

Pengesanan Denyutan Pi2 Hampir Masa Nyata Berdasarkan Model Pembelajaran Mesin Terautomatik

(Near-Real Time Detection of Pi2 Pulsations Based on an Automated Machine Learning Model)

Khairul Adib Yusof^{a,b,*}, Mardina Abdullah^{b,c}, Nurul Shazana Abdul Hamid^{b,d} & Suaidi Ahadi^e

^aDepartment of Physics, Faculty of Science, Universiti Putra Malaysia, Seri Kembangan 43400, Malaysia

^bSpace Science Center, Institute of Climate Change, Universiti Kebangsaan Malaysia (UKM), Bangi 43600, Malaysia

^cDepartment of Electrical, Electronic and Systems Engineering, Faculty of Engineering and Built Environment, UKM, Bangi 43600, Malaysia.

^dDepartment of Applied Physics, Faculty of Science and Technology, UKM, Bangi 43600, Malaysia

^eDepartment of Geophysics, Indonesian Agency for Meteorology, Climatology and Geophysics, Jakarta 10610, Indonesia

*Corresponding author: adib.yusof@upm.edu.my

Received 25 June 2023, Received in revised form 12 February 2024

Accepted 15 March 2024, Available online 30 May 2024

ABSTRAK

Impak ribut kecil geomagnet terhadap cuaca angkasa dan pelbagai teknologi di Bumi mewujudkan keperluan untuk mengesan fenomena tersebut. Walau bagaimanapun, tempoh kejadian ribut kecil yang singkat menyukarkan ia sukar dikesan secara terus, justeru, ia pengesanan secara proksi dalam bentuk petanda perlu dilakukan. Denyutan Pi2, iaitu isyarat teredam tidak sekata pada komponen mendatar medan geomagnet (Hz), dilihat sebagai petanda ribut kecil yang andal. Penghentian indeks yang sebelum ini merupakan petunjuk utama pengesanan denyutan Pi2 memangkin kajian ini untuk memperkenalkan kaedah pengesanan yang baharu dengan menerapkan pembelajaran mesin. Dalam kajian ini, beberapa sifat dicari daripada data medan geomagnet berdasarkan parameter statistik, metrik impulsif dan isyarat untuk digunakan dalam pembangunan model pengelasan. Pembangunan ini dilakukan berdasarkan pendekatan pembelajaran mesin terautomatik (AutoML) yang menentukan algoritma terbaik dan mengoptimumkan hiperparameter secara automatik. AutoML mencuba pelbagai algoritma seperti ensembel, rangkaian neural, Naïve Bayes, mesin vektor sokongan, jiran terhampir k dan pepohon keputusan perduaan yang dioptimumkan oleh Algoritma Menyeparuh Berturutan Tak Segerak. Sebuah model pengelasan jenis ensembel ditentukan sebagai model berprestasi terbaik dengan ketepatan 99.15%. Ia seterusnya diguna pakai dalam pembangunan sistem pengesanan hampir masa nyata. Sistem tersebut menstrim data medan geomagnet berterusan dan melaporkan pengesanan Pi2 di repositori awan yang terbuka. Sistem pengesanan ini boleh dijadikan salah satu sumber rujukan untuk mengetahui kejadian denyutan Pi2 di serata dunia.

Kata Kunci: Denyutan Pi2; Medan Geomagnet; Pembelajaran Mesin Terautomatik; Sistem Pengesanan Hampir Masa Nyata

ABSTRACT

The impact of geomagnetic substorms on space weather and various technology on the Earth urges the need to detect the phenomenon. However, the brief period of substorm occurrences makes it more difficult to directly detect them, therefore, their detection by proxy in the form of precursors need to be done. Pi2 pulsations, which is a damped irregular signal in the horizontal component of the geomagnetic field ($6.7 - 25.0 \times 10^{-3}$ Hz), is seen as a reliable precursor to substorms. The cessation of W_p index that previously was the main detection indicator of Pi2 pulsations motivates this study to introduce a new detection method by implementing machine learning. In this study, several features were extracted from geomagnetic field data based on several statistical parameters, impulsive and signal metrics to be used in the development of classification model. The development was performed based on the automated machine learning (AutoML) approach that determines the best algorithm and optimizes the hyperparameters automatically. AutoML tries various algorithms like ensemble, neural network, Naïve Bayes, support vector machine, k-nearest neighbour and binary decision tree that are optimized by Asynchronous Successive Halving Algorithm. An ensemble classification model was determined to be the best performing model with an accuracy of 99.15%. It was then used in the development of near-real time detection system. The system streams the geomagnetic field data continuously and reports the Pi2 detections on an open cloud repository, The detection system is envisioned to be one of the reference sources to inquire the occurrences of Pi2 pulsations globally.

Keywords: Pi2 Pulsations; Geomagnetic Field; Automated Machine Learning; Near-Real Time Detection System

PENGENALAN

Ribut kecil geomagnet merupakan suatu fenomena yang membebaskan tenaga elektromagnet di kawasan ekor magnetosfera sebelah malam ke dalam kawasan ionosfera. Pembebasan ini melibatkan tenaga yang besar namun berlaku dalam masa yang singkat sekitar beberapa minit (Ghamry et al. 2020). Hal ini menjadikan ribut kecil antara fenomena cuaca angkasa yang paling sukar diperhatikan (Henderson 2022). Kejadiannya mengakibatkan pelbagai jenis fenomena lain termasuklah pemisahan aurora, pembebasan zarah bertenaga ke dalam magnetosfera, pendwikutuban medan magnet, aliran plasma yang meledak ke dalam lapisan plasma serta penyambungan semula magnet di dalam ekor magnetosfera. Ia juga menyebabkan gangguan kepada medan geomagnet di permukaan Bumi terutamanya di rantau latitud tinggi dan tengah (Park et al. 2020). Walaupun kecil secara relatif, ribut kecil membawa impak yang sangat besar terhadap Bumi dan penduduknya seperti sistem telekomunikasi, pengemudian dan pengagihan kuasa elektrik (Orr et al. 2021). Oleh yang demikian, pengesanan awal ribut kecil adalah sangat penting. Salah satu petanda awal ribut kecil yang boleh dipertimbangkan adalah denyutan Pi2 (Takahashi et al. 2022).

Denyutan Pi2 merupakan sejenis ayunan medan geomagnet yang teredam dan mempunyai bentuk gelombang tidak sekata. Kajian terdahulu mendapati bahawa Pi2 boleh diperhatikan dengan jelas di pancaran frekuensi ultrarendah yang mempunyai kala dalam julat

antara 40 – 150 s (Takahashi et al. 2022). Fenomena ini adalah berkait rapat dengan ribut dari magnetosfera dan gangguan lain yang berasal daripada ekor magnetosfera berhampiran Bumi. Magnetometer dasar yang terletak di latitud rendah ($L < 2$) digunakan untuk mengesan Pi2 pada semua masa tempatan (Kim et al. 2019). Hal ini menjadikan Pi2 sebagai petanda kepada permulaan ribut kecil geomagnet yang sangat berkesan. Sejak tahun 2005, pengesanan Pi2 bergantung kepada indeks W_p yang dibangunkan oleh Institute for Space-Earth Environmental Research melalui projek Substorm Swift Search (Nosé et al. 2012). Namun, projek ini berhenti menerbitkan data indeks tersebut untuk kegunaan awam pada hujung tahun 2019. Oleh yang demikian, terdapat keperluan untuk mengesan Pi2 melalui alternatif lain.

Pengaplikasian pembelajaran mesin (*machine learning*, ML) bagi mengesan denyutan Pi2 ialah suatu prospek yang sangat berpotensi untuk diterokai. Penggunaan ML walau bagaimanapun bukanlah suatu tugas yang mudah terutamanya apabila melibatkan data siri masa. Beberapa tahun kebelakangan ini, terdapat perkembangan terkini dalam ML bagi mengautomasikan prosedur-prosedur rutin pembangunan model berasaskan ML. Pengautomasian ini dikenali sebagai pembelajaran mesin terautomatik (*automated machine learning*, AutoML). Sesebuah kerangka kerja AutoML meliputi kejuruteraan sifat, pemilihan algoritma dan penalaan hiperparameter yang dijalankan berlelaran secara automatik dan dibantu oleh sebuah pengoptimum (Hutter et al. 2019). Penggunaan AutoML bakal menjimatkan masa kerana ia mencuba

beberapa algoritma dan kombinasi hiperparameter yang sesuai dalam memperoleh model terbaik tanpa penglibatan manusia (He, Zhao, & Chu, 2021).

Tiga faktor memangkin pelaksanaan kajian ini, iaitu: (i) keperluan pengesanan awal ribut kecil geomagnet melalui pengesanan denyutan Pi2, (ii) penghentian penerbitan indeks W_p yang sebelum ini digunakan sebagai petanda Pi2, dan (iii) perkembangan terkini dalam pengautomasian ML dalam bentuk AutoML. Dalam kajian ini, dengan menggunakan data medan geomagnet beresolusi tinggi dan indeks W_p yang pernah diterbitkan, model pengelasan dibangunkan melalui pendekatan AutoML. Kajian ini turut mengambil selangkah ke hadapan dengan membangunkan sistem pengesanan Pi2 hampir masa nyata sumber terbuka yang praktikal dengan kos yang rendah. Ia diharapkan mampu berperanan sebagai sumber maklumat kejadian Pi2 kepada komuniti penyelidik dalam memacu kajian-kajian dalam bidang berkaitan.

DATA DAN KAEDAH

Perihal dan Penyediaan Data Dua jenis data diguna pakai dalam kajian ini, iaitu indeks W_p dan medan geomagnet. Indeks W_p merupakan petunjuk yang berkesan dalam mengenal pasti kejadian ribut kecil, sama seperti indeks AE dan ASY. Indeks W_p diterbitkan berdasarkan data magnetometer di 11 buah stesen yang terletak di latitud rendah dan tengah ($\pm 20^\circ$ hingga $\pm 50^\circ$) seperti dalam Jadual 1. Maka, ia lebih peka terhadap gangguan yang berlaku di latitud rendah dan tengah dan merupakan pengukur kuasa Pi2 yang baik (Nosé et al. 2012). Ciri ini adalah berbeza berbanding indeks AE dan ASY yang lebih peka terhadap gangguan di latitud tengah dan tinggi (Mendes, Adhikari, Domingues, Echer, & Takeshi, 2022). Dalam kajian ini, data indeks W_p (tempoh pensampelan 1 minit) diperolehi daripada pangkalan data dalam talian www.isee.nagoya-u.ac.jp/~nose.masahito/s-cubed. Data yang diperolehi adalah terhad kepada tempoh projek tersebut berjalan iaitu dari 1 Januari 2005 sehingga 31 Disember 2019. Indeks ini mempunyai nilai antara 0 hingga 10, yang mana lebih daripada 99% daripada titik data berada dalam lingkungan nilai kurang daripada 0.50. Justeru, kajian ini mengguna pakai nilai ambang $W_p > 0.50$ bagi mengenal pasti 18202 kejadian denyutan Pi2 sepanjang tempoh indeks tersebut tersedia (tahun 2005 – 2019).

JADUAL 1. Senarai stesen magnetometer yang digunakan dalam penerbitan indeks W_p . Koordinat yang dinyatakan adalah geografi diikuti geomagnet (Nosé et al. 2012)

Lokasi	Kod	Latitud ($^\circ$)	Longitud ($^\circ$)
Tucson	TUC	32.2/39.7	249.3/316.2
Honolulu	HON	21.3/21.5	202.0/269.8
Canberra	CNB	-35.3/-42.5	149.4/226.9
Kakioka	KAK	36.2/27.2	140.2/208.8
Learmonth	LRM	-22.2/-32.3	114.1/186.5
Urumqi	WMQ	43.8/33.4	87.7/162.2
Izmir	IZN	40.5/37.6	29.7/109.5
Fürstfeldbruck	FUR	48.2/48.2	11.3/94.6
Ebro	EBR	40.8/43.0	0.5/81.3
Tristan da Cunha	TDC	-37.3/-31.4	347.5/53.5
San Juan	SJG	18.1/28.2	293.9/6.1

Bagi medan geomagnet, data berfrekuensi 1 Hz diperolehi daripada rangkaian magnetometer dunia INTERMAGNET yang menerbitkan data secara sumber terbuka (www.imag-data.bgs.ac.uk). Data tersebut mengandungi tiga komponen iaitu ke utara (N), ke timur (E) dan mencancang ke bawah (Z) (St-Louis, INTERMAGNET Operations Committee, & INTERMAGNET Executive Council, 2020). Bagi setiap denyutan Pi2, set data geomagnet berdurasi 10 minit dibentuk yang merangkumi komponen mendatar ($H = \sqrt{N^2 + E^2}$) dari setiap stesen indeks W_p (Jadual 1) yang berada pada waktu malam tempatan (jam 22:00 hingga 02:00). Kewujudan isyarat Pi2 dalam data geomagnet ditentukan melalui algoritma berikut: (i) penurasan data dalam julat frekuensi $6.7 - 25.0 \times 10^{-3}$ Hz, (ii) pengesanan puncak isyarat dalam tempoh 150 s pertama yang mencapai amplitud minimum 1 nT, (iii) sekiranya pengesanan berjaya, denyutan Pi2 dianggap wujud. Set data ini seterusnya dikelaskan sebagai data 'Pi2'. Sebaliknya jika pengesanan puncak gagal, set data ini dikelaskan sebagai 'bukan Pi2'.

Dua kelas ini diperlukan bagi mencapai matlamat kajian iaitu pengelasan perdua. Sebanyak 1657 set data 'Pi2' berjaya dibentuk, manakala set data 'bukan Pi2' adalah lebih sedikit. Bagi menambah data 'bukan Pi2', tempoh cerapan dianjak 20 saat ke belakang atau hadapan sehinggalah data ketika $W_p \leq 0.5$ berdurasi 10 minit ditemui. Pencarian ini dihadkan kepada waktu malam tempatan hari yang sama sahaja bagi meminimumkan sebarang pengaruh harian atau musim. Walaupun selepas penambahan, hanya 755 set data 'bukan Pi2' dapat dibentuk, justeru suatu teknik penokokan data diguna pakai bagi mencapai keseimbangan kelas selepas penyarian sifat dijalankan.

PENYARIAN DAN PEMILIHAN SIFAT

Dalam proses penyarian sifat, setiap set data yang mengandungi 600 titik data (10 minit \times 60 saat) dirangkumkan menjadi 10 sifat berdasarkan parameter statistik, metrik impulsif dan metrik isyarat, seperti yang disenaraikan dalam Jadual 2 (The MathWorks Inc., 2023a). Bagi menyeimbangkan kedua-dua kelas, Teknik Pensampelan Lebih Minoriti Sintetik (*Synthetic Minority Over-sampling Technique*, SMOTE) diguna pakai bagi menokok data kelas minoriti, iaitu 'bukan Pi2' dalam kajian ini. Melalui SMOTE, jiran-jiran terhampir titik data sedia ada dihubungkan, seterusnya titik data sintetik dicipta (x_s, y_s) sepanjang garisan penghubung ini. Algoritma ini diperihalkan oleh persamaan berikut:

$$\begin{aligned}(\Delta x, \Delta y) &= (x_1, y_1) - (x_2, y_2) \\ (x_s, y_s) &= (x_1, y_1) + r(\Delta x, \Delta y)\end{aligned}$$

Dalam persamaan tersebut, (x_1, y_1) dan (x_2, y_2) adalah dua titik data sedia ada yang berjiran dan $(\Delta x, \Delta y)$ adalah jarak antara dua titik tersebut. Bagi menerapkan kerawakan pada data yang dicipta, nombor rawak r berjulat antara 0 hingga 1 digunakan (Wongvorachan et al. 2023). Proses penokokan data ini menghasilkan set data yang seimbang kelas bersaiz 3314×10 , iaitu 1657 set setiap kelas.

JADUAL 2. Sifat-sifat berdasarkan parameter statik (*), metrik impulsif (†) dan metrik isyarat (°)

Sifat	Rumus
Faktor bentuk*	$x_{SF} = \frac{x_{RMS}}{ x }$
Kurtosis*	$x_{\kappa} = \frac{\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^4}{\sigma^4}$
Pencongan*	$x_{\gamma} = \frac{\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^3}{\sigma^3}$
Nilai puncak†	$x_p = \max_i x_i $
Faktor impuls†	$x_{IF} = \frac{x_p}{ x }$
Faktor kemuncak†	$x_{CF} = \frac{x_p}{x_{RMS}}$

bersambung ...

... sambungan

$$\text{Faktor kelegaan}^\dagger \quad x_{CL} = \frac{x_p}{\left(\frac{1}{N} \sum_{i=1}^N \sqrt{|x_i|}\right)^2}$$

$$\text{Nisbah isyarat kepada hingar}^\circ \quad x_{SNR} = \frac{P_s}{P_n}$$

$$\text{Jumlah herotan harmonik}^\circ \quad x_{THD} = \sqrt{P_d}$$

$$\text{Nisbah isyarat kepada hingar dan herotan}^\circ \quad x_{SINAD} = \frac{P_s + P_n + P_d}{P_n + P_d}$$

Rumus tambahan yang berkaitan

Punca min kuasa dua,

$$x_{RMS} = \sqrt{\frac{1}{N} \sum_{i=1}^N x_i^2}$$

Min nilai mutlak,

$$|\bar{x}| = \frac{1}{N} \sum_{i=1}^N |x_i|$$

Sisihan piawai,

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2}$$

Kuasa herotan,

$$P_d = \sqrt{\frac{\sum_{i=2}^N P_{s,i}}{P_{s,1}}}$$

Dalam kalangan 10 sifat yang dicari, terdapat sifat yang perlu lebih diutamakan berbanding yang lain, justeru, pemilihan sifat dilakukan menggunakan algoritma Kelewahan Minimum Kerelevanan Maksimum (*Minimum Redundancy Maximum Relevance*, MRMR). Algoritma ini mencari satu set sifat optimum yang paling saling berbeza (meminimumkan kelewahan) dan dapat mewakili pemboleh ubah bergerak balas dengan efektif (memaksimumkan kerelevanan). Bagi setiap sifat x , ia mengira kelewahan, W_x dan kerelevanannya, V_x menggunakan maklumat salingan sifat, I , yang ditakrifkan oleh rumus-rumus berikut:

$$V_x = I(x, y)$$

$$W_x = \frac{1}{|S|} \sum_{z \in S} I(x, z)$$

$$I(X, Z) = \sum_{i,j} P(X = x_i, Z = z_j) \log \frac{P(X = x_i, Z = z_j)}{P(X = x_i) P(Z = z_j)}$$

$$MIQ_x = \frac{V_x}{W_x}$$

Nilai hasil bahagi maklumat salingan (MIQ_x) kemudiannya dikira sebagai parameter akhir yang menentukan keutamaan sesuatu sifat (Xie et al. 2023). Melalui algoritma ini, 3 daripada 10 sifat dipilih untuk digunakan dalam pelatitan model.

PEMBANGUNAN MODEL PENGELASAN

Data input yang berdimensi 3314 × 3 dibahagikan kepada 75% data pelatitan dan 25% data pengujian secara rawak. Bagi pengoptimuman model, pengimplementasian AutoML melalui fungsi terbina MATLAB diguna pakai. Fungsi tersebut mampu melatih pelbagai model pengelasan secara lelaran. Model yang boleh dilatih adalah ensembel (ES; pembelajar lemah: jiran terhampir k , analisis pembeza layan dan pepohon keputusan), rangkaian neural (*neural network*, NN), Naïve Bayes (NB), mesin vektor sokongan (*support vector machine*, SVM), jiran terhampir k (*k-nearest neighbour*, KNN) dan pepohon keputusan perduaan (*binary decision tree*, BDT) dengan pelbagai nilai hiperparameter yang ditentukan oleh Algoritma Menyeparuh Berturutan Tak Segerak (*Asynchronous Successive Halving Algorithm*, ASHA) (The MathWorks Inc., 2023b). ASHA memilih kelompok pertama algoritma secara rawak dan melatih algoritma-algoritma tersebut dengan sebuah subset data pelatitan. Prestasi model yang dilatih dinilai menggunakan ralat pengelasan pengesahsahihan silang. Penilaian ini membolehkan ASHA memutuskan sama ada untuk menambah baik model tersebut dengan melatihnya semula menggunakan set data pelatitan yang lebih besar, atau memilih model dan nilai hiperparameter baharu menggunakan carian rawak (Li et al. 2020). Pengesahsahihan silang juga dilakukan bagi mengelakkan penyuaian lampau model daripada berlaku.

Di akhir proses pengoptimuman, model terbaik diperolehi. Model ini boleh diuji menggunakan pelbagai metrik penilaian prestasi seperti ketepatan, kepersisan, kepekaan, skor F_1 dan ketentuan, yang diberikan oleh:

$$\text{Ketepatan} = \frac{TP + TN}{TP + FP + TN + FN}$$

$$\text{Kepersisan} = \frac{TP}{TP + FP}$$

$$\text{Kepekaan} = \frac{TP}{TP + FN}$$

$$\text{Skor } F_1 = \frac{TP}{TP + \frac{1}{2}(FP + FN)}$$

$$\text{Ketentuan} = \frac{TN}{TN + FP}$$

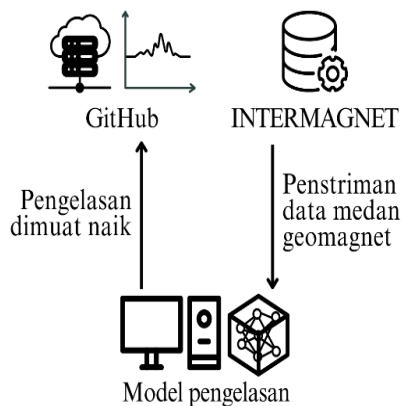
yang mana TP , FP , TN dan FN masing-masing merujuk kepada positif dan negatif benar dan palsu. Memandangkan pengelasan 'Pi2' dan 'bukan Pi2' yang tepat adalah sama penting, metrik tambahan yang lebih komprehensif dalam menilai prestasi model diperlukan. Bagi tujuan tersebut, pekali korelasi Matthews, ρ_M turut dikira:

$$\rho_M = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

Nilai ρ_M menandakan peramalan sempurna bagi kedua-dua kelas, walaupun salah satu kelasnya lebih atau kurang diwakili (Chicco & Jurman, 2020). Kajian ini menormalkan ρ_M (yakni $\tilde{\rho}_M = (\rho_M + 1)/2$) bagi mengubah nilai julat ρ_M daripada $[-1, 1]$ kepada $[0, 1]$ agar perbandingan dengan metrik lain dapat dilakukan dengan setara.

SISTEM PENGESANAN PI2 HAMPIR MASA NYATA

Sebagai kesinambungan kepada pembangunan model yang dilatih, sistem pengesanan hampir masa nyata dibangunkan. Sistem ini bergantung kepada tiga komponen utama iaitu: (i) pangkalan data INTERMAGNET, (ii) model pengelasan yang dijalankan oleh sebuah komputer, dan (iii) repositori awan sumber terbuka yang disediakan oleh GitHub (www.github.com). Data antara INTERMAGNET distrim kepada komputer melalui Antara Muka Pengaturcaraan Aplikasi (*Application Programming Interface*, API). Manakala data antara komputer dan GitHub dihubungkan oleh protokol Kerangka Selamat (*Secure Shell*, SSH) yang membenarkan pindahan data tersulit dua hala. Pindahan data antara ketiga-tiga komponen digambarkan dalam Rajah 1. Prinsip kerja sistem ini diulas lebih lanjut dalam bahagian Hasil dan Perbincangan.

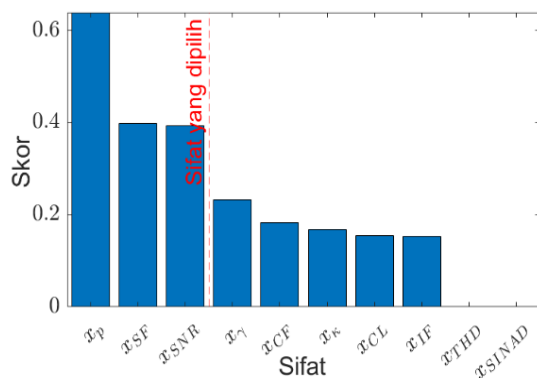


RAJAH 1. Pindahan data antara komponen sistem pengelasan hampir masa nyata

HASIL DAN PERBINCANGAN

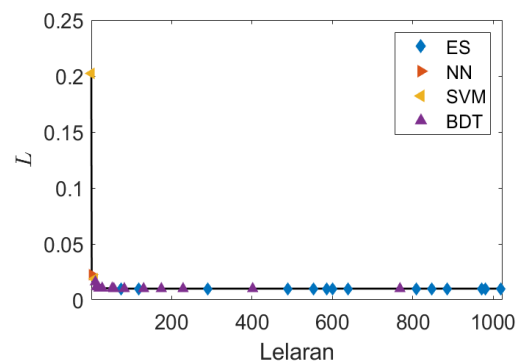
PEMBANGUNAN MODEL PENGELASAN

Model dilatih menggunakan sifat-sifat yang dipilih berdasarkan algoritma MRMR. Rajah 2 menunjukkan kedudukan sifat berdasarkan skor yang dikira menggunakan algoritma MRMR. Dalam rajah tersebut, tiga sifat yang berada di sebelah kiri garisan merah putus-putus merupakan sifat yang dipilih. Tiga sifat tersebut adalah: x_p (0.64), x_{SF} (0.40) dan x_{SNR} (0.39), masing-masing mewakili setiap satu jenis sifat, yakni metrik impulsif, parameter statistik dan metrik isyarat. Sifat x_p dan x_{SF} mencatatkan amplitud dan bentuk isyarat pepaku yang berdurasi singkat, manakala x_{SNR} menunjukkan kejelasan isyarat secara keseluruhan berbanding latar belakang. Dapat dirumuskan bahawa sifat-sifat yang dipilih adalah deskriptif kerana memberi pemerihalan menyeluruh terhadap isyarat Pi2 dari aspek amplitud, bentuk dan kejelasan.



RAJAH 2. Kedudukan sifat berdasarkan skor algoritma MRMR

Sifat-sifat ini digunakan untuk pelatihan model yang berlelar sebanyak 1023 kali dengan bantuan ASHA sebagai pengoptimum. Matlamat pelatihan adalah untuk mencari suatu kombinasi algoritma dan hiperparameter optimum, yakni yang memberikan nilai kerugian penentusahihan, L minimum. Evolusi pelatihan ditunjukkan dalam Rajah 3 yang mana garisan hitam adalah nilai L minimum yang dicatatkan setakat lelaran ke- i . Setiap kali sesuatu algoritma berjaya menurunkan nilai L , lelaran tersebut ditandakan dengan bentuk yang berbeza. Berdasarkan rajah tersebut, hanya empat algoritma pernah menurunkan nilai L , iaitu ES, NN, SVM dan BDT. Dapat diperhatikan bahawa SVM dan NN menurunkan L secara mendadak yang menunjukkan bahawa sifat-sifat yang dipilih menyimpan maklumat yang sangat jelas dalam membezakan antara kedua-dua kelas. Walau bagaimanapun, ASHA memilih untuk melatih SVM (segi tiga kiri kuning) dan NN (segi tiga kanan jingga) hanya di peringkat awal pelatihan yang mungkin dilakukan secara rawak. Pemerhatian ini menggambarkan bahawa kedua-dua algoritma tidak dapat dioptimumkan lebih lanjut bagi memperolehi model terbaik. Sebaliknya ASHA akan mengutamakan algoritma yang menghasilkan nilai L terendah pada lelaran-lelaran terdahulu dan hanya membuat penalaan hiperparameter secara halus kepada algoritma yang berpotensi sahaja. Dapat dilihat daripada Rajah 3 bahawa ES (berlian biru tua) dan BDT (segi tiga atas ungu) merupakan algoritma yang lebih berpotensi kerana kedua-duanya memonopoli pelatihan selepas lelaran ke-11.

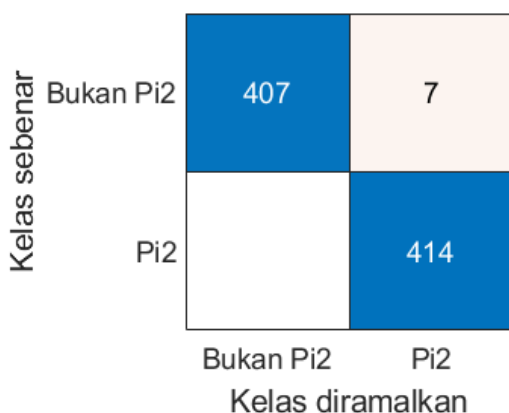


RAJAH 3. Nilai L melawan lelaran semasa pelatihan model. Algoritma model yang menurunkan L ditandakan dengan bentuk dan warna berbeza.

Di hujung pelatihan, ES menghasilkan model terbaik ($L = 0.0101$) dengan nilai hiperparameter yang dioptimumkan seperti dalam Jadual 3. Matriks kekeliruan model diilustrasikan dalam Rajah 4 yang mana petak biru menunjukkan peramalan tepat bagi kelas 'Pi2' ($TP = 414$) dan 'bukan Pi2' ($TN = 407$), manakala petak merah yang sangat cair menunjukkan peramalan salah bagi kelas 'Pi2' yang sangat rendah ($FP = 7$). Tiada peramalan salah bagi kelas 'bukan Pi2' yang dicatatkan ($FN = 0$).

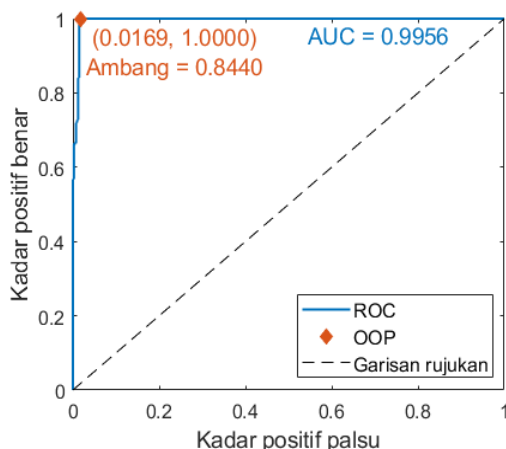
JADUAL 3. Nilai hiperparameter model ES yang dilatih

Hiperparameter	Nilai
Saiz set pelatihan	1989
Kaedah pengagregatan	Butstrap
Bilangan kitaran pembelajaran	215
Saiz dedaun minimum	4



RAJAH 4. Matriks kekeliruan model yang dilatih

Berdasarkan kadar positif benar (*true positive rate*, TPR) dan palsu (*false positive rate*, FPR), lengkung ciri pengoperasian penerima (*receiver operating characteristic*, ROC) diplotkan dalam Rajah 5 dan luas di bawah lengkung (*area under the curve*, AUC) dikira untuk mengkaji prestasi model pada semua nilai ambang pengelasan. Titik pengoperasian optimum (*optimal operating point*, OOP) ditandai bentuk berlian jingga, yang mana model mencapai kombinasi TPR dan FPR terbaik pada ambang pengelasan 0.8440. Pada titik ini, 100% peramalan kelas positif ('Pi2') adalah betul dan hanya 1.69% adalah salah.



RAJAH 5. Lengkung ROC dengan nilai AUC model. Tanda berlian menunjukkan nilai OOP dan ambang pengelasannya

Prestasi model ES tersebut dinilai menggunakan beberapa metrik yang diterangkan sebelum ini. Seperti yang diperihalkan dalam Jadual 4, semua metrik selain $\tilde{\rho}_M$ mencatatkan nilai yang sangat baik iaitu antara 0.9915 – 0.9917. Dapatan ini menunjukkan bahawa model yang diperolehi adalah seimbang dalam mengenal pasti kedua-dua kelas ('Pi2' dan 'bukan Pi2') dengan tepat. Ketepatan dan keseimbangan ini dirangkumkan oleh metrik yang komprehensif yakni $\tilde{\rho}_M$ dengan nilai yang sangat baik iaitu mencapai 0.9832. Prestasi model adalah lebih baik berbanding model-model terdahulu, misalnya model berasaskan jelmaan wavelet diskrit dengan ketepatan 0.7638 (Rabie et al. 2021), model pembelajaran mendalam dengan ketepatan 0.9111 (Rabie et al. 2021) dan model rangkaian neural buatan dengan kadar jaya 0.9000 (Sutcliffe, 1997).

JADUAL 4. Metrik penilaian model yang dilatih

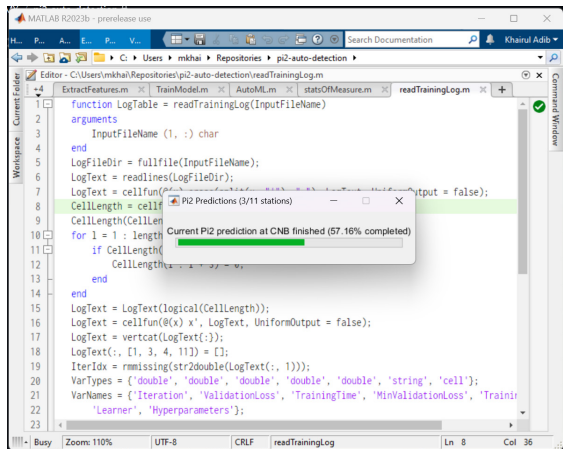
Metrik	Nilai
Ketepatan	0.9915
Kepersisan	0.9917
Kepekaan	0.9915
Skor F_1	0.9916
Ketentuan	0.9915
$\tilde{\rho}_M$	0.9832

SISTEM PENGESANAN PI2 HAMPIR MASA NYATA

Dalam seksyen ini, sistem pengesanan Pi2 hampir masa nyata dibincangkan. Di bahagian belakang sistem terdapat sebuah komputer terhubung Internet yang menjalankan perisian MATLAB dan sentiasa memperoleh data medan geomagnet berfrekuensi 1 Hz daripada pangkalan data INTERMAGNET. Terdapat sela masa sekitar 5 jam atau lebih antara waktu pencerapan dan penerbitan data. Sela ini adalah disebabkan prosedur prapemprosesan dan penentuan data oleh pihak INTERMAGNET (St-Louis et al. 2020). Disebabkan sela ini, penstriman data dan prosedur-prosedur seterusnya dilakukan dalam blok masa 6 jam, 12 jam selepas permulaan blok tersebut. Sebagai contoh, pengesanan bagi blok masa jam 17:00 hingga 23:00 pada 1 Jun 2023 (berdurasi 6 jam) dilakukan pada 05:00 2 Jun 2023, yakni 12 jam selepas itu. Data yang terlibat adalah semua stesen yang berada pada waktu malam tempatan (22:00 – 02:00) yang kemudiannya dipenggal menjadi beberapa segmen data berdurasi 10 minit.

Seterusnya, proses penyarian sifat yang sama dilakukan sebelum sifat-sifat ini dimasukkan ke dalam model pengelasan bagi mengesan kejadian denyutan Pi2.

Sekiranya denyutan dikesan, gambar rajah isyarat medan geomagnet pada julat frekuensi Pi2 dihasilkan dan dimuat naik ke GitHub bersama butiran waktu dan stesen. Dengan mengelog setiap pengesanan, sistem ini dapat menyemak kali terakhir pengesanan dilakukan. Ciri ini membolehkannya menyambung pengesanan yang terganggu sejauh 30 hari ke belakang. Rajah 6 menunjukkan antara muka MATLAB yang sedang menjalankan pengesanan pada data stesen Canberra (CNB).



RAJAH 6. Bahagian belakang sistem yang merangkumi komputer, perisian MATLAB, model pengelasan dan beberapa subrutin yang mengautomatiskan proses-proses berkaitan.

GitHub merupakan penyedia sistem kawalan versi Git berasaskan awan. Melalui protokol SSH yang terbina dalam sistem Git, pengesanan denyutan Pi2 dalam bentuk grafik dan teks dapat dimuat naik secara automatik tanpa memerlukan pengendalian manusia. Sebuah repositori awan dibuka di GitHub (<https://github.com/DataScienceMY/pi2-auto-detection>) yang boleh dicapai oleh orang awam (Rajah 7). Repositori ini memaparkan gambar rajah isyarat denyutan Pi2 yang dikesan serta waktu dan stesen. Harus ditekankan bahawa penggunaan Git dan GitHub sebagai bahagian hadapan sistem bukanlah penyelesaian elegan berbanding mempunyai laman sesawang rasmi yang khusus. Namun ia boleh dianggap penyelesaian paling menjimatkan kerana tidak melibatkan sebarang kos serta masih mencapai matlamat yang sama.

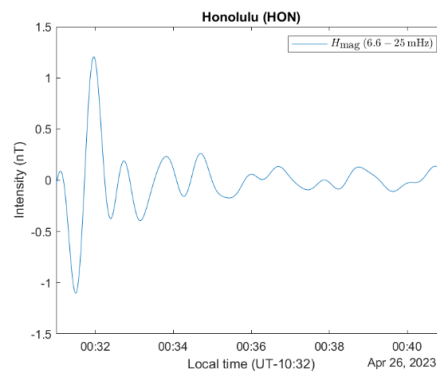
Sistem ini turut memperoleh indeks AE dalam bentuk plot grafik daripada World Data Center for Geomagnetism, Kyoto (https://wdc.kugi.kyoto-u.ac.jp/ae_realtime) dan memaparkannya bersama gambar rajah Pi2. Tujuan plot indeks AE tersebut adalah untuk menentusahkan pengesanan Pi2 kerana ia juga merupakan petanda ribut kecil geomagnet. Namun, hanya penentusahan tidak langsung dapat dilakukan kerana perbezaan rantau antara kedua-duanya, iaitu latitud tinggi (AE) berbanding latitud rendah dan tengah (Pi2) (Mendes et al. 2022). Pemerhatian

secara kasar mendapati pengesanan Pi2 majoritinya selari dengan peningkatan nilai AE, kecuali bagi beberapa kes.



RAJAH 7. Bahagian hadapan sistem merangkumi pelantar GitHub yang melaporkan pengesanan secara terbuka

Rajah 8 menunjukkan contoh pengesanan denyutan Pi2 pada jam 00:31 waktu tempatan 26 April 2023 di stesen Honolulu (HON). Isyarat tersebut dihasilkan daripada komponen H medan geomagnet yang dituras jalur dalam julat frekuensi Pi2 iaitu $6.7 - 25.0 \times 10^{-3}$ Hz. Ia menunjukkan puncak awal beramplitud 1.2 nT dengan kala 62.4 s, diikuti dengan isyarat menyusul yang teredam dan mempunyai amplitud antara 0.1 – 0.4 nT. Pengesanan yang dilaporkan oleh sistem ini adalah terbuka kepada orang awam, khususnya komuniti penyelidik cuaca angkasa bagi membolehkan mereka mengkaji mekanisme, ragam, implikasi dan aspek-aspek lain denyutan Pi2 secara lebih mendalam.



RAJAH 8. Contoh pengesanan denyutan Pi2 yang dilaporkan oleh sistem

KESIMPULAN

Denyutan Pi2 merupakan petanda ribut kecil geomagnet yang efektif terutamanya di latitud rendah dan tengah. Kajian ini dilaksanakan bagi mengusulkan kaedah dan sistem pengesanan Pi2 melalui pendekatan pembelajaran mesin terautomatik (AutoML) sebagai alternatif kepada indeks W_p yang telah dihentikan. Dalam kajian ini, penyarian dan pemilihan sifat dilakukan kepada data medan geomagnet INTERMAGNET. Melalui kerangka kerja AutoML, pelatihan pelbagai algoritma dan penalaan hiperparameter dijalankan secara automatik. Sebuah model pengelasan ensemble diperolehi yang seterusnya diuji dengan pelbagai metrik penilaian. Prestasi model tersebut adalah sangat baik dengan nilai antara 0.9832 – 0.9917. Ia seterusnya diguna pakai dalam sistem pengesanan Pi2 hampir masa nyata yang dibangunkan di repositori awan GitHub. Ia bergantung kepada penstriman data semasa daripada INTERMAGNET secara berterusan. Sistem ini walau bagaimanapun terbatas dari aspek sela masa pengesanan yang mencapai sehingga 6 jam. Keterbatasan ini berpunca daripada sela masa antara waktu data dicerap dan waktu ia diterbitkan oleh INTERMAGNET. Melalui penyingkatan sela masa ini, sistem ini akan dapat ditambah baik bagi mencapai pengesanan masa nyata sebenar.

PENGHARGAAN

Kajian ini ditaja oleh Universiti Putra Malaysia melalui Geran Putra: Inisiatif Putra Muda (Nombor Vot: 9710900). Data medan geomagnet diperolehi daripada pangkalan data sumber terbuka INTERMAGNET, manakala indeks W_p daripada Institute for Space-Earth Environmental Research. Penghargaan turut ditujukan kepada para ahli kumpulan projek 553 International Space Science Institute (ISSI) Bern yang berjudul “CSES and Swarm Investigation of the Generation Mechanisms of Low Latitude Pi2 Waves”, diketuai oleh Essam Ghamry dan Zeren Zhima atas perbincangan dan buah fikiran mereka.

PENGISYTIHARAN KEPENTINGAN BERSAING

Tiada kepentingan bersaing.

RUJUKAN

- Chicco, D., & Jurman, G. 2020. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics* 21(1): 6.
- Ghamry, E., Marchetti, D., Yoshikawa, A., Uozumi, T., Santis, A. de, Perrone, L., et al. 2020. The first Pi2 pulsation observed by China seismo-electromagnetic satellite. *Remote Sensing* 12(14): 2300.
- He, X., Zhao, K., & Chu, X. 2021. AutoML: A survey of the state-of-the-art. *Knowledge-Based Systems* 212: 106622.
- Henderson, M. G. 2022. Association of mesoscale auroral structures and breakups with energetic particle injections at geosynchronous orbit. *Frontiers in Astronomy and Space Sciences* 9.
- Hutter, F., Kotthoff, L., & Vanschoren, J. 2019. *Automated machine learning* (Vol. 1). Cham, Switzerland: Springer International Publishing.
- Kim, K.-H., Park, J.-H., Lee, D.-H., Lysak, R., Kwon, H.-J., & Hwang, J. 2019. Magnetic field oscillations observed by Swarm satellites in the nightside upper ionosphere during low-latitude Pi2 pulsations. *Journal of Geophysical Research: Space Physics* 124(8): 6596–6612.
- Li, L., Jamieson, K., Rostamizadeh, A., Gonina, E., Hardt, M., Recht, B., & Talwalkar, A. 2020. A system for massively parallel hyperparameter tuning. *arXiv*, from <http://arxiv.org/pdf/1810.05934v5>.
- The MathWorks Inc. 2023a. *Predictive Maintenance Toolbox user's guide* (Vol. 1). Massachusetts, USA: The MathWorks Inc.
- The MathWorks Inc. 2023b. *Statistics and Machine Learning Toolbox user's guide* (Vol. 1). Massachusetts, USA: The MathWorks Inc.
- Mendes, O., Adhikari, B., Domingues, M. O., Echer, E., & Takeshi, R. S. 2022. Interrelationships of similar magnetic effects at low and high latitudes during high-intensity long-duration auroral activity events: Case studies. *Brazilian Journal of Physics* 52(5).
- Nosé, M., Iyemori, T., Wang, L., Hitchman, A., Matzka, J., Feller, M., et al. 2012. W_p index: A new substorm index derived from high-resolution geomagnetic field data at low latitude. *Space Weather* 10(8): S08002.
- Orr, L., Chapman, S. C., Gjerloev, J. W., & Guo, W. 2021. Network community structure of substorms using SuperMAG magnetometers. *Nature Communications* 12(1): 1842.
- Park, J.-H., Kim, K.-H., Kwon, H.-J., Jee, G., & Hwang, J. 2020. A statistical study of Pi2 pulsations observed in the upper ionosphere using swarm magnetic field data. *Journal of Geophysical Research: Space Physics* 125(1).

- Rabie, E., Hafez, A. G., Saad, O. M., El-Sayed, A.-H. M., Abdelrahman, K., & Al-Otaibi, N. 2021. Geomagnetic micro-pulsation automatic detection via deep learning approach guided with discrete wavelet transform. *Journal of King Saud University - Science* 33(1): 101263.
- St-Louis, B., INTERMAGNET Operations Committee, & INTERMAGNET Executive Council 2020. *INTERMAGNET Technical Reference Manual, Version 5.0.0*. Edinburgh, UK: GFZ Data Services.
- Sutcliffe, P. R. 1997. Substorm onset identification using neural networks and Pi2 pulsations. *Annales Geophysicae* 15(10): 1257–1264.
- Takahashi, K., Lysak, R., & Vellante, M. 2022. Statistical analysis of Pi2 pulsations observed by Van Allen probes. *Journal of Geophysical Research: Space Physics* 127(9).
- Wongvorachan, T., He, S., & Bulut, O. 2023. A comparison of undersampling, oversampling, and SMOTE methods for dealing with imbalanced classification in educational data mining. *Information* 14(1): 54.
- Xie, S., Zhang, Y., Lv, D., Chen, X., Lu, J., & Liu, J. 2023. A new improved maximal relevance and minimal redundancy method based on feature subset. *The Journal of Supercomputing* 79(3): 3157–3180.