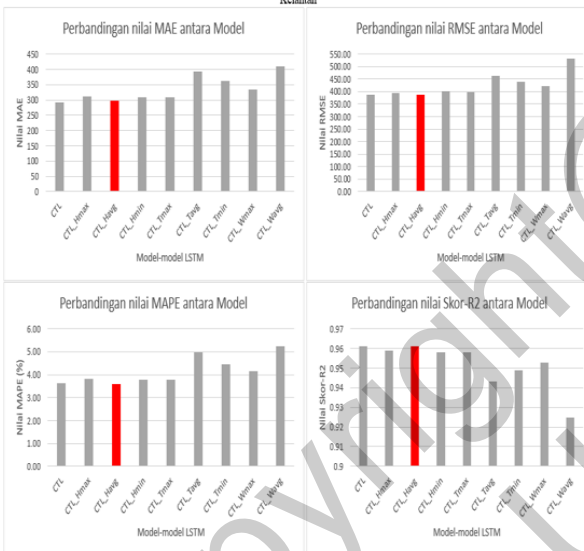
Jobor



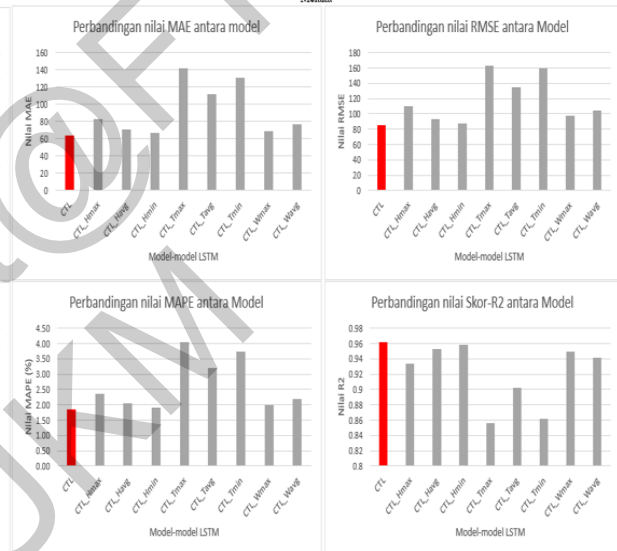
Kedah



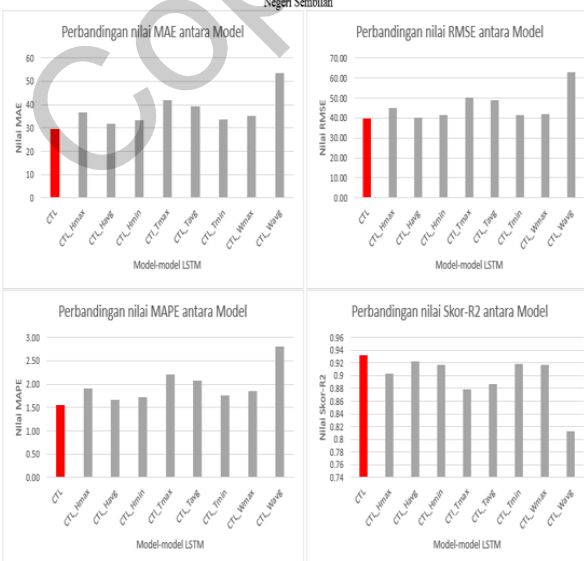
Kelantan



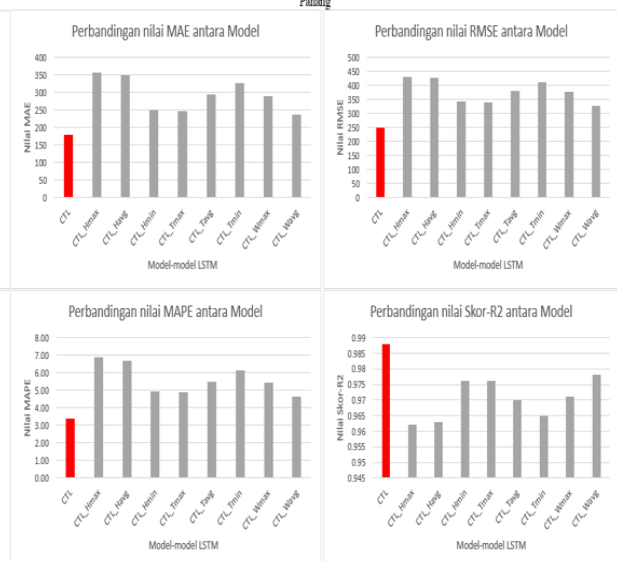
Melaka



Negeri Sembilan



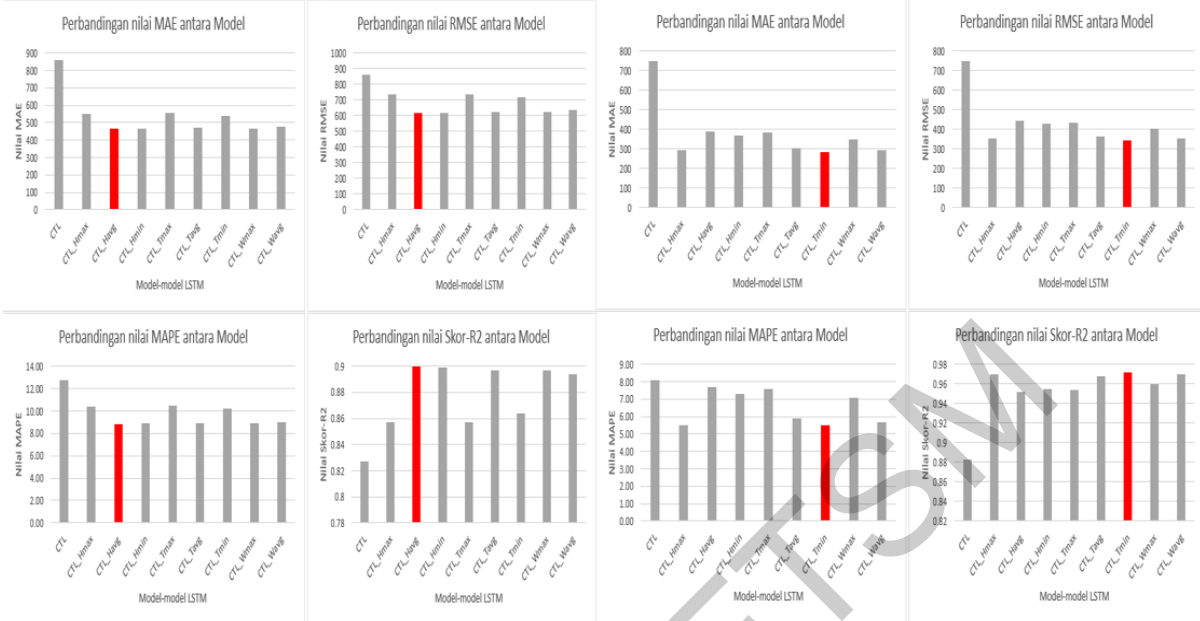
Pahang





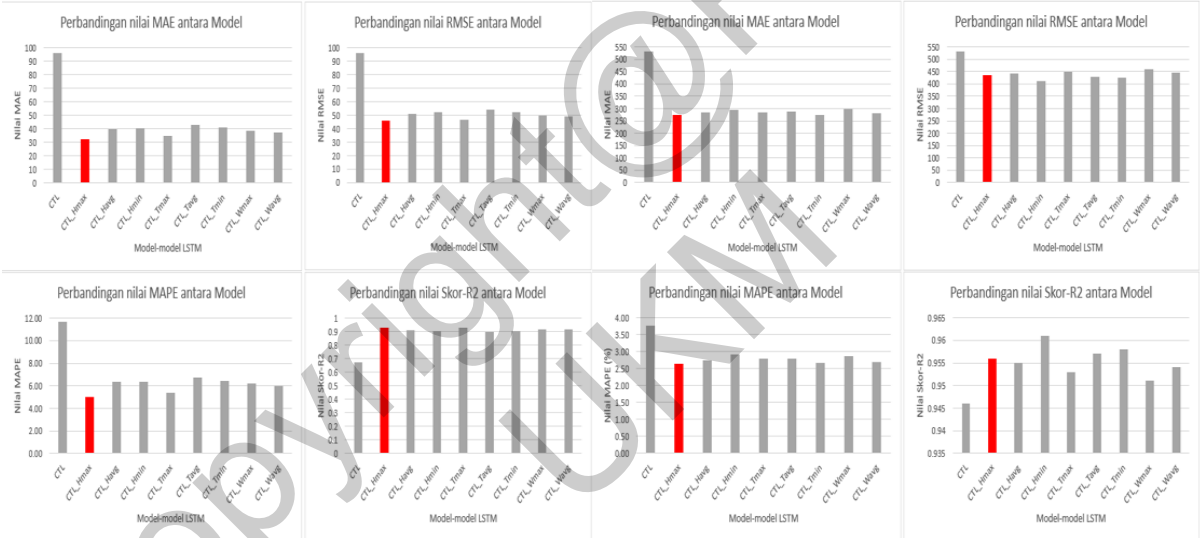
Pulau Pinang

Perak



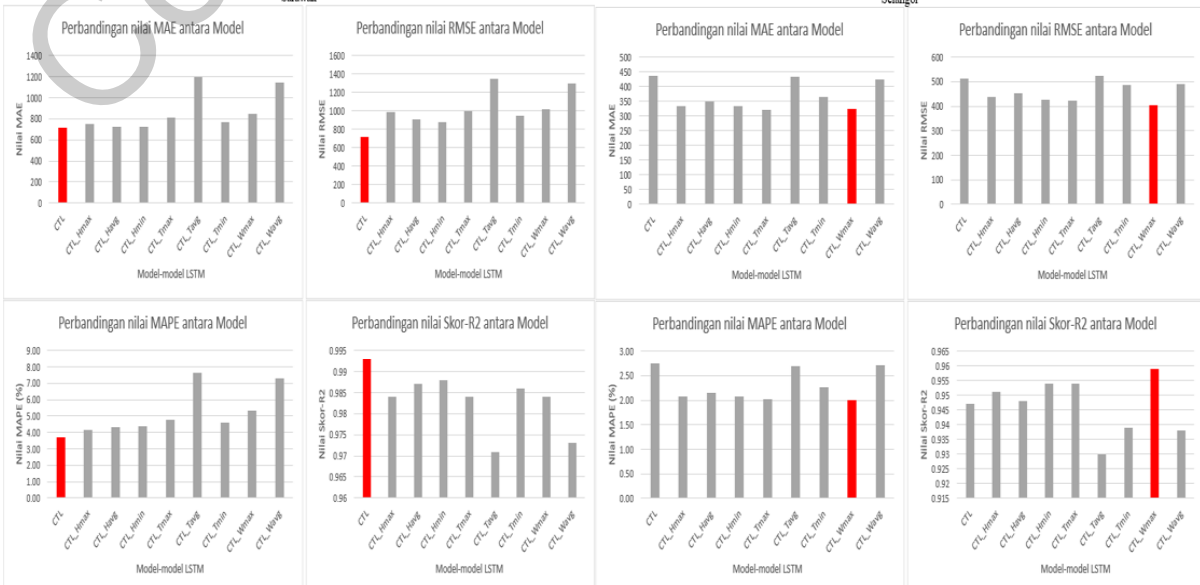
Perlis

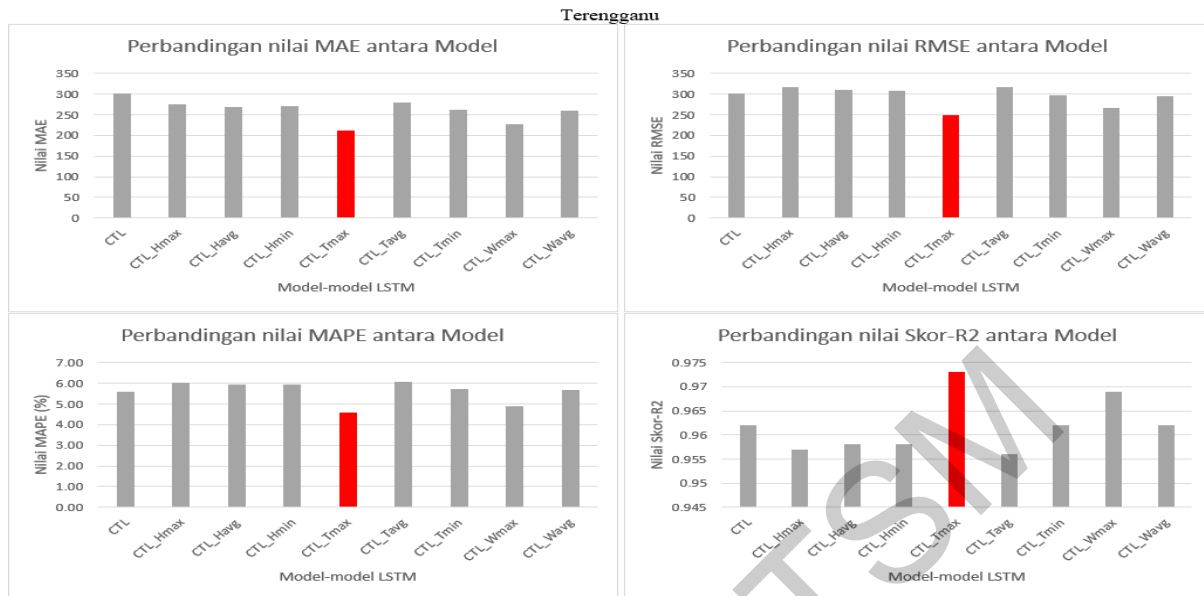
Sabah



Sarawak

Selangor





Rajah 5.4: Keputusan Prestasi Model Ramalan Mengikut Negeri

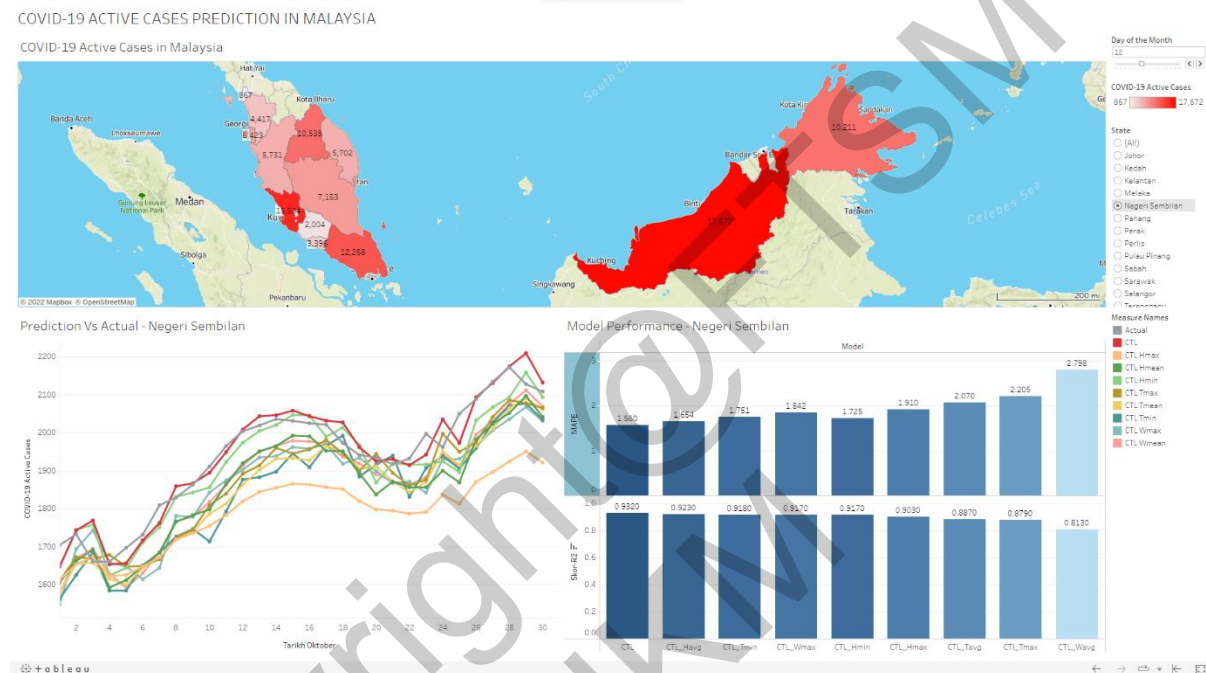
Berdasarkan jadual 5.2 di bawah, boleh mendapati bahawa negeri Johor, Melaka, Negeri Sembilan, Pahang dan Sarawak, model LSTM Univariat berprestasi lebih tinggi daripada model LSTM Multivariat bagi negeri-negeri tersebut. Model LSTM Multivariat dapat mencapai prestasi yang lebih tinggi berbanding dengan model LSTM Univariat bagi negeri-negeri yang lain. Seterusnya, Negeri Perak, Perlis dan Selangor dapat mencapai keputusan yang agak sama dengan menggunakan atribut meteorologi lebih daripada satu. Oleh itu, boleh mendapati bahawa korelasi antara atribut meteorologi dengan kes aktif harian COVID-19 bagi negeri Johor, Melaka, Negeri Sembilan, Pahang dan Sarawak adalah tidak kuat manakala korelasi antara atribut meteorologi dengan kes aktif harian COVID-19 bagi negeri-negeri yang lain adalah lebih kuat daripada negeri yang dinyatakan atas.

Jadual 5.2: Model ramalan LSTM terbaik bagi setiap negeri

Negeri	Model Terbaik	Negeri	Model Terbaik
Johor	CTL	Perak	CTL_Tmin
Kedah	CTL_Tmax	Perlis	CTL_Hmax
Kelantan	CTL_Havg	Sabah	CTL_Hmax
Melaka	CTL	Sarawak	CTL
Negeri Sembilan	CTL	Selangor	CTL_Wmax
Pahang	CTL	Terengganu	CTL_Tmax
Pulau Pinang	CTL_Havg		

## 5.5 Papan Pemuka

Rajah 5.5 di bawah telah menunjukkan papan pemuka yang dihasilkan dalam kajian tersebut. Di sebelah kanan papan pemuka, terdapat penapis-penapis yang boleh diinteraksi oleh pengguna. Pengguna boleh memilih tarikh untuk mendapati kes aktif harian COVID-19 yang sebenar bagi setiap negeri di Malaysia. Pengguna juga boleh memilih negeri untuk mendapat keputusan ramalan model-model LSTM bagi negeri tersebut dan keputusan model yang berprestasi tertinggi melalui penapis tersebut.



Rajah 5.5: Papan Pemuka Ramalan Kes Aktif COVID-19 di Malaysia

## 6 KESIMPULAN

Projek kajian ini adalah menggunakan algoritma-algoritma pembelajaran mesin untuk meramalkan kes COVID-19 melalui data meteorologi. Pembangunan model ramalan telah menepati matlamat perniagaan yang telah dinyatakan pada perancangan awal projek. Set data yang digunakan dalam projek ini iaitu set data harian kes COVID-19 dan set data meteorologi telah menjalankan pemprosesan pada peringkat awal untuk menjamin kualiti data dan memastikan tiada ralat pada data yang akan mempengaruhi keputusan ramalan. Keputusan ramalan kes COVID-19 melalui data meteorologi telah dihasilkan melalui perbandingan pencapaian antara modal ramalan dengan menggunakan penilaian metrik yang sama. Hasil kajian yang didapati dalam kajian ini adalah model LSTM Univariat dapat meramalkan kes aktif COVID-19 dengan baik bagi kebanyakan negeri di Malaysia. Model LSTM Multivariat dapat meningkatkan prestasi ramalan kes aktif COVID-19 dengan menggunakan atribut-atribut

seperti suhu tertinggi, suhu terendah, kelembapan tertinggi, kelembapan min, dan kelajuan angin tertinggi. Keputusan ramalan yang diperoleh diharapkan dalam membantu pihak-pihak untuk menjalankan pencegahan bagi mengelakkan penyebaran virus COVID-19 secara wabak.

## 7 RUJUKAN

- Valsamatzi-Panagiotou, A., Penchovsky, R. 2021. Environmental factors influencing the transmission of the coronavirus 2019: a review. *Environ Chem Lett* 20, 1603–1610. <https://doi.org/10.1007/s10311-022-01418-91>.
- Nick, H. 2020. CRISP-DM. Data Science Process Alliance. <https://www.datascience-pm.com/crisp-dm-2/>.
- Pal, M., Berhanu, G., Desalegn, C., & Kandi, V. 2020. Severe Acute Respiratory Syndrome Coronavirus-2 (SARS-CoV-2). <https://doi.org/10.7759/cureus.7423>.
- Bashir, M. F., Ma, B., Bilal, Komal, B., Bashir, M. A., Tan, D., Bashir, M. 2020. Correlation between climate indicators and COVID-19 pandemic in New York, USA. *The Science of the total environment*, 728, 138835. <https://doi.org/10.1016/j.scitotenv.2020.138835>.
- Bloom-Feshbach K, Alonso WJ, Charu V, Tamerius J, Simonsen L, Miller MA, Viboud C. 2013. Latitudinal variations in seasonal activity of influenza and respiratory syncytial virus (RSV): a global comparative review. doi: 10.1371/journal.pone.0054445. Epub 2013 Feb 14. PMID: 23457451; PMCID: PMC3573019.
- Miguel-Hurtado, Oscar & Guest, Richard & Stevenage, Sarah & Neil, Greg & Black, Sue. 2016. Comparing Machine Learning Classifiers and Linear/Logistic Regression to Explore the Relationship between Hand Dimensions and Demographic Characteristics. *PLoS ONE*. 11. e0165521. 10.1371/journal.pone.0165521.
- Leonenko, V. N. , Bochenina, K. O. , & Kesarev, S. A. 2017. Influenza peaks forecasting in Russia: Assessing the applicability of statistical methods. *Procedia Computer Science*, 108, 2363–2367. 10.1016/j.procs.2017.05.196.
- Modes of transmission of virus causing COVID-19: implications for IPC precaution recommendations. 2020. World Health Organization. <https://www.who.int/news->

room/commentaries/detail/modes-of-transmission-of-virus-causing-covid-19-implications-for-ipc-precaution-recommendations.

- P. Sowa. 2020. Cross Industry Standard for Data Mining. [https://hpi.de/fileadmin/user\\_upload/fachgebiete/rabl/Lectures/PDE\\_Poster/PDE\\_Patricia\\_Sowa.pdf](https://hpi.de/fileadmin/user_upload/fachgebiete/rabl/Lectures/PDE_Poster/PDE_Patricia_Sowa.pdf) .
- Manu Rastogi. 2020. Tutorial on LSTMs: A Computational Perspective. <https://towardsdatascience.com/tutorial-on-lstm-a-computational-perspective-f3417442c2cd#0d00>.
- Bhimala, K.R., Patra, G.K., Mopuri, R., Mutheneni, S.R. Prediction of COVID-19 cases using the weather integrated deep learning approach for India. 2021. *Transbound Emerg Dis.* 2021 Apr 10:10.1111/tbed.14102. doi: 10.1111/tbed.14102. Epub ahead of print. PMID: 33837675; PMCID: PMC8250893.
- Shrivastav L.K., Jha S.K. 2021. A gradient boosting machine learning approach in modeling the impact of temperature and humidity on the transmission rate of COVID-19 in India. <https://doi.org/10.1007/s10489-020-01997-6>
- Muhammad Rendana, Wan Mohd Razi Idris. New COVID-19 variant (B.1.1.7). 2021. Forecasting the occasion of virus and the related meteorological factors, *Journal of Infection and Public Health*, Volume 14, Pages 1320-1327, ISSN 1876-0341. <https://doi.org/10.1016/j.jiph.2021.05.019>.
- “Investopedia”. 2022. What Does Autoregressive Mean? <https://www.investopedia.com/terms/a/autoregressive.asp#:~:text=A%20statistical%20model%20is%20autoregressive,based%20on%20its%20past%20performance>.
- Phi, M. 2020. Illustrated guide to LSTM's and GRU's: A step by step explanation. Medium. from <https://towardsdatascience.com/illustrated-guide-to-lstms-and-gru-s-a-step-by-step-explanation-44e9eb85bf21>.
- “Medium”. 2022. What are the best metrics to evaluate your regression model? <https://towardsdatascience.com/what-are-the-best-metrics-to-evaluate-your-regression-model-418ca481755b> .
- Sammut, C., Webb, G.I. 2011. Mean Absolute Error. *Encyclopedia of Machine Learning*. Springer, Boston, MA. [https://doi.org/10.1007/978-0-387-30164-8\\_525](https://doi.org/10.1007/978-0-387-30164-8_525).
- Stephanie Glen. 2022. RMSE: Root Mean Square Error. <https://www.statisticshowto.com/probability-and-statistics/regression-analysis/rmse-root-mean-square-error/> .

- Stephanie Glen. 2022. Mean Absolute Percentage Error (MAPE) From StatisticsHowTo.com: Elementary Statistics for the rest of us! <https://www.statisticshowto.com/mean-absolute-percentage-error-mape/> .
- “CFI”. 2022. R Squared. <https://corporatefinanceinstitute.com/resources/knowledge/other/r-squared/> .
- Vijaysinh L. 2021. A Guide to Different Evaluation Metrics for Time Series Forecasting Models. <https://analyticsindiamag.com/a-guide-to-different-evaluation-metrics-for-time-series-forecasting-models/> .
- Stephen A., 2021. What is good MAPE Score? <https://stephenallwright.com/good-mape-score/> .
- Abu Bakar, Mohd Aftar, Mohd ariff, Noratiqah, Mohd Nadzir, Mohd Shahrul, Wen, Ong, Suris, Fatin Nur Afiqah. 2022. Prediction of Multivariate Air Quality Time Series Data using Long Short-Term Memory Network. Malaysian Journal of Fundamental and Applied Sciences. 18. 52-59. 10.11113/mjfas.v18n1.2393.

Ng Wei Shen (A176186)  
Azuraliza Abu Bakar  
Fakulti Teknologi & Sains Maklumat,  
Universiti Kebangsaan Malaysia