



**FAKULTI TEKNOLOGI DAN SAINS MAKLUMAT**

**BORANG PENYERAHAN LAPORAN ILMIAH**

SEM 2 SESI 2020 / 2021

**Bahagian A: Maklumat Diri Pelajar**

*Part A: Student's Details*

No. Matrik ( <i>Matric Number</i> )	A170894
Nama ( <i>Name</i> )	Tan Wei Ming
Program pengajian ( <i>Programme</i> )	Sarjana Muda Sains Komputer dengan Kejurian
No. Telefon ( <i>Telephone Number</i> )	011-36008592
Emel ( <i>Email</i> )	a170894@siswa.ukm.edu.my

Tajuk Projek (*Project Title*):

Ramalan Jumlah Kandungan Elektron GPS Menggunakan Ingatan Jangka Pendek yang Panjang dan Unit Berulang Berpagar

Tandatangan (*Signature*): Ming

Tarikh (*Date*): 10/7/2021

**Bahagian B: Perakuan Penyelia**

*Part B: Supervisor's Approval*

Saya peraku laporan ini telah disemak dan dibaiki, dan **menyokong / tidak menyokong\*** penyerahan laporan ilmiah ini.

*I certify that this report has been reviewed and amended, and **approved / rejected\*** the report submission.*

Tandatangan (*Signature*):

Tarikh (*Date*):

13/7/21

Cap Rasmi :

(*Official Stamp*)

# **RAMALAN JUMLAH KANDUNGAN ELEKTRON GPS MENGUNAKAN INGATAN JANGKA PENDEK YANG PANJANG DAN UNIT BERULANG BERPAGAR**

Tan Wei Ming  
Zalinda Binti Othman

Fakulti Teknologi dan Sains Maklumat, Universiti Kebangsaan Malaysia

## **ABSTRAK**

Cuaca angkasa merupakan keadaan di angkasa yang diakibatkan oleh matahari. Selama ini, cuaca angkasa telah mengakibatkan banyak masalah di bumi seperti menjejaskan isyarat radio, mematikan sistem elektrik, malah keadaan yang lebih serius akan menyebabkan manusia terdedah kepada radiasi. Projek ini bertujuan untuk meramalkan jumlah kandungan elektron (TEC) untuk membantu penyelidik cuaca angkasa dalam ramalan cuaca angkasa. Dalam projek ini, kajian hanya fokus pada permukaan ionosfera bumi. Dengan meramalkan TEC dalam gelombang radio, secara merentak keputusan tersebut dapat digunakan oleh penyelidik cuaca angkasa untuk meramalkan cuaca angkasa. Dalam kajian ini, data yang digunakan adalah data jumlah kandungan electron secara menegak (VTEC) berbentuk siri masa dan ukuran bagi VTEC ialah TECu. Selain itu, metodologi yang akan digunakan dalam kajian ini ialah perlombongan data siri masa. Terdapat enam fasa dalam perlombongan data iaitu fasa mengenalpasti masalah, fasa pengumpulan data, fasa prapemprosesan data, fasa pembinaan model, fasa penilaian corak dan fasa representasi pengetahuan. Seterusnya, dalam fasa pemodelan, kaedah pembelajaran mesin yang akan digunakan ialah ingatan jangka pendek yang panjang (LSTM) dan unit berulang berpagar (GRU). Selepas model dibina dan dilatih, kedua-dua model akan dinilai dan dibanding dengan menggunakan teknik penilaian model untuk memilih model yang mendapat ketepatan yang tinggi dalam ramalan. Akhirnya, model yang dipilih merupakan LSTM dan digunakan untuk meramal purata harian VTEC pada hari depan.

## **1 PENGENALAN**

Orang ramai biasa dengan perubahan cuaca di bumi tetapi tidak mengenali bahawa 'cuaca' juga berlaku di angkasa. Sama seperti mendorong cuaca ke Bumi, matahari juga bertanggungjawab atas gangguan di persekitaran ruang bumi kita. Oleh itu, cuaca angkasa dikenali sebagai keadaan atau aktiviti pada matahari (Zell, 2020). Cuaca angkasa dapat dipantau oleh jumlah kandungan electron (TEC) dalam sepanjang jalur gelombang radio di inonsfera bumi. Sementara itu, alat yang digunakan untuk merekod TEC ialah Sistem Penentududukan Global (GPS). TEC adalah jumlah elektron yang terdapat di sepanjang jalan antara pemancar dan

penerima radio (Weather, 2016). Oleh itu, ramalan terhadap TEC dapat membantu penyelidik dalam meramalkan cuaca angkasa. Pada masa yang sama, ramalan TEC memerlukan kaedah pembelajaran mendalam dengan ketepatan yang tinggi.

Kajian ini akan menggunakan ingatan jangka pendek yang panjang (LSTM) dan unit berulang berpagar (GRU) sebagai model pembelajaran mendalam untuk meramalkan TEC. Pembelajaran mendalam adalah satu subset daripada pembelajaran mesin dalam kecerdasan buatan yang mempunyai rangkaian yang mampu belajar tanpa pengawasan dari data yang tidak berstruktur atau tidak berlabel (Hargrave, 2021). Selain itu, kedua-dua model ini juga merupakan jenis model yang berbeza berasaskan dari rangkaian neural berulang (RNN). RNN merupakan sejenis rangkaian neural buatan (ANN) yang menggunakan data berurutan atau data siri masa (Education, 2020). Di samping itu, perbandingan antara kedua-dua model dijalankan selepas dilatih untuk memilih model yang memperolehi ketepatan yang tinggi dalam ramalan TEC dan kemudian menggunakan model yang dipilih meramalkan TEC pada hari depan. Selepas ramalan TEC dilakukan, ia dapat membantu penyelidik meramalkan cuaca angkasa pada hari depan.

## **2 PENYATAAN MASALAH**

Secara umumnya, cuaca angkasa akan mendatangkan kesan kepada sistem teknologi berasaskan angkasa dan bumi seperti gangguan pada penyebaran radio gelombang pendek dan grid kuasa elektrik, kerosakan orbit satelit, bahaya radiasi untuk satelit dan angkasawan dalam beberapa fasa misi angkasa (Weather, 2016). Selain itu, model pembelajaran mesin yang biasa digunakan dalam ramalan siri masa atau TEC merupakan purata pergerakan bersepadu regresif auto (ARIMA). Sebab model tersebut tidak digunakan dalam kajian ini disebabkan terdapat beberapa kelemahan model iaitu terhad dengan meramalkan nilai yang melampau jika data yang digunakan mempunyai titik terpencil (Gorgan, 2020). Sementara itu, model pembelajaran mesin sukar mengenali corak data yang kompleks dan tidak sesuai untuk ramalan jangka panjang (Pra, 2020).

Cadangan untuk menyelesaikan masalah ini adalah meramalkan TEC dan output tersebut

akan diambil oleh penyelidik untuk meramal cuaca angkasa. Oleh itu, ramalan TEC secara tidak langsung juga membantu penyelidik cuaca angkasa meramal cuaca angkasa. Selain itu, model LSTM dan GRU dapat mengatasi masalah dan kelemahan model ARIMA. Model LSTM dan GRU dapat menjalankan ramalan dengan data yang kompleks dan mempunyai kehilangan data, disebabkan kaedah pembelajaran mendalam tidak serupa dengan kaedah pembelajaran mesin iaitu ia dapat mempelajari dari data yang tidak lengkap atau mempunyai kehilangan data dan meramalkan dengan ketepatan yang tinggi. Seterusnya, pembelajaran mendalam berfungsi dengan mengambil ketepatan yang boleh diterima dari pengulangan berterusan dalam algoritmanya menerap dan mempelajari transformasi yang tidak linear dan corak yang berbeza (Brush, 2020).

### **3 OBJEKTIF KAJIAN**

- a) Mengkaji corak GPS TEC di seluruh Malaysia untuk mengenali corak TEC yang direkod dan dikumpul oleh penyelidik malaysia
- b) Pembangunan model LSTM dan GRU dalam ramalan GPS TEC
- c) Penilaian dan perbandingan prestasi model LSTM dan GRU dalam ramalan GPS TEC

### **4 METODOLOGI KAJIAN**

Metodologi kajian yang digunakan ialah perlombongan data siri masa. Perlombongan data merupakan proses untuk mengubah data mentah menjadi maklumat berguna (Twin, 2020). Selain itu, perlombongan data dibahagi kepada 6 fasa. Fasa-fasa tersebut adalah fasa mengenalpasti masalah, fasa pengumpulan data, fasa prapemprosesan data, fasa pemodelan, fasa penilaian corak dan fasa representasi pengetahuan. Dalam fasa prapemprosesan data, juga dibahagi kepada 4 fasa iaitu fasa penyatuan data, fasa pembersihan data, fasa pengurangan data dan fasa transformasi data. Di samping itu, fasa visualisasi data dijalankan ketika fasa prapemprosesan data. Perisian yang menjalankan perlombongan data adalah Excel dan *Jupyter Notebook*.

#### **4.1 Fasa Mengenalpasti Masalah**

Tujuan kajian ini adalah meramalkan VTEC pada hari depan dengan menggunakan pendekatan perlombongan data siri masa untuk membantu penyelidik dalam ramalan cuaca angkasa. Matlamat perlombongan data siri masa ini adalah membuat satu model yang dapat meramal purata harian VTEC pada hari depan dengan ketepatan yang tinggi. Oleh itu, model pembelajaran mendalam iaitu model LSTM dan GRU digunakan dalam perlombongan data siri masa. Kedua-dua model dipilih disebabkan mereka dapat mengenali corak dan urutan set data yang rumit dan kehilangan sesetengah data.

## 4.2 Fasa Pengumpulan Data

Dalam kajian ini, set data yang diperolehi adalah dari Pusat Sains Angkasa IPI UKM. Selain itu, set data tersebut berbentuk siri masa dan mengenai VTEC dalam harian dari Januari 2011 hingga Disember 2012. Di samping itu, set data tersebut mengandungi 4 atribut iaitu “Stesen & Tarikh (UKMBxxx-yyyy-mm-dd.std)”, “masa (dalam ukuran saat)”, “nombor satelit” dan “VTEC (dalam ukuran TECu)”. Keempat-empat atribut tersebut merupakan stesen serta tarikh ketika merekod VTEC, tempoh masa ketika merekod VTEC, nombor satelit Ketika merekod VTEC dan VTEC. Jumlah bilangan set data tersebut adalah sebanyak 4,062,812 bagi kedua-dua tahun dalam 645 fail yang berbeza tetapi selepas fasa pembersihan data, bilangan set data akan menjadi 731 dan bilangan atribut juga tinggal 2 sahaja. Selain itu, satu laporan data dijanakan dengan menggunakan kod Python dalam *Jupyter Notebook* selepas fasa pembersihan data. Rajah 1 menunjukkan laporan kualiti data.

	Hari	Purata_VTEC
<b>count</b>	731.000000	731.000000
<b>mean</b>	366.000000	28.597620
<b>std</b>	211.165812	7.885373
<b>min</b>	1.000000	11.880000
<b>25%</b>	183.500000	22.360000
<b>50%</b>	366.000000	27.990000
<b>75%</b>	548.500000	34.730000
<b>max</b>	731.000000	56.390000

Rajah 1 laporan kualiti data

## 4.3 Fasa Prapemprosesan Data

Urutan fasa ini bermula dari fasa penyatuan data, fasa pembersihan data, fasa pengurangan data dan fasa transformasi data.

### a) Fasa Penyatuan Data

Fasa ini akan menggabungkan set data dari sumber yang berbeza menjadi satu. Set data yang diperolehi daripada fail yang sebanyak 360 bagi tahun 2011 dan 285 bagi tahun 2012, setiap fail representasi satu hari. Oleh itu, terdapat beberapa hari yang hilang dalam data tersebut tetapi sebelum menyelesaikan hari yang hilang, penyatuan data akan dijalankan sebelum ini disebabkan data yang dikumpul di fail yang berbeza. Penyatuan data dijalankan mengikut tahun yang berbeza terlebih dahulu dalam Excel.

#### b) Fasa Pembersihan Data

Secara umumnya, fasa ini menguruskan set data yang hilang. Sebelum menjalankan fasa ini, seluruh VTEC dalam sehari dikira untuk mendapat purata VTEC bagi setiap hari dalam kedua-dua tahun masing-masing menggunakan alat *Pivot Table* dalam Excel. Sementara itu, dua set data yang baru akan dibuat mengikut tahun yang berbeza melalui *Pivot Table* dan mempunyai dua atribut iaitu “Label Baris” dan “Purata bagi Kolum 3”. Atribut “Label Baris” mengandungi stesen dan tarikh merekod VTEC manakala atribut “Purata bagi Kolum 3” mengandungi purata VTEC dalam harian mengikut tarikh dan tahun masing-masing. Terdapat nilai hilang dalam kedua-dua set data baru. Penggantian nilai dibuat dengan menggantikan purata bulanan atau tahunan VTEC pada nilai yang hilang.

#### c) Fasa Pengurangan Data

Fasa ini menjalankan pengurangan secara dimensi terhadap set data yang digunakan iaitu akan membuang atribut yang tidak berkaitan dengan matlamat kajian ini tetapi kedua-dua set data baru sudah mengurangkan sesetengah atribut semasa pembinaan set data baru menggunakan *Pivot Table* dalam fasa pembersihan data. Oleh itu, fasa ini hanya menerangkan atribut yang telah dan akan dibuang. Pertamanya, atribut yang telah dibuang adalah “nombor satelit” dan “masa (dalam ukuran saat)”. Hal ini kerana tujuan projek ini adalah meramal purata harian VTEC bukan nombor satelit dan masa. Semenjak itu, kedua-dua set data baru pada tahun 2011 dan 2012 akan digabungkan menjadi satu set data. Selain itu, atribut yang akan dibuang adalah “Label Baris”, ia juga tidak berguna dalam ramalan kemudian. Seterusnya, nama atribut “Purata bagi kolum 3” akan diubah menjadi “Purata\_VTEC”. Kemudian, set data baru ini diimport ke dalam *Jupyter Notebook*. Rajah 2 menunjukkan set data baru yang lengkap dan bersih.

Purata_VTEC	
0	15.93
1	15.47
2	16.27
3	16.68
4	14.88
...	...
726	19.96
727	19.96
728	19.96
729	19.96
730	19.96

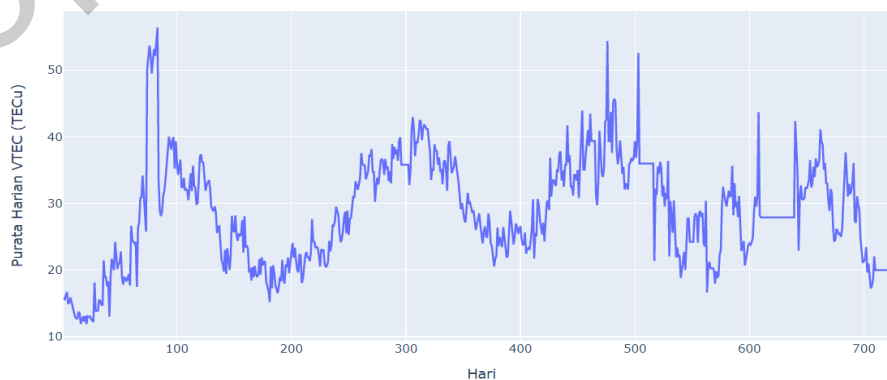
Rajah 2 Set data baru yang lengkap dan bersih

#### d) Fasa Transformasi Data

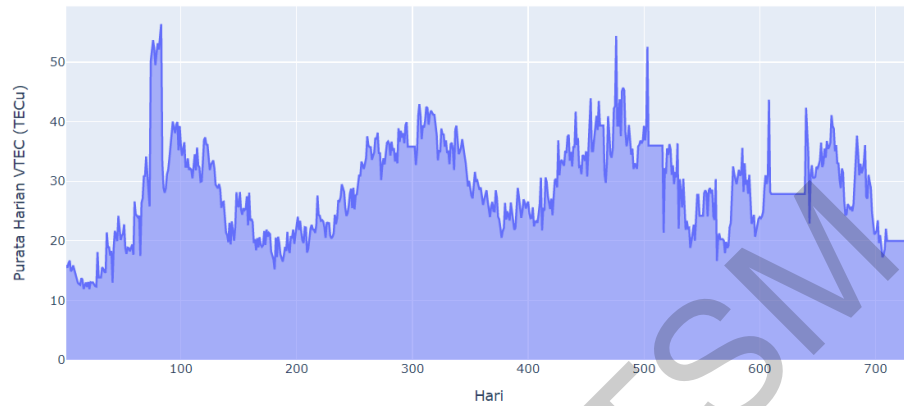
Set data diubah dengan susunan atau format yang sesuai untuk melombong data dalam fasa ini. Transformasi data melibatkan langkah-langkah seperti melicinkan, pengumpulan, kebijaksanaan, pembinaan atribut, normalisasi dan generalisasi. Normalisasi data dijalankan dalam fasa ini dan jenis normalisasi yang dijalankan adalah normalisasi min-maks dengan julat antara 0 dan 1.

#### 4.4 Fasa Visualisasi Data

Fasa ini menjalankan visualisasi set data dengan menggunakan representasi graf. Fasa ini penting disebabkan ramalan siri masa akan mengikut corak dan variasi data. Jenis graf yang digunakan adalah graf garis carta kawasan bertumpuk, carta bar dan graf berselerak. Rajah 3, 4, 5 dan 6 menunjukkan graf garis, carta kawasan bertumpuk, carta bar dan graf berselerak untuk set data baru.



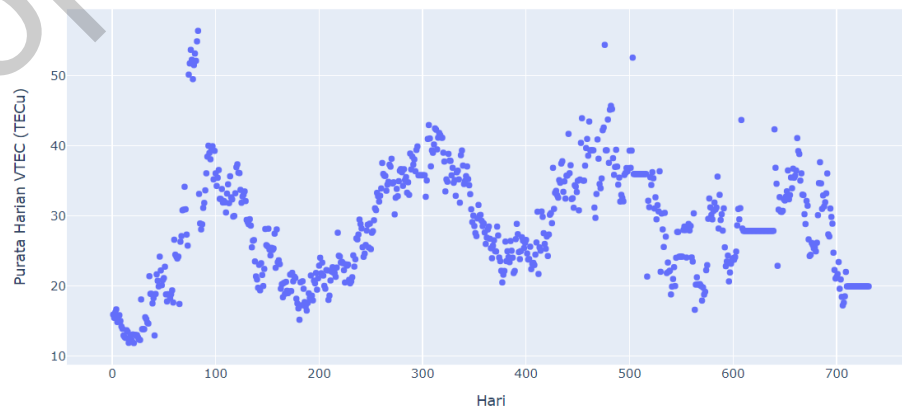
Rajah 3 Graf garis untuk purata harian VTEC dari tahun 2011 hingga 2012



Rajah 4 Carta kawasan bertumpuk untuk purata harian VTEC dari tahun 2011 hingga 2012



Rajah 5 Carta bar untuk purata harian VTEC dari tahun 2011 hingga 2012



Rajah 6 Graf berselerak untuk purata harian VTEC dari tahun 2011 hingga 2012

#### 4.5 Fasa Pemodelan



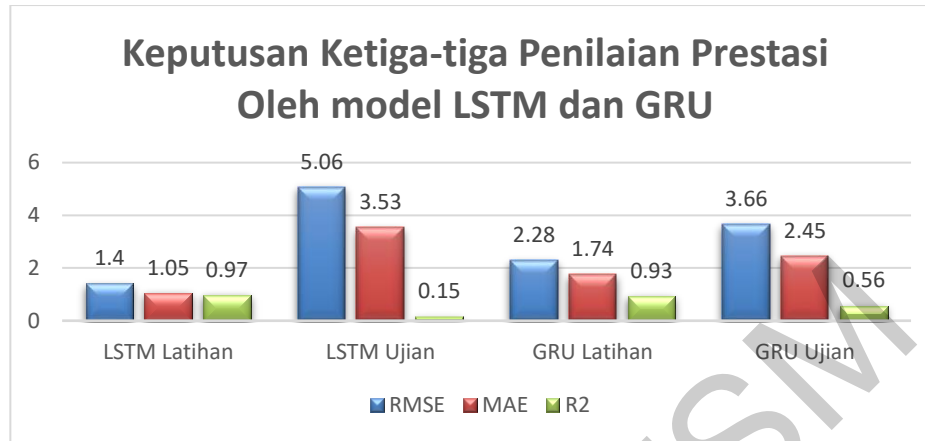
Fasa ini bertujuan untuk membina model dengan menggunakan kaedah pembelajaran mesin atau mendalam yang sesuai untuk ramalan, Model pembelajaran mendalam digunakan dan model-model tersebut merupakan model LSTM dan GRU. Sebelum fasa ini, set data dipecah kepada set data latihan dan ujian dengan peratusan 0.7 kepada set data latihan manakala 0.3 kepada set data ujian. Seterusnya, fungsi bernama *create\_dataset()* dibuat untuk mengubah susunan set data latihan dan set data ujian dengan *look\_back*. Dalam fungsi ini, susunan set data latihan dan ujian diubah kepada urutan dalam satu baris berdasarkan bilangan *look\_back* yang ditentukan. Bilangan *look\_back* yang ditentukan adalah 15. Sementara itu, satu set data baru juga akan dibuat untuk memperolehi nilai selepas nilai yang terakhir dalam urutan set data latihan dan ujian. Kemudian, set data latihan dan ujian dibentuk semula untuk menyesuaikan model LSTM dan GRU. Selepas ini, Kedua-dua model dibina dan dilatihkan dengan menggunakan set data latihan.

Secara umumnya, model LSTM yang dibina mempunyai 1 lapisan input dengan 50 neuron LSTM, 2 lapisan tersembunyi dengan 50 neuron LSTM dan 1 lapisan output yang membuat satu ramalan. Fungsi kerugian yang digunakan merupakan purata ralat kuasa dua (MSE). Model LSTM disesuaikan dengan set data latihan selepas dibina dan dilatih selama 200 kali.

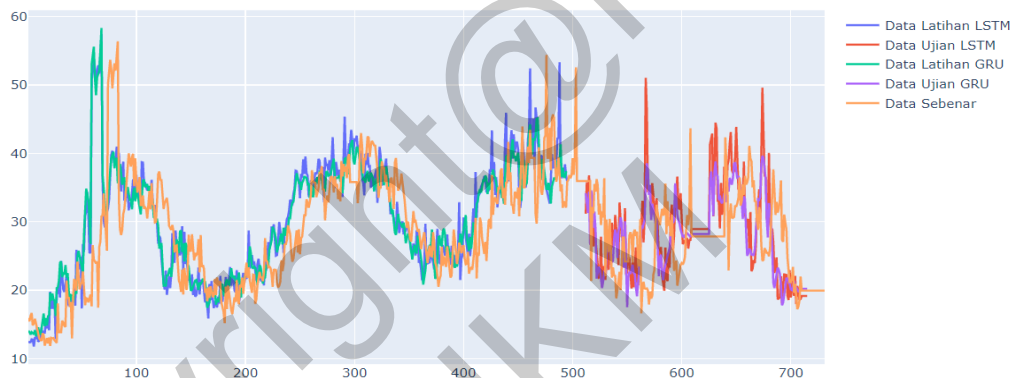
Kemudian, model GRU yang dibina mempunyai 1 lapisan input dengan 50 neuron GRU dan dropout digunakan adalah pecahan dalam 0.2, 1 lapisan tersembunyi dengan 50 neuron GRU dan dropout digunakan adalah pecahan dalam 0.2, dan 1 lapisan output yang membuat satu ramalan. Fungsi kerugian yang digunakan merupakan purata ralat kuasa dua. Semenjak itu, model GRU disesuaikan dengan set data latihan selepas dibina dan dilatih selama 200 kali.

#### **4.6 Fasa Penilaian Corak**

Fasa ini menjalankan penilaian dan pengujian prestasi model yang digunakan dengan menggunakan metrik penilaian. Jenis metrik penilaian yang digunakan adalah punca kuasa dua purata ralat kuasa dua yang dilakukan (RMSE), purata ralat mutlak (MAE), skor kuasa dua R ( $R^2$ ). Penilaian prestasi model ini dijalankan terhadap set data latihan dan ujian. Rajah 7 dan 8 menunjukkan carta bar untuk keputusan ketiga-tiga penilaian prestasi oleh model LSTM dan GRU dan graf garis untuk membandingkan nilai sebenar dengan ramalan.



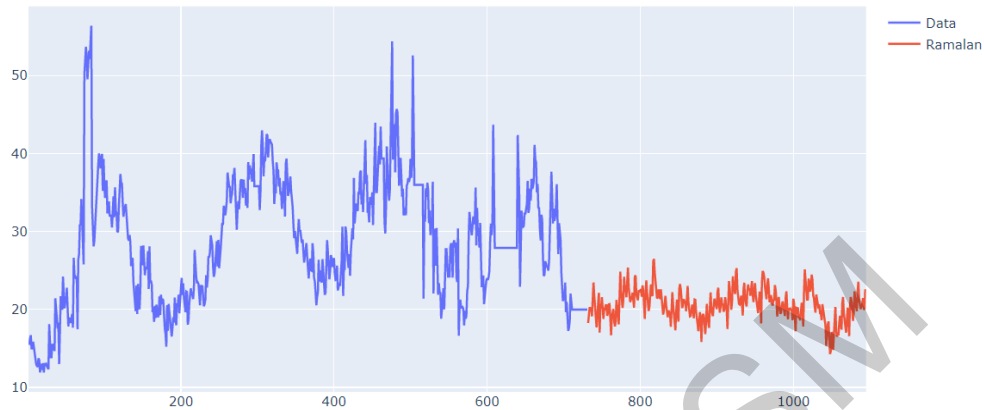
Rajah 7 Carta bar untuk keputusan ketiga-tiga penilaian prestasi oleh model LSTM dan GRU



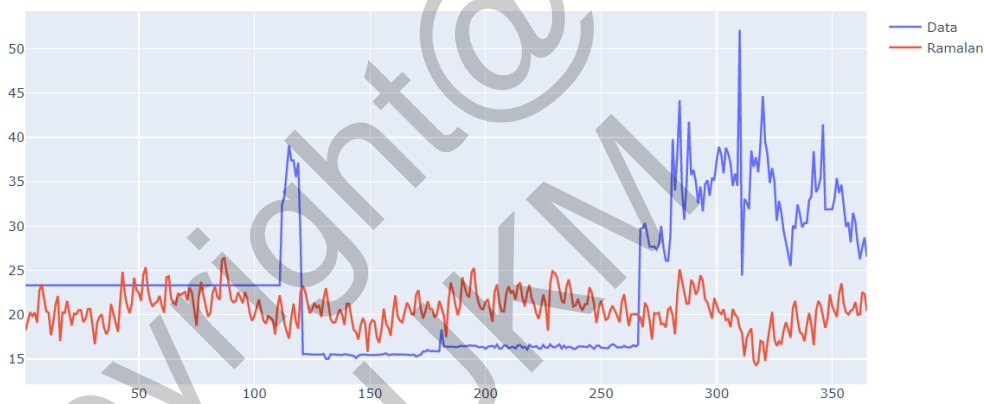
Rajah 8 Graf garis untuk membandingkan nilai sebenar dengan ramalan

#### 4.7 Fasa Representasi Pengetahuan

Fasa ini menjalankan representasi set data ramalan dengan menggunakan graf, antara muka pengguna dan jadual. Graf garis dan antara muka pengguna grafik digunakan dalam fasa ini. Selain daripada itu, model yang dipilih untuk meramalkan purata harian VTEC pada hari depan adalah model LSTM disebabkan model LSTM mendapat prestasi yang bagus dalam penilaian terhadap set data latihan. Bilangan hari yang diramalkan ialah 365 iaitu setahun selepas 31 Disember 2012. Sementara itu, satu set data yang baru yang berbilang 15 dibuat dari set data sebelum ini untuk sebagai data input ramalan pada hari depan. Nilai dalam set data input tersebut adalah urutan dari hari ke-695 hingga hari ke-710. Selepas ramalan dijalankan, graf garis diplotkan. Rajah 9 dan 10 menunjukkan graf garis untuk ramalan purata harian VTEC pada 365 hari depan selepas 31 Disember 2012 dan perbandingan nilai sebenar dan ramalan pada tahun 2013.



Rajah 9 Graf garis untuk ramalan purata harian VTEC pada 365 hari depan selepas 31 Disember 2012



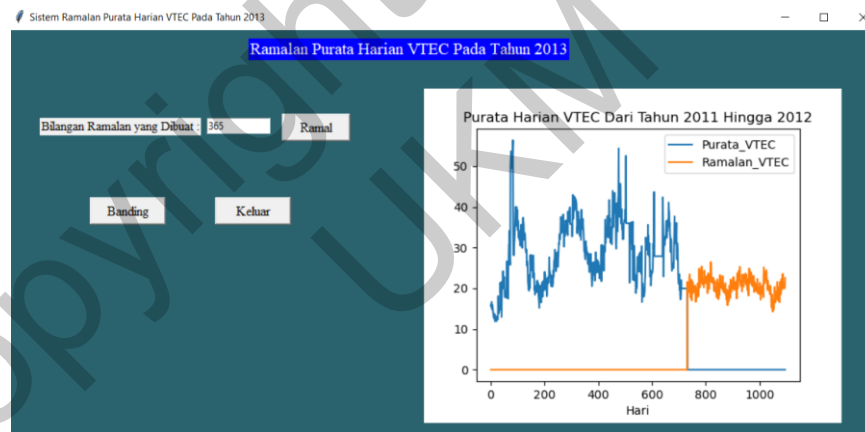
Rajah 10 Graf garis untuk perbandingan nilai sebenar dan ramalan pada tahun 2013

Seterusnya, satu antara muka pengguna grafik dibuat dengan menggunakan GUI Python. Dalam antara muka pengguna tersebut, mempunyai satu tajuk yang bernama “Ramalan Purata Harian VTEC pada Tahun 2013”, satu teks yang bernama “Bilangan Ramalan yang Dibuat :”, satu kotak teks, dan tiga button bernama “Ramal”, “Banding” dan “Keluar”. Secara umumnya, kotak teks tersebut digunakan untuk menentukan bilangan hari yang dibuat. Selain itu, butang-butang tersebut juga mempunyai kegunaan masing-masing. Bagi butang “Ramal” digunakan meramalkan purata harian VTEC berdasarkan bilangan ramalan yang ditentukan dalam kotak teks dan mempaparkan graf garis mengenai ramalan tersebut selepas ditekan manakala bagi butang “Banding”, ia dapat membandingkan nilai yang diramalkan dengan nilai sebenar pada

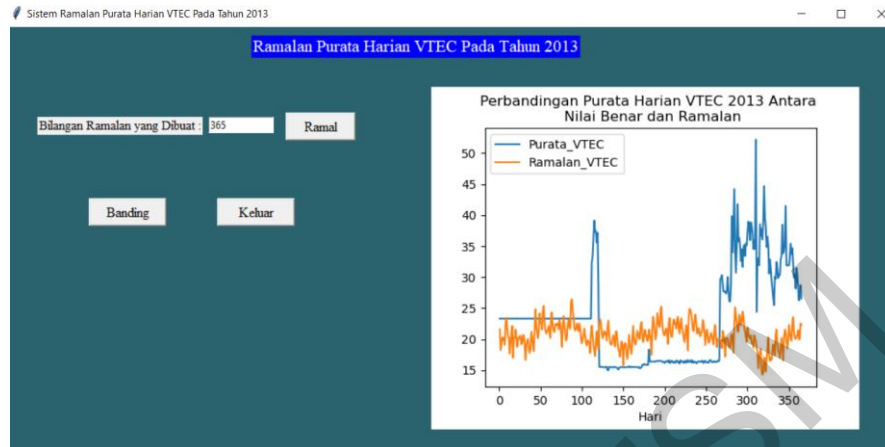
tahun 2013 dengan memaparkan graf garis selepas ditekan. Lebih-lebih lagi, kegunaan butang “keluar” adalah dapat terus menutup antara muka pengguna tersebut selepas ditekan. Rajah 11, 12 dan 13 menunjukkan antara muka pengguna grafik untuk ramalan purata harian VTEC dan perubahannya selepas menjalankan tindakan yang berbeza.



Rajah 11 Antara muka pengguna grafik untuk ramalan purata harian VTEC



Rajah 12 Antara muka pengguna grafik untuk ramalan purata harian VTEC selepas bilangan ramalan ditentu kepada 365 dan butang “Ramal” ditekan



Rajah 13 Antara muka pengguna grafik untuk ramalan purata harian VTEC selepas butang “Ramal” ditekan

## 5 KESIMPULAN

Kesimpulannya, kajian ini berjaya meramalkan purata harian VTEC pada hari depan. Model LSTM telah meramalkan 365 hari selepas 31 Disember 2012 dan ramalan yang dibuat menepati corak set data purata harian VTEC dari tahun 2011 hingga 2012. Selain itu, ramalan yang dibuat juga menepati corak pada tahun 2013 apabila dibandingkan dengan nilai sebenar pada tahun 2013. Set data ramalan ini dapat digunakan oleh penyelidik untuk meramalkan cuaca angkasa. Dengan itu, cuaca angkasa secara langsung dapat diramalkan. Cuaca angkasa merupakan bencana alam yang sukar diabaikan di angkasa. Dengan melakukan ramalan cuaca angkasa, sebarang kerugian, kerosakan dan bahaya dapat dikurangkan dalam peratusan yang rendah.

## 6 RUJUKAN

- Brush, K. 2020. *deep learning*. SearchEnterpriseAI:  
<https://searchenterpriseai.techtarget.com/definition/deep-learning-deep-neural-network>
- Education, I. C. 2020. *What are recurrent neural networks?* IBM Cloud Learn Hub:  
<https://www.ibm.com/cloud/learn/recurrent-neural-networks>
- Gorgan, M. 2020. *Limitations of ARIMA: Dealing with Outliers*. towards data science:  
<https://towardsdatascience.com/limitations-of-arima-dealing-with-outliers-30cc0c6ddf33>
- Hargrave, M. 2021. *Deep Learning*. Investopedia:  
<https://www.investopedia.com/terms/d/deep-learning.asp>

Pra, M. D. 2020. *Time Series Forecasting with Deep Learning and Attention Mechanism*. towards data science: <https://towardsdatascience.com/time-series-forecasting-with-deep-learning-and-attention-mechanism-2d001fc871fc>

Twin, A. 2020. *Investopedia, Business. Marketing Essentials*. What Is Data Mining?: <https://www.investopedia.com/terms/d/datamining.asp#:~:text=Data%20mining%20is%20a%20process,increase%20sales%20and%20decrease%20costs>

Weather, S. S. 2016. *Why is forecasting space weather important?* SANSa Space Weather: <https://spaceweather.sansa.org.za/space-weather-information/frequently-asked-questions/264-why-is-forecasting-space-weather-important>

Zell, H. 2020. Space Weather definition : [https://www.nasa.gov/mission\\_pages/rbsp/science/rbsp-spaceweather.html](https://www.nasa.gov/mission_pages/rbsp/science/rbsp-spaceweather.html)

Copyright@FTSM  
UKM